بسم الله الرحمن الرحیم

آمار و احتمال مهندسی

مدرس: دکتر محمد مهدی نایبی

دانشگاه صنعتی شریف

فصل ۱: تاریخچه و مفهوم احتمال

Section 1.1

تاریخچه و مفهوم احتمال:

در واقع شانس و عدم قطعیت تاریخچه‌ای به درازای تمدن بشریت دارد.

گفته می‌شود که شواهدی از قماربازی در ۳۵۰۰ سال قبل از میلاد به‌دست آمده (در مصر و ...) و تاسی شبیه تاس کنونی در مصر (۲۰۰۰ سال قبل از میلاد) به دست آمده است. متأسفانه قماربازی و تاس نقش مهمی در توسعهٔ تئوری احتمال داشته است.

تئوری احتمال به طور ریاضی توسط پاسکال (و فرما) در قرن ۱۷ آغاز شد که سعی در حل و به‌دست آوردن احتمال دقیق در برخی مسائل قماربازی به طور ریاضی داشتند. البته قبل از آنها نیز کاردان و گالیله (قرن ۱۶) به حل چنین مسائلی (به طور عددی) پرداخته‌اند.

از قرن هفدهم مرتباً تئوری احتمال توسعه یافت و در رشته‌های مختلف به‌کار گرفته شد. امروزه احتمال در اغلب زمینه‌های مهندسی و علوم مدیریت ابزار مهمی است و حتی استفاده از آن در پزشکی، رفتارشناسی، حقوق و ... مطرح است!

پاسکال: فوق‌العاده است که این علم در آغاز برای بررسی بازیهای شانس ابداع شده بود، ولی امروزه باید به عنوان مهمترین دانش بشری درآید.

با وجود کاربرد وسیع احتمال و علیرغم اینکه چنین مفاهیمی را دائماً در زندگی روزمره استفاده می‌کنیم، تعریف علمی یگانه‌ای برای احتمال وجود ندارد و در طول تاریخ رشد تئوری احتمال تعاریف مختلفی از احتمال شده است که هر یک بعداً مورد انتقاد دیگران قرار گرفته است.

۱. **تعریف کلاسیک احتمال** (توسط پاسکال در قرن ۱۷)

اگر در یک آزمایش تصادفی (بعد دقیق تعریف می‌کنیم)، تعداد کل نتایج ممکنه N باشد، احتمال واقعهٔ A عبارت است از:

$$P(A)=\frac{N_A}{N}=\frac{\text{تعداد نتایج مطلوب}}{\text{تعدادکل نتایج ممکنه}}$$

۲. **تعریف فراوانی (فرکانس) نسبی (تعریف آماری)**

این تعریف اولین بار در قرن جاری (۱۹۵۷) توسط Von Mises برای اصلاح تعریف کلاسیک معرفی شد.

اگر آزمایش تصادفی را n بار انجام دهیم، برای n بزرگ داریم: $P(A)\simeq\frac{n_A}{n}$، یا به تعبیر بهتر: $P(A)=\lim\limits_{n\to+\infty}\frac{n_A}{n}$.

۳. **تعریف ذهنی (Subjective)**

نگرش به احتمال به عنوان معیاری از میزان اعتقاد به یک امر است.

مثلاً وقتی می‌گوییم فلان متهم به احتمال ۷۰٪ مجرم است، در اینجا احتمال بیانگر میزان اعتقاد ما به حقیقت یک امر می‌باشد. البته این بسیار به قضاوت کننده بستگی دارد (Subjective است) و ممکن است با همان دلایل و مدارک شخص دیگری بگوید به احتمال ۹۰٪ مجرم است.

تعاریف گذشته اشکالاتی دارند که از آنها به عنوان مبنای یک تئوری ریاضی نمی‌توان استفاده کرد.

۴. تعریف اصولی (Axiomatic Definition)

این تعریف توسط ***کولموگروف*** در سال ۱۹۳۳ ارائه شد (البته سالها طول کشید تا مورد توجه قرار گیرد). در اینجا احتمال بر مبنای تئوری اندازه ارائه می‌شود و به هر واقعه عددی (که احتمال آن واقعه نامیده می‌شود و باید در اصول موضوعهٔ سه‌گانه صدق کند) نسبت داده می‌شود. اینکه چه عددی به هر واقعه نسبت داده شود (با فرض ارضاء شرایط اصول موضوعه) دلخواه است و ممکن است تطابق کامل با واقعیت نداشته باشد. اما با فرض صحت این احتمالات مفروض برای واقعه‌ها، با استفاده از تئوری احتمال می‌توانیم احتمال وقایع دیگر مورد نظرمان را به‌دست آوریم.

یعنی سه مرحله (در مدلسازی احتمالاتی) وجود دارد:

۱. با پروسه‌ای مواجهیم که به دلیل پیچیدگی، وقوع وقایع را به صورت احتمالی می‌خواهیم مدل کنیم (مثل شیر یا خط آمدن سکه و یا پیش‌بینی وضع هوا، در واقع پروسه‌های دترمینیستیک(Deterministic) و تابع قوانین فیزیکی) یا اینکه ذاتاً احتمالاتی است (مثل مکانیک کوانتومی). لذا برای واقعه‌های A_i، احتمالات $P(A_i)$ را در نظر می‌گیریم.

۲. با فرض اینکه $P(A_i)$ها اصول موضوعهٔ معینی را ارضاء کنند، با استفاده از منطق و استدلال احتمال وقایع B_i را $P(B_i)$ محاسبه می کنیم.

۳. پیش‌بینی می‌کنیم که در عمل وقایع B_i به احتمال $P(B_i)$ اتفاق بیفتند.

مراحل ۱ و ۳ با جهان خارج سر و کار دارند، در حالی که مرحلهٔ دوم کاملاً مفهومی است و ما در آن با مدل احتمالاتی که از جهان خارج ساخته‌ایم سر و کار داریم. در مرحلهٔ دوم هیچ شک و شبهه و گمانی نیست. همه چیز بر مبنای استدلالات کامل و دقیق منطقی است. یعنی به این ترتیب احتمال نیز علمی کاملاً دقیق و استدلالی خواهد بود.

در این درس ما از تعریف اصولی احتمال استفاده می‌کنیم. البته در مراحل ۱ و ۳ که نسبت دادن احتمالات به جهان خارج است، استفاده از سایر تعاریف احتمال مفید واقع می‌شود. ممکن است گفته شود چه فایده که نتایج لزوماً با نتایج واقعی تطابق نخواهد داشت. ولی در هر علم دیگری نیز همین‌طور است.

ما مدلهایی را که اغلب خیلی ساده‌تر از جهان خارج هستند را فرض کرده و آنالیز می‌کنیم و نتیجهٔ آنالیز را با تقریب برای آنچه در جهان خارج اتفاق می‌افتد پیش‌بینی می‌کنیم. مثلاً در تئوری مدار داریم: $R = \frac{V}{I}$. مقاومت واقعی با آنچه ما در آنالیز ایده‌آل خود مدل کرده‌ایم متفاوت است و لذا پاسخ مداری که در جهان خارج داریم کاملاً با پاسخی که تئوری مدار می‌دهد مطابقت نمی‌کند. ولی به عنوان تقریبی از آن قابل استفاده است.

نکتهٔ دیگر اینکه ممکن است تصور شود استفاده از احتمالات همواره ناشی از جهل ما نسبت به پدیده‌ها و قوانین حاکم بر آنها است و لذا آنالیز احتمالاتی را پیش می‌گیریم. ولی در واقع باید دانست که در بسیاری از موارد وقتی از آنالیز دترمینیستیک استفاده می‌کنیم، مسأله را آن‌قدر ساده کرده‌ایم که به مراتب میزان جهل و کنار گذاشتن اطلاعات در آن بیشتر است.

مثلاً مقدار مقاومت 2Ω دقیقاً 2Ω نیست، بلکه توزیعی حول و حوش 2Ω دارد. وقتی دو مقاومت 2Ω را سری می‌کنیم اگر فقط متوسطها را به‌کار بریم، می‌گوییم مقاومت حاصله 4Ω است. در حالی که آن هم یک رنج مقادیر و توزیع خاصی دارد.

یا گلولهٔ توپ وقتی پرتاب می‌شود، در آنالیز سادهٔ دترمینیستیک یک نقطه برای محل برخورد آن با زمین پیش‌بینی می‌شود، در حالی که در آنالیز احتمالاتی یک محوطه با تعیین میزان چگالی احتمال نقاط مختلف به‌دست می‌آید.

فصل ۲: مفاهیم اساسی احتمال

Chapter 2, Section 1.2

۱. یادآوری تئوری مجموعه‌ها
۲. فضای احتمال و تعریف اصولی احتمال
۳. تعیین احتمال واقعه‌ها و تعاریف دیگر احتمال
۴. مروری بر آنالیز ترکیبی
۵. مثالهایی از تعیین احتمال واقعه‌ها در آزمایشهای با نتایج هم احتمال
۶. احتمال شرطی
۷. قضیهٔ احتمال کل
۸. قضیهٔ بیز
۹. وقایع مستقل

ابتدا می‌خواهیم تئوری مجموعه‌ها را بررسی کنیم. ولی اول آشنا شویم که چرا در احتمال با تئوری مجموعه‌ها سر و کار (فراوان) می‌یابیم. اصولاً در احتمال با مشاهدات یا آزمایشهای تصادفی سر و کار داریم.

آزمایش تصادفی (Random Experiment) آزمایشی است که نتیجهٔ آن از پیش معلوم نیست، مثلاً انداختن تاس یا سکه.

طبق تعریف **فضای نمونه‌ها (Sample Space)** عبارت است از مجموعهٔ کلیهٔ نتایج (Outcome) ممکنه برای یک آزمایش تصادفی که آن را با Ω (یا S) نمایش می‌دهیم (توجه کنید که ممکن است آزمایشی فقط یک بار انجام شود، ولی در عین حال آزمایش تصادفی باشد) و احتمال برای هر زیر مجموعه‌ای از Ω تعریف می‌شود.

$$\Omega\ \boxed{\begin{matrix} \bullet\omega_1 & \bullet\omega_2 & \\ \bullet & \bullet & \bullet \\ \bullet & \cdots & \bullet\omega_N \end{matrix}}$$

ω_i : نقاط نمونه

مثلاً در آزمایش انداختن سکه داریم: $\Omega=\{H,T\}$ که ۲ عضو دارد.

یا در آزمایش انداختن تاس داریم: $\Omega=\{f_1,f_2,f_3,f_4,f_5,f_6\}$ که ۶ عضو دارد.

هر زیر مجموعه از فضای نمونه را **واقعه (Event)** می‌نامند. احتمال برای واقعه‌ها تعریف می‌شود.

مثلاً واقعهٔ زوج آمدن عدد تاس عبارت است از:

$A=\{f_i :$ زوج باشد $i\}=\{f_2, f_4, f_6\}$

به همین لحاظ به تئوری مجموعه‌ها نیازمندیم.

یادآوری تئوری مجموعه‌ها:

مجموعه (Set): دسته‌ای از اشیاء را مجموعه گویند. مانند: $\{a, b, 2,$ توپ$\}$.

هر عضو مجموعه را یک **عنصر (Element)** گویند. تعداد اعضای مجموعه می‌تواند محدود، نامحدود قابل شمارش (تناظر یک به یک با اعداد طبیعی) یا غیرقابل شمارش باشد. ترتیب در اعضای مجموعه‌ها مهم نیست.

مجموعهٔ A را **زیرمجموعهٔ** مجموعهٔ B گویند: $A \subset B$، اگر و تنها اگر هر عضو A متعلق به B نیز باشد.

مجموعهٔ A را **مساوی** مجموعهٔ B گویند، اگر و تنها اگر $A \subset B$ و $B \subset A$.

مجموعهٔ شامل تمام عناصر (المانهای) مورد نظر را **مجموعهٔ مرجع** Ω گویند.

مجموعهٔ فاقد عضو را **تهی** گویند: $\{\ \}$ یا $\varnothing$.

اجتماع (اتحاد) (Union):

مجموعهٔ $A\bigcup B$ (یا $A+B$)، مجموعهٔ عناصری است که در A یا در B باشند (یا در A یا در B یا در هر دو).

اشتراک (Intersection):

مجموعهٔ $A\bigcap B$ (یا AB)، مجموعهٔ عناصری است هم در A و هم در B باشند.

دو مجموعه را **جدااز هم (Disjoint)** گویند، اگر و تنها اگر $A\bigcap B=\varnothing$، یعنی عضو مشترکی نداشته باشند.

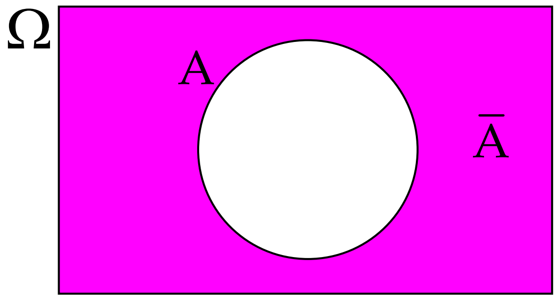
مکمل یک مجموعه (Complement):

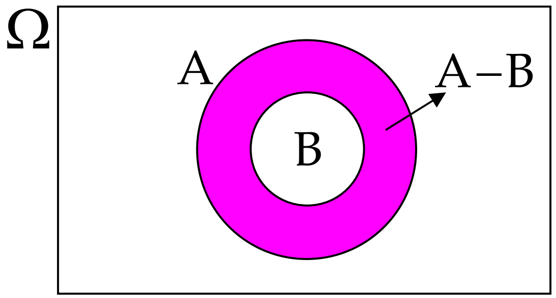
مکمل مجموعهٔ A، مجموعه‌ای است شامل تمام اعضای مجموعهٔ مرجع که در A نباشند و آن را با A^c یا $\bar{A}$ نمایش می‌دهیم.

طبق تعریف تفاضل دو مجموعهٔ A و B برابر است با:

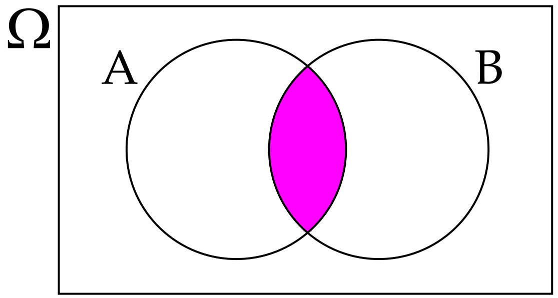
$$A-B=A\bigcap \bar{B}$$

دیاگرام وِن (Venn Diagram):

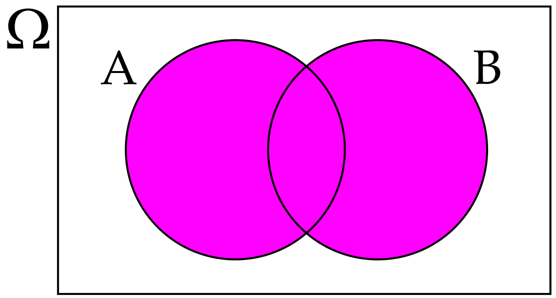




$B \subset A$



$A \cap B$



$A \cup B$

از تعاریف فوق (به سادگی) می‌توان نتیجه گرفت:

1) $A\cup\Omega=\Omega$ 2) $A\cap\Omega=A$ 3) $A\cup\varnothing=A$ 4) $A\cap\varnothing=\varnothing$

5) $B\subset A\Rightarrow A\cup B=A$ 6) $B\subset A\Rightarrow A\cap B=B$

7) $A\subset B, B\subset C\Rightarrow A\subset C$: قانون تعدّی 8) $A\cup B=B\cup A, A\cap B=B\cap A$: قانون جابجایی

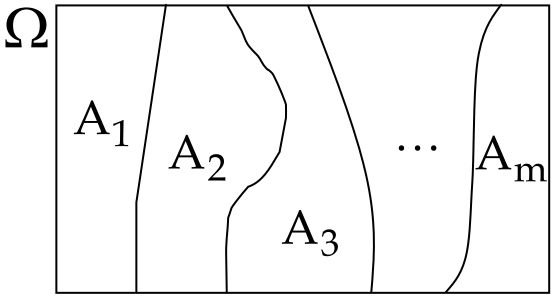
9) $\begin{cases}(A\cup B)\cup C=A\cup(B\cup C)\\(A\cap B)\cap C=A\cap(B\cap C)\end{cases}$: قانون شرکت‌پذیری (یعنی پرانتز لزومی ندارد)

10) $\begin{cases}(A\cup B)\cap C=(A\cap C)\cup(B\cap C)\\(A\cap B)\cup C=(A\cup C)\cap(B\cup C)\end{cases}$: قانون توزیع‌پذیری

11) $\begin{cases}\overline{(A\cup B)}=\bar{A}\cap\bar{B}\\\overline{(A\cap B)}=\bar{A}\cup\bar{B}\end{cases}$: قوانین دمورگان

افراز:

اگر مجموعه‌های غیرتهی $A_1, A_2, \ldots, A_m$ آن‌چنان باشند که:



$$\forall i \neq j : A_i \cap A_j = \varnothing \quad , \quad \bigcup_{i=1}^{m} A_i = \Omega$$

گوییم A_iها افرازی از Ω هستند.

حاصلضرب دکارتی:

حاصلضرب دکارتی مجموعهٔ A (با عناصر α_i) در مجموعهٔ B (با عناصر β_j) عبارت است از مجموعهٔ تمام زوج مرتب‌های به صورت (α_i, β_j) و به صورت $C = A \times B$ نشان داده می‌شود.

اگر A، m عضو و B، n عضو داشته باشد، $A \times B$، mn عضو خواهد داشت (می‌دانید که ترتیب در زوج مرتب مهم است).

مثال: حاصلضرب دکارتی مجموعهٔ $A = \{H, T\}$ در خودش برابر است با:

$$C = A \times A = \{(H,H), (H,T), (T,H), (T,T)\}$$

حال می‌توانیم تعریف اصولی احتمال را مطرح کنیم.

فضای احتمال:

فضای احتمال یا مدل احتمالاتی یک آزمایش از عوامل زیر تشکیل می‌شود:

۱. مجموعهٔ Ω شامل کلیهٔ نتایج ممکنهٔ ω_i برای آزمایش

۲. زیرمجموعه‌های Ω که واقعه نامیده می‌شوند.

۳. عدد $P(A)$ که به هر یک از واقعه‌ها (طبق اصول موضوعه) نسبت داده می‌شود.

فضای نمونه‌ها $\mathbf{(Sample\ Space)}$ $\Omega =$: مجموعهٔ کلیهٔ نتایج ممکنه برای یک آزمایش تصادفی

در آزمایش انداختن سکه:

دارای ۲ عضو : $\Omega=\{H,T\}$

در آزمایش انداختن دو سکه:

دارای ۴ عضو : $\Omega=\{(H,H),(H,T),(T,H),(T,T)\}$

در آزمایش انداختن دو تاس:

دارای ۳۶ عضو : $\Omega=\{(f_i, f_j) : i,j=1,2,\ldots,6\}$

در مداری با چهار سوئیچِ ON/OFF (0 یا 1):

دارای ۱۶=۲^۴ عضو (اعداد چهار بیتی باینری تصادفی) : $\Omega=\{0000,0001,\ldots,1110,1111\}$

طول عمر یک المان الکتریکی در مدار:

$$\Omega=\{T:0\leq T<+\infty\}$$

ولتاژ نویز روی یک مقاومت:

$$\Omega=\{V:-\infty<V<+\infty\}$$

واقعه (Event): هر زیرمجموعهٔ فضای نمونه‌ها را واقعه گویند.

در آزمایش انداختن تاس:

$$A=\{f_i:i\leq 4\}=\{f_1,f_2,f_3,f_4\}$$

در آزمایش انداختن دو سکه:

$A=\{$حداقل یکی از سکه‌ها شیر باشد$\}=\{(H,H),(H,T),(T,H)\}$

در مورد طول عمر یک المان الکترونیکی:

$A=\{$عمر المان بیش از ۵ ساعت نباشد$\}=\{T:0\leq T\leq 5h\}$

گوییم واقعهٔ A اتفاق افتاده است، هر گاه نتیجهٔ آزمایش یکی از اعضای A باشد.

توجه کنید که در یک بار انجام آزمایش تصادفی (Trial)، یک نتیجه (پیشامد) (Outcome) حاصل می‌شود (یکی از اعضای Ω)، ولی همزمان واقعه‌های مختلفی اتفاق افتاده‌اند.

از جملهٔ زیرمجموعه‌های Ω (واقعه‌ها)، خود Ω و $\varnothing$ هستند.

Ω را **واقعهٔ حتمی (Sure Event)** گویند، زیرا نتیجهٔ آزمایش مسلماً عضو Ω است. پس Ω اتفاق می‌افتد.

$\varnothing$ را **واقعهٔ ناممکن** یا **واقعهٔ خنثی (Null Event)** می‌گویند که هرگز اتفاق نمی‌افتد، زیرا نتیجهٔ آزمایش نمی‌تواند عضوی از اعضای $\varnothing$ باشد!

دو واقعهٔ A و B را **ناسازگار** (**مانعه الجمع** یا **غیرمتلاقی**) گویند، هر گاه مجموعه‌های A و B جدااَزهم باشند، یعنی:

$$A \cap B = \varnothing$$

واقعه‌ای را که تنها یک عضو داشته باشد، **واقعۀ ساده** گویند. مثلاً واقعۀ اینکه در آزمایش انداختن تاس عدد ۲ بیاید:

$A=\{f_2\}$

توجه کنید که f_2 (Outcome) با $\{f_2\}$ (Event) فرق دارد.

(در بعضی کتابها، Event، پیشامد ترجمه شده است.)

البته نحوۀ تعریف واقعه و نتیجه به نظر و مشخصات مورد توجه ما بستگی دارد.

مثلاً ممکن است بگوییم: $\Omega=\{f_1, f_2, f_3, f_4, f_5, f_6\}$ که در این صورت { زوج } یک واقعۀ ۳ عنصری است.

یا ممکن است بگوییم: { فرد , زوج }=Ω که در این صورت { زوج } یک واقعۀ ساده خواهد بود.

از طرف دیگر ممکن است محل قرار گرفتن تاس روی میز هم مورد نظر ما باشد که در این صورت دیگر $\{f_2\}$ نیز یک واقعۀ مرکب متشکل از بی‌نهایت عنصر خواهد بود.

حال که فضای نمونه و واقعه معلوم شد، قسمت سوم مدل، نسبت دادن احتمال به واقعه‌ها است.

تعریف Axiomatic:

به هر واقعهٔ A عدد $P(A)$ نسبت داده می‌شود به طوری که (اصول کولموگروف):

اصل (۱): $P(A)\geq 0$؛

اصل (۲): $P(\Omega)=1$؛

اصل (۳): اگر واقعه‌های A و B ناسازگار باشند $(A\cap B=\emptyset)$، آنگاه: $P(A\cup B)=P(A)+P(B)$.

اصل ۳ با تکرار آن برای هر تعداد محدودی از وقایع قابل بیان است، ولی نه برای تعداد نامحدود. اگر عناصر فضای نمونه نامحدود باشند، باید به جای اصل ۳، اصل قوی‌تری را جایگزین کرد.

اصل ('۳): اگر A_1، A_2 و ... دو به دو ناسازگار باشند، آنگاه:

$$P(A_1\cup A_2\cup\cdots)=P(A_1)+P(A_2)+\cdots$$

این را اصل جمع‌پذیری نامحدود (قابل شمارش) گویند (اصل ۳ حالت خاصی از اصل '۳ است).

[یا معادلاً می‌توانیم در کنار اصل ۳، اصل دیگری را اضافه کنیم که پیوستگی احتمال است. اگر B_1، B_2 و ... یک سری مجموعهٔ همگرا (صعودی یا نزولی) باشند، آنگاه:

$$P(\lim_{n\to+\infty} B_n)=\lim_{n\to+\infty} P(B_n)$$

به راحتی با مفروض گرفتن اصل '۳ می‌توان پیوستگی را ثابت کرد (کتاب راس، صفحهٔ ۴۸).]

حال با استفاده از همین اصول می‌توانیم احتمال واقعه‌های مختلف را از روی احتمال واقعه‌های داده شده به دست آوریم.

قضیهٔ ۱: $P(\bar{A})=1-P(A)$؛

زیرا:

$$\Omega=A\cup\bar{A}$$

$$P(\Omega)=P(A\cup\bar{A})$$

$$\downarrow \text{اصل ۱} \qquad \downarrow \text{اصل ۲}$$

$$1 \quad =P(A)+P(\bar{A})\Rightarrow P(\bar{A})=1-P(A)$$

قضیهٔ ۲: $P(\varnothing)=0$؛

زیرا:

$$\varnothing=\bar{\Omega}$$

$$P(\varnothing)=P(\bar{\Omega})=1-P(\Omega)=1-1=0$$

خواهیم دید که اگر چه احتمال واقعهٔ ناممکن صفر است، اما هر چه احتمالش صفر باشد را نمی‌گوییم ناممکن! و به همین ترتیب اگر چه احتمال واقعهٔ حتمی یک است $(P(\Omega)=0)$، ولی هر چه احتمالش یک باشد، حتمی نیست!

قضیهٔ ۳: $P(A)\leq 1$؛

زیرا طبق قضیهٔ ۱ داریم:

$$P(A)=1-P(\bar{A}) \rightarrow P(\bar{A})\geq 0 \Rightarrow P(A)\leq 1$$

(پس $0\leq P(A)\leq 1$ است.)

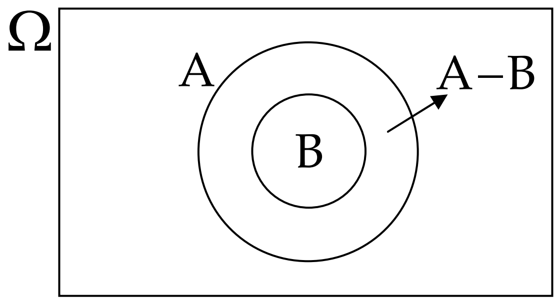
قضیهٔ ۴: اگر $B\subset A$ باشد، آنگاه: $P(B)\leq P(A)$؛

زیرا:

$$A=B\cup(\underbrace{A\cap\bar{B}}_{A-B})$$

چون B و $A\cap\bar{B}$ جداازهم هستند، طبق اصل ۳ داریم:

$$P(A)=P(B)+P(A\cap\bar{B}) \rightarrow P(A\cap\bar{B})\geq 0 \Rightarrow P(B)\leq P(A)$$



$B\subset A$

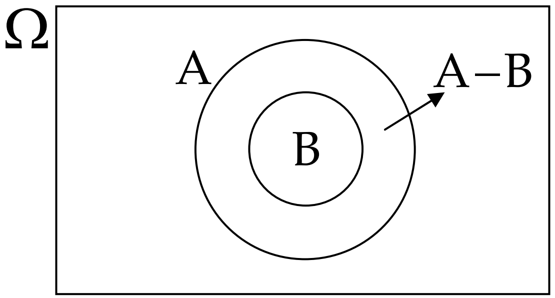
قضیهٔ ۵: اگر $B\subset A$ باشد، آنگاه: $P(A-B)=P(A)-P(B)$؛

زیرا:

$$A=B\cup(A-B)$$

چون B و $A-B$ جداازهم هستند، طبق اصل ۳ داریم:

$$P(A)=P(B)+P(A-B) \Rightarrow P(A-B)=P(B)-P(A)$$



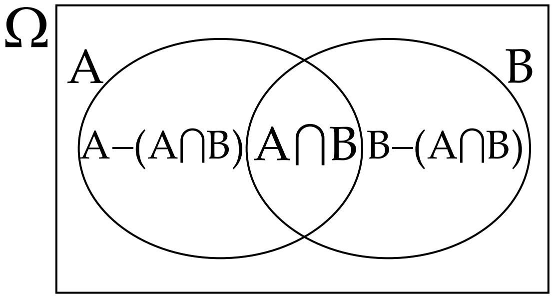
$B\subset A$

قضیهٔ ۶: برای هر دو مجموعهٔ A و B (نه لزوماً ناسازگار) داریم:

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)$$

اثبات:

$$\left.\begin{cases}P(A\cup B)=P(A\cup(\bar{A}\cap B))=P(A)+P(\bar{A}\cap B)\\ B=(A\cap B)\cup(\bar{A}\cap B)\Rightarrow P(B)=P(A\cap B)+P(\bar{A}\cap B)\end{cases}\right\}\Rightarrow P(A\cup B)=P(A)+P(B)-P(A\cap B)$$



یا به روش دیگر:

این سه مجموعه جداازهم هستند : $A\cup B=(B-(A\cap B))\cup(A\cap B)\cup(A-(A\cap B))$

$$\begin{aligned}P(A\cup B)&=P(B-(A\cap B))+P(A\cap B)+P(A-(A\cap B))\\&=P(B)-P(A\cap B)+P(A\cap B)+P(A)-P(A\cap B)\\&=P(B)+P(A)-P(A\cap B)\end{aligned}$$

تعیین احتمال واقعه‌ها:

گفتیم که برای مدل کردن یک آزمایش تصادفی باید احتمال کلیهٔ واقعه‌ها را تعیین کنیم (یک مجموعهٔ N عضوی، 2^N زیرمجموعه دارد). ولی با توجه به اصول موضوعه (و قضایا) نیازی نیست که به هر واقعه احتمالی نسبت دهیم. مثلاً اگر $P(A)$ را معلوم کنیم، $P(\bar{A})=1-P(A)$ خودبه‌خود معلوم خواهد بود. لذا با مشخص کردن احتمال یک تعداد حداقل واقعه، احتمال همهٔ واقعه‌ها معلوم خواهد بود. مثلاً اگر Ω متشکل از N نقطهٔ نمونهٔ $\omega_1,\omega_2,\ldots,\omega_N$ باشد، کافی است احتمال وقایع سادهٔ $\{\omega_1\},\{\omega_2\},\ldots,\{\omega_N\}$ را بدانیم. در این صورت اگر مجموعهٔ A شامل $\omega_{k_1},\omega_{k_2},\ldots,\omega_{k_r}$ باشد، طبق اصل ۳ خواهیم داشت:

$$P(A)=P\{\omega_{k_1}\}+P\{\omega_{k_2}\}+\cdots+P\{\omega_{k_r}\}$$

اگر $P\{\omega_i\}$ را P_i بنامیم، طبق اصل ۱ باید $P_i\geq 0$ بوده و طبق اصل ۲ باید $\sum_{i=1}^{N}P_i=1$ باشد. ولی P_iها از هر حیث دیگر اختیاری هستند (در تعریف اصولی).

حال اگر تعداد عناصر Ω نامحدود، ولی قابل شمارش باشد، باز هم بحث فوق صادق است و با تعیین $P_i=P\{\omega_i\}$ها که $P_i\geq 0$ و $\sum_{i=1}^{+\infty}P_i=1$، فضای احتمال مشخص می‌شود (مثلاً $P_1=\frac{1}{2}$، $P_2=\frac{1}{4}$، $P_3=\frac{1}{8}$ و ...).

ولی اگر تعداد عناصر Ω نامحدود غیرقابل شمارش باشد، مثلاً فضای نمونهٔ زمان شروع یک مکالمهٔ تلفنی (یا زمان خراب شدن یک المان الکترونیکی)، در اینجا هر فاصلهٔ $\{t_1 \le t \le t_2\}$ یک واقعه است و $\Omega=\{0\le t\le +\infty\}$.

اغلب در چنین مواردی احتمال واقعهٔ سادهٔ $P\{t=t_i\}$ برابر صفر است، اگر چه Ω اجتماع این واقعه‌های ساده است.

(این مسئله تناقضی با اصل ۳ یا اصل ′۳ ندارد، زیرا این اصول برای حالت غیرقابل شمارش نبودند.)

(اگر چه احتمال واقعهٔ ناممکن صفر است، اما هر چه احتمالش صفر باشد ناممکن تلقی نمی‌شود.)

در اینجا فضای احتمال را نمی‌توان با احتمال واقعه‌های ساده مشخص کرد. در عوض باید احتمال بازه‌ها را معین کرد. برای این منظور تابعی را که بیانگر چگالی احتمال است معرفی می‌کنیم:

$$P\{t_1 \le t \le t_1 + dt\} = \alpha(t_1)dt$$

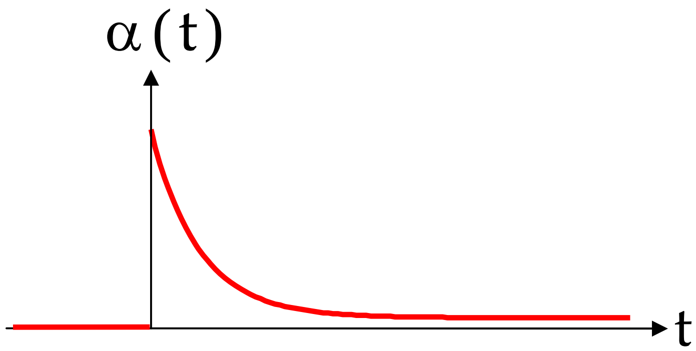
یا:

$$P\{t_1 \le t \le t_2\} = \int_{t_1}^{t_2} \alpha(t)dt$$

چون احتمال هیچ فاصله‌ای نمی‌تواند منفی شود، پس: $\alpha(t)\ge 0$ و چون احتمال Ω یک است، باید داشته باشیم:

$$\int_{\Omega} \alpha(t)dt = 1$$

در مثال بالا که $\{0 \le t \le +\infty\}$ بود، خواهیم داشت:



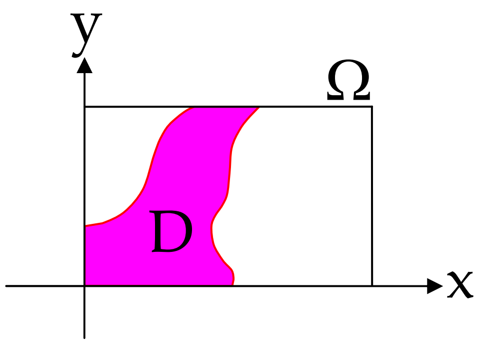
$$\int_{0}^{+\infty} \alpha(t)dt = 1$$

چون کلیهٔ واقعه‌های Ω به صورت اجتماع و اشتراک فواصل $[t_1, t_2]$ قابل بیان هستند، به این ترتیب احتمال کلیهٔ واقعه‌ها تعیین شده است.

به همین ترتیب در فضای دوبُعدی نیز داریم:

$$P\{(x,y) \in D\} = \iint_{D} \alpha(x,y)\,dx\,dy$$

که $\alpha(x,y)$ یک رویه است و باید داشته باشیم:



$$\alpha(x,y) \ge 0 \quad , \quad \iint_{\Omega} \alpha(x,y)\,dx\,dy = 1$$

مثلاً موقعیت یک شیء.

ما احتمال را بر مبنای تعریف اصولی پی‌ریزی می‌کنیم، ولی از تعاریف دیگر با توجه به اشکالاتشان نه به عنوان تعریف بلکه در مراحل ۱ و ۳ مدل‌سازی برای ایجاد ارتباط مدل با جهان خارج استفاده می‌نماییم.

تعریف ذهنی (شخصی) (Subjective): احتمال به عنوان معیاری از میزان اعتقاد به یک امر

مثلاً با قرائن و شواهدی که در آسمان می‌بینیم، می‌گوییم به احتمال ۸۰٪ فردا باران می‌بارد. احتمال‌هایی که در چنین جمله‌ای بدان اشاره می‌شود، اندازهٔ اعتقاد فرد گوینده است. روشن است که این تعریف وابسته‌به‌فرد، نمی‌تواند مبنای یک تئوری ریاضی و مستحکم قرار گیرد. اما می‌توانیم در مواردی از آن برای نسبت دادن معقول $P(A_i)$ها در تعریف Axiomatic استفاده کنیم. به نظر منطقی می‌رسد که احتمال به عنوان اندازهٔ اعتقاد نیز باید اصول احتمال را رعایت کند. مثلاً اگر کسی می‌گوید ۷۰٪ مطمئن هستیم که کتاب *تجرید الاعتقاد* نوشتهٔ *خواجهٔ طوسی* می‌باشد و ۲۰٪ مطمئن هستیم که آن را *ابوریحان* نوشته است، پس منطقی است که بگوییم به اعتقاد ما به احتمال ۹۰٪ یکی از این دو نفر این کتاب را نوشته است. در بحث تخمین از این تعبیر احتمال سود خواهیم برد.

تعریف فراوانی (فرکانس) نسبی (تعریف آماری یا تجربی):

اگر یک آزمایش تصادفی را به کرّات (n بار) انجام دهیم و در این n بار، n_A بار واقعهٔ A اتفاق افتد، علی‌الاصول «برای n بزرگ، $P(A)$ عددی نزدیک به $\frac{n_A}{n}$ خواهد بود»، یعنی: $P(A) \simeq \frac{n_A}{n}$. این دیگر Subjectiv نیست و قابل بررسی تجربی است.

مثلاً در آزمایش پرتاب سکه: $\Omega=\{H,T\}$ است. اگر $A=\{H\}$ باشد، برای محاسبهٔ $P(A)$، سکه‌ای را به کرّات پرتاب می‌کنیم:

آزمایش Pearson:

$$\frac{n_A}{n}=\frac{\text{تعداد شیرهای حاصله در پرتاب‌ها}}{\text{تعداد پرتاب‌ها}}=\frac{12012}{24000}=0.5005$$

یا در هر تولد: $\Omega=\{m,f\}$ است. اگر $A=\{m\}$ باشد، برای محاسبهٔ $P(A)$، داریم:

تعداد متولدین ذکور در سال ۱۹۶۰ در آمریکا:

$$\frac{n_A}{n}=\frac{2179708}{4257850}=0.5121$$

(مطالعات فراوان نشان داده که در واقع این عدد بیشتر از ۰/۵ است.)

ولی این تعریف خیلی شهودی و غیرریاضی است: «برای n بزرگ»، «عددی نزدیک به» و ...

پس این هم به عنوان تعریف احتمال قابل استفاده نیست. ولی در مراحل ۱ و ۳ برای ایجاد ارتباط بین $P(A)$ در مدل با نسبت تجربی $\frac{n_A}{n}$ و ربط دادن مدل به جهان خارج مفید است. حتی می‌توانیم برای اصول موضوعه و قضایای احتمال، تعبیر فرکانسی بیاوریم تا محسوس بودن آنها را نشان دهد.

مثلاً Ω در هر واقعه‌ای اتفاق می‌افتد، پس در n بار تکرار آزمایش، $n_\Omega = n$، یعنی: $P(\Omega) \simeq \frac{n_\Omega}{n} = 1$.

یا مثلاً اگر واقعه‌های A و B ناسازگار باشند (مجموعه‌های A و B عنصر مشترکی ندارند، پس A و B توأماً اتفاق نمی‌افتند). پس در n بار تکرار آزمایش، اگر واقعهٔ A، n_A بار و واقعهٔ B، n_B بار اتفاق بیفتد، آنگاه:

$$n_{A\cup B} = n_A + n_B$$

$$P(A\cup B) \simeq \frac{n_{A\cup B}}{n} = \frac{n_A}{n} + \frac{n_B}{n} \simeq P(A) + P(B)$$

یعنی با این تعریف دیگر نیازی به اصول موضوعه نیست و آنها قابل اثبات هستند.

تعریف بهتر:

$$P(A) = \lim_{n\to+\infty} \frac{n_A}{n}$$

ولی چه دلیل ریاضی دارید که حد فوق وجود دارد؟ (و به چه دلیل اگر تکرار شود، دوباره همان حد می‌شود؟)

به علاوه چون n و n_A اعدادی هستند که از آزمایش به دست می‌آیند، هرگز نمی‌توانند نامحدود باشند. پس به این وسیله نیز مشکل این تعریف حل نشده است. مگر اینکه $P(A)$ را یک مفهوم تئوریک بدانیم (تعریف اصولی) و $\frac{n_A}{n}$ فقط برای تخمین صحت آن استفاده شود (در فصل تخمین در این باره دقیقاً صحبت خواهیم کرد). به علاوه در برخی آزمایش‌های تصادفی، امکان تکرار یا تکرار زیاد آزمایش وجود ندارد.

(طرفداران این نظریه می‌گویند که وجود این حد یک فرض یا اصل است، ولی این فرضی پیچیده و دور از ذهن است. ولی ما با استفاده از اصول کولموگروف که اصول ساده‌تری هستند، می‌توانیم این موضوع را که $\frac{n_A}{n}$ به سمت $P(A)$ میل می‌کند، ثابت کنیم که این اثبات را بعداً در قانون اعداد بزرگ خواهیم دید.)

تعریف کلاسیک:

اگر در یک آزمایش تصادفی، N نتیجهٔ ممکنه (N outcomes) وجود داشته باشد و N_A تا از آنها مطلوب باشند (تعداد اعضای واقعهٔ A، N_A باشد)، آنگاه:

$$P(A)=\frac{N_A}{N}$$

مثلاً در آزمایش انداختن تاس، احتمال اینکه عدد حاصل کوچکتر از 3 باشد، $\frac{2}{6}$ است، زیرا:

$\Omega=\{1,2,3,4,5,6\}$, $A=\{1,2\}$ (از تقارن تاس، متساوی الاحتمال بودن نتایج را فرض کرده‌ایم)

البته باید به تفاوت N و n و نیز تفاوت N_A و n_A دقت شود. n تعداد آزمایش‌های انجام شده و n_A تعدادی از این آزمایش‌ها که واقعهٔ A اتفاق افتاده بود. در حالی که N تعداد نتایج ممکنه در یک آزمایش است و N_A تعدادی از این نتایج که عضو A باشند.

با این تعریف دیگر نیازی به اصول موضوعه نخواهد بود و همهٔ آنها قابل اثبات هستند. مثلاً اگر A و B ناسازگار باشند و واقعهٔ A، N_A عضو و واقعهٔ B، N_B عضو داشته باشد، $A\cup B$ دارای N_A+N_B عضو خواهد بود. لذا:

$$P(A\cup B)=\frac{N_{A\cup B}}{N}=\frac{N_A}{N}+\frac{N_B}{N}=P(A)+P(B)$$

اما تعریف کلاسیک، اگر نتایج آزمایش تصادفی متساوی الاحتمال (Equally Likely) نباشند، دیگر درست نخواهد بود.

مثلاً در آزمایش تصادفی تولد یک کودک که $\Omega=\{m,f\}$ است، وقایع $A=\{m\}$ و $B=\{f\}$ متساوی الاحتمال نیستند و لذا $P\{m\}=\frac{1}{2}$ صحیح نیست.

یا در آزمایش تصادفی رأی‌گیری دوحزبی که $\Omega=\{\alpha,\beta\}$ است، وقایع $A=\{\alpha\}$ و $B=\{\beta\}$ متساوی الاحتمال نیستند و لذا $P\{\alpha\}=\frac{1}{2}$ صحیح نیست.

به علاوه نتایج آزمایش ممکن است به گونه‌های مختلف تعبیر شود که لزوماً متساوی الاحتمال نباشند.

مثال: دو تاس را می‌اندازیم. احتمال این واقعه را می‌خواهیم که مجموع اعداد دو تاس مساوی ۷ باشد.

الف) نتایج ممکنه را می‌توانیم به صورت جمع‌های مختلف ممکنه بیان کنیم، یعنی اعداد ۲، ۳، ...، ۱۲. پس:

$$\Omega=\{2,3,\ldots,12\} \quad , \quad A=\{7\}$$

پس احتمال $\frac{1}{11}$ است!

ب) ممکن است بگوییم ۲۱ نتیجه داریم (بدون تمایز بین تاس اول و تاس دوم). پس:

$$\Omega=\{1-1,1-2,1-3,\ldots,5-6,6-6\} \quad , \quad A=\{1-6,2-5,3-4\}$$

پس احتمال $\frac{3}{21}$ است!

ج) ممکن است بگوییم ۳۶ نتیجهٔ ممکنه داریم (تمایز بین تاس اول و تاس دوم). پس:

$$\Omega=\{(i,j):i,j=1,2,\ldots,6\} \quad , \quad A=\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

پس احتمال $\frac{6}{36}=\frac{1}{6}$ است.

که فقط در حالت آخر که نتایج متساوی الاحتمال هستند، جواب درست است.

به علاوه معمولاً متساوی الاحتمال بودن را صرفاً از روی نبودن دلیلی برای ترجیح بیان می‌کنیم که در مورد سکه یا تاس درست درمی‌آید، ولی دیدیم که در مورد تولد بچه درست درنمی‌آید.

به این ترتیب ما از تعریف کلاسیک نه به عنوان تعریف، بلکه در مواردی که متساوی الاحتمال بودن معقول باشد، برای انتخاب $P(A_i)$ها از آن استفاده می‌کنیم.

اصل ناکافی بودن دلیل: اگر یک آزمایش تصادفی دارای N نتیجهٔ ممکنهٔ $\omega_1, \omega_2, \ldots, \omega_N$ باشد و ما هیچ اطلاعی در مورد نحوهٔ وقوع آنها نداشته باشیم، باید احتمال آنها را مساوی فرض کنیم، یعنی:

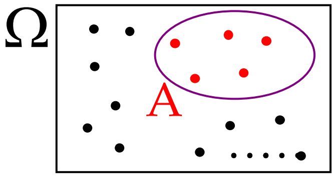
$$P\{\omega_1\}=P\{\omega_2\}=\cdots=P\{\omega_N\}=\frac{1}{N}$$

تفاوت این اصل با تعریف کلاسیک این است که در تعریف کلاسیک می‌گوییم: می‌دانیم که نتایج متساوی الاحتمال هستند. ولی اینجا می‌گوییم: چون احتمال‌ها را نمی‌دانیم و هیچ دلیلی بر برتری و محتمل بودن یکی بر دیگری نداریم، آنها را متساوی الاحتمال فرض می‌کنیم.

در آزمایش با نتایج هم احتمال $P\{\omega_1\}=P\{\omega_2\}=\cdots=P\{\omega_N\}=\frac{1}{N}$ می‌باشد. از تعریف اصولی هم داریم:

$$P(A)=\sum_{i=1}^{N_A}\frac{1}{N}=\frac{N_A}{N}$$

این در واقع تعمیم اصل ۳ است، زیرا A اجتماع N_Aتا مجموعهٔ جدااز‌هم با احتمال $\frac{1}{N}$ است.



در چنین مواردی برای معین کردن احتمال واقعه‌ها، کسب مهارت محاسبهٔ تعداد نقاطی از Ω که خاصیت معینی دارند لازم است و این کار به وسیلهٔ آنالیز ترکیبی (Combinational Analysis) صورت می‌گیرد.

مروری بر آنالیز ترکیبی:

ترتیب (جایگشت) (Permutation): تعداد نحوهٔ مرتب کردن mشیء از Nشیء

فرض کنید Nشیء (متمایز) داریم و می‌خواهیم mتا $(m \le N)$ از آنها را انتخاب کرده و در یک خط بچینیم (ترتیب مهم است).
تعداد حالات ممکن برابر است با:

$$\begin{array}{ccccccc} 1 & & 2 & & \cdots & & m \\ \text{------} & & \text{------} & & \cdots & & \text{------} \\ N & \times & (N-1) & \times & \cdots & \times & (N-m+1) \end{array}$$

$$P_m^N = N \times (N-1) \times \cdots \times (N-m+1) = \frac{N!}{(N-m)!} : 0 \le m \le N$$

نتیجه:

$$1)\ P_m^N = (N-m+1) P_{m-1}^N$$

۲) ترتیب Nشیء (نحوهٔ مرتب کردن کلیهٔ Nشیء):

$$P_N^N = N \times (N-1) \times \cdots \times 1 = N!$$ (طبق تعریف: $0! = 1$)

در ترتیب (Permutation) فرض می‌کنیم تکرار مجاز نیست، یعنی وقتی شیئی را در جایگاه ۱ گذاشتیم، دیگر همان شیء نمی‌تواند در جایگاه دیگری هم باشد. ولی اگر تکرار مجاز باشد، تعداد حالات چقدر می‌شود؟

پاسخ N^m است.

مثلاً با اعداد ۱ تا ۹ چند عدد سه رقمی می‌توان ساخت؟

$$\underline{\quad 9 \quad} \times \underline{\quad 9 \quad} \times \underline{\quad 9 \quad} \rightarrow 9^3 = 729$$

در حالی که:

$$P_3^9 = 9 \times 8 \times 7 = 504$$

این تعداد اعداد سه رقمی است که رقم تکراری نداشته باشند.

مثال: سه تاس را پرتاب می‌کنیم. تعداد حالات ممکنه چندتا است؟

$$\Omega = \{(f_1, f_1, f_1), (f_1, f_1, f_2), \ldots, (f_6, f_6, f_6)\}$$

یعنی $6^3 = 216$ حالت خواهیم داشت.

پس اگر آزمایشی با N نتیجهٔ ممکنه را mبار انجام دهیم، N^m حالت خواهیم داشت.

در حالت کلی‌تر اگر m آزمایش که تعداد نتایج ممکنهٔ هر یک $N_1, N_2, \ldots, N_m$ باشد، انجام شوند، تعداد نتایج ممکنهٔ کل m آزمایش برابر $N_1 N_2 \cdots N_m$ می‌باشد (تعداد عناصر حاصلضرب دکارتی $\Omega_1 \times \Omega_2 \times \cdots \times \Omega_m$) (اصل شمارش).

ترکیب (Combination):

اگر N شیء (متمایز) داشته باشیم و بخواهیم یک گروه mتایی از میان آنها انتخاب کنیم (ترتیب دیگر مهم نیست)، داریم:

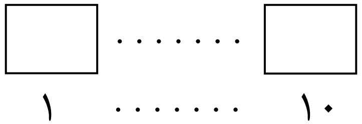
$$C_m^N = \binom{N}{m} = \frac{P_m^N}{m!} = \frac{N!}{m!(N-m)!} : 0 \le m \le N$$

با توجه به رابطه روشن است که: $C_m^N = C_{N-m}^N$، زیرا هر زمان که m شیء را از Nتا انتخاب می‌کنیم، $N-m$تا را باقی می‌گذاریم.

مثال: کلاسی ۲۰ نفر دانشجو دارد. ۳ نفر برای صحبت با رئیس دانشگاه انتخاب می‌شوند. چند گروه نماینده متصور است؟

$$\text{تعداد حالات} = \binom{20}{3} = \frac{20 \times 19 \times 18}{3!} = 1140$$

مثال: ۷ توپ سفید و ۳ توپ قرمز داریم. به چند طریق می‌توان این توپها را در ده جعبه قرار داد؟



درست است که توپهای سفید از هم متمایز نیستند و توپهای قرمز نیز از هم متمایز نیستند، ولی جعبه‌ها از هم متمایزند. در واقع می‌توانیم از میان ده جعبه، ۳تا را برای قرار گرفتن توپهای قرمز انتخاب کنیم (یا معادلاً ۷تا را برای قرار گرفتن توپهای سفید انتخاب کنیم) و ترتیب جعبه‌های انتخابی مهم نیستند.

$$\text{تعداد حالات} = \binom{10}{3} = \frac{10\times 9\times 8}{3!} = 120$$

کلاً اگر N شیء تشکیل دو گروه بدهند، یک گروه mتایی و گروه دیگر $N-m$تایی، قرار دادن اینها در N جعبه (یا به خط کردن آنها) به $\binom{N}{m}$ طریق ممکن است.

مثال: تعداد اعداد باینری Nبیتی که m رقم 1 و $N-m$ رقم 0 داشته باشند؟

مانند مثال قبلی داریم:

$$\text{تعداد حالات} = \binom{N}{m}$$

در Permutation دیدیم که اگر تکرار را مجاز می‌دانستیم، تعداد حالات به جای P_m^N می‌شود: N^m. ولی در ترکیب چه؟

مثلاً m توپ غیرمتمایز داریم که می‌خواهیم آنها را در N جعبه قرار دهیم:

- اگر در هر جعبه فقط یک توپ بتوان قرار داد؛

$$\text{تعداد حالات} = \binom{N}{m} : m \le N$$

- ولی اگر در هر جعبه به تعداد دلخواه (حتی کلیهٔ m توپ را) بتوان قرار داد، تعداد حالات چقدر خواهد بود؟ (در اینجا لازم نیست داشته باشیم: $m \le N$)

		....	

دیواره‌های میانی را هم به عنوان شیء در نظر بگیرید. گروه ۱ توپها هستند که تعداد آنها mتا است و گروه ۲ دیواره‌ها هستند که تعداد آنها $N-1$تا است. می‌خواهیم این $N+m-1$ شیء را در $N+m-1$ مکان قرار دهیم.

مثلاً اگر $N = ۳$ و $m = ۲$ باشد، گویی دو تا w و دو تا b داریم و می‌خواهیم آنها را در چهار مکان قرار دهیم:

bwbw
wbwb
bwwb
bbww
wbbw
wwbb

لذا با توجه به آنچه قبلاً (در مورد دو گروه) گفتیم، داریم:

$$\text{تعداد حالات} = \binom{N+m-1}{m}$$

پس در مثال فوق داریم:

$$\text{تعداد حالات} = \binom{3+2-1}{2} = \binom{4}{2} = 6$$

ضمناً نتیجه می‌گیریم به تعداد $\binom{N+m-1}{m}$ بردار N عنصری $(k_1,\ldots,k_N)$ با عناصر صحیح نامنفی k_i با شرط $k_1+k_2+\cdots+k_N=m$ وجود دارد.

سؤال: اگر الزام بداریم که هیچ یک از ظرفها نباید خالی بماند $(m \geq N)$، چند حالت خواهیم داشت؟

اکنون m توپ داریم. $m-1$ تا w را باید از میان $N-1$ تا فاصلهٔ بین توپها انتخاب کرد تا بین هر دو w حداقل یک b باشد. پس تعداد حالات برابر می‌شود با: $\binom{m-1}{N-1}$.

$$\overbrace{b\square b\square b\square\cdots\square b}^{\text{m-1 جای خالی}}$$
$$\uparrow\ \uparrow\ \uparrow\cdots\uparrow$$
$$\underbrace{ww\cdots w}_{\text{N-1 تا w}}$$

این معادل است با:

به تعداد $\binom{m-1}{N-1}$ بردار N عنصری $(k_1, k_2, \ldots, k_N)$ با عناصر صحیح مثبت k_i ($i = 1, 2, \ldots, N$ و $k_i > 0$) که $k_1 + k_2 + \cdots + k_N = m$ $(m \geq N)$ وجود دارد.

قضیهٔ دوجمله‌ای:

$$(x+y)^N = \sum_{k=0}^{N} \binom{N}{k} x^k y^{N-k}$$

دلیل وجود ضریب $\binom{N}{k}$

(عین مثال عدد باینری N رقمی)

یا اثبات با استقراء و با استفاده از: $1 \leq k \leq n$: $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ (کتاب راس، ص ۸ و ۹)

حالت خاص: اگر $x=y=1$ باشد، داریم:

$$2^N = \sum_{k=0}^{N} \binom{N}{k}$$

مثال: دیدیم که تعداد اعداد باینری N بیتی که m رقم 1 داشته باشند، $\binom{N}{m}$ است.

پس تعداد کل اعداد باینری Nبیتی برابر است با:

$$\sum_{m=0}^{N}\binom{N}{m}=2^{N}$$

(2^N را از راه دیگری هم می‌توانید به دست آورید.)

مثال: تعداد زیرمجموعه‌های یک مجموعۀ N عضوی:

$$\binom{N}{0}+\binom{N}{1}+\binom{N}{2}+\cdots+\binom{N}{N}=\sum_{k=0}^{N}\binom{N}{k}=2^{N}$$

تعداد زیرمجموعه‌هایی که عنصر خاصی از Ω را شامل باشند (تعداد واقعه‌هایی که نتیجۀ خاصی را شامل باشند) 2^{N-1} است. یعنی در یک آزمایش یک نتیجه اتفاق می‌افتد، ولی 2^{N-1} واقعه اتفاق می‌افتند و 2^{N-1} واقعه اتفاق نمی‌افتند.

ترکیب تعمیم‌یافته:

در ترکیب از N شیء mتا را انتخاب می‌کردیم. در واقع N شیء را به دو گروه mتایی و $N-m$تایی تقسیم می‌کردیم و $C_m^N = \dfrac{N!}{m!(N-m)!}$ عبارت بود از تعداد حالات ممکنه برای تقسیم‌بندی به دو گروه. حال اگر N شیء داشته باشیم و بخواهیم آنها را به r گروه A_1، ... و A_r که به ترتیب m_1، ... و m_r عضو داشته باشند $(m_1+\cdots+m_r=N)$ تقسیم کنیم، تعداد حالات ممکنه برابر است با:

$$\binom{N}{m_1,m_2,\ldots,m_r} = C^N_{m_1,m_2,\ldots,m_r} = \frac{N!}{m_1!m_2!\ldots m_r!}$$

ایدهٔ اثبات در کتاب:

$$\binom{N}{m_1}\binom{N-m_1}{m_2}\cdots\binom{N-m_1-m_2-\cdots-m_{r-1}}{m_r} = \frac{N!}{m_1!m_2!\ldots m_r!}$$

در واقع برای $r = ۲$ همان فرمول ترکیب قبلی را خواهیم داشت. اثبات دیگر با نگاه به صورت و مخرج کسر مشخص است. همان‌طور که ترکیب را می‌توانستیم برای وقتی که اشیاء دو گروه را می‌خواهیم بچینیم استفاده کنیم، در اینجا نیز وقتی N شیء ما r گروه باشند قابل استفاده است.

مثال: ۱۰ توپ داریم. ۲تا قرمز، ۳تا سفید و ۵تا سیاه. به چند طریق می‌توان آنها را در یک خط چید (یا در ۱۰ جعبه قرار داد)؟

$$\binom{10}{2,3,5}=\frac{10!}{2!3!5!}$$

قضیهٔ چندجمله‌ای:

$$(x_1+x_2+\cdots+x_r)^N=\sum_{\substack{k_1,k_2,\ldots,k_r=0\\ k_1+k_2+\cdots+k_r=N}}^{N}\binom{N}{k_1,k_2,\ldots,k_r}x_1^{k_1}x_2^{k_2}\cdots x_r^{k_r}$$

مثال: کلاسی ۲۰ نفر دانشجو دارد. باید ۳ نفر برای روابط عمومی، ۷ نفر برای همکاری با امور دانشجویی (و ۱۰ نفر برای همکاری با امور آموزشی) انتخاب شوند. تعداد حالات ممکنه برابر است با:

$$\binom{20}{3,7,10}=\frac{20!}{10!7!3!}$$

تا اینجا دیدیم که چگونه تعداد نقاط را حساب کنیم. حال در مثالهایی احتمال را حساب می‌کنیم (آزمایشهای با نتایج هم‌احتمال).

مثالهایی از تعیین احتمال واقعه‌ها در آزمایشهای با نتایج هم‌احتمال

مثال ۱: اگر یک سکه را ۱۰ بار پرتاب کنیم، احتمال اینکه حداکثر ۳ بار شیر بیاید چقدر است؟

مجموعهٔ Ω، مجموعهٔ ۱۰تایی‌های مرتبی است که هر عضو این ۱۰تایی‌های مرتب می‌تواند H یا T باشد.

$$\Omega=\{\mathrm{HHH\cdots H},\ldots,\mathrm{TTT\cdots T}\}$$

Ω دارای ۲^۱۰ عضو است. فرض می‌کنیم این ۲^۱۰ عضو متساوی الاحتمال باشند (سکه سالم باشد و پرتابها مستقل باشند).

کل حالات ممکنه = ۲^۱۰

واقعهٔ A: حداکثر ۳ بار شیر بیاید.

این واقعه اجتماع چهار واقعهٔ غیرمتلاقی است:

واقعهٔ $\mathrm{A}_۳$: دقیقاً ۳ بار شیر بیاید. واقعهٔ $\mathrm{A}_۲$: دقیقاً ۲ بار شیر بیاید. واقعهٔ $\mathrm{A}_۱$: دقیقاً ۱ بار شیر بیاید. واقعهٔ A_0: اصلاً شیر نیاید.

تعداد عناصر $\mathrm{A}_۳$ = $\binom{10}{3}$ و به همین ترتیب تعداد عناصر $\mathrm{A}_۲$، $\mathrm{A}_۱$ و A_0 نیز به دست می‌آید:

$$P(A)=P(A_0)+P(A_1)+P(A_2)+P(A_3)=\frac{\binom{10}{0}+\binom{10}{1}+\binom{10}{2}+\binom{10}{3}}{2^{10}}=\frac{176}{2^{10}}=0.1719$$

مثال ۲: به طور کاملاً تصادفی m توپ را در n جعبه قرار می‌دهیم $(n \geq m)$. (احتمال قرار گرفتن هر یک از توپها در جعبه‌های مختلف مساوی است و حتماً هم در یکی از جعبه‌ها قرار می‌گیرد.)

احتمال اینکه این m توپ در m جعبهٔ مورد نظر (یکی در هر جعبه) قرار بگیرند چیست؟

الف) اگر هر جعبه گنجایش فقط یک توپ را داشته باشد؛

فرقی نمی‌کند چه توپها را متمایز بگیریم و چه غیرمتمایز (فلسفتاً چه فرقی می‌کند؟)

۱. توپها غیرمتمایز: کل حالات ممکنهٔ قرار گرفتن m توپ در n جعبه: $\binom{n}{m}$ متساوی الاحتمال

حالت مورد نظر m توپ در m جعبهٔ خاص: ۱

$$P = \frac{1}{\binom{n}{m}} = \frac{m!(n-m)!}{n!}$$

۲. توپها متمایز: حالات ممکنه: P_m^n متساوی الاحتمال (ترتیب: حالات ممکن برای اینکه از n شیء متمایز mتا را انتخاب کنیم با مهم بودن ترتیب)

حالات مورد نظر: $m!$

$$P = \frac{m!}{P_m^n} = \frac{m!(n-m)!}{n!}$$

ب) اگر گنجایش جعبه محدود نباشد (در هر جعبه هر تعداد توپ ممکن است قرار گیرد)؛

باز هم فرقی نمی‌کند توپها را متمایز بگیریم یا غیرمتمایز؛ ولی دقت کنید که چه چیزی را متساوی الاحتمال می‌گیرید.

۱. توپها متمایز: حالات ممکنه: n^m متساوی الاحتمال (برای هر توپ n حالت داریم)

حالات مورد نظر: $m!$

$$P = \frac{m!}{n^m}$$

۲. توپها غیرمتمایز: حالات ممکنه (ترتیب مهم نیست، تکرار مجاز است): $\binom{n+m-1}{m}$

حالات مورد نظر: ۱

$$P = \frac{1}{\binom{n+m-1}{m}}$$

این جواب با مفروضات مسأله (که احتمال قرار گرفتن هر توپ در جعبه‌های مختلف مساوی است) غلط است، زیرا $\binom{n+m-1}{m}$ حالت متساوی الاحتمال نیستند. مثلاً اگر $n = ۲$ و $m = ۲$، گر چه سه حالت داریم: 11 و 02 و 20؛ ولی این سه حالت متساوی الاحتمال نیستند و احتمال وقوع 11، $\frac{2}{4}$ است و نه $\frac{1}{3}$.

مگر اینکه صورت مسأله طوری باشد که این حالات متساوی الاحتمال فرض شده باشند.

(پس از متمایز گرفتن هیچ وقت ضرر نمی‌کنید.)

مثال ۳: جعبه‌ای شامل ۶۰ توپ قرمز و ۴۰ توپ سیاه است. ۲۰ توپ از این جعبه به طور کاملاً تصادفی انتخاب می‌کنیم (بدون جایگزینی). احتمال اینکه ۱۵تا از این توپها قرمز و ۵تا سیاه باشند چیست؟

نتایج آزمایش تصادفی: کلیهٔ روشهای ممکن در انتخاب ۲۰ توپ

تعداد حالات ممکنه (تعداد نقاط Ω): $\binom{100}{20}$ متساوی الاحتمال

تعداد حالات مطلوب: $\binom{60}{15}\binom{40}{5}$

$$P = \frac{\binom{60}{15}\binom{40}{5}}{\binom{100}{20}} = 0.065$$

متمایز یا غیرمتمایز گرفتن توپها فرقی ایجاد نمی‌کند. Ω را طوری تعریف کنید که نقاط متساوی الاحتمال باشند.

(راه غلط: تعداد حالات ممکنه: ۲۱ حالت، تعداد حالات مطلوب: ۱ حالت. ولی این ۲۱ حالت متساوی الاحتمال نیستند.)

اگر توپها را متمایز بگیریم، خواهیم داشت:

تعداد حالات ممکنه: $P_{20}^{100} = \dfrac{100!}{80!}$

تعداد حالات مطلوب: $\dfrac{20!}{15!5!}\dfrac{40!}{35!}\dfrac{60!}{45!} = \binom{20}{5} P_5^{40} P_{15}^{60} = 20!\binom{60}{15}\binom{40}{5}$

که حاصل تقسیم همان عدد قبلی خواهد بود.

برای محاسبهٔ تقریبی فاکتوریل‌های بزرگ می‌توانید از فرمول استرلینگ (Stirling's Formula) استفاده کنید:

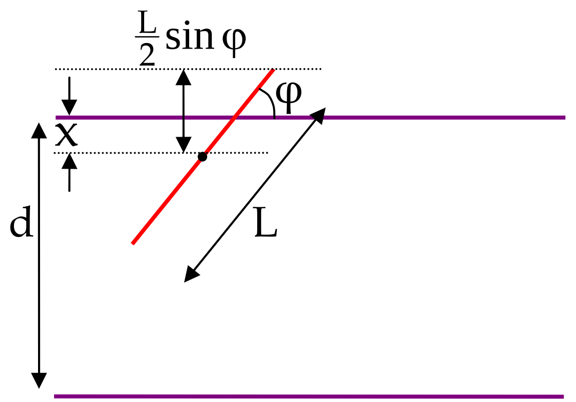
برای n بزرگ : $n! \simeq n^n e^{-n}\sqrt{2\pi n}$

(برای ۱۰ = n خطا کمتر از ۱٪ است)

سرانجام مثالی از حالت Ω با نقاط غیرقابل شمارش می‌آوریم.

مثال ۴: مسألهٔ سوزن Buffon (1777)

سوزنی به طول L را روی کاغذی خط‌کشی شده که فاصلهٔ خطوط آن $d > L$ است می‌اندازیم. احتمال اینکه سوزن خطی را قطع کند چیست؟ با فرض اینکه کلیهٔ مکانهای مرکز سوزن و کلیهٔ جهتهای قرار گرفتن سوزن متساوی الاحتمال باشند.



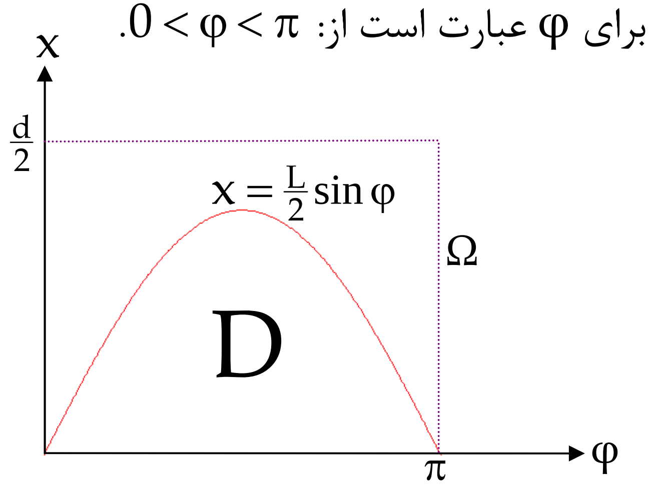
فرض کنید که فاصلهٔ مرکز سوزن از نزدیکترین خط x باشد و φ زاویه با این خط باشد.

در صورت تقاطع خواهیم داشت: $x<\frac{L}{2}\sin\varphi$؛

که مقادیر ممکنه (و متساوی الاحتمال) برای x عبارت است از: $0<x<\frac{d}{2}$؛

(برای $x>\frac{d}{2}$ خط دیگر نزدیکتر خواهد بود و مسأله عیناً تکرار می‌شود) و مقادیر ممکنه برای φ عبارت است از: $0<\varphi<\pi$.

یعنی احتمال را باید روی این سطح حساب کنیم:



$$P=\iint_D \alpha(x,\varphi)\,dx\,d\varphi \quad \left(=\frac{\text{تعداد نقاط مطلوب}}{\text{تعداد کل نقاط}}\right)$$

$$\alpha(x,\varphi)=\frac{1}{\pi\frac{d}{2}}$$

$$\Rightarrow P=\iint_D \frac{1}{\pi\frac{d}{2}}\,dx\,d\varphi=\frac{1}{\pi}\frac{1}{\frac{d}{2}}\int_0^{\pi}\frac{L}{2}\sin\varphi\,d\varphi=\frac{2L}{\pi d}$$

جالب است که برای L و d مشخص به این ترتیب می‌توان به روش آماری π را تخمین زد! (روش مونت کارلو)

احتمال شرطی:

اگر M واقعه‌ای در Ω باشد که $P(M) \neq 0$، داریم:

$$P(A|M) \triangleq \frac{P(A \cap M)}{P(M)}$$

تعبیر تجربی (فراوانی نسبی): اگر آزمایش تصادفی nبار انجام شود، داریم:

$$\frac{P(A \cap M)}{P(M)} = \frac{\frac{n_{A \cap M}}{n}}{\frac{n_M}{n}} = \frac{n_{A \cap M}}{n_M}$$

به این معنی که در n_Mباری که واقعهٔ M اتفاق افتاده، فراوانی نسبی (میزان تکرار نسبی) وقوع A چقدر است. یعنی اگر کلیهٔ آزمایشهایی که M در آنها اتفاق نیفتاده را دور بریزیم، فراوانی نسبی A اینقدر بوده است.

$P(A|M)$ را به صورت نسبت دو احتمال تعریف کردیم. اما از کجا معلوم که این هم یک احتمال باشد؟ (اگر چه از نظر تعبیر فرکانس نسبی احتمال است، اما باید توسط اصولمان این امر را نشان دهیم) برای این منظور باید نشان دهیم که در اصول سه‌گانه صدق می‌کند.

۱)

$$P(A|M)=\frac{P(A\cap M)}{P(M)}\geq 0$$

۲)

$$P(\Omega|M)=\frac{P(\Omega\cap M)}{P(M)}=\frac{P(M)}{P(M)}=1$$

۳) اگر $A\cap B=\varnothing$، داریم:

$$P(A\cup B|M)=\frac{P[(A\cup B)\cap M]}{P(M)}=\frac{P[(A\cap M)\cup(B\cap M)]}{P(M)}=\frac{P(A\cap M)}{P(M)}+\frac{P(B\cap M)}{P(M)}$$
$$=P(A|M)+P(B|M)$$

لذا احتمالهای شرطی همهٔ خصوصیات یک احتمال معمولی را دارا می‌باشند و در تمام قضایایی که اثبات شد صدق می‌کنند.

مثلاً داریم: $P(A^c|M)=1-P(A|M)$

اگر $P(A) \neq 0$ باشد، از تعریف احتمال شرطی داریم:

$$P(AB) = P(A)P(B|A)$$

حال با تعمیم آن، به شرط اینکه $P(AB) \neq 0$ باشد، داریم:

$$P(ABC) = P(A)P(B|A)P(C|AB)$$

به همین ترتیب خواهیم داشت:

قاعدهٔ زنجیری: اگر $P(A_1A_2\cdots A_{n-1}) \neq 0$ باشد، آنگاه:

$$P(A_1A_2\cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$$

مثال ۱: در آزمایش انداختن تاس اگر بدانیم که تاس زوج آمده است، احتمال اینکه کوچکتر از ۴ باشد چیست؟

A = { کوچکتر از ۴ } = {۱, ۲, ۳}

B = { زوج } = {۲, ۴, ۶}

$\Rightarrow A \bigcap B = \{۲\}$

$\Rightarrow P(A\,|\,B) = \dfrac{P(A \bigcap B)}{P(B)} = \dfrac{\frac{1}{6}}{\frac{3}{6}} = \dfrac{1}{3} \rightarrow$ فضای نمونه کاهش یافته

در حالی که $P(A) = \dfrac{3}{6} = \dfrac{1}{2}$، یعنی $P(A\,|\,B)$ کمتر از $P(A)$ است (واقعهٔ B اطلاعات منفی در مورد وقوع A دارد).

مثال ۲: احتمال اینکه خرابی دیود پس از لحظهٔ t اتفاق بیفتد $e^{-\alpha t^2}$ است. احتمال اینکه بین لحظات t_1 و t_2 خراب شود چه خواهد بود اگر بدانیم که تا لحظهٔ T که $T < t_1 < t_2$ درست کار می‌کرده است؟

آزمایش تصادفی: خراب شدن دیود

$$\Omega = \{t : 0 \le t < \infty\} \rightarrow \quad t : \text{زمان خرابی}$$

$$P(\{\tau \le t < \infty\}) = e^{-\alpha\tau^2}$$

$$A = \{t_1 < t < t_2\} = \{t_1 < t < \infty\} - \{t_2 \le t < \infty\}$$

$$M = \{T \le t < \infty\}$$

$$P(A \mid M) = \frac{P(A \cap M)}{P(M)} = \frac{P(A)}{P(M)} = \frac{e^{-\alpha t_1^2} - e^{-\alpha t_2^2}}{e^{-\alpha T^2}}$$

یعنی $P(A \mid M)$ بزرگتر از $P(A)$ است (واقعهٔ M اطلاعات مثبت در مورد وقوع واقعهٔ A دارد).

مثال ۳: مؤسسه‌ای که در آن آقای x کار می‌کند یک مهمانی شام برای کارمندانی که حداقل یک پسر داشته باشند ترتیب داده است. اگر بدانیم که آقای x دو فرزند دارد و به مهمانی دعوت شده است، احتمال اینکه هر دو فرزند او پسر باشند چقدر است؟

$$\Omega = \{(b,b),(b,g),(g,b),(g,g)\}$$

فرزند اول   فرزند دوم

E : هر دو فرزند پسر

F : حداقل یک فرزند پسر

$$P(E|F) = \frac{P(EF)}{P(F)} = \frac{P\{(b,b)\}}{P\{(b,b),(b,g),(g,b)\}} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{3}$$

نه $\frac{1}{2}$ (فرزند دیگر یا پسر است یا دختر). در حالی که سه پیشامد هم‌شانس داریم: (g,b) , (b,g) , (b,b)

مثال ۴: ظرفی دارای ۸ توپ قرمز و ۴ توپ سفید است. دو توپ (بدون جایگذاری) انتخاب می‌کنیم. اگر انتخاب هر یک از توپها هم‌شانس باشد احتمال اینکه هر دو توپ انتخابی قرمز باشند چیست؟

قبلاً به دست آورده بودیم: $\dfrac{\binom{8}{2}}{\binom{12}{2}}$ یا $\dfrac{P_2^8 + P_2^8}{P_2^{12}}$؛

نقاط Ω را توپهای داخل ظرف می‌گیریم.

R_1 : قرمز بودن توپ اول $\qquad \Rightarrow P(R_1) = \dfrac{8}{12}$

R_2 : قرمز بودن توپ دوم $\qquad \Rightarrow P(R_2 \mid R_1) = \dfrac{7}{11}$

$$\Rightarrow P(R_1 R_2) = \frac{8}{12} \times \frac{7}{11} = 0.424$$

احتمال شرطی را به عنوان ابزاری برای محاسبهٔ احتمال پیشامدها می‌توان استفاده کرد.

قضیهٔ احتمال کل (Total Probability Theorem):

اگر $i = 1, 2, \ldots, m$ و B_iها افرازی از Ω باشند، برای هر واقعهٔ دلخواه A از Ω داریم:

$$P(A) = \sum_{i=1}^{m} P(A \mid B_i) P(B_i)$$

اثبات:

$$P(A) = P(A \cap \Omega) = P(A \cap (\bigcup_{i=1}^{m} B_i)) = P[\bigcup_{i=1}^{m} (A \cap B_i)] = \sum_{i=1}^{m} P(A \cap B_i)$$

$$= \sum_{i=1}^{m} P(A \mid B_i) P(B_i)$$

از این قضیه برای محاسبهٔ احتمال یک واقعه ($P(A)$) که احتمال‌های مشروط آن ($P(A \mid B_i)$) داده شده استفاده می‌کنیم.
یعنی $P(A)$ متوسط وزن‌داده‌شده‌های از $P(A \mid B_i)$ها است که مقدار وزن $P(B_i)$ است.

مثال: دو جعبه داریم. جعبهٔ اول شامل ۲تا ترانزیستور خراب و ۸تا سالم است. جعبهٔ دوم شامل ۹تا ترانزیستور خراب و ۶تا سالم است. به طور تصادفی یکی از جعبه‌ها را انتخاب می‌کنیم و از جعبهٔ انتخاب شده به طور تصادفی یک ترانزیستور برمی‌داریم. احتمال اینکه ترانزیستور انتخاب شده خراب باشد چقدر است؟

نقاط Ω عبارت است از ترانزیستورهای موجود در دو جعبه

واقعهٔ B_1 عبارت است از ۱۰ ترانزیستور جعبهٔ اول — واقعهٔ B_2 عبارت است از ۱۵ ترانزیستور جعبهٔ دوم

واقعهٔ خراب (D) عبارت است از ۱۱ ترانزیستور خراب — واقعهٔ سالم شامل ۱۴ ترانزیستور سالم است

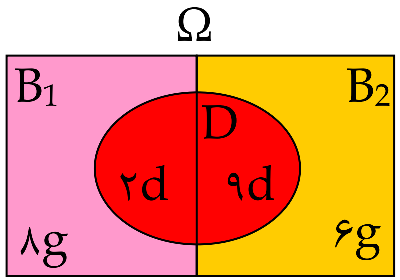
طبق مفروضات مسأله داریم:

$$P(B_1) = P(B_2) = \frac{1}{2}$$

$$P(D \mid B_1) = \frac{2}{10}, P(D \mid B_2) = \frac{9}{15}$$

باید $P(D)$ را به دست آوریم:

$$P(D) = P(B_1)P(D \mid B_1) + P(B_2)P(D \mid B_2) = \frac{1}{2} \times \frac{2}{10} + \frac{1}{2} \times \frac{9}{15} = 0.4$$



در برخی موارد $P(A \mid B_i)$ها داده شده و ما طالب $P(B_i \mid A)$ها هستیم. در این صورت از **قضیهٔ بِیز** کمک می‌گیریم.

قضیهٔ بیز (Bayes Theorem):

اگر $i = 1, 2, \ldots, m$ و B_iها افرازی از Ω باشند، برای هر واقعهٔ دلخواه A از Ω داریم:

$$P(B_k \mid A) = \frac{P(A \mid B_k)P(B_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{m} P(A \mid B_i)P(B_i)}$$

اثبات:

$$P(B_k \mid A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{P(A)} = \frac{P(A \mid B_k)P(B_k)}{\sum_{i=1}^{m} P(A \mid B_i)P(B_i)}$$

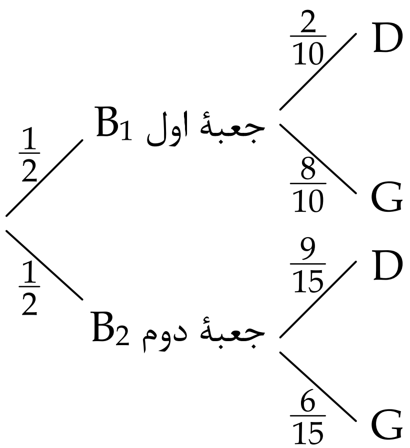
قضیهٔ بیز کاربرد بسیار زیادی در آمار، تخمین و آشکارسازی دارد و در تحلیل نحوه یا علت وقوع واقعه‌های انجام شده از آن استفاده می‌شود.

مثال ۱: در مثال پیش اگر ترانزیستور انتخاب شده خراب باشد احتمال اینکه متعلق به جعبۀ دوم باشد چقدر است؟

$P(D)$ را قبلاً حساب کردیم. پس داریم:

$$P(B_2 \mid D) = \frac{P(D \mid B_2)P(B_2)}{P(D)} = \frac{\frac{9}{15} \times \frac{1}{2}}{0.4} = \frac{3}{4}$$

درخت احتمال (دیاگرام درختی):



$$P(B_2 \mid D) = \frac{P(D \mid B_2)P(B_2)}{P(D)} = \frac{\frac{9}{15} \times \frac{1}{2}}{\frac{9}{15} \times \frac{1}{2} + \frac{2}{10} \times \frac{1}{2}} = \frac{3}{4}$$

اصطلاحاً $P(B_2)$ را احتمال پیشین (A Priory Probability) جعبۀ دوم و $P(B_2 \mid D)$ را احتمال پسین (A Posteriory Probability) آن گویند.

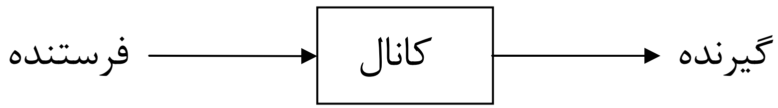
ضمناً توجه کنید که: $P(B_1 | D) = \frac{1}{4}$ (یعنی $1 - \frac{3}{4}$).

اصولاً اگر $i = 1, 2, \ldots, m$ و B_iها افرازی از Ω باشند، داریم: $\sum_{i=1}^{m} P(B_i | A) = 1$.

توجه کنید که اگر جای شرط و مشروط را عوض کنید این رابطه دیگر صادق نیست.

چه در قضیهٔ احتمال کل و چه در قضیهٔ بیز اگر B_iها افرازی از واقعهٔ A یا واقعهٔ شامل واقعهٔ A باشند نیز قضایا صادق خواهند بود.

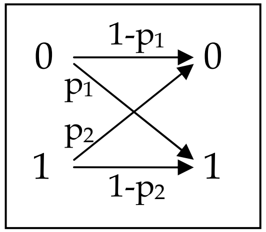
مثال۲: کانال مخابراتی دیجیتال



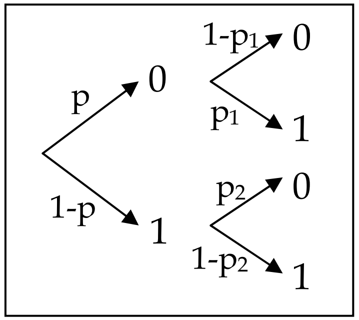
پیغام به صورت باینری کد شده و ارسال می‌شود و احتمال ارسال صفر p (و احتمال ارسال یک $1-p$ است).

ولی به دلیل وجود نویز بعضی 0ها به 1 و بعضی 1ها به 0 تبدیل می‌شوند.

احتمال خطای نوع اول p_1 و احتمال خطای نوع دوم p_2 است.



دیاگرام احتمال گذر (Transition Probability Diagram):



دیاگرام درختی:

نقاط فضای نمونه ترکیب ارسال و دریافتهای مختلف ((T,R)) است. ولی این نقاط هم‌احتمال نیستند:

$$\Omega = \{(0,0),(0,1),(1,0),(1,1)\}$$

اگر واقعهٔ T_0، واقعهٔ ارسال 0 باشد، داریم: $T_0 = \{(0,0),(0,1)\}$ و $P(T_0) = p$؛

اگر واقعهٔ T_1، واقعهٔ ارسال 1 باشد، داریم: $T_1 = \{(1,0),(1,1)\}$ و $P(T_1) = 1-p$؛

اگر واقعهٔ R_0، واقعهٔ دریافت 0 باشد، داریم: $R_0 = \{(0,0),(1,0)\}$؛

اگر واقعهٔ R_1، واقعهٔ دریافت 1 باشد، داریم: $R_1 = \{(0,1),(1,1)\}$؛

و می‌دانیم که:

$$P(R_1|T_0)=p_1 \quad , \quad P(R_0|T_1)=p_2$$

۱. احتمال خطا، یعنی $P(E)$ چقدر است؟

۲. اگر صفر دریافت شده باشد احتمال اینکه واقعاً صفر ارسال شده باشد، یعنی $P(T_0 | R_0)$ چیست؟

۳. اگر یک دریافت شده باشد احتمال اینکه واقعاً یک ارسال شده باشد، یعنی $P(T_1 | R_1)$ چیست؟

$$\begin{aligned} P_{error} &= P(E) = P(E | T_0)P(T_0) + P(E | T_1)P(T_1) \\ &= P(R_1 | T_0)P(T_0) + P(R_0 | T_1)P(T_1) = p \times p_1 + (1-p) \times p_2 \end{aligned}$$

$$P(T_0 | R_0) = \frac{P(T_0)P(R_0 | T_0)}{P(T_0)P(R_0 | T_0) + P(T_1)P(R_0 | T_1)} = \frac{p \times (1-p_1)}{p \times (1-p_1) + (1-p) \times p_2}$$

$$P(T_1 | R_1) = \frac{P(T_1)P(R_1 | T_1)}{P(T_1)P(R_1 | T_1) + P(T_0)P(R_1 | T_0)} = \frac{(1-p)\times(1-p_2)}{(1-p)\times(1-p_2) + p\times p_1}$$

مثلاً اگر $p_1 = p_2 = p_e = 0.05$ و $p = 0.9$ باشد، خواهیم داشت:

$$P_{error} = p\times p_e + (1-p)\times p_e = p_e = 0.05$$

$$P(T_1 | R_1) = \frac{0.1\times 0.95}{0.1\times 0.95 + 0.9\times 0.05} = 0.68$$

ولی اگر $p = 0.6$ باشد، آنگاه خواهیم داشت:

$$P(T_1 | R_1) = \frac{0.4\times 0.95}{0.4\times 0.95 + 0.6\times 0.05} = 0.926$$

در یک کانال خوب p_e باید حدود 10^{-6} باشد.

اصولاً اثر مخابره آن است که احتمال پیشین پیغام را به احتمال پسین آن تغییر دهد، مثلاً با دریافت 0، $P(T_0)$ به $P(T_0 | R_0)$ تغییر می‌یابد.

مثال ۳: سه سکه داریم. سکهٔ C_1 هر دو طرفش شیر، سکهٔ C_2 هر دو طرفش خط و سکهٔ C_3 یک طرفش شیر و طرف دیگرش خط است. یکی از این سه را به تصادف انتخاب کرده و پرتاب می‌کنیم. اگر طرف هویدای این سکه شیر باشد، احتمال اینکه طرف دیگرش خط باشد چیست؟

$$P(C_3 \mid H) = \frac{P(H \mid C_3)P(C_3)}{P(H \mid C_1)P(C_1) + P(H \mid C_2)P(C_2) + P(H \mid C_3)P(C_3)}$$

$$= \frac{\frac{1}{2} \times \frac{1}{3}}{1 \times \frac{1}{3} + 0 \times \frac{1}{3} + \frac{1}{2} \times \frac{1}{3}} = \frac{1}{3}$$

وقایع مستقل (Independent Events):

دو واقعهٔ A و B را مستقل گویند، هرگاه:

$$P(A\cap B)=P(A)P(B)$$

از طرفی (اگر $P(A)$ و $P(B)$ صفر نباشند) داریم: $P(A\cap B)=P(A)P(B|A)=P(B)P(A|B)$. پس اگر A و B مستقل باشند، داریم:

$$P(A|B)=P(A) \quad , \quad P(B|A)=P(B)$$

تعبیر تجربی: $P(A|B)\simeq\frac{n_{A\cap B}}{n_B}\simeq\frac{n_A}{n}\simeq P(A)$؛ یعنی فرکانس نسبی وقوع A در کل n آزمایش با فرکانس نسبی آن در n_B آزمایش که B در آنها رخ داده یکسان است (موافق با شهود ما از مفهوم استقلال).

قضیه: اگر A و B مستقل باشند، $\bar{A}$ و B نیز مستقلند.

اثبات: صفحهٔ ۵۵ کتاب.

از اعمال مجدد همین قضیه داریم:

قضیه: اگر A و B مستقل باشند، $\bar{A}$ و $\bar{B}$ نیز مستقلند.

اما اگر A از B و A از C مستقل باشد، A از BC یا ترکیبهای دیگر B و C مستقل نخواهد بود (راس، صفحهٔ ۸۲). لذا تعریف استقلال سه واقعه باید فراتر از استقلال دو به دوی آنها باشد.

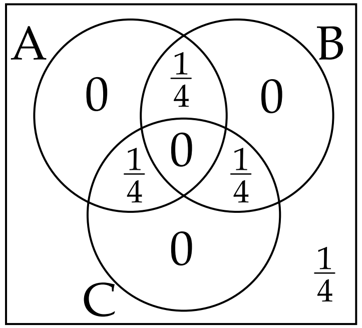
سه واقعهٔ A و B و C را مستقل گویند، هرگاه:

$$P(AB)=P(A)P(B) \quad P(AC)=P(A)P(C) \quad P(BC)=P(B)P(C) \quad P(ABC)=P(A)P(B)P(C)$$

یعنی اگر سه واقعه مستقل باشند، دو به دو نیز مستقلند؛ ولی عکس آن لزوماً صحیح نیست.

در تمرین نشان خواهید داد که اگر A و B و C مستقل باشند، A از $\overline{BC}$ و $B+C$ (و اصولاً از هر ترکیب آنها) مستقل است.

اصولاً هیچ یک از این چهار رابطه از سه رابطهٔ دیگر نتیجه نمی‌شود:



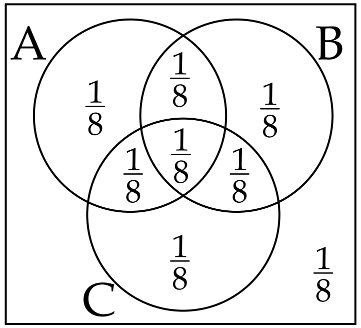
☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(B)$

☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(C)$

☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(B)P(C)$

☒ $0\neq\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=P(A)P(B)P(C)$

دو به دو مستقل



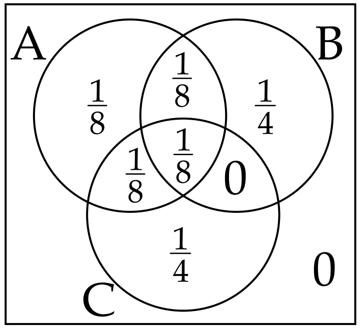
☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(B)$

☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(C)$

☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(B)P(C)$

☑ $\frac{1}{8}=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=P(A)P(B)P(C)$

مستقل



☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(B)$

☑ $\frac{1}{4}=\frac{1}{2}\times\frac{1}{2}=P(A)P(C)$

☒ $\frac{1}{8}\neq\frac{1}{2}\times\frac{1}{2}=P(B)P(C)$

☑ $\frac{1}{8}=\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=P(A)P(B)P(C)$

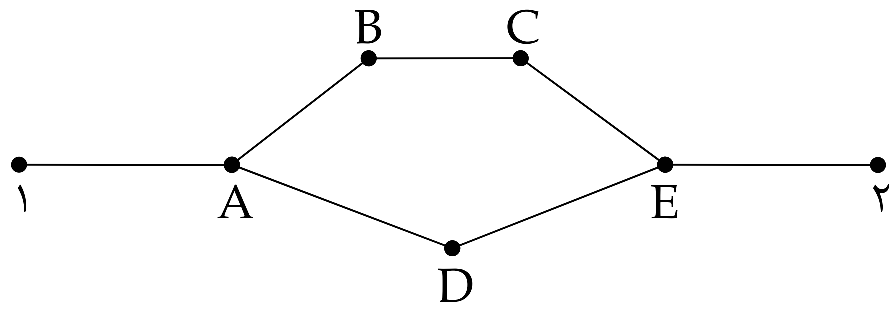
A و B مستقل، A و C مستقل

n واقعهٔ A_1، A_2، ... و A_n را مستقل گویند، هرگاه برای هر دسته اعداد صحیح $k_1,k_2,\ldots,k_r$ (که $r\leq n$) داشته باشیم:

$$P(A_{k_1}A_{k_2}\cdots A_{k_r})=P(A_{k_1})P(A_{k_2})\cdots P(A_{k_r})$$

یعنی $2^n-(n+1)$ رابطه باید صادق باشند.

مثال: یک لینک مخابراتی بین نقاط ۱ و ۲ از ایستگاه‌های A و B و C و D و E استفاده می‌کند (از D وقتی استفاده می‌شود که B یا C از کار بیفتد). تمام ایستگاه‌های رله مثل همند و بررسی‌های آماری نشان داده است که احتمال سالم بودن آنها در طی زمان T برابر با p است. احتمال اینکه لینک بین ۱ و ۲ در این زمان برقرار باشد چقدر است (با فرض استقلال خرابی رله‌ها)؟



احتمال اینکه خط بالایی مسیر موازی کار کند p^2 است (بنا بر استقلال خرابی دو ایستگاه):

$$P(B \cap C) = P(B)P(C) = p^2$$

اگر چه این نتیجه صحیح است، ولی مفهوم آن بر مبنای اصول احتمال چیست؟ اصلاً $B \cap C$ یعنی چه؟

اگر آزمایشهای تصادفی خراب شدن ایستگاههای مختلف را جدا در نظر بگیریم اشتراک بین آنها مفهوم ندارد.

در اینجا Ω حاصلضرب دکارتی فضای نمونهٔ آزمایشهای تصادفی خراب شدن ایستگاهها است و 2^5 نقطه دارد (غیرمتساوی الاحتمال):

$$\Omega = \{(F,F,F,F,F),(F,F,F,F,T),\ldots,(T,T,T,T,T)\}$$

$$\begin{matrix} \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ A & B & C & D & E \end{matrix}$$

خراب بودن A عبارت است از کلیهٔ ۵تایی‌های مرتب که عنصر اول آنها F باشد.

$$
\begin{aligned}
\text{احتمال برقرار بودن لینک} &= P(A(BC+D)E) \\
&= P(A)P(BC+D)P(E) \\
&= P(A)[P(BC)+P(D)-P(BCD)]P(E) \\
&= P(A)[P(B)P(C)+P(D)-P(B)P(C)P(D)]P(E) \\
&= p(p^2+p-p^3)p = p^3(1+p-p^2)
\end{aligned}
$$

مثلاً اگر $p = 0.9$ باشد، احتمال فوق 0.795 خواهد بود.

فصل ۳: آزمایشهای تکراری

Chapter 3

۱. **تعبیر مفهومی آزمایشهای تکراری**

۲. **آزمایش برنولی**

۳. **قضیهٔ دموآور-لاپلاس**

۴. **قضیهٔ پواسن**

۵. **نقاط پواسن**

وقتی یک آزمایش تصادفی را تحت شرایط یکسانی تکرار می‌کنیم، دو تعبیر موجود است.

یکی تعبیر تجربی که در آن احتمال واقعهٔ A در فضای Ω حدود $\frac{n_A}{n}$ است.

دیگری تعبیر مفهومی است. در واقع با تکرار آزمایش به جای فضای Ω قبلی یک فضای جدید Ω_n داریم که: $\Omega_n = \underbrace{\Omega\times\cdots\times\Omega}_{n \text{ بار}}$.

نقاط این فضای نمونهٔ جدید nتایی‌های مرتبی هستند که هر عنصر آن عضوی از Ω قبلی است.

مثال: پنج بار سکه‌ای را پرتاب می‌کنیم. در این سکه داریم:

$$P\{H\} = p$$

$$P\{T\} = q = 1 - p$$

الف) احتمال واقعهٔ B که در آن فقط دو بار اول شیر بیاید (و بقیه خط) چیست؟

ب) احتمال واقعهٔ D که در آن دو بار شیر بیاید (با هر ترتیبی برای آمدن دو شیر و سه خط) چیست؟

Ω مجموعهٔ کلیهٔ پنج‌تایی‌های مرتبی است که هر عضو این پنج‌تایی‌های مرتب H یا T است:

$$\Omega = \{(HHHHH),\ldots,(TTTTT)\}$$

یعنی $۲^۵$ عنصر دارد که لزوماً متساوی الاحتمال نیستند (چون p لزوماً $\frac{1}{2}$ نیست).

$$B = \{HHTTT\}$$

$$B=\{\text{بار اول شیر بیاید}\}\cap\{\text{بار دوم شیر بیاید}\}\cap\{\text{بار سوم خط بیاید}\}\cap\{\text{بار چهارم خط بیاید}\}\cap\{\text{بار پنجم خط بیاید}\}$$

کلیهٔ پنج‌تایی‌های مرتبی که عنصر اول آنها H است

با توجه به فرض استقلال آزمایشها (یعنی احتمال وقایع در یک آزمایش در اثر نتیجهٔ آزمایش دیگر تغییر نمی‌کند) داریم:

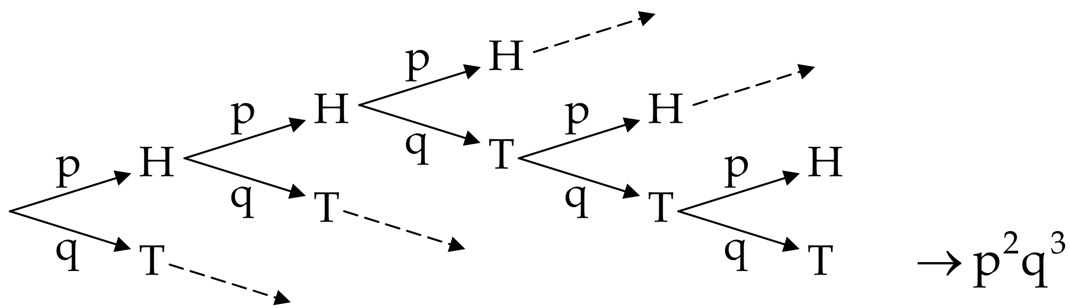
$$P(B)=P(H_1)P(H_2)P(T_3)P(T_4)P(T_5)=p^2q^3$$

اصولاً n آزمایش را مستقل گویند اگر واقعه‌های A_1، A_2، ... و A_n و ... مستقل باشند که A_i واقعه‌ای است که نتیجهٔ آن در ارتباط با آزمایش iام حاصل می‌شود.

اگر بدون توجه به تعریف Ω احتمال این وقایع را در هم ضرب کنیم، بی‌معنی است. چون اشتراکی ندارند.

ضمناً اگر $p=q=\frac{1}{2}$ بود، همان $\frac{1}{2^5}$ می‌شد که قبل از این به دست می‌آوردیم (با تقسیم تعداد نقاط نمونهٔ مطلوب بر کل تعداد نقاط نمونه).

با استفاده از درخت احتمال نیز داریم:



حال برای واقعهٔ D داریم:

$$D = \{\text{دو تا شیر بیاید (و سه خط)}\}$$

کلیهٔ وقایع ساده مانند B که دارای دو تا H و سه تا T (با ترتیب مشخص) هستند دارای احتمال p^2q^3 می‌باشند و این وقایع از هم جدا هستند و D در واقع اجتماع چنین وقایعی است. پس کافی است تعداد این وقایع را به دست آوریم.

به چند طریق می‌توان دو تا H و سه تا T را در ۵ جایگاه قرار داد؟ $\binom{5}{2}$؛

پس داریم:

$$P(B) = \binom{5}{2} p^2 q^3$$

که برای $p = \frac{1}{2}$ همان $\frac{\binom{5}{2}}{2^5}$ می‌شود.

حالت کلی این مثال را **آزمایش برنولی** گویند.

آزمایش برنولی (Bernoulli Trials):

واقعهٔ A در فضای Ω را در نظر بگیرید. اگر $P(A)=p$ و $P(\bar{A})=1-p=q$ باشد، در هر بار انجام آزمایش یا A یا $\bar{A}$ اتفاق می‌افتد و آزمایشها در شرایط یکسان تکرار شده و از هم مستقل می‌باشند.

حال اگر واقعهٔ B در فضای $\Omega_n=\Omega\times\cdots\times\Omega$ این باشد که واقعهٔ A، kبار با ترتیب خاصی اتفاق افتد، مثلاً در kبار اول A اتفاق بیفتد (و در n-kبار بعدی $\bar{A}$) خواهیم داشت:

$$\Omega_n=\{(00\cdots00),(00\cdots01),\ldots,(11\cdots11)\}$$

$$B=\{(\underbrace{11\ldots1}_{\text{بار } k}\underbrace{00\ldots0}_{\text{بار } n\text{-}k})\}$$

A_i: مجموعهٔ کلیهٔ nتایی‌های مرتبی که عنصر iام آنها یک است

$$P(B)=P(A_1A_2\cdots A_k\bar{A}_{k+1}\cdots\bar{A}_n)\underset{\text{استقلال آزمایشها}}{=}P(A_1)P(A_2)\cdots P(A_k)P(\bar{A}_{k+1})\cdots P(\bar{A}_n)=\underbrace{pp\cdots p}_{\text{بار } k}\underbrace{qq\cdots q}_{\text{بار } n\text{-}k}$$

$$=p^kq^{n-k}$$

حال اگر واقعهٔ D در Ω_n این باشد که واقعهٔ A، kبار (با هر ترتیبی) اتفاق افتد، $P(D)=P_n(k)$ چقدر است؟

تعداد وقایع ساده‌ای که در آن A، kبار اتفاق می‌افتد $\binom{n}{k}$ است و همگی احتمال $p^k q^{n-k}$ دارند. پس (طبق اصل ۳) داریم:

$$P_n(k)=\binom{n}{k}p^k q^{n-k}$$

روشن است که:

$$\sum_{k=0}^{n} P_n(k)=1$$

آزمایش برنولی تعمیم‌یافته:

در آزمایش برنولی فقط دو حالت داشتیم: وقوع A یا وقوع $\bar{A}$. در حالت کلی اگر $i = 1, 2, \ldots, r$ و A_iها مجموعهٔ Ω را افراز کنند و احتمال هر یک از آنها p_i $(i = 1, 2, \ldots, r)$ باشد (و آزمایشها در شرایط یکسان تکرار شده و از هم مستقل باشند)، احتمال واقعهٔ B (در Ω_n) که در n آزمایش، A_iها هر یک k_iبار $(k_1 + k_2 + \cdots + k_r = n)$ به ترتیب مشخصی اتفاق افتند برابر است با:

$$p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$

احتمال واقعهٔ D (در Ω_n) که A_iها هر یک k_iبار (با هر ترتیبی) اتفاق افتند برابر است با:

$$C^n_{k_1, k_2, \ldots, k_r} P_1^{k_1} P_2^{k_2} \cdots P_r^{k_r}$$

یعنی:

$$P_n(k_1, k_2, \ldots, k_r) = \frac{n!}{k_1! k_2! \cdots k_r!} p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$$ (توزیع چند جمله‌ای)

مثال: جعبه‌ای شامل N ترانزیستور است که Mتا از آنها $(M \leq N)$ خرابند. به طور تصادفی ترانزیستوری را از جعبه برداشته و آزمایش می‌کنیم و دوباره به جعبه برمی‌گردانیم. اگر این کار را nبار انجام دهیم (n انتخاب با جایگزینی) احتمال اینکه kتا از این n ترانزیستور خراب بوده باشند چیست؟

چون ترانزیستور را پس از تست به جعبه برمی‌گردانیم، شرایط آزمایش تغییری نمی‌کند و لذا یک آزمایش برنولی است.

واقعهٔ A مورد نظر در Ω اصلی (تک آزمایش) واقعهٔ خراب بودن ترانزیستور است که داریم:

احتمال خراب بودن: $p = \dfrac{M}{N}$

لذا احتمال این واقعهٔ در Ω_n که kتا از n ترانزیستور خراب باشند (مجموعهٔ کلیهٔ زوجهای مرتبی که k عنصر آنها یک است) برابر است با:

$$p_1 = \binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}$$

اگر انتخاب بدون جایگزینی بود، این مثال را (مثال ۳، فصل ۲، ص ۱۵) قبلاً داشتیم:

$$\frac{\underbrace{\binom{60}{15}}^{\text{توپهای قرمز}}\underbrace{\binom{40}{5}}^{\text{توپهای سیاه}}}{\binom{100}{20}}.$$

پس داریم:

$$p_2 = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

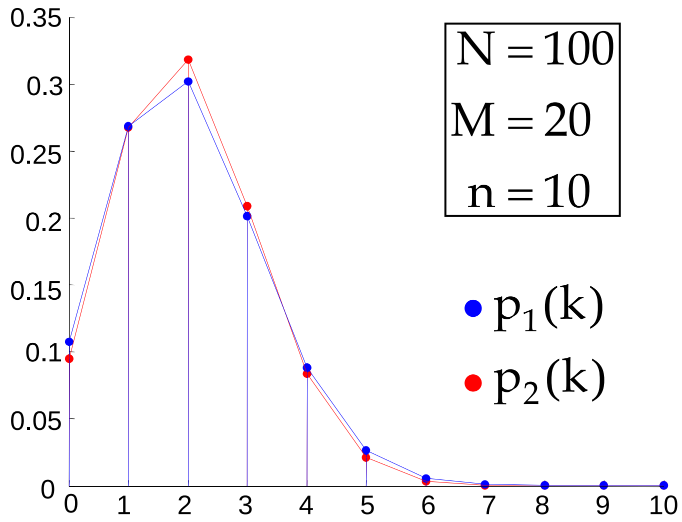
حال اگر بخواهیم مشابه $P_n(k)$ اینها را نیز بر حسب p بیان کنیم، داریم: $p = \frac{M}{N}$ و لذا:

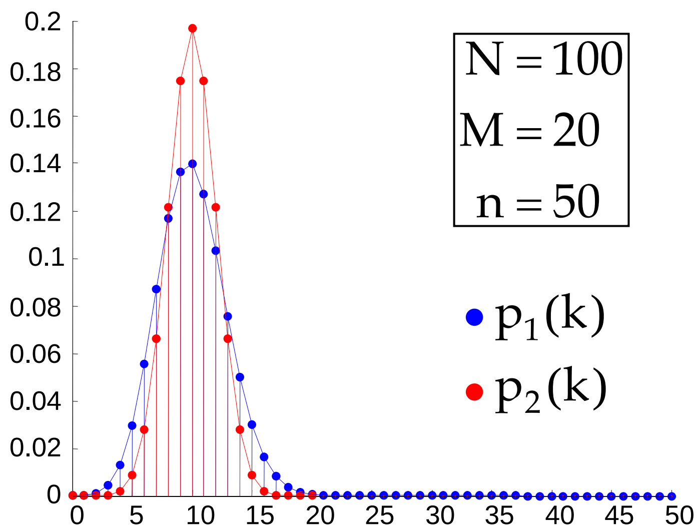
سری فوق هندسی (Hypergeometric): $$p_2(k) = \frac{\binom{pN}{k}\binom{N-pN}{n-k}}{\binom{N}{n}}$$

در حالی که:

(دوجمله‌ای): $$p_1(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

اگر $M \ll k$ و $N \ll n$ باشد، داریم: $p_1 \simeq p_2$.





مثال: برای یک آزمایش احتیاج به مقاومت ۱۰ Ω دقیق داریم. تنها ۱٪ = p از مقاومتهای موجود در بازار در بازهٔ مورد نظر ما هستند. چند تا مقاومت بخریم تا به احتمال ۹۵٪ = P مطمئن باشیم که لااقل یکی در بازهٔ مورد نظر ما خواهد بود.

آزمایش تصادفی: انتخاب یک مقاومت ۱۰ Ω از میان مقاومتهای موجود در بازار

Ω: کلیهٔ مقادیر ممکنه برای مقدار یک مقاومت ۱۰ Ω (یا اینکه دو نقطه بگیریم: بودن و نبودن در بازهٔ مورد نظر)

A: مقاومت مورد نظر در بازهٔ مطلوب باشد (۱٪ = p)

Ω_n: کلیهٔ مقادیر ممکنه برای n مقاومت (یا اگر Ω را دو نقطه در نظر بگیریم، Ω_n، 2^n نقطه خواهد داشت)

C: واقعهٔ مورد نظر در Ω_n؛ در n آزمایش لااقل یک بار A اتفاق افتد

حال می‌خواهیم داشته باشیم: $P(C) = P$؛

$$P(C) = \sum_{k=1}^{n} P_n(k) = \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} = 1 - P_n(0) = 1 - q^n = 1 - (1-p)^n$$

$$\Rightarrow 1 - (1-p)^n = P \Rightarrow 1 - P = (1-p)^n \Rightarrow n = \frac{\log(1-P)}{\log(1-p)}$$

$$\begin{cases} P = 0.95 \\ p = 0.01 \end{cases} \Rightarrow n = 296$$

مثال: احتمال اینکه نوعی دیود قبل از ۱۰۰۰ ساعت کار خراب شود ۰/۱۵ است. احتمال اینکه در میان ۱۰۰تا از این دیودها، ۹۳تا یا بیشتر قبل از ۱۰۰۰ ساعت سالم باشند (حداکثر ۷تا خراب شوند) چیست؟

Ω: زمان خراب شدن یک دیود

Ω_n: زمانهای خراب شدن n دیود

A: واقعهٔ مورد نظر در Ω؛ خرابی در فاصلهٔ $(0, 1000)$ با احتمال $p =$ ۰/۱۵

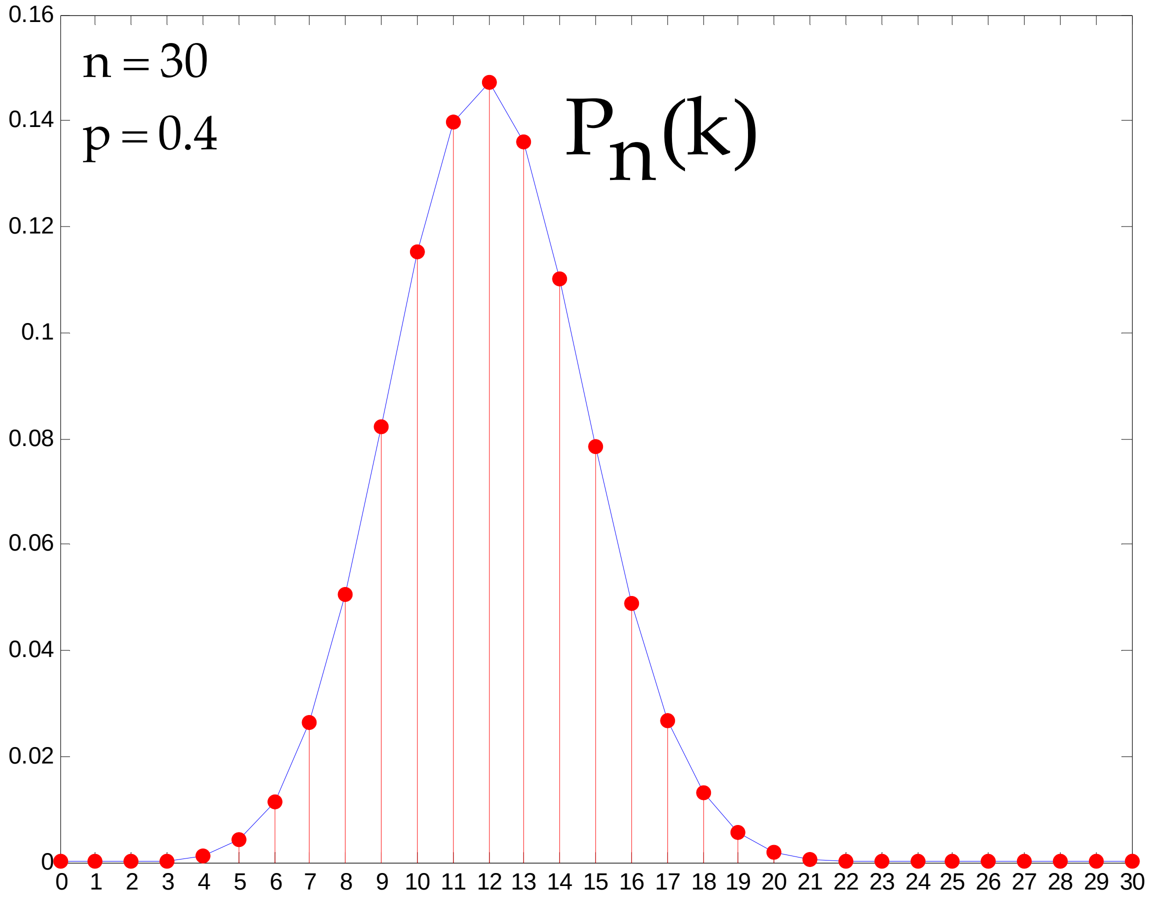
B: واقعهٔ مورد نظر در Ω_n ($n =$ ۱۰۰)؛ حداکثر ۷ دیود خراب شوند (در ۱۰۰ آزمایش، در هفت آزمایش یا کمتر واقعهٔ A اتفاق بیفتد)

$$P(B) = \sum_{k=0}^{7} \binom{100}{k} (0.15)^k (0.85)^{100-k} = 0.012165$$

اگر خرابی یک دیود سبب خرابی (یا افزایش احتمال خرابی) دیود دیگر شود دیگر محاسبهٔ فوق معتبر نخواهد بود. چون آزمایشها وابسته هستند و مقدار احتمال واقعهٔ مورد نظر در یک آزمایش با نتیجهٔ آزمایش قبلی تغییر می‌کند.

مثلاً احتمال اینکه در میان ۹ حرف یک متن انگلیسی دوتا u وجود داشته باشد در حالی که بدانیم کلاً احتمال وقوع u در حروف انگلیسی p است، $\binom{9}{2}p^2q^7$ نخواهد بود. چون بسته به اینکه حرف قبلی چه باشد احتمال u فرق می‌کند. اگر q باشد احتمال وقوع u تقریباً یک خواهد بود. از سوی دیگر اگر حرف قبلی i باشد احتمال وقوع u بسیار کم خواهد بود. یا به همین ترتیب احتمال s به دنبال t یا w زیاد است، ولی احتمال آن پس از j یا f خیلی کم است (رجوع شود به درس رمزنگاری).

برای nهای بزرگ محاسبهٔ $P_n(k)$ مشکل می‌شود و بدتر از آن وقتی است که $\sum$ با تعداد زیاد روی چنین $P_n(k)$هایی داشته باشیم.



که مقدار ماکزیمم آن در حدود np اتفاق می‌افتد و توضیح دقیق آن در کتاب آمده است.

(اگر $(n+1)p$ صحیح نباشد، $k_m = \lceil (n+1)p \rceil$ و اگر $(n+1)p$ صحیح باشد، $\begin{cases} k_1 = (n+1)p-1 \\ k_2 = (n+1)p \end{cases}$)

تقریب $P_n(k)$ برای nهای بزرگ:

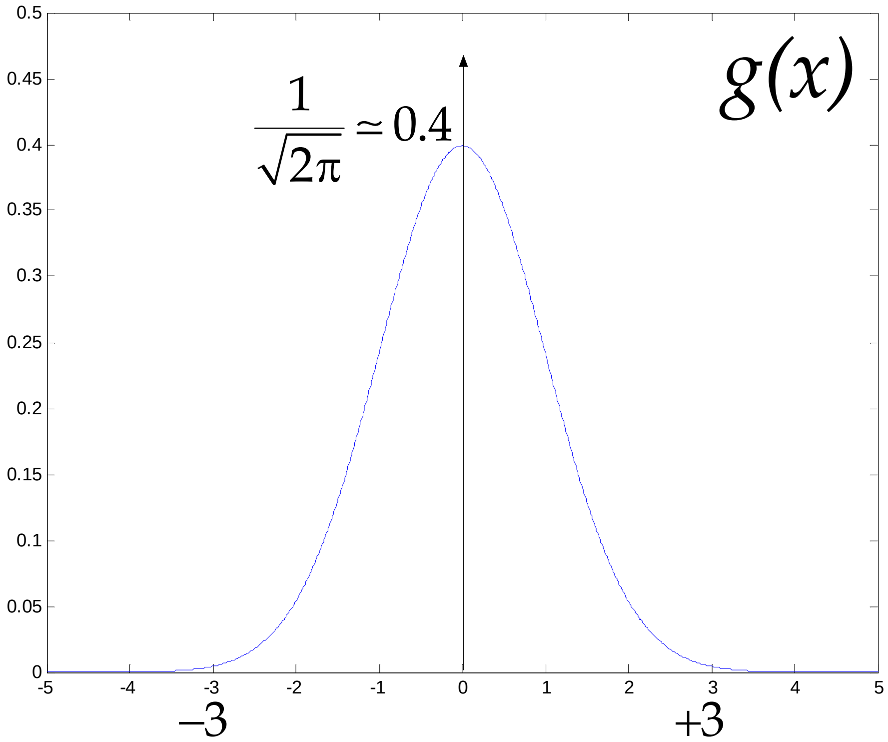
برای تقریب‌زدن $P_n(k)$ برای nهای بزرگ می‌توان از منحنی نرمال استفاده کرد (شیفت یافته و scale شده).

منحنی نرمال (گوسی):

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

$$g(x) = g(-x)$$

$$\int_{-\infty}^{+\infty} g(x)dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$



مقدار $g(x)$ برای خارج بازهٔ (-3, 3) بسیار کوچک است و نقاط عطف آن 1 و 1- هستند.

انتگرال $g(x)$:

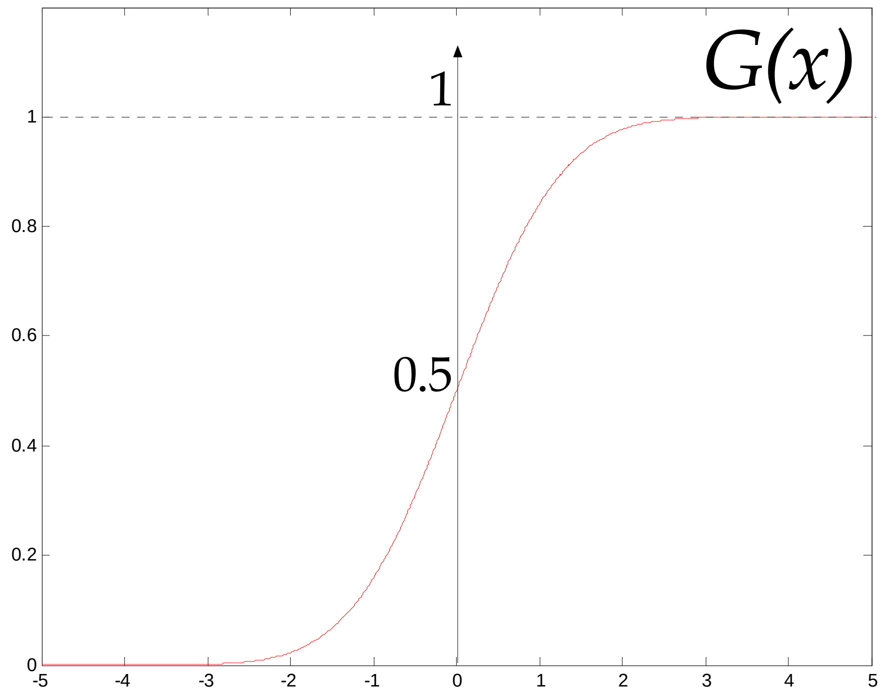
$$G(x) = \int_{-\infty}^{x} g(u)du = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

$$G(-\infty) = 0$$

$$G(+\infty) = 1$$

$$G(0) = \tfrac{1}{2}$$

$$G(-x) = 1 - G(x)$$



$G(x)$ در جداول برای $-3 \le x \le 3$ قید شده است.

برای $x > 3$، $G(x)$ خیلی به 1 نزدیک است و داریم:

$$G(x) \simeq 1 - \frac{1}{x} g(x) : x > 3$$

قضیهٔ دموآور-لاپلاس (De Moivre - Laplace Theorem):

برای nهای بزرگ داریم:

$$P_n(k) \simeq \frac{1}{\sigma} g(\frac{k-\eta}{\sigma})$$

که در آن: $\begin{cases} \eta = np \\ \sigma = \sqrt{npq} \end{cases}$ است.

یعنی:

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi}\sqrt{npq}} e^{-\frac{(k-np)^2}{2npq}}$$

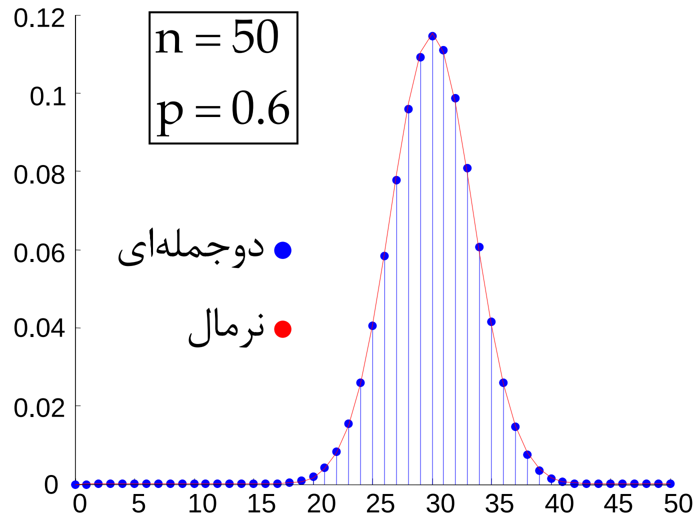
به شرط اینکه: $npq \gg 1$ و $0 < np - 3\sqrt{npq} < k < np + 3\sqrt{npq} < n$.

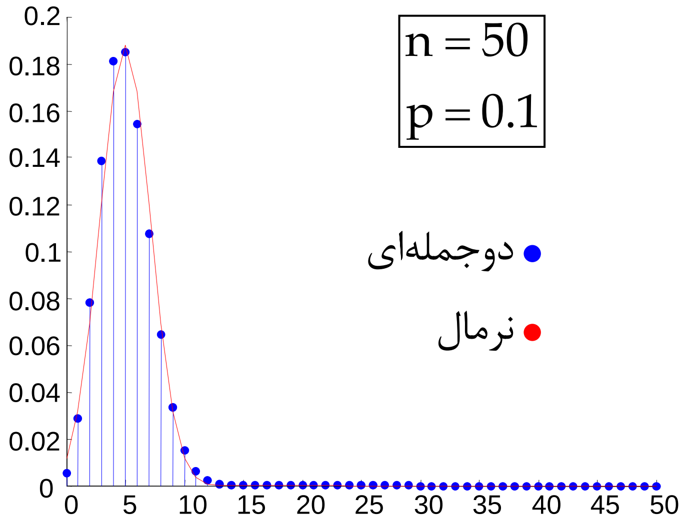
(یعنی کل گسترش ... $P_n(k)$ باید در محدودهٔ 0 و n باشد و نیز k مورد نظر خارج ناحیهٔ 3σ نباشد، چون مقدار تابع ناچیز شده و خطای نسبی تقریب زیاد می‌شود.)

این قضیه را بدون اثبات می‌پذیریم (اثبات در کتاب Beckmann، ص ۳۵).

اصولاً می‌توان نشان داد که حد نسبت اینها یک می‌شود.

این تقریب برای pهای نزدیک به 0.5 بهتر است و هر چه p به صفر یا یک نزدیک شود، تقریب خرابتر می‌شود و n بزرگتری برای خوبی تقریب لازم است. چون تقارن منحنی دوجمله‌ای از بین می‌رود، در حالی که منحنی نرمال متقارن است (بعداً تقریب پواسون را خواهیم دید).





قضیهٔ انتگرال دموآور-لاپلاس:

$$\sum_{k=k_1}^{k_2}\binom{n}{k}p^k q^{n-k} \simeq G\left(\frac{k_2-np}{\sqrt{npq}}\right)-G\left(\frac{k_1-np}{\sqrt{npq}}\right)$$

به شرط اینکه: $(npq \gg 1)$ و $(k_2 < np+3\sqrt{npq} < n)$ یا $(0 < np-3\sqrt{npq} < k_1)$.

(اگر یکی از k_1 یا k_2 خارج از محدوده باشد، جمله‌های متناظر با خارج محدودهٔ 3σ در مقابل سایر جمله‌ها قابل صرف‌نظر است. لذا باز تقریب قابل استفاده است.)

تقریب بهتری هم در کتاب آمده است (ص ۷۴)، یعنی با در نظر گرفتن $k_1 - 0.5$ و $k_2 + 0.5$ که خصوصاً وقتی تفاوت k_1 و k_2 کم باشد، بهتر است.

برای $k_1 = 0$ چون از شرط دوم داریم: $np > 3\sqrt{npq}$، جملهٔ دوم G عددی کوچکتر از -3 می‌شود و لذا تقریباً صفر است.
یعنی:

$$\sum_{k=0}^{k_2}\binom{n}{k}p^k q^{n-k} \simeq G\left(\frac{k_2-np}{\sqrt{npq}}\right)$$

به شرط اینکه: $npq \gg 1$ و $0 < np-3\sqrt{npq} < k_2 < np+3\sqrt{npq} < n$.

مثال: برای مثالی که در صفحهٔ ۱۲ داشتیم (خرابی دیودها)، خواهیم داشت:

$$\begin{cases} p = 0.15 \\ n = 100 \\ k_2 = 7 \end{cases} \rightarrow \sum_{k=0}^{7} P_n(k) = ?$$

$$\begin{cases} np = 15 \\ \\ npq = 12.75 \Rightarrow \sqrt{npq} = 3.57 \Rightarrow \begin{cases} np - 3\sqrt{npq} = 15 - 3(3.57) = 4.3 \\ np + 3\sqrt{npq} = 15 + 3(3.57) = 25.7 \end{cases} \end{cases}$$

$$\rightarrow 12.75 \gg 1 : \boxtimes$$

$$\rightarrow 0 < 4.3 < 7 < 25.7 < 100 : \boxtimes$$

$$\Rightarrow \sum_{k=0}^{7} \binom{100}{7} (0.15)^k (0.85)^{100-k} \simeq G\left(\frac{7-15}{3.57}\right) = G(-2.24) = 0.0125$$

(جواب دقیق 0.0122 بود.)

تعمیم قضیهٔ دموآور-لاپلاس:

$$P_n(k_1,k_2,\ldots,k_r)=\binom{n}{k_1,k_2,\ldots,k_r}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}\simeq\frac{e^{-\frac{1}{2}\sum_{i=1}^{r}q_i x_i^2}}{(2\pi)^{\frac{r-1}{2}}\sqrt{p_1p_2\cdots p_r}\sqrt{n^{r-1}}}:x_i=\frac{k_i-np_i}{\sqrt{np_iq_i}}$$

با استفاده از قضیهٔ دموآور-لاپلاس (که حالت خاصی از قضیهٔ حد مرکزی است) می‌توانیم حالت خاصی از **قانون اعداد بزرگ** را که بعداً خواهیم دید اثبات کنیم.

قانون اعداد بزرگ (The Law of Large Numbers):

به وسیلهٔ قضیهٔ لاپلاس دیدیم که اگر واقعهٔ A در n مرتبه تکرار آزمایشی kبار اتفاق افتد، داریم:

$$P\{k_1\le k\le k_2\}\simeq G\left(\frac{k_2-np}{\sqrt{npq}}\right)-G\left(\frac{k_1-np}{\sqrt{npq}}\right)$$

به این وسیله نشان می‌دهیم که برای nهای بزرگ، $\frac{k}{n}$ نزدیک به p خواهد بود. این همان چیزی است که در تعبیر تجربی احتمال داشتیم و حالا مفهوم دقیق ریاضی برای آن بیان می‌کنیم.

قضیه: اگر n_0 به قدر کافی بزرگ انتخاب شود، برای هر $n > n_0$ داریم:

$$P\{|\frac{k}{n} - p| \le \varepsilon\} \ge 1 - \delta$$

به عبارت دیگر برای هر $\varepsilon > 0$ داده شده داریم:

$$P\{|\frac{k}{n} - p| \le \varepsilon\} \xrightarrow{n\to\infty} 1$$

اثبات:

$$P\{|\frac{k}{n} - p| \le \varepsilon\} = P\{(p-\varepsilon)n \le \frac{k}{n} \le (p+\varepsilon)n\} = G(\frac{(p+\varepsilon)n - np}{\sqrt{npq}}) - G(\frac{(p-\varepsilon)n - np}{\sqrt{npq}})$$

$$= G(\varepsilon\sqrt{\frac{n}{pq}}) - G(-\varepsilon\sqrt{\frac{n}{pq}}) = 2G(\varepsilon\sqrt{\frac{n}{pq}}) - 1$$

پس برای هر $\varepsilon > 0$ داده شده داریم: $\varepsilon\sqrt{\frac{n}{pq}} \xrightarrow{n\to\infty} \infty$ و لذا خواهیم داشت: $G(\varepsilon\sqrt{\frac{n}{pq}}) \xrightarrow{n\to\infty} 1$؛

در نتیجه داریم:

$$P\{|\frac{k}{n} - p| \le \varepsilon\} \xrightarrow{n\to\infty} 1$$

مثال: یک سکهٔ سالم را چند بار بیندازیم تا به احتمال ۹۹/۹٪، $\frac{k}{n}$ در بازهٔ (0.49, 0.51) باشد.

$$P\{|\frac{k}{n}-p|\le \varepsilon\}=2G(\varepsilon\sqrt{\frac{n}{pq}})-1$$

$$P\{|\frac{k}{n}-\frac{1}{2}|\le 0.01\}=2G(0.02\sqrt{n})-1=0.999 \Rightarrow G(0.02\sqrt{n})=0.9995$$

$$\Rightarrow 0.02\sqrt{n}=3.291 \Rightarrow n=27077$$

وقایع نادر:

در آزمایش برنولی اگر واقعهٔ A نادر الوقوع باشد، یعنی $p \ll 1$ باشد، در صورتی که n آن‌قدر بزرگ باشد که $np \ll 1$، می‌توان از قضیهٔ لاپلاس استفاده کرد. ولی اگر np در حدود 1 باشد، شرط صحت قضیهٔ لاپلاس دیگر برقرار نخواهد بود.

قضیهٔ پواسُن (Poisson Theorem):

$$P_n(k) \simeq e^{-\eta} \frac{\eta^k}{k!}$$

یعنی:

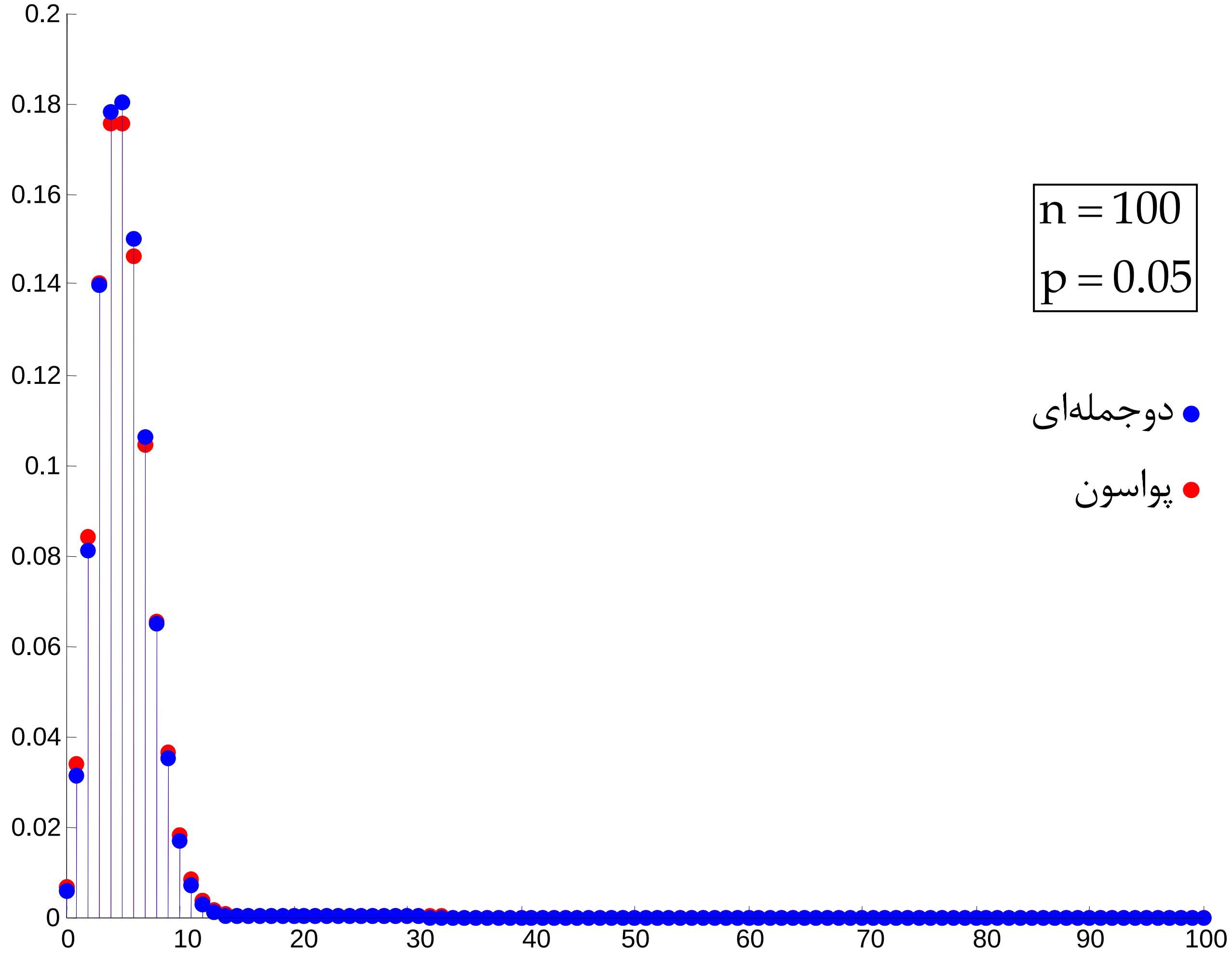
$$\binom{n}{k} p^k q^{n-k} \simeq e^{-np} \frac{(np)^k}{k!}$$

در صورتی که: $p \ll 1$ و $n \ll 1$ و k در حدود np باشد (و np عددی مثلاً در حدود ۱ تا ۱۰ باشد).

(اثبات در کتاب، ص ۷۸)

در واقع می‌توان نشان داد (در کتاب Beckmann، ص ۳۹) که:

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np\to\eta}} P_n(k) = e^{-\eta} \frac{\eta^k}{k!}$$


$n = 100$
$p = 0.05$
دوجمله‌ای
پواسون

مثال ۱: در مثالی که قبلاً داشتیم (خرابی دیودها)، $n = ۱۰۰$ بود. حال اگر $p = ۰/۰۳$ باشد، $np = ۳$ خواهد بود و داریم:

$$\sum_{k=0}^{7}\binom{100}{k}(0.03)^k(0.97)^{100-k} \simeq \sum_{k=0}^{7} e^{-3}\frac{3^k}{k!} = e^{-3}(1+3+\frac{3^2}{2!}+\frac{3^3}{3!}+\frac{3^4}{4!}+\frac{3^5}{5!}+\frac{3^6}{6!}+\frac{3^7}{7!}) = 0.988$$

جواب دقیق 0.989 می‌باشد.

تعمیم قضیهٔ پواسن:

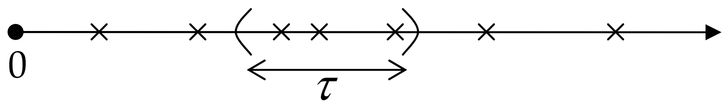
$$P_n(k_1,k_2,\ldots,k_r) = \frac{n!}{k_1!k_2!\cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r} \simeq \prod_{i=1}^{r-1} e^{-\eta_i}\frac{\eta_i^{k_i}}{k_i!} : \eta_i = np_i$$

با فرض اینکه برای هر $i = 1,2,\ldots,r-1$ داشته باشیم: $p_i \ll 1$.

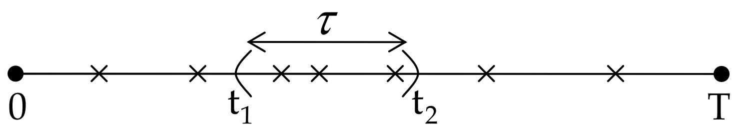
برای η و k بزرگ استفاده از تقریب پواسن مشکل پیدا می‌کند (زیرا $e^{-\eta}\eta^k$ حاصلضرب عددی خیلی کوچک در عددی خیلی بزرگ می‌شود) و می‌توان نشان داد که پواسن را می‌توان با نرمال تقریب زد. یعنی دوباره به استفاده از قضیهٔ لاپلاس برمی‌گردیم (با فرض برقراری شرایط قضیهٔ لاپلاس).

نقاط پواسن:

وقتی با نقاط تصادفی در زمان یا مکان سر و کار داریم توزیع پواسن استفادهٔ زیادی دارد. مثلاً تعداد مکالمات تلفنی در یک مرکز سوئیچ در مدت مشخصی از زمان، تعداد نقاط خرابی روی طول مشخصی از یک نوار مغناطیسی یا پارچه، تعداد وقوع زلزله یا جنگ در مدت زمانی معین، تعداد اشتباهات چاپی روی یک صفحه، تعداد فوت‌شدگان بیمهٔ عمر در یک سال، تعداد ذرات اتمی منتشر شده از یک مادهٔ رادیواکتیو که به هدف خاصی در طی زمان مشخصی برخورد کنند و یا تعداد الکترونهای منتشر شده از یک مادهٔ فوتوالکتریک در اثر تابش نور در طی مدت زمانی مشخص.



در این آزمایشها: $n \to \infty$ و $p \to 0$، ولی: $np \to \lambda$.



برای n محدود آزمایش تصادفی تکراری ما انتخاب n نقطه در $(0, T)$ است.
یعنی همهٔ n نقطه در فاصلهٔ $(0, T)$ خواهند بود، اما کجای آن را نمی‌دانیم.

احتمال این را می‌خواهیم که kتا از این n نقطه در فاصلهٔ (t_1, t_2) که $t_2 - t_1 = \tau$ است، قرار گیرند. یعنی واقعهٔ $A = \{t \in (t_1, t_2)\}$ در nبار تکرار آزمایش kبار اتفاق افتد.

$$P\{\tau \text{ نقطه در فاصلهٔ زمانی } k\} = \binom{n}{k} p^k q^{n-k} : p = \frac{\tau}{T}$$

اگر n خیلی بزرگ و $p = \frac{\tau}{T}$ خیلی کوچک باشد، با استفاده از تقریب پواسن داریم:

$$\text{تعداد نقاط در واحد زمان} = \lambda = \frac{n}{T} : \quad P\{\tau \text{ نقطه در فاصلهٔ زمانی } k\} \simeq e^{-\frac{n\tau}{T}} \frac{(\frac{n\tau}{T})^k}{k!} = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!}$$

اگر $T \to \infty$ (یعنی معادلاً $p \to 0$) و $n \to \infty$، ولی λ را ثابت نگه داریم، تقریب فوق دقیق شده و داریم:

$$P\{\tau \text{ نقطه در فاصلهٔ زمانی } k\} = e^{-\lambda\tau} \frac{(\lambda\tau)^k}{k!}$$

$$P\{\text{ نقطه در واحد زمان } k\} = e^{-\lambda} \frac{\lambda^k}{k!}$$

از جمله خواص نقاط پواسن این است که احتمال وقوع یک نقطه در هر بازه در صورتی که بازه کوچک باشد، متناسب با طول بازه است، زیرا:

$$P\{\tau \text{ یک نقطه در فاصلهٔ زمانی }\} = e^{-\lambda\tau}\lambda\tau \overset{\tau \text{ کوچک}}{\simeq} \lambda\tau$$

به عبارت دیگر:

$$\lim_{\tau\to 0}\frac{P\{\tau \text{ یک نقطه در فاصلهٔ زمانی }\}}{\tau} = \lim_{\tau\to 0}\lambda e^{-\lambda\tau} = \lambda$$

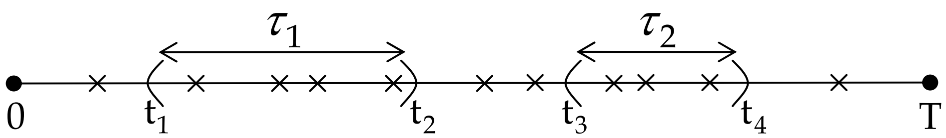
توجه کنید که λ عدد ثابتی است و $\lambda(t)$ نیست. یعنی فرایند ایستان است و نرخ متوسط وقوع پدیده‌ها همیشه یکسان است.

همچنین می‌توان نشان داد که:

$$\lim_{\tau\to 0}\frac{P\{\tau \text{ بیش از یک نقطه در فاصلهٔ زمانی }\}}{\tau} = 0$$

(چون احتمال موجود در صورت کسر عبارت است از: $e^{-\lambda\tau}\dfrac{\lambda^2\tau^2}{2!}$)

ویژگی دیگر نقاط پواسن این است که اگر دو بازهٔ غیرمتلاقی را در نظر بگیریم، احتمالها مستقل خواهند بود (اثبات در ص ۸۰ کتاب).



یعنی اگر (t_1, t_2) و (t_3, t_4) دو بازهٔ غیرمتلاقی باشند و $\tau_1 = t_2 - t_1$ و $\tau_2 = t_4 - t_3$ باشد، داریم:

$$P\{k_1 \text{ نقطه در } \tau_1 \text{ و } k_2 \text{ نقطه در } \tau_2\} = P\{k_1 \text{ نقطه در } \tau_1\} P\{k_2 \text{ نقطه در } \tau_2\}$$

اصولاً می‌توان (با استفاده از معادلات دیفرانسیل یا با استفاده از توزیع دوجمله‌ای - راس، ص ۱۶۶) نشان داد که اگر بخواهیم تعداد وقوع پدیده‌ای در طی مدت مشخصی از زمان در شرایط زیر صدق کند، توزیع پواسن خواهیم داشت:

۱. احتمال وقوع پدیده در هر فاصلهٔ زمانی در حد (فاصلهٔ زمانی بسیار کوتاه) متناسب با طول فاصله باشد:

$$P\{\text{یک نقطه در فاصلهٔ زمانی } \tau\} = \lambda\tau + O(\tau)$$

$O(\tau)$ تابعی است که برای آن $\lim\limits_{\tau\to 0}\dfrac{O(\tau)}{\tau} = 0$ باشد، یعنی با کاهش τ سریعتر از τ به سمت صفر رود، مانند τ^{α} وقتی که $\alpha > 1$ باشد.

۲. احتمال اینکه پدیده در یک فاصلهٔ زمانی بیش از یک بار اتفاق افتد در حد (فاصلهٔ زمانی کوتاه) صفر باشد:

$$P\{\text{بیش از یک نقطه در فاصلهٔ زمانی } \tau\} = O(\tau)$$

۳. تعداد وقوع پدیده در دو بازهٔ زمانی غیرمتلاقی مستقل باشد.

با صرف این سه شرط می‌توان نشان داد که به توزیع پواسن می‌رسیم.

توجه کنید که شرط صحت به کار بردن دوجمله‌ای و تقریبهای آن از جمله پواسن، استقلال آزمایشها بود. البته اگر وابستگی ضعیف باشد نیز تقریب مناسب خواهد بود.

مثلاً در مسألهٔ تطابق، اگر واقعهٔ A_i، قرار گرفتن نامهٔ iام در پاکت خود باشد، داریم:

$$P(A_i)=\frac{1}{n},P(A_iA_j)=\frac{1}{n(n-1)}$$

یعنی A_iها مستقل نیستند. ولی بستگی آنها برای nهای بزرگ خیلی کم است.

لذا منطقی است که تعداد موفقیتها به طور تقریبی دارای توزیع پواسن با پارامتر $np=\frac{n}{n}=1$ باشد.

یعنی برای nهای بزرگ، احتمال هیچ موفقیت برابر با $\frac{1}{e}$ و احتمال k موفقیت برابر با $\frac{e^{-1}}{k!}$ است.

فصل ۴: متغیر تصادفی Random Variable (RV)

Chapter 4

Chapter 5: Section 5.2

۱. تعریف متغیر تصادفی

۲. تابع pmf

۳. تابع CDF

۴. تابع pdf

۵. برخی توزیع‌های خاص

۶. تابع یک متغیر تصادفی

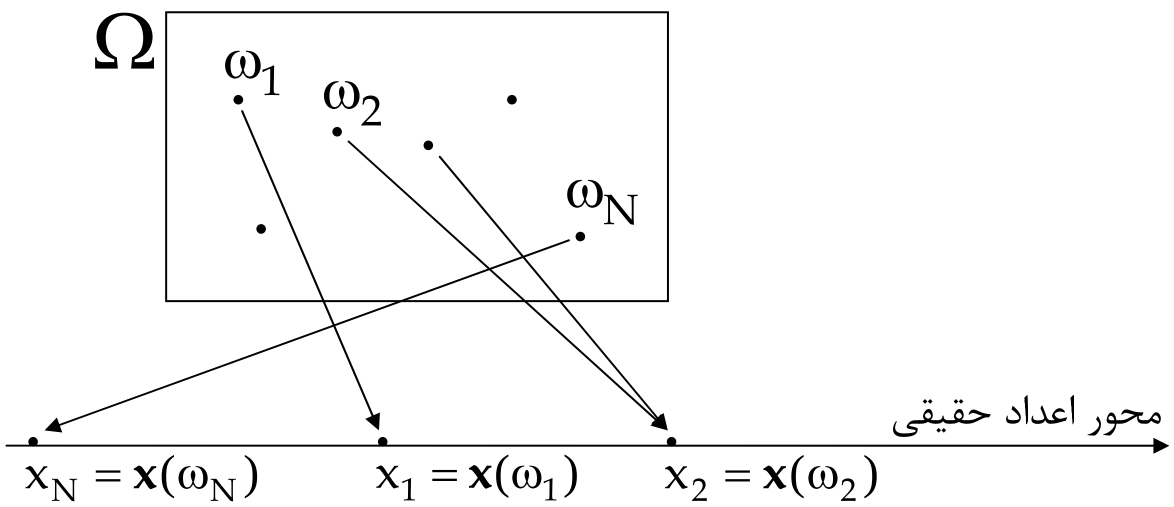
۷. میانگین و واریانس یک متغیر تصادفی

۸. میانگین و واریانس برخی توزیع‌های خاص

۹. گشتاورهای یک متغیر تصادفی

متغیر تصادفی تابعی است از نقاط فضای نمونه مثل $\mathbf{x}(\omega)$ که به هر یک از نقاط فضای نمونه عددی را نسبت می‌دهد.

دامنهٔ (Domain) این تابع مجموعهٔ Ω است و برد (Range) آن مجموعه‌ای از اعداد حقیقی است.



(متغیر تصادفی حقیقی که مورد بحث ما است، عددی حقیقی را نسبت می‌دهد. متغیر تصادفی مختلط، عددی مختلط را نسبت می‌دهد.)

چون خیلی اوقات به جای نتایج آزمایش، تابعی از نتایج مورد توجه ما است. مثلاً در پرتاب دو تاس مقدار حاصلجمع اعداد دو تاس یا در پرتاب سکه‌ها، مجموع تعداد شیرهای ظاهر شده ممکن است مد نظر باشد.

مثال: در آزمایش پرتاب سه سکۀ سالم داریم:

$$\Omega = \{\underbrace{HHH}_{\omega_1},\ldots,\underbrace{TTT}_{\omega_8}\}$$

مثلاً تعداد شیرها یک متغیر تصادفی است:

$\mathbf{x}(\omega) =$ تعداد شیرها در واقعۀ ω

ω	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$\mathbf{x}(\omega)$	۳	۲	۲	۱	۲	۱	۱	۰

حال در مثال فوق احتمال اینکه $\mathbf{x}(\omega) \le 1$ باشد چیست؟ احتمال روی واقعه‌ها تعریف می‌شود. احتمال $\mathbf{x}(\omega) \le 1$ یعنی احتمال مجموعۀ آن ωهایی که برای آنها $\mathbf{x}(\omega) \le 1$ باشد:

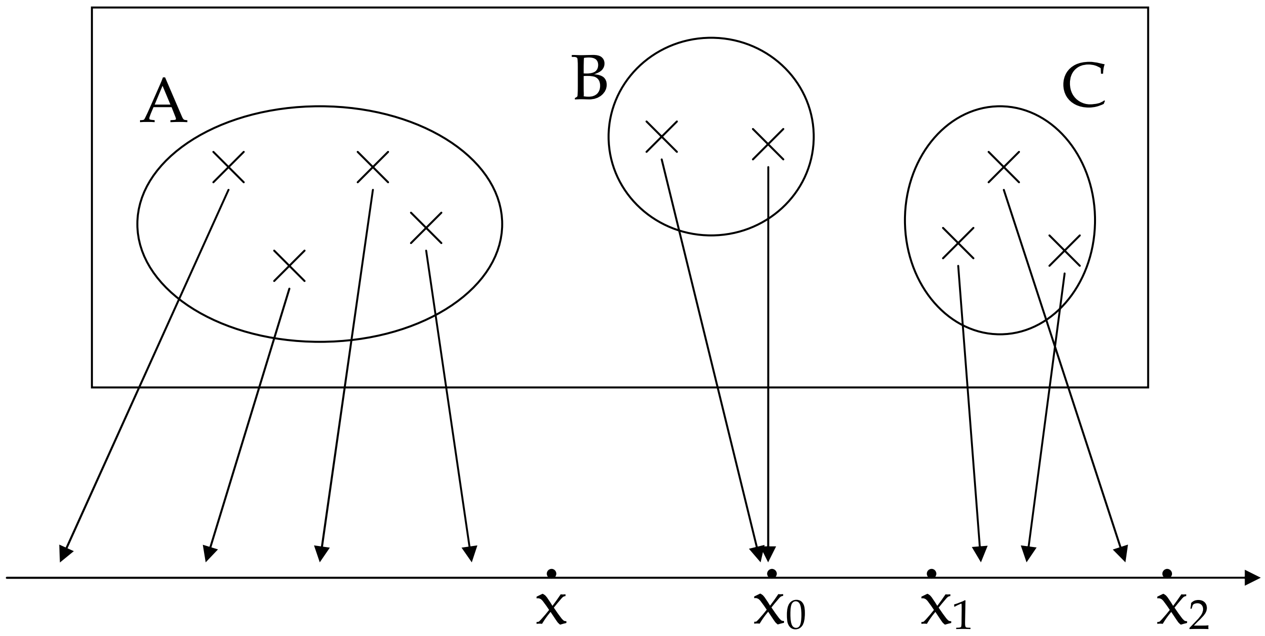
$$P\{\mathbf{x} \le 1\} = P\{\omega : \mathbf{x}(\omega) \le 1\} = P\{HTT, THT, TTH, TTT\} = \frac{4}{8}$$

به طور خلاصه می‌نویسیم: $\{\mathbf{x} \le x\}$، ولی این مجموعه‌ای از اعداد نیست بلکه یعنی: $\{\omega : \mathbf{x}(\omega) \le x\}$.

احتمال اینکه $\mathbf{x}(\omega) = 2$ باشد چیست؟

$$P\{\mathbf{x} = 2\} = P\{\omega : \mathbf{x}(\omega) = 2\} = P\{HHT, HTH, THH\} = \frac{3}{8} = \binom{3}{2}(\frac{1}{2})^2(\frac{1}{2})^{3-2}$$

یعنی:



$$A = \{\mathbf{x} \le x\}$$
$$B = \{\mathbf{x} = x_0\}$$
$$C = \{x_1 < \mathbf{x} \le x_2\}$$

حال مثلاً اگر احتمال $\{\mathbf{x} \le x\}$ را برای هر x بدانیم، احتمال همهٔ واقعه‌های مورد نظر را خواهیم داشت و لزومی به دانستن فضای احتمال Ω نیست.

$-\infty$ یا $+\infty$ اعداد حقیقی نیستند، لذا $\mathbf{x}(\omega)$ نباید $-\infty$ یا $+\infty$ شود (مگر آنکه احتمال آن ω صفر باشد). یعنی اجازه می‌دهیم که $\mathbf{x}(\omega)$ برای برخی ωها $-\infty$ یا $+\infty$ شود مشروط بر آنکه احتمال آن ωها صفر باشد:

$$P\{\mathbf{x} = +\infty\} = P\{\mathbf{x} = -\infty\} = 0$$

(تعریف: متغیر تصادفی حقیقی $\mathbf{x}(\omega)$ تابعی است حقیقی از نقاط فضای نمونهٔ $\omega \in \Omega$ به طوری که برای هر عدد حقیقی x، مجموعهٔ $\{\mathbf{x}(\omega) \le x\}$ یک واقعه باشد و $P\{\mathbf{x} = +\infty\} = P\{\mathbf{x} = -\infty\} = 0$)

متغیر تصادفی را گسسته گویند هرگاه مقادیری که $\mathbf{x}(\omega)$ می‌تواند اختیار کند (برد تابع) قابل شمارش باشد.

در حالت گسسته مثل مثالی که داشتیم ساده‌ترین راه برای مشخص کردن احتمال واقعه‌ها، مشخص کردن احتمال واقعه‌های $\{\mathbf{x} = x_i\}$ برای x_iهای ممکنه است (نسبت به مشخص کردن احتمال $\{\mathbf{x} \le x\}$ برای هر x).

تابع جرمی احتمال pmf (Probability Mass Function) یا تابع احتمال یا تابع فراوانی:

تعریف: اگر $\mathbf{x}$ فقط مقادیر ممکن x_1، x_2، x_3 و ... (قابل شمارش) را با احتمالهای p_1، p_2، p_3 و ... اختیار کند، تابع احتمال متغیر تصادفی $\mathbf{x}$ به صورت زیر تعریف می‌شود:

آرگومان تابع

$$\underset{\substack{\downarrow \\ \text{بیانگر نوع تابعیت}}}{P_{\mathbf{x}}}(\overset{\substack{\text{آرگومان تابع} \\ \uparrow}}{x}) = \text{Prob}\{\mathbf{x} = x\} = \begin{cases} p_i & x = x_i \\ 0 & \text{otherwise} \end{cases}$$

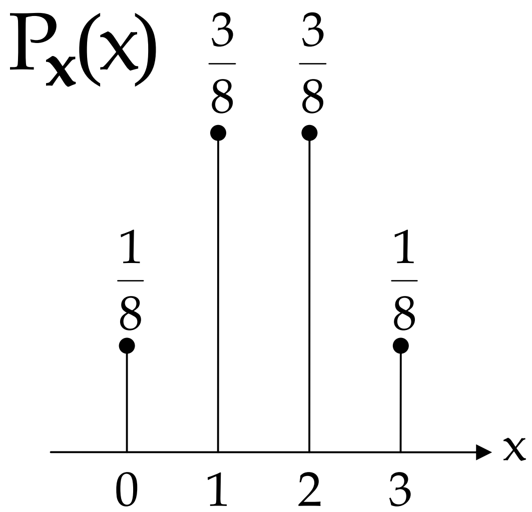
بیانگر نوع تابعیت

(اگر فقط با یک متغیر تصادفی سروکار داشته باشیم اندیس $\mathbf{x}$ را می‌توان حذف کرد: $P(x_i)$)

در مثالی که داشتیم، برای $i = 0, 1, 2, 3$ داریم:

$$P_{\mathbf{x}}(i) = \text{Prob}\{\mathbf{x} = i\} = \binom{3}{i}(\frac{1}{2})^{i}(\frac{1}{2})^{3-i} = \frac{\binom{3}{i}}{8} = \begin{cases} \frac{1}{8} & i=0 \\ \frac{3}{8} & i=1 \\ \frac{3}{8} & i=2 \\ \frac{1}{8} & i=3 \end{cases}$$

و برای سایر xها داریم: $P_{\mathbf{x}}(x) = 0$.



تابع $P_{\mathbf{x}}(x)$ است که فقط در یک سری نقاط محدود یا حداکثر قابل شمارش مقدار غیرصفر دارد.

به طور کلی روشن است که $P(x_i)$ها اعدادی بین صفر و یک هستند (چون احتمال واقعه‌های $\{\mathbf{x} = x_i\}$اند) و داریم:

$$\sum_{i=1}^{\infty} P(x_i) = 1$$

اگر احتمال هر واقعه را بخواهیم، با داشتن $P(x_i)$ها قابل محاسبه است. مثلاً در مثال فوق داریم:

$$P\{\mathbf{x} \le 1.1\} = P\{\mathbf{x} = 0\} + P\{\mathbf{x} = 1\} = \frac{1}{8} + \frac{3}{8} = \frac{4}{8}$$

و به طور کلی برای هر زیرمجموعه از اعداد حقیقی مثل A داریم:

$$\text{Prob}\{\mathbf{x} \in A\} = \sum_{x_i \in A} P(x_i)$$

تابع احتمال برای متغیر تصادفی گسسته تعریف می‌شود.

(وقتی متغیر تصادفی پیوسته باشد یعنی مقادیری که می‌تواند اختیار کند غیرقابل شمارش باشد، تعیین احتمال $\{\mathbf{x} = x_i\}$ها کفایت نمی‌کند. اغلب همه صفرند. در اینجا می‌توانیم تابع توزیع انباشته را مطرح کنیم.)

در مهندسی برق خیلی اوقات با متغیرهای تصادفی پیوسته مثل ولتاژ، توان و امثالهم سروکار داریم.

تابع توزیع انباشته CDF (Cumulative Distribution Function):

طبق تعریف داریم:

آرگومان تابع

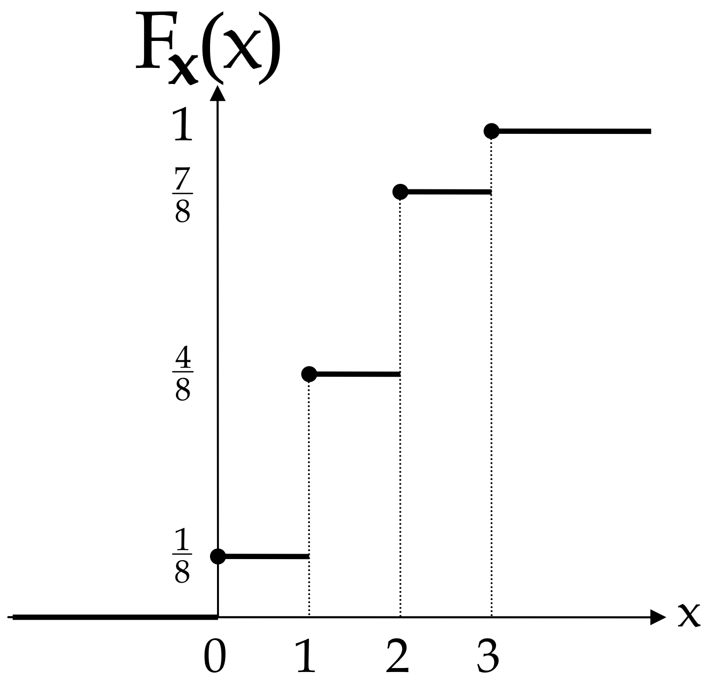
$$\underset{\text{بیانگر نوع تابعیت}}{F_{\mathbf{x}}}(\overset{\text{}}{x}) = \text{Prob}\{\mathbf{x} \le x\}$$

بیانگر نوع تابعیت

(اگر فقط با توزیع یک متغیر تصادفی سروکار داشته باشیم، $F(x)$ کفایت می‌کند.)

چه برای متغیر تصادفی پیوسته و چه گسسته اگر احتمال واقعه‌های $\{\mathbf{x} \le x\}$ را برای هر x بدانیم، احتمال همهٔ واقعه‌ها مشخص خواهد شد. پس $F(x)$ به طور کامل می‌تواند متغیر تصادفی را توصیف کند.

مثال ۱: در مثالی که داشتیم:



$$F_{\mathbf{x}}(-0.001) = P\{\mathbf{x} \le -0.001\} = 0$$
$$F_{\mathbf{x}}(0) = P\{\mathbf{x} \le 0\} = \tfrac{1}{8}$$
$$F_{\mathbf{x}}(0.001) = P\{\mathbf{x} \le 0.001\} = \tfrac{1}{8}$$
$$F_{\mathbf{x}}(0.999) = P\{\mathbf{x} \le 0.999\} = \tfrac{1}{8}$$
$$F_{\mathbf{x}}(1) = P\{\mathbf{x} \le 1\} = \tfrac{4}{8}$$
$$F_{\mathbf{x}}(1.0001) = P\{\mathbf{x} \le 1.0001\} = \tfrac{4}{8}$$
$$F_{\mathbf{x}}(10) = P\{\mathbf{x} \le 10\} = 1$$

در حالت گسسته $F_{\mathbf{x}}(x)$ به صورت پلکانی است.

می‌دانیم که در مورد RV گسسته: $P\{\mathbf{x} \in A\} = \sum_{x_i \in A} P(x_i)$؛ لذا داریم:

$$F(x) = P\{\mathbf{x} \le x\} = \sum_{x_i \le x} P(x_i) = \sum_{i} P(x_i) u(x - x_i)$$

که:

$$u(x) = \begin{cases} 1 & x \ge 0 \\ 0 & x < 0 \end{cases}$$

و همچنین:

$$P(x_k) = x_k \text{ در } F(x) \text{ مقدار پرش } = F(x_k) - F(x_k^-)$$

(به طور ریاضی نیز این را نشان خواهیم داد.)

مثال ۲: یک مکالمهٔ تلفنی در زمانی کاملاً تصادفی بین $[0,T]$ واقع می‌شود.

$$\Omega = \{t : 0 \le t \le T\}$$

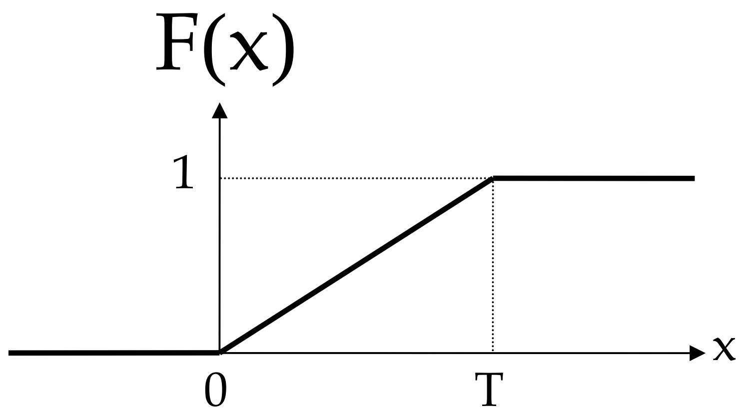
به طوری که برای هر $0 \le t_1 \le t_2 \le T$:

$$P\{t_1 \le t \le t_2\} = \frac{t_2 - t_1}{T}$$

اگر متغیر تصادفی ما، خود زمان وقوع مکالمهٔ تلفنی باشد، داریم:

$$\mathbf{x}(t) = t : 0 \le t \le T$$

$$F_{\mathbf{x}}(x) = P\{\mathbf{x}(t) \le x\} = \begin{cases} P(\Omega) = 1 & x > T \\ P\{0 \le t \le x\} = \dfrac{x}{T} & 0 \le x \le T \\ P(\varnothing) = 0 & x < 0 \end{cases}$$



تابع توزیع انباشتهٔ Ramp برای توزیع یکنواخت حاصل می‌شود.

توجه کنید که عناصر Ω اعداد بین 0 تا T بودند، ولی $F(x)$ برای همهٔ xها تعریف شده است (مثلاً xهای منفی متناظر با مجموعهٔ $\varnothing$ می‌شوند و احتمال صفر دارند).

خواص تابع توزیع انباشته (CDF):

$$1)\ F(-\infty)=0$$

زیرا:

$$F(-\infty)=P\{\mathbf{x}=-\infty\}=0$$

$$2)\ F(+\infty)=1$$

زیرا:

$$F(+\infty)=P\{\mathbf{x}\le+\infty\}=1$$

$3)\ x_1<x_2\Rightarrow F(x_1)\le F(x_2):$ تابعی غیرنزولی است $F(x)$

زیرا:

$$x_1<x_2\Rightarrow\{\omega:\mathbf{x}(\omega)\le x_1\}\subset\{\omega:\mathbf{x}(\omega)\le x_2\}\Rightarrow P\{\mathbf{x}\le x_1\}\le P\{\mathbf{x}\le x_2\}\Rightarrow F(x_1)\le F(x_2)$$

از اینجا نتیجه می‌شود که: $0\le F(x)\le 1$، زیرا دیدیم که در $-\infty$، صفر و در $+\infty$، یک است.

$$4)\ P\{\mathbf{x} > x\} = 1 - F(x)$$

زیرا:

$$\{\omega : \mathbf{x}(\omega) > x\} = \{\omega : \mathbf{x}(\omega) \le x\}^c \Rightarrow P\{\mathbf{x} > x\} = 1 - P\{\mathbf{x} \le x\} = 1 - F(x) = F^c(x)$$

اغلب در مخابرات با $P\{\mathbf{x} > x\}$ سروکار داریم (بزرگتر شدن از یک سطح آستانه) که مکمل $F(x)$ می‌شود.

$$5)\ P\{x_1 < \mathbf{x} \le x_2\} = F(x_2) - F(x_1)$$

زیرا:

$$\{\mathbf{x} \le x_2\} = \{\mathbf{x} \le x_1\} \cup \{x_1 < \mathbf{x} \le x_2\}$$

دو مجموعهٔ جدا از هم

$$\Rightarrow \underbrace{P\{\mathbf{x} \le x_2\}}_{F(x_2)} = \underbrace{P\{\mathbf{x} \le x_1\}}_{F(x_1)} + P\{x_1 < \mathbf{x} \le x_2\}$$

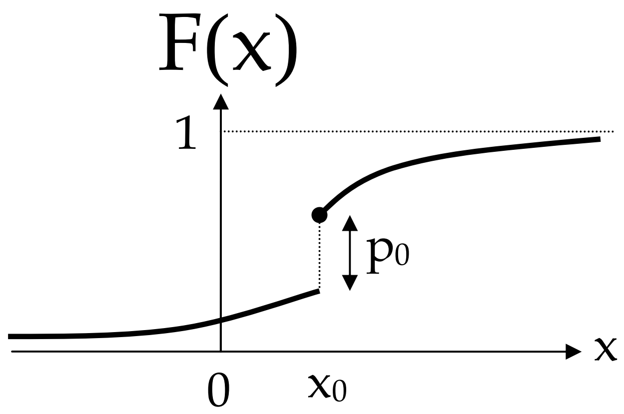
$$\Rightarrow P\{x_1 < \mathbf{x} \le x_2\} = F(x_2) - F(x_1)$$

از همین جا به ازای $x_2 = x$ و $x_1 = x - \varepsilon$ و میل دادن ε به سمت صفر نتیجه می‌شود که:

$$P\{\mathbf{x} = x\} = F(x) - F(x^-)$$

اگر پرش در $F(x)$ داشته باشیم، $F(x)$ با $F(x^-)$ به اندازهٔ $P(x)=P\{\mathbf{x}=x\}$ متفاوت خواهد بود، ولی $F(x)$ به هر حال از راست پیوسته است، یعنی:

$$F(x_0)=F(x_0^+)=P\{\mathbf{x}\le x_0\}$$
$$F(x_0^-)=P\{\mathbf{x}<x_0\}$$
$$p_0=F(x_0^+)-F(x_0^-)=F(x_0)-F(x_0^-)$$



(ضمناً با توجه به $P\{\mathbf{x}=x\}=F(x)-F(x^-)$ داریم:

$P\{x_1\le\mathbf{x}\le x_2\}=F(x_2)-F(x_1^-)$)

6) $F(x)=F(x^+)$

(قابل اثبات به طور دقیق ریاضی، کتاب DeGroot، ص ۱۱۰)

حال که $F(x)$ را فهمیدیم می‌توانیم گسسته یا پیوسته بودن متغیر تصادفی را طور دیگری نیز تعریف کنیم:

متغیر تصادفی را گسسته (Discrete) گویند اگر $F(x)$ به صورت پلکانی باشد، مانند مثالی که برای سه سکه داشتیم.

در این حالت همان طور که قبلاً دیدیم:

$$P\{\mathbf{x}=x_i\}=F(x_i)-F(x_i^-)=p_i=P(x_i)$$

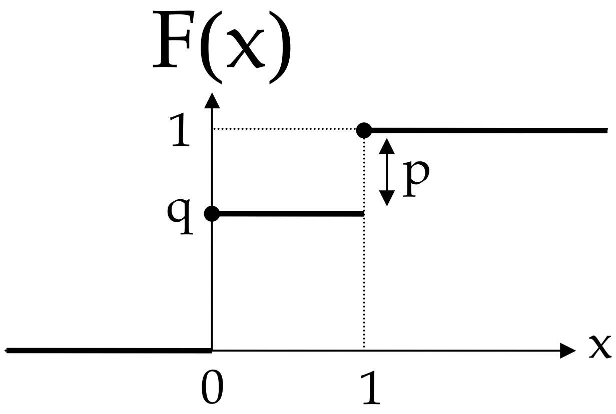
تعریف فوق با این تعریف که مقادیر ممکنه برای $\mathbf{x}$ قابل شمارش هستند معادل است.

توجه کنید که گر چه اگر فضای نمونه (Ω) قابل شمارش باشد، هر $\mathbf{x}$ای روی آن گسسته خواهد بود، ولی روی فضای Ω پیوسته نیز می‌توان $\mathbf{x}$ گسسته تعریف کرد.

مثلاً اگر A واقعه‌ای در فضای نمونهٔ غیرقابل شمارش Ω باشد و تعریف کنیم:

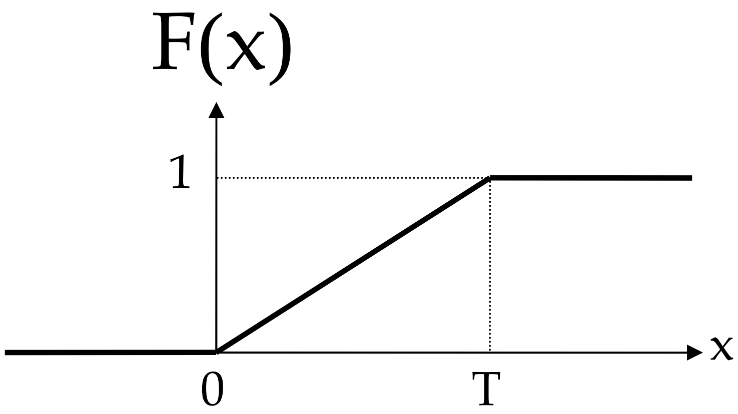
متغیر تصادفی یک-صفر (Zero-One RV) متناظر با واقعهٔ A: $\mathbf{x}(\omega)=\begin{cases}1 & \omega\in A\\ 0 & \omega\notin A\end{cases}$

در این صورت اگر $P(A)=p$ و $P(\bar{A})=1-p=q$ باشد، داریم:



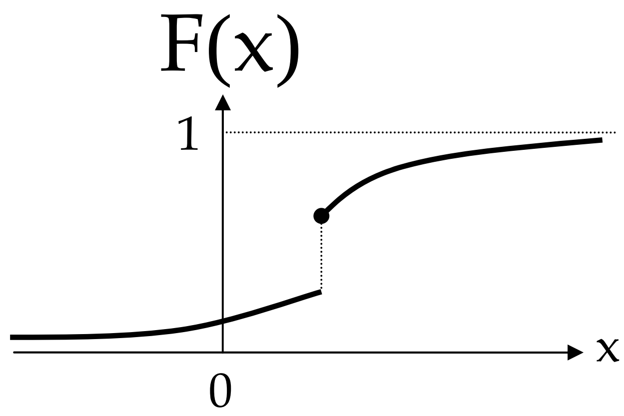
متغیر تصادفی را پیوسته (Continuous) گویند اگر $F(x)$ برای هر x پیوسته باشد.

مانند مثالی که داشتیم (تلفن):

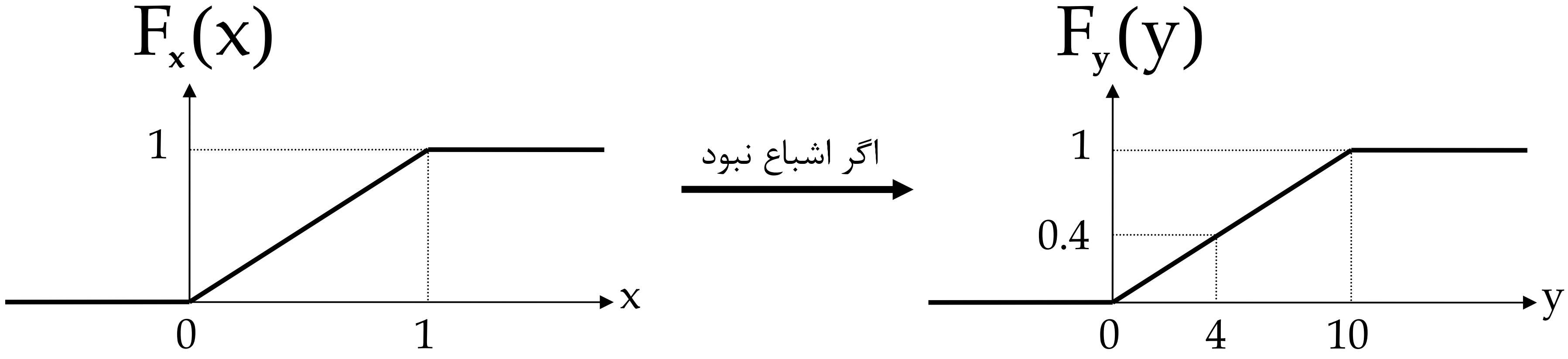


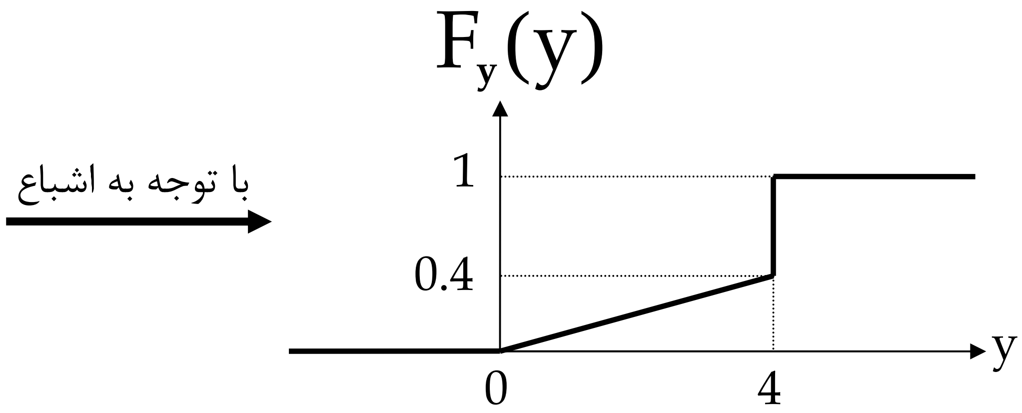
با توجه به ویژگی ۵ برای متغیر تصادفی پیوسته، نتیجه می‌گیریم که برای هر x داریم: $P\{\mathbf{x} = x\} = 0$.

متغیر تصادفی را از نوع مخلوط (Mixed) گویند اگر F(x) دارای ناپیوستگی باشد، ولی پلکانی نباشد.
(تعریفی که قبلاً برای متغیر تصادفی پیوسته کردیم دو نوع اخیر را با هم در برمی‌گرفت، چون در هر دو تعداد مقادیر ممکنه برای **x** غیرقابل‌شمارش است.)



مثال: ولتاژ ورودی تقویت‌کننده‌ای، ولتاژی کاملاً تصادفی بین 0 و 1 (احتمال یکنواخت) است. بهرهٔ تقویت‌کننده ۱۰ است، ولی حداکثر ولتاژ خروجی ۴ V است و برای بیشتر از آن روی ۴ V اشباع می‌شود.





$\{\mathbf{x}=4\}$ احتمالی برابر با ۰/۶ دارد و سایر $\{\mathbf{x}=x\}$ها احتمالشان صفر است.

صدک‌ها (نقاط درصد) (Percentiles):

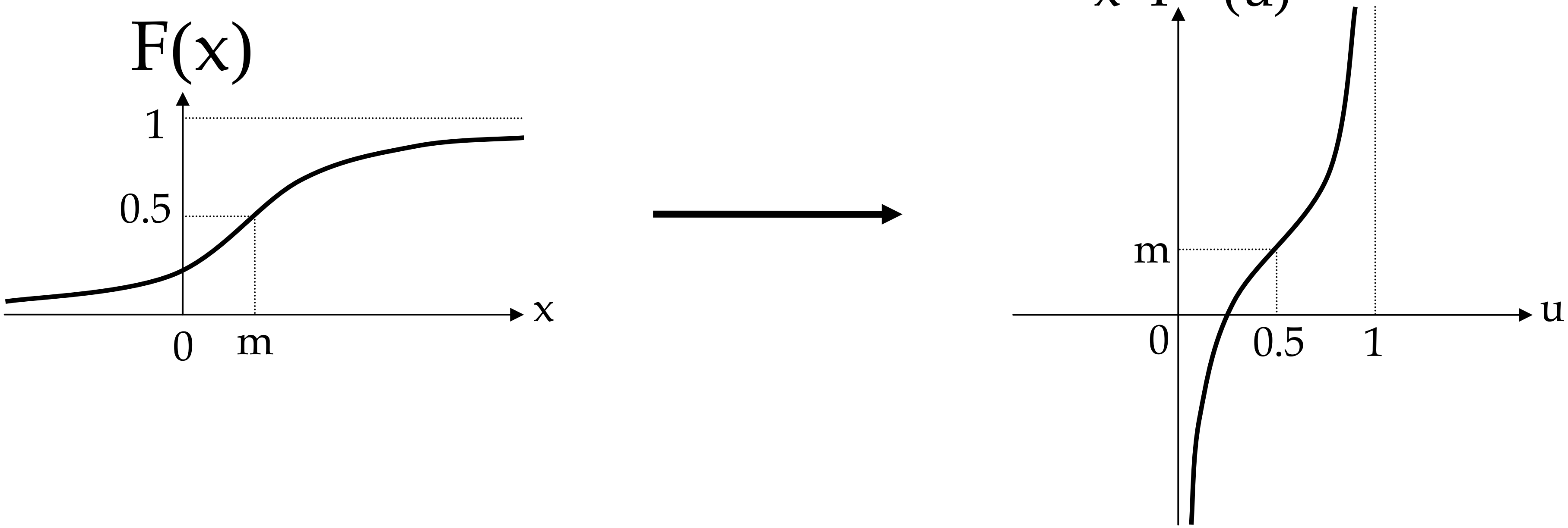
وقتی می‌گوییم $F_{\mathbf{x}}(x) = u$ یعنی به احتمال u متغیر تصادفی $\mathbf{x}$ از مقدار x بزرگتر نمی‌شود (کوچکتر یا مساوی x است). خیلی اوقات با عکس مسأله مواجهیم. مثلاً می‌خواهیم ببینیم متغیر تصادفی $\mathbf{x}$ به احتمال ۰/۹۵ کوچکتر یا مساوی کدام x است؟

$$F(x) = P\{\mathbf{x} \le x\} = u$$

$0 \le u \le 1$ داده شده و $-\infty \le x \le +\infty$ مطلوب است:

تابعی از u: $x_u = F^{-1}(u)$

x_u را صدک u-۱۰۰ام متغیر تصادفی $\mathbf{x}$ گویند. مثلاً $x_{0.95}$، صدک نود و پنجم است.



صدک دهم را دهک (Decile) اول، صدک بیستم را دهک دوم و ... گویند.

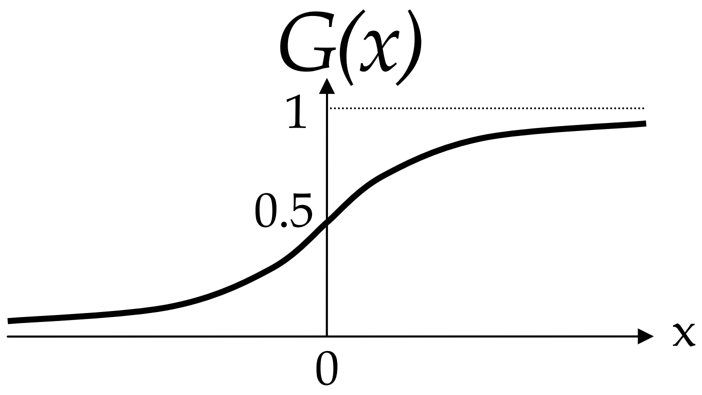
صدک ۲۵ام را چارک (Quartile) اول یا پایین و صدک ۷۵ام را چارک سوم یا بالا گویند.

مهمتر از همه اینکه صدک پنجاهم را چارک دوم یا میانه (Median) گویند.

پس میانهٔ (m) متغیر تصادفی **x** مقداری است که به ازای آن داریم:

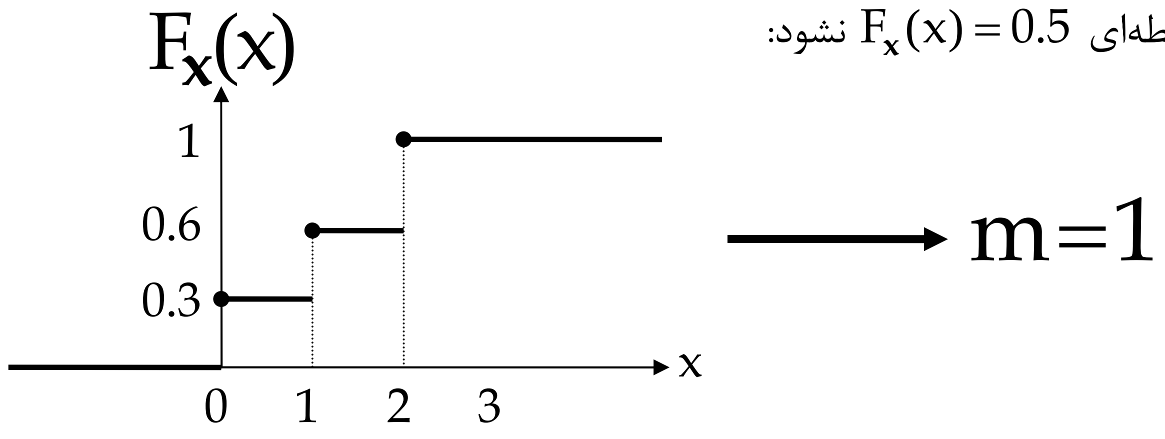
$$F_{\mathbf{x}}(m) = 0.5$$

مثال: $F_{\mathbf{x}}(x) = G(x)$؛

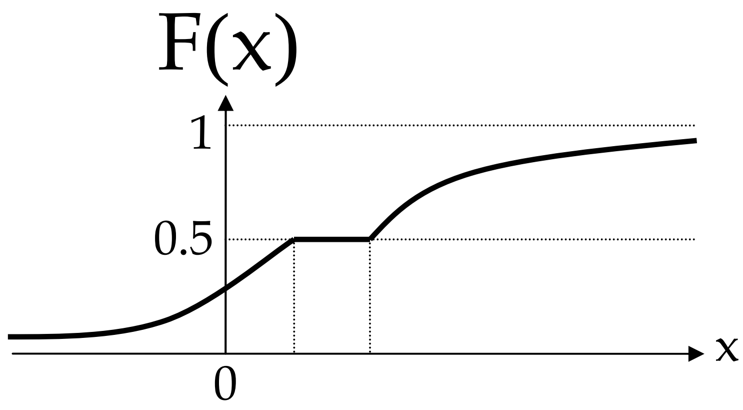


$$\longrightarrow \quad m = 0$$

ولی ممکن است F پرش داشته باشد و لذا در هیچ نقطه‌ای $F_x(x) = 0.5$ نشود:



یا ممکن است یک بازه میانه باشد:

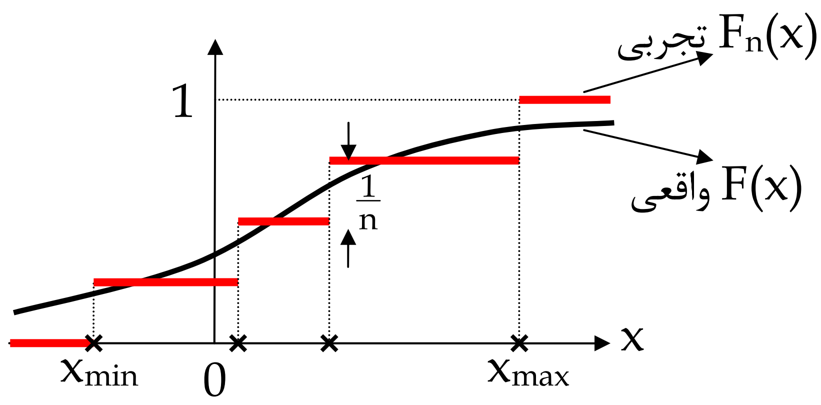


تعریف میانه: برای متغیر تصادفی **x**، میانهٔ m مقداری است که برای آن داشته باشیم:

$$P\{\mathbf{x} \geq m\} \geq \frac{1}{2}, P\{\mathbf{x} \leq m\} \geq \frac{1}{2}$$

به دست آوردن تجربی F(x):

آزمایش تصادفی را nبار تکرار می‌کنیم و مقادیر حاصله برای متغیر تصادفی **x** در این nبار $(x_i : i = 1, 2, \ldots, n)$ را مرتب می‌کنیم.



مثلاً با قرائت مکرر ولتاژ نویزی که توزیع آن را نمی‌دانیم می‌توانیم F(x) تقریبی آن را به دست آوریم:

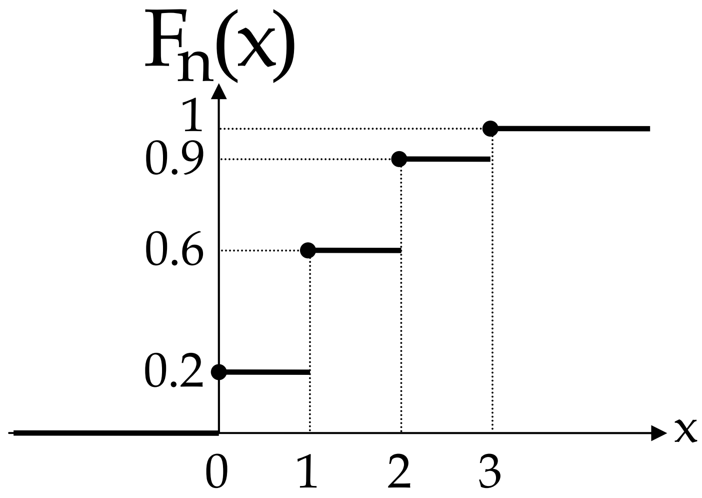
n_x تعداد x_iهایی است که برای آنها $x_i \leq x$ است: $F(x) \simeq F_n(x) = \frac{n_x}{n}$

برای n بزرگ منحنی تجربی به واقعی میل می‌کند.

اگر یک مقدار بیش از یک بار حاصل شود (در مورد متغیر تصادفی گسسته معمولاً این طور است)، پله به جای $\frac{1}{n}$، $\frac{k}{n}$ خواهد بود.

مثلاً در مثال سه سکه که داشتیم اگر در ۱۰ بار تکرار آزمایش، این نتایج حاصل شده باشند، داریم:

تعداد شیرها	صفر	۱	۲	۳
تعداد وقوع	دو بار	چهار بار	سه بار	یک بار



مقادیر واقعی:

$$F(0)=\tfrac{1}{8} \quad F(1)=\tfrac{4}{8} \quad F(2)=\tfrac{7}{8} \quad F(3)=1$$

(به صورت مشابهی منحنی صدک نیز به طور تجربی قابل تخمین است.)

تابع چگالی احتمال (Probability Density Function) pdf:

برای متغیر تصادفی پیوسته راه دیگر برای مشخص کردن احتمال واقعه‌هایی که متغیر تصادفی **x** را تعریف می‌کنند آن است که (مشابه آنچه مستقیماً برای احتمال خود واقعه‌های Ω با تعریف $\alpha(x)$ کردیم) تابعی که بیانگر چگالی احتمال است را داشته باشیم.

(به وسیلهٔ pdf دید بیشتری (در مقایسه با CDF) نسبت به میزان محتمل بودن مقادیر مختلف پیدا می‌کنیم.)

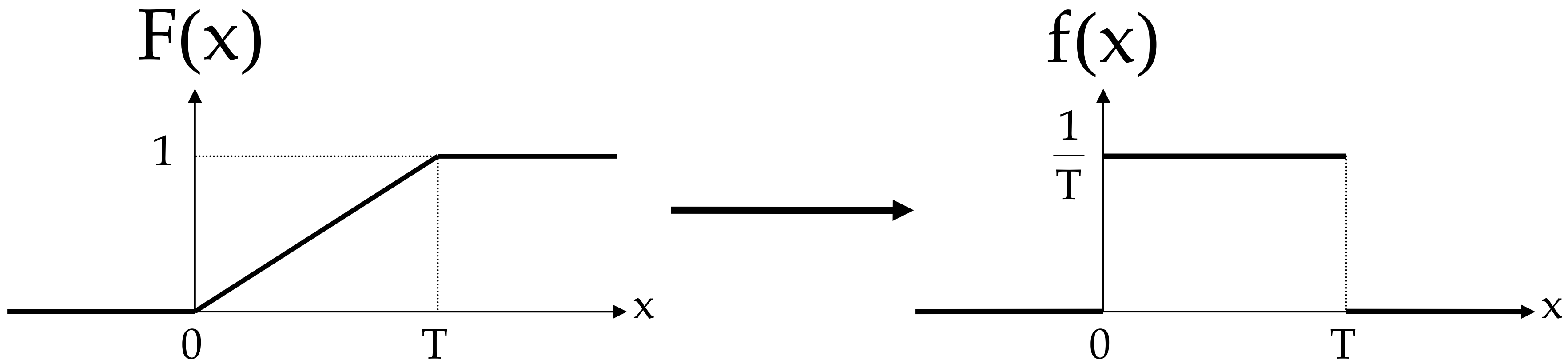
pdf را به صورت زیر تعریف می‌کنیم:

$$f_{\mathbf{x}}(x) = \frac{dF_{\mathbf{x}}(x)}{dx}$$

در کتاب نتاسیون (Notation) آن با تابع احتمال یکی است.

گاهی CDF و گاهی pdf را تابع توزیع (بیشتر CDF) یا توزیع (بیشتر pdf) متغیر تصادفی **x** گویند.

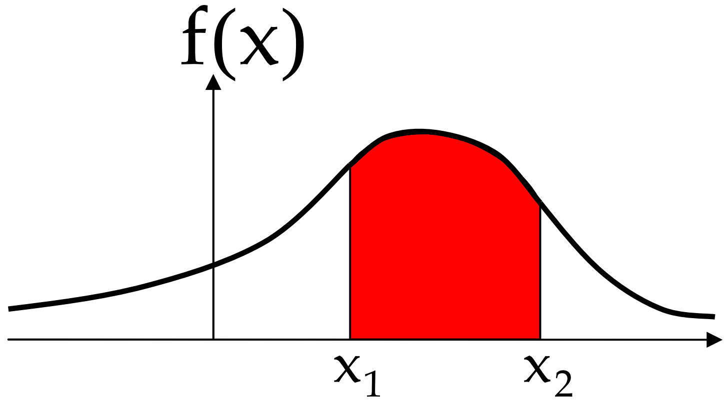
مثال: در مثال مکالمهٔ تلفنی کاملاً تصادفی بین $[0,T]$ دیدیم که:



خواص تابع چگالی احتمال (pdf):

$$1)\ P\{x_1 \le \mathbf{x} \le x_2\} = \int_{x_1}^{x_2} f_{\mathbf{x}}(x)\,dx$$

(یعنی واقعاً تابع چگالی احتمال است)



اثبات: می‌دانیم که:

$$P\{x_1 \le \mathbf{x} \le x_2\} = F(x_2) - F(x_1)$$

و از تعریف pdf داریم:

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)\,dx$$

پس حکم ثابت است (توجه کنید که متغیر تصادفی را پیوسته فرض کردیم).

ضمناً اگر در رابطهٔ فوق $x_1 = -\infty$ قرار دهیم، داریم:

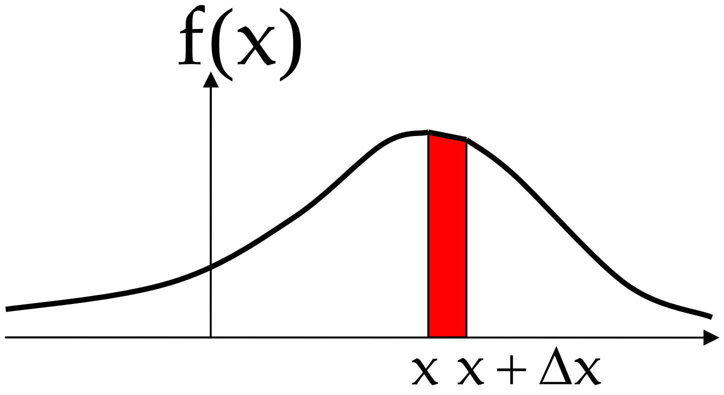
$$F(x) = \int_{-\infty}^{x} f(u)\,du$$

همچنین از خاصیت فوق وقتی $\Delta x \to 0$ داریم:

$$P\{x \le \mathbf{x} \le x+\Delta x\} = f(x)\Delta x$$ (مفهوم چگالی)

یا:

$$f(x) = \lim_{\Delta x \to 0} \frac{P\{x \le \mathbf{x} \le x+\Delta x\}}{\Delta x}$$



(این در واقع $f(x^+)$ است، چون CDF یک متغیر تصادفی پیوسته اگر چه پیوسته است، ولی لزوماً مشتق‌پذیر نیست. ممکن است مشتق راست و چپ برابر نباشند، در این صورت ناپیوستگی در $f(x)$ خواهیم داشت، مانند مثال بالا. در چنین نقاطی که مشتق وجود ندارد هر مقداری را می‌توانید به $f(x)$ نسبت دهید، مثلاً مشتق راست را.)

2) $f(x) \ge 0$

چون $F(x)$ غیرنزولی است.

(f چگالی احتمال مانند P تابع احتمال نامنفی است، ولی f چگالی احتمال می‌تواند بزرگتر از یک هم بشود و حتی می‌تواند به سمت بی‌نهایت برود، مثلاً $f(x) = \frac{2}{3}x^{-\frac{1}{3}}u(x)$)

$$3)\ \int_{-\infty}^{+\infty} f(x)\,dx = 1$$

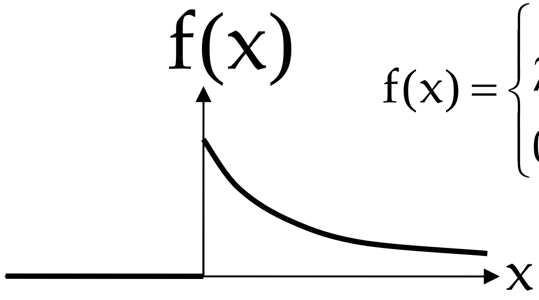
چون $F(+\infty) = 1$ است (مشابه خاصیت $\sum_{i=1}^{\infty} P(x_i) = 1$ برای تابع احتمال).

مثال: مدت زمان کارکرد یک کامپیوتر (بر حسب روز) قبل از خرابی یک متغیر تصادفی پیوسته با تابع چگالی زیر است:

الف) احتمال اینکه کامپیوتر بین ۵۰ تا ۱۵۰ روز کار کند چقدر است؟

ب) احتمال اینکه کمتر از ۱۰۰ روز کار کند چقدر است؟

$$f(x)=\begin{cases}\lambda e^{-\frac{x}{100}} & x\geq 0\\ 0 & x<0\end{cases}$$



الف)

$$\int_{-\infty}^{+\infty} f(x)\,dx=1\Rightarrow \lambda\int_{0}^{+\infty} e^{-\frac{x}{100}}\,dx=1\Rightarrow -\lambda(100)e^{-\frac{x}{100}}\Big|_{0}^{+\infty}=100\lambda=1\Rightarrow \lambda=\frac{1}{100}$$

$$P\{50<\mathbf{x}<150\}=\int_{50}^{150}\frac{1}{100}e^{-\frac{x}{100}}\,dx=-e^{-\frac{x}{100}}\Big|_{50}^{150}=e^{-\frac{1}{2}}-e^{-\frac{3}{2}}=0.384$$

ب)

$$P\{\mathbf{x}<100\}=\int_{0}^{100}\frac{1}{100}e^{-\frac{x}{100}}\,dx=-e^{-\frac{x}{100}}\Big|_{0}^{100}=1-e^{-1}=0.632$$

یعنی در ۶۳٪ اوقات کامپیوتر قبل از اینکه ۱۰۰ روز کار کند از کار می‌افتد.

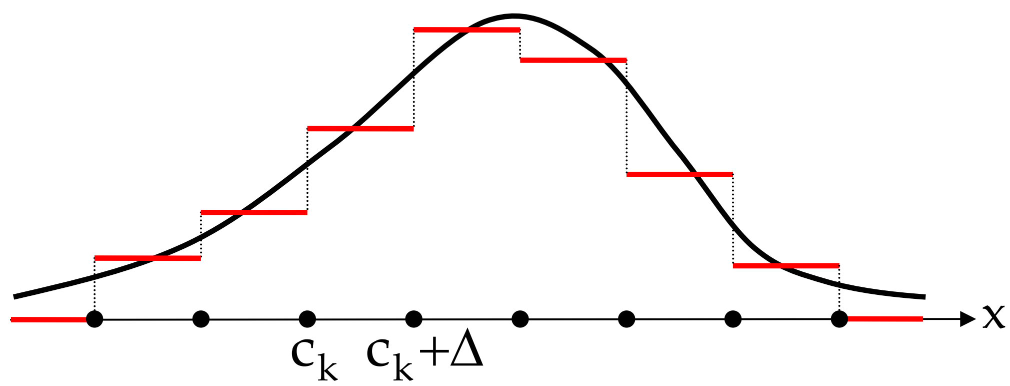
به دست آوردن تجربی $f(x)$:

آزمایش را nبار تکرار می‌کنیم و مقدار متغیر تصادفی $\mathbf{x}$ حاصله از هر آزمایش را ثبت می‌کنیم. محور xها را به فواصلی به طول Δ تقسیم می‌کنیم. اگر تعداد x_iهایی که در فاصلهٔ kام، یعنی $[c_k, c_k+\Delta)$ قرار دارند را با n_k نشان دهیم، داریم:

$$f(x) \simeq f_n(x) \text{ هیستوگرام } = \frac{n_k}{n\Delta} : c_k \le x < c_k + \Delta$$

به این ترتیب:

$$f(c_k)\Delta \simeq P\{c_k \le \mathbf{x} < c_k + \Delta\} \simeq \frac{n_k}{n} = f_n(x)\Delta$$

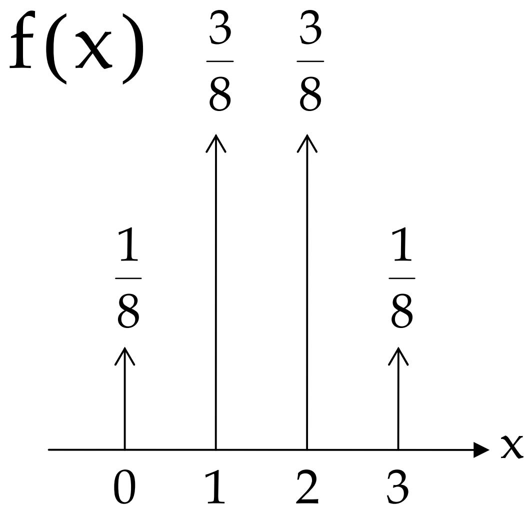


pdf برای متغیرهای تصادفی غیر پیوسته:

در مورد متغیر تصادفی غیرپیوسته یا مخلوط چون CDF دارای ناپیوستگی است، مشتق بی‌مفهوم خواهد بود (بی‌نهایت می شود). ولی با استفاده از تابع (تابع تعمیم یافتهٔ) δ می‌توانیم آن را بیان کنیم.

مثلاً برای متغیر تصادفی گسسته در مثال پرتاب سه سکه داریم:

$$f(x)=\sum_i P_i\,\delta(x-x_i)$$

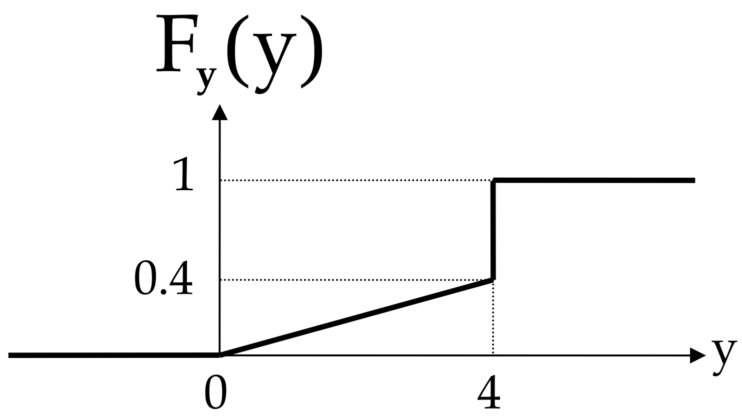


همان طور که داشتیم:

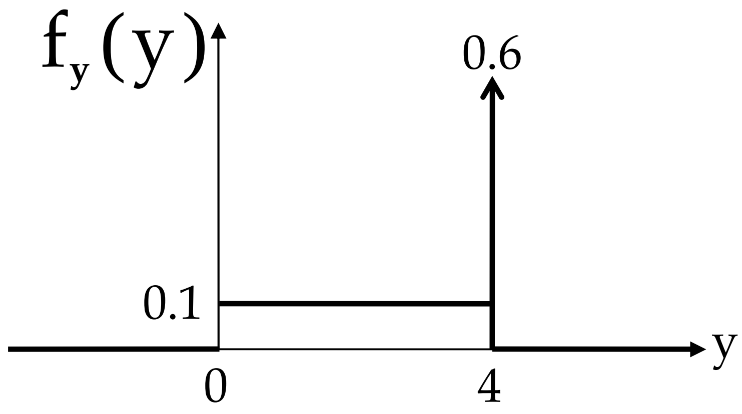
$$F(x)=\sum_i P_i\,u(x-x_i)$$

(یعنی نقاط تابع احتمال به دلتاهایی با آن وزنه‌ها تبدیل شد)

در مثال تقویت‌کننده‌ای که اشباع می شود نیز دیدیم:



لذا:



وقتی اجازه دهیم که $f(x)$ حاوی دلتا باشد، خواهیم داشت:

$$P\{x_1 \le \mathbf{x} \le x_2\} = \int_{x_1^-}^{x_2^+} f(x)dx$$

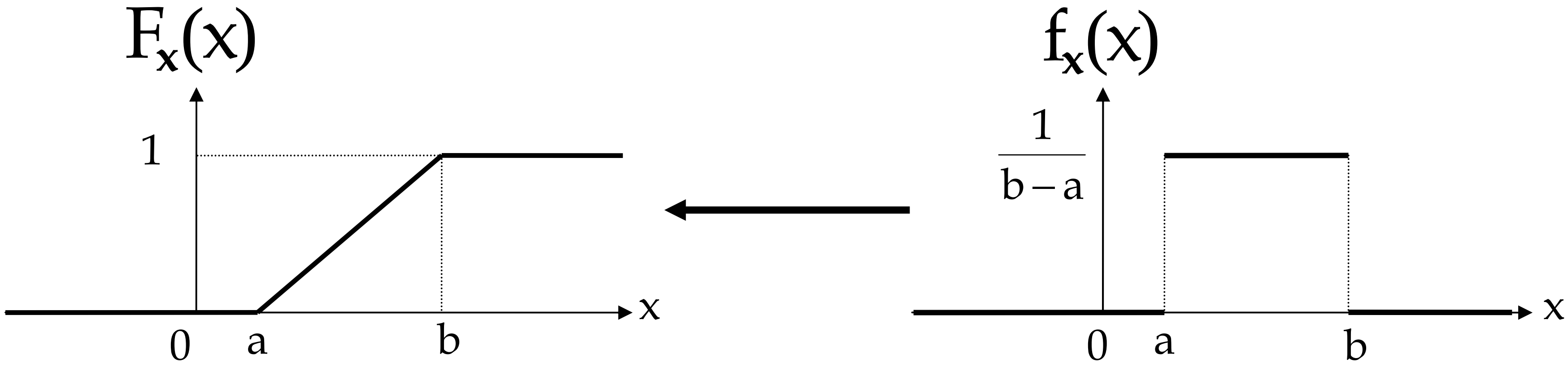
$$P\{x_1 < \mathbf{x} \le x_2\} = \int_{x_1^+}^{x_2^+} f(x)dx$$

برخی توزیع های خاص:

به سادگی می‌توان نشان داد که برای هر تابع غیرنزولی (و از راست پیوسته) که در بی‌نهایت، یک و در منهای بی‌نهایت صفر باشد، متغیر تصادفی‌ای وجود دارد که این تابع، تابع توزیع انباشتهٔ آن باشد (قضیهٔ وجود). ولی برخی توزیعها کاربرد بیشتری دارند.

توزیع یکنواخت (Uniform):

گوییم x دارای توزیع یکنواخت بین a و b است؛ $\mathbf{x} \sim u(a,b)$، اگر:



مثلاً اغلب فاز سیگنال دریافتی، یکنواخت بین 0 و 2π فرض می‌شود. مواقعی که هیچ اطلاع پیشینی در مورد نحوهٔ توزیع نداریم نیز با توجه به اصل ناکافی بودن دلیل از توزیع یکنواخت استفاده می‌شود.

توزیع نرمال (Normal) یا گوسی (Gaussian):

کاربرد بسیار زیادی دارد و در بسیاری موارد عملی تقریب خوبی برای pdf متغیر تصادفی مورد نظر عمل می‌کند. مثلاً در یک جامعهٔ همگن، توزیع قد و وزن افراد نرمال است. توزیع مشخصات قطعات تولیدی یک کارخانه معمولاً نرمال است. طبق **قضیهٔ حد مرکزی** مجموع متغیرهای تصادفی، ولو نرمال نباشند، به نرمال میل می‌کند. نشان داده شده که نویز حرارتی، ؟؟؟ نویز و ... گوسی هستند. از نظر ریاضی نیز کار با آن ساده است.

نرمال استاندارد؛ $\mathbf{x} \sim N(0,1)$، یعنی همان $g(x)$ای که معرفی کردیم:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = g(x)$$

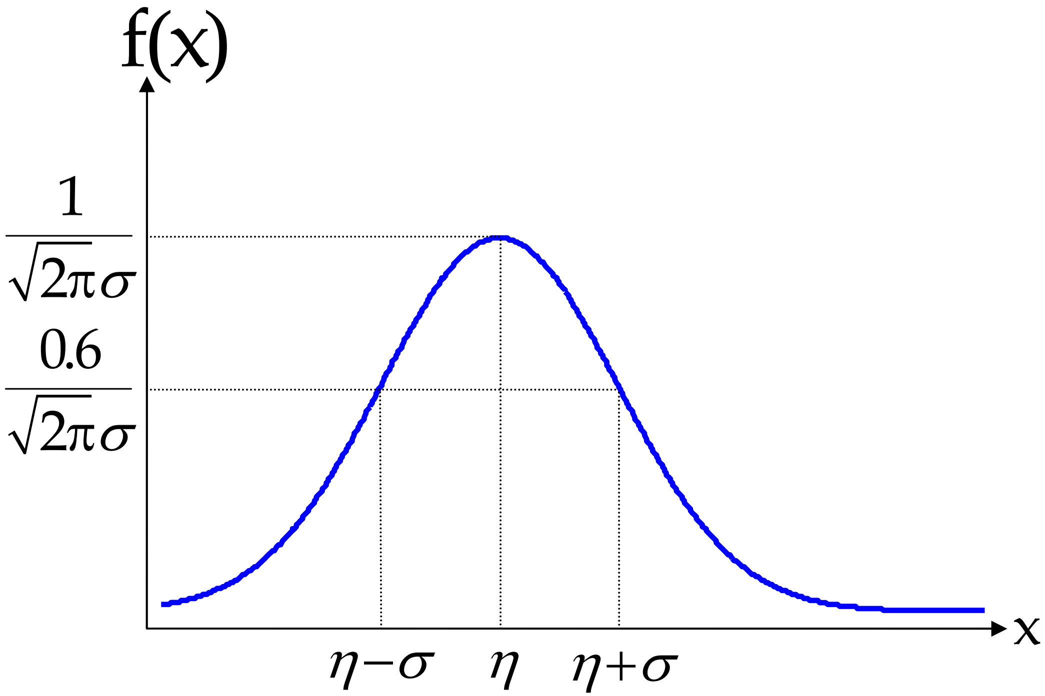
$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}} du = G(x)$$

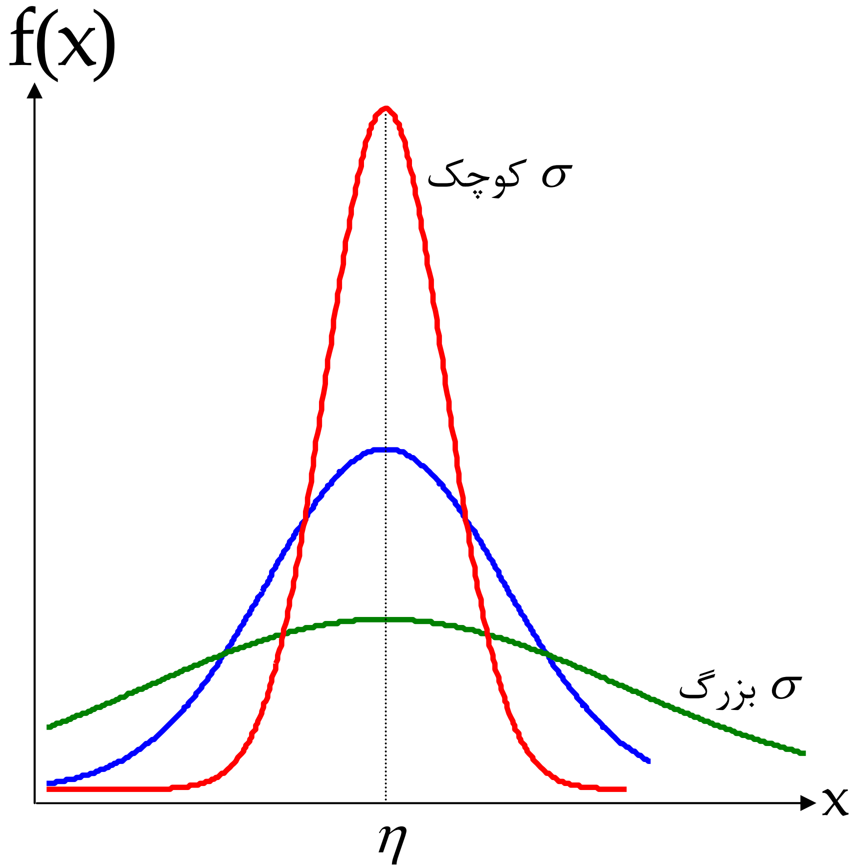
متغیر تصادفی نرمال استاندارد خیلی اوقات با **z** نشان داده می‌شود.

در حالت کلی (برای $\sigma > 0$)، اگر $\mathbf{x} \sim N(\eta,\sigma)$، داریم:

$$f_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\eta)^2}{2\sigma^2}} = \frac{1}{\sigma} g\left(\frac{x-\eta}{\sigma}\right)$$

$$F_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-\frac{(u-\eta)^2}{2\sigma^2}} du = G\left(\frac{x-\eta}{\sigma}\right)$$





اصولاً اگر $\mathbf{x}\sim N(\eta,\sigma)$ باشد، $a\mathbf{x}+b\sim N(a\eta+b,a\sigma)$ خواهد بود و لذا $\dfrac{\mathbf{x}-\eta}{\sigma}\sim N(0,1)$ می‌شود.

$$\mathbf{y}=a\mathbf{x}+b$$

$$F_{\mathbf{y}}(y)=P\{\mathbf{y}\le y\}=P\{a\mathbf{x}+b\le y\}=P\{\mathbf{x}\le \frac{y-b}{a}\}=\int_{-\infty}^{\frac{y-b}{a}}\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\eta)^2}{2\sigma^2}}\,du$$

$$=\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}\,a\sigma}e^{-\frac{[t-(a\eta+b)]^2}{2a^2\sigma^2}}\,dt=\int_{-\infty}^{y}f_{\mathbf{y}}(t)\,dt$$

$$P\{x_1\le\mathbf{x}\le x_2\}=P\{\frac{x_1-\eta}{\sigma}\le\underbrace{\frac{\mathbf{x}-\eta}{\sigma}}_{\mathbf{z}}\le\frac{x_2-\eta}{\sigma}\}=F_{\mathbf{z}}(\frac{x_2-\eta}{\sigma})-F_{\mathbf{z}}(\frac{x_1-\eta}{\sigma})=G(\frac{x_2-\eta}{\sigma})-G(\frac{x_1-\eta}{\sigma})$$

$$P\{\eta-\sigma\le\mathbf{x}\le\eta+\sigma\}\simeq 0.683$$
$$P\{\eta-2\sigma\le\mathbf{x}\le\eta+2\sigma\}\simeq 0.954$$
$$P\{\eta-3\sigma\le\mathbf{x}\le\eta+3\sigma\}\simeq 0.997$$

خوب است به خاطر داشته باشید که در توزیع نرمال، در خیلی از کتابها $G(x)$ با $\Phi(x)$ نشان داده می‌شود:

Kreyszig, Tables A8 and A9.

در کاربردهای مهندسی برق خیلی اوقات $F^c(x) = P\{\mathbf{x} > x\}$ نرمال را لازم داریم:

$$Q(x) = 1 - G(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

بعضاً نیز بر حسب تابع معروف Error Function (تابع خطا) بیان می‌کنند:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-u^2} du$$

$$G(x) = \frac{1}{2} + \frac{1}{2} erf(\frac{x}{\sqrt{2}})$$

مثال: ولتاژ نویزی دارای توزیع نرمال با $\eta = 0$ و $\sigma = 0.75$ است. اگر $\mathbf{x}$ نشان دهندهٔ ولتاژ نویز باشد؛

الف) احتمال اینکه $|\mathbf{x}| \leq 1.5\,\text{Volts}$ باشد چقدر است؟

ب) ولتاژ نویز $\mathbf{x}$ از چه سطح ولتاژی در ٪۹۹ موارد تجاوز نمی‌کند؟

ج) $|\mathbf{x}|$ از چه سطح ولتاژی در ٪۹۹ موارد تجاوز نمی‌کند؟

الف)

$$P\{|\mathbf{x}| \leq 1.5\} = P\{-1.5 \leq \mathbf{x} \leq 1.5\} = G\left(\frac{1.5-0}{0.75}\right) - G\left(\frac{-1.5-0}{0.75}\right) = 2G\left(\frac{1.5}{0.75}\right) - 1 = 0.954$$

چون این همان محدودهٔ $\eta \pm 2\sigma$ است، از قبل هم می‌دانستیم ٪۹۵ می شود.

ب)

$P\{\mathbf{x} \leq x\} = 0.99 \rightarrow$ یعنی دنبال صدک ۹۹ام هستیم

$$G\left(\frac{x-0}{0.75}\right) = 0.99$$

از جدول A9 می‌توانیم G^{-1} را حساب کنیم:

$$\frac{x-0}{0.75} = G^{-1}(0.99) = 2.326 \Rightarrow x = 1.745 :$$ صدک نود و نهم

ج)

$$P\{|\mathbf{x}| \le x\} = 0.99$$

$$2G(\frac{x-0}{0.75}) - 1 = 0.99 \Rightarrow \frac{x}{0.75} = G^{-1}(0.995) \Rightarrow x = 1.932$$

یا از جدول A9 مستقیماً D^{-1} را داریم:

$$D(z) = G(z) - G(-z)$$

$$\frac{x-0}{0.75} = 2.576 \Rightarrow x = 1.932$$

توزیع نمایی (Exponential):

$$f_{\mathbf{x}}(x)=\lambda e^{-\lambda x} : x \geq 0$$

یا می‌نویسیم:

$$f_{\mathbf{x}}(x)=\lambda e^{-\lambda x}u(x) \Rightarrow F_{\mathbf{x}}(x)=(1-e^{-\lambda x})u(x)$$

این توزیع کاربرد بسیار زیادی دارد. مثلاً فاصلهٔ بین دو نقطهٔ تصادفی دارای توزیع نمایی است.

مکالمات تلفنی:

تعداد این نقاط در یک بازهٔ زمانی مشخص دارای توزیع پواسن است، اما فاصلهٔ بین آنها توزیع نمایی دارد.

t می‌تواند مقادیر بین صفر تا بی‌نهایت را اختیار کند.

$$f_{\mathbf{t}}(\tau)d\tau = P\{\tau < \mathbf{t} \leq \tau + d\tau\}$$

t وقتی بین τ و $\tau+d\tau$ خواهد بود که تا قبل از τ مکالمه‌ای اتفاق نیفتاده باشد و بین τ و $\tau+d\tau$ یک مکالمه اتفاق افتد.

$$P\{\tau \text{ در } k\text{نقطه}\} = e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}$$

$$f_t(\tau)d\tau = P\{\tau\ \text{عدم وقوع در فاصلهٔ}\}P\{d\tau\ \text{یک بار وقوع در}\} = e^{-\lambda\tau}\frac{(\lambda\tau)^0}{0!}e^{-\lambda d\tau}\frac{(\lambda d\tau)^1}{1!}$$

$$f_t(\tau)d\tau = \lambda e^{-\lambda(\tau+d\tau)}d\tau$$

$$d\tau \to 0 \Rightarrow f_t(\tau) = \lambda e^{-\lambda\tau} : \tau \geq 0$$

یا به عبارت بهتر:

$$F_t(\tau) = P\{t \leq \tau\} = P\{\tau\ \text{لااقل یک نقطه در فاصلهٔ}\} = P\{k \geq 1\} = 1 - P\{k = 0\} = 1 - e^{-\lambda\tau}\frac{(\lambda\tau)^0}{0!} = 1 - e^{-\lambda\tau} : \tau \geq 0$$

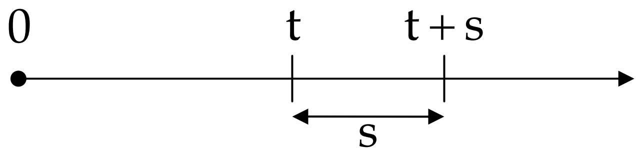
فاصلهٔ بین دو زمین‌لرزهٔ متوالی و مستقل، فاصلهٔ بین دو بار خراب شدن یک دستگاه و ... (با فرض استقلال از هم و متساوی الاحتمال بودن در تمام زمانها) توزیع نمایی دارند.

لذا این توزیع برای تصادف اتومبیل و طوفانهای خورشیدی قابل استفاده نیست.

دیده شد که طول مکالمهٔ تلفنی هم توزیع نمایی دارد.

از خواص جالب توزیع نمایی این است که بدون حافظه است، یعنی:

$$P\{\mathbf{x} > t+s \mid \mathbf{x} > t\} = P\{\mathbf{x} > s\}$$



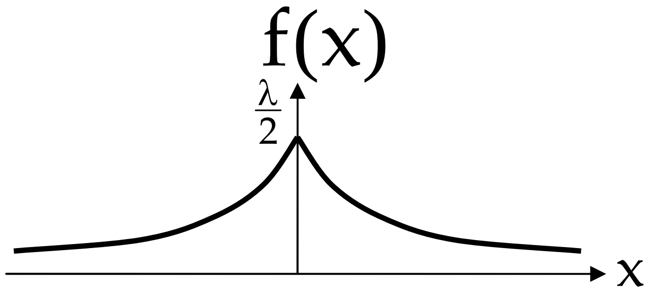
زیرا:

$$P\{\mathbf{x} > t+s \mid \mathbf{x} > t\} = \frac{1-F_{\mathbf{x}}(t+s)}{1-F_{\mathbf{x}}(t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = 1-F_{\mathbf{x}}(s) = P\{\mathbf{x} > s\}$$

می توان نشان داد که تنها تابع توزیعی که دارای این خاصیت می‌باشد تابع نمایی است.

توزیع لاپلاس (Laplace):

$$f(x)=\frac{\lambda}{2}e^{-\lambda|x|}$$



این توزیع در مورد نویزهای اسپایکی مدل خوبی است.

توزیع گاما (Gamma):

$$f(x) = A\,x^{r-1}\,e^{-\lambda x}\,u(x) : r > 0, \lambda > 0$$

از این توزیع نیز درکاربردهای مختلفی استفاده می‌شود.

ضریب ثابت A را باید طوری به دست آوریم که:

$$\int_{-\infty}^{+\infty} f(x)\,dx = \int_{0}^{+\infty} Ax^{r-1}e^{-\lambda x}\,dx = 1$$

$$y = \lambda x \Rightarrow \frac{A}{\lambda^r}\int_0^{+\infty} y^{r-1}e^{-y}\,dy = 1$$

از طرفی می‌دانیم که **تابع گاما** برابر است با:

$$\Gamma(r) = \int_0^{+\infty} y^{r-1}e^{-y}\,dy$$

در واقع $\Gamma(\bullet)$، فاکتوریل تعمیم یافته است. چون $\Gamma(r) = (r-1)\Gamma(r-1)$، برای r صحیح و مثبت داریم: $\Gamma(r) = (r-1)!$.

$$\frac{A}{\lambda^r}\Gamma(r)=1 \Rightarrow A=\frac{\lambda^r}{\Gamma(r)}$$

$$f_{\mathbf{x}}(x)=\frac{\lambda(\lambda x)^{r-1}e^{-\lambda x}}{\Gamma(r)}u(x)$$

برای نشان دادن این توزیع می‌نویسیم $\mathbf{x} \sim \mathrm{Gamma}(r,\lambda)$ است.

شکل صفحهٔ ۸۵ کتاب Helstrom برای $\lambda=1$ و rهای مختلف.

حالت خاص: اگر $r=1$ باشد، داریم:

توزیع نمایی: $f(x)=\lambda e^{-\lambda x}u(x)$

حالت خاص: اگر $\lambda=\frac{1}{2}$ و $r=\frac{n}{2}$ که n عددی صحیح است، داریم:

$$f(x)=\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}x^{(\frac{n}{2}-1)}e^{-\frac{x}{2}}u(x)$$

حالت خاص: اگر $r=n$ که n عددی صحیح است، داریم:

$$f(x)=\frac{\lambda^n}{(n-1)!}x^{n-1}e^{-\lambda x}u(x)$$

این توزیع مدت زمان لازم برای وقوع n واقعهٔ کاملاً تصادفی (حاصل جمع n نمایی مستقل از هم) است که **توزیع ارلانگ (Erlang)** نام دارد و در تئوری صف و ترافیک (مکالمات تلفنی)، تشعشع رادیواکتیو و ... استفاده می‌شود.

توزیع χ^2 (Chi-Square) با n درجه آزادی (n degrees of freedom):

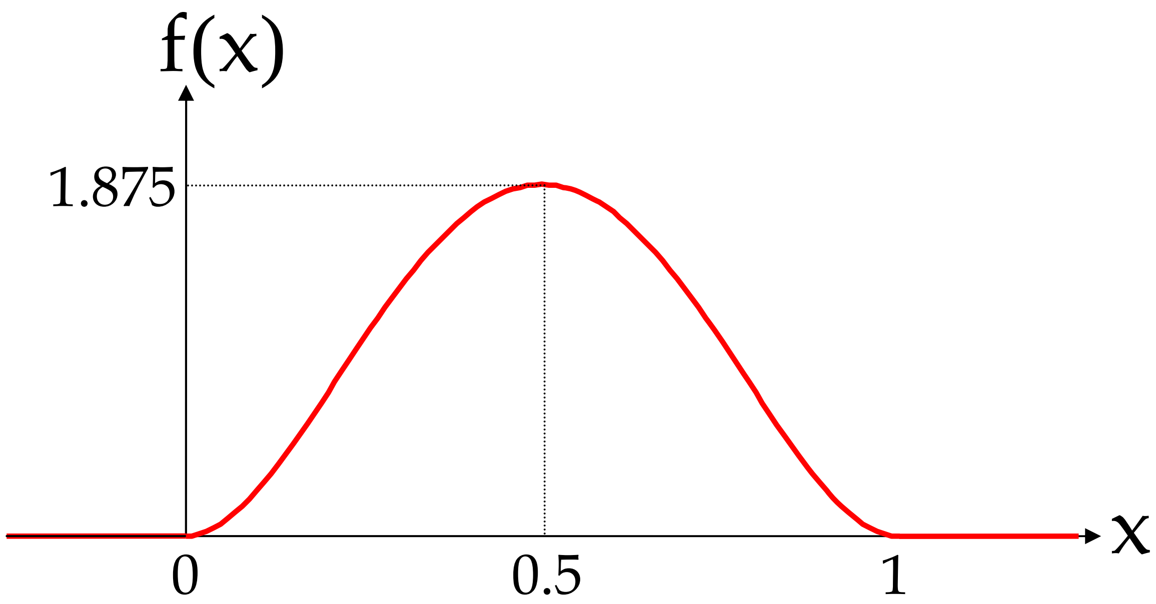
اگر $\mathbf{y} = \mathbf{x}_1^2 + \mathbf{x}_2^2 + \cdots + \mathbf{x}_n^2$ که $\mathbf{x}_i$ها $N(0,1)$ بوده و مستقل باشند، $\mathbf{y}$ دارای توزیع $\chi^2(n)$ خواهد بود. این توزیع در مهندسی برق و آمار استفادهٔ زیادی دارد.

توزیع بتا (Beta):

$$f(x)=\begin{cases}\dfrac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1} & 0\le x\le 1\\ 0 & \text{otherwise}\end{cases}$$

اگر $a=b=1$ باشد، همان توزیع یکنواخت روی $[0,1]$ خواهد بود.

برای $a=b=3$، نمودار توزیع به این صورت خواهد بود:



یادآوری: تابع بتا برابر است با:

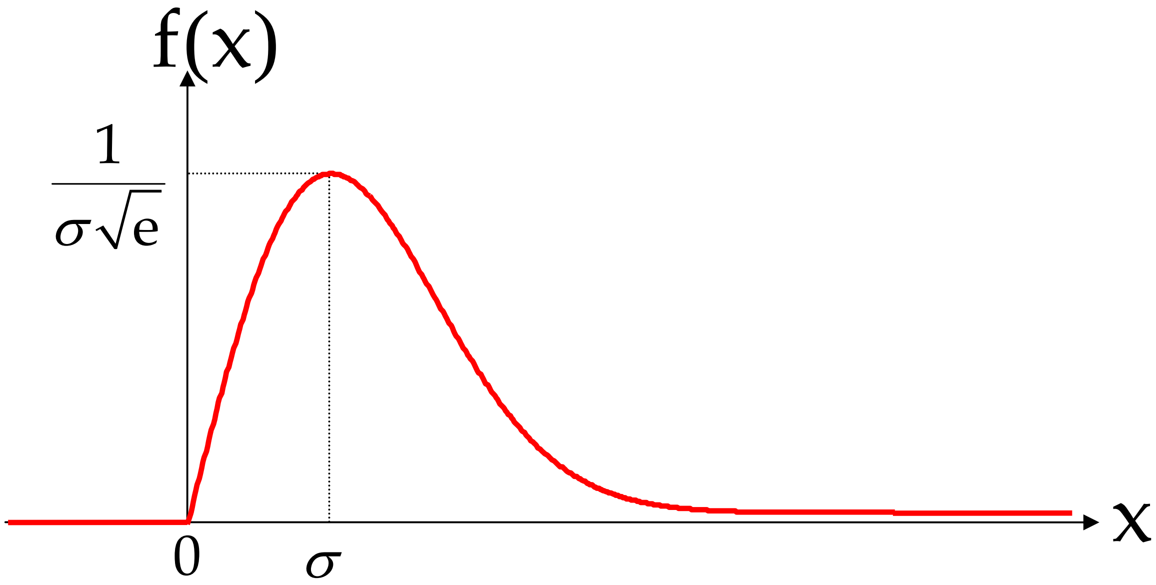
$$\beta(a,b)=\int_0^1 x^{a-1}(1-x)^{b-1}dx=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

پس:

$$f(x)=\begin{cases}\dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}x^{a-1}(1-x)^{b-1} & 0\le x\le 1\\ 0 & \text{otherwise}\end{cases}$$

توزیع رایلی (Rayleigh):

$$f(x)=\frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}u(x)$$

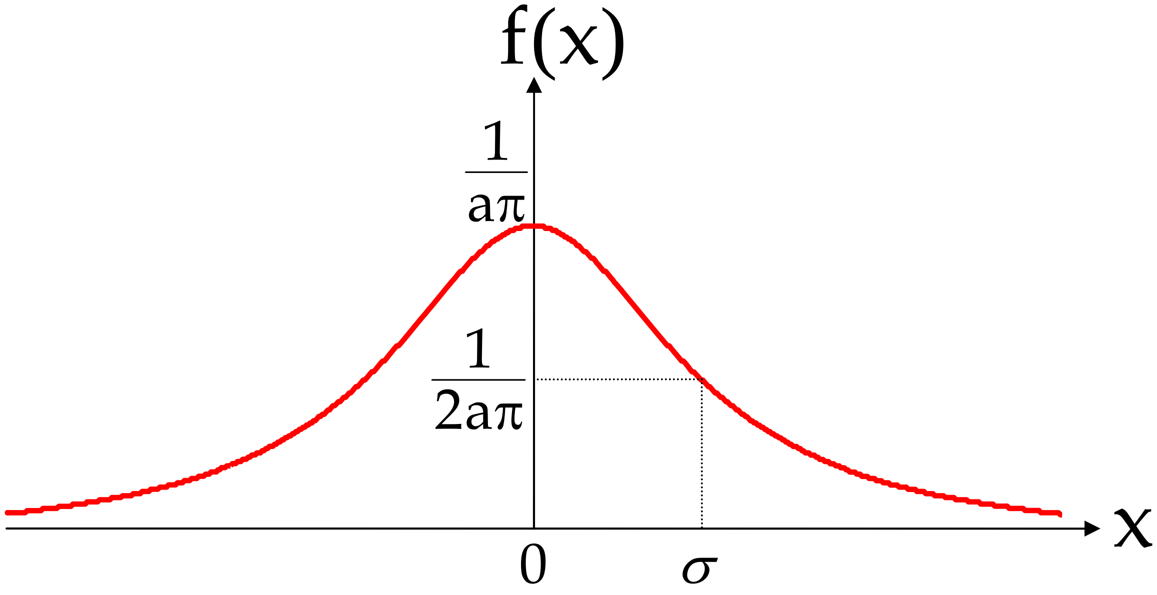


اگر فرض کنیم $\mathbf{y}=\sqrt{\mathbf{x}_1^2+\mathbf{x}_2^2}$ باشد و $\mathbf{x}_1$ و $\mathbf{x}_2$ مستقل و دارای توزیع نرمال $N(0,\sigma)$ باشند، $\mathbf{y}$ دارای توزیع رایلی خواهد بود.
مثلاً اگر $\mathbf{z}=\mathbf{x}+j\mathbf{y}$ باشد، $|\mathbf{z}|$ دارای توزیع رایلی خواهد بود.

این توزیع نیز کاربرد بسیار زیادی در مهندسی برق دارد، مانند پوش نویز گوسی.

توزیع کوشی (Cauchy):

$$f(x)=\frac{\frac{a}{\pi}}{x^2+a^2}$$



اگر θ دارای توزیع یکنواخت $[-\frac{\pi}{2},\frac{\pi}{2}]$ باشد، $a\,tg\theta$ دارای این توزیع خواهد بود.

توزیع دوجمله‌ای:

این یک توزیع گسسته است. اگر در یک آزمایش تصادفی احتمال وقوع واقعهٔ A برابر $P(A)=p$ باشد و برای آزمایش حاصل از nبار تکرار این آزمایش، متغیر تصادفی $\mathbf{x}$ را به این صورت تعریف کنیم که: $\mathbf{x}$ = تعداد وقوع واقعهٔ A در n آزمایش، $\mathbf{x}$ دارای توزیع دوجمله‌ای خواهد بود:

$$\mathbf{x} \sim \text{Binomial}(n,p)$$

$$P(k)=P\{\mathbf{x}=k\}=\binom{n}{k}p^k q^{n-k} : k=0,1,2,\ldots,n$$

و برای سایر مقادیر x داریم: $P(x)=0$.

مثال: در جعبه‌ای N دیود قرار دارند که $K \le N$تا از آنها خرابند. N نمونه با جایگزینی برداشته می‌شود. اگر متغیر تصادفی $\mathbf{x}$ به صورت: $\mathbf{x}$ = تعداد خرابها، تعریف شود، $P(x)$ را به دست آورید.

$$P(x)=\begin{cases}\binom{n}{x}(\frac{K}{N})^x(1-\frac{K}{N})^{n-x} & x=0,1,2,\ldots,n\\ 0 & \text{otherwise}\end{cases}$$

حالت خاص: اگر $n=1$ باشد، یعنی آزمایش تصادفی را فقط یک بار انجام می‌دهیم. پس **x** فقط می‌تواند صفر یا یک باشد (صفر با احتمال q و یک با احتمال p)، پس داریم:

$$P(k)=p^k q^{1-k} : k=0,1$$

یعنی $P(0)=q$ و $P(1)=p$ می‌باشد.

این حالت خاص را **توزیع برنولی** گویند.

با توجه به قضیهٔ لاپلاس، برای n بزرگ توزیع (گسستهٔ) دوجمله‌ای را می‌توان با توزیع (پیوستهٔ) نرمال تقریب زد:

$$P(x)=\binom{n}{k}p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}}e^{-\frac{(k-np)^2}{2npq}}=N(np,\sqrt{npq})$$

$$F(x)\simeq G(\frac{x-np}{\sqrt{npq}})$$

و با توجه به قضیهٔ پواسون، برای n بزرگ و p کوچک می‌توان توسط توزیع پواسون تقریب زد:

$$\binom{n}{k}p^k q^{n-k} \simeq e^{-np}\frac{(np)^k}{k!}=\text{Poisson}(np)$$

توزیع فوق‌هندسی (Hypergeometric):

در مثال قبل اگر انتخاب بدون جایگزینی باشد، داریم:

$$P(x)=\frac{\text{تعداد انتخابهای مورد نظر}}{\text{تعداد کل انتخابها}}=\begin{cases}\dfrac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} & \max(0,n-N+K)\le k\le \min(n,K)\\ 0 & \text{otherwise}\end{cases}$$

$\max(0,n-N+K)$ بدین خاطر است که تعداد سالم‌های انتخاب‌شده n-k بوده و نمی‌تواند از تعداد کل سالم‌ها که N-K است، بیشتر شود، پس: $n-k\le N-K$؛

$\min(n,K)$ نیز بدین خاطر است که تعداد خرابهای انتخاب‌شده نمی‌تواند از تعداد کل خرابها بیشتر شود.

یا اگر بگیریم: $p=\frac{K}{N}$، خواهیم داشت:

$$P(k)=\frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}}$$

توزیع دوجمله‌ای منفی (پاسکال):

فرض کنید با مسئلهٔ زیر مواجهیم. سکه‌ای را آنقدر پرتاب می‌کنیم که rبار شیر بیاید. اگر تعداد کل پرتابها را $\mathbf{y}$ بنامیم، احتمال $\mathbf{y}=n$ را به دست آورید $(n=r, r+1, \dots)$.

در حالت کلی اگر یک آزمایش برنولی را آنقدر تکرار کنیم تا واقعهٔ A که $P(A)=p$، rبار اتفاق افتد و متغیر تصادفی $\mathbf{y}$ به صورت:
$\mathbf{y}$ = تعداد کل آزمایشهای انجام شده تا rبار وقوع A تعریف شود، داریم:

احتمال rامین موفقیت که حتماً باید در nامین آزمایش باشد

$$P\{\mathbf{y}=n\}=p\underbrace{\binom{n-1}{k-1}p^{r-1}q^{(n-1)-(r-1)}}_{\text{احتمال r-1 موفقیت در میان بقیهٔ آزمایشها (n-1 آزمایش)}}$$

یعنی:

$$P\{\mathbf{y}=n\}=\binom{n-1}{r-1}p^{r}q^{n-r} : n=r, r+1, r+2, \dots$$

در بعضی کتابها این را توزیع دوجمله‌ای منفی می‌گویند.

اما اگر فرض کنیم: $\mathbf{y}=\mathbf{x}+r$، که $\mathbf{x}$ تعداد شکستها و $\mathbf{y}$ تعداد کل آزمایشها است، واقعهٔ $\mathbf{x}=k$ با واقعهٔ $\mathbf{y}=k+r$ معادل بوده و احتمال یکسانی دارد، یعنی:

$$P\{\mathbf{x}=k\}=\binom{n-1}{r-1}p^r q^{n-r} \overset{n=k+r}{=} \binom{r+k-1}{r-1}p^r q^k=\binom{r+k-1}{k}p^r q^k : k=0,1,2,\dots$$

در تمرین‌ها داشتیم که:

$$\binom{r+k-1}{k}=(-1)^k\binom{-r}{k}$$

لذا:

$$P\{\mathbf{x}=k\}=P_{\mathbf{x}}(k)=\binom{-r}{k}p^r(-q)^k : k=0,1,2,\dots$$

این را **توزیع دوجمله‌ای منفی** گویند.

توزیع هندسی (Geometric):

حالت خاص توزیع دوجمله‌ای منفی است که در آن $r=1$ باشد، یعنی:

$$P\{\mathbf{x}=k\}=P_{\mathbf{x}}(k)=pq^{k}:k=0,1,2,\dots$$

در اینجا، $\mathbf{x}$، تعداد شکستها در تکرار آزمایش تا حصول اولین موفقیت است.

این توزیع در واقع یک دنبالۀ هندسی است و داریم:

$$\sum_{k=0}^{+\infty}pq^{k}=\frac{p}{1-q}=1$$

در بعضی کتابها حالت $r=1$ از آنچه ما با $P(\mathbf{y})$ نشان دادیم را توزیع هندسی گویند. برای $r=1$ داریم:

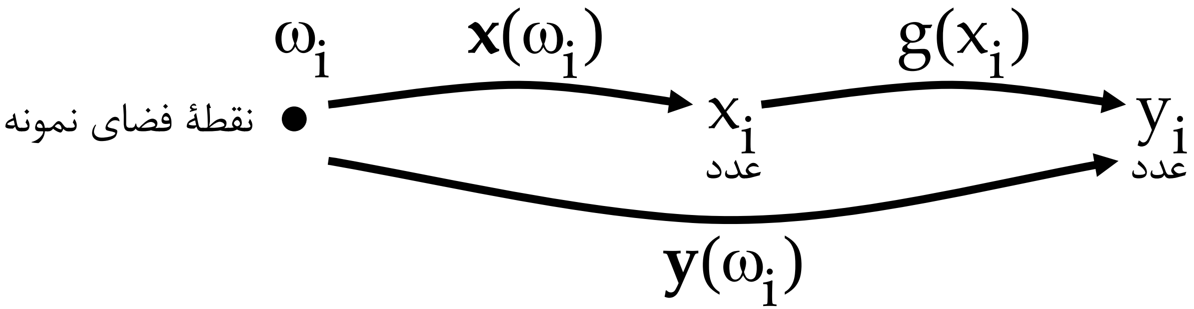
$$P\{\mathbf{y}=n\}=pq^{n-1}:n=1,2,3,\dots$$

در اینجا، $\mathbf{y}$، تعداد کل آزمایشها تا حصول اولین موفقیت است.

تابع یک متغیر تصادفی:

در بسیاری موارد با فرض دانستن تابع توزیع یک متغیر تصادفی، با توجه به اینکه در سیستم مورد نظر عملیاتی روی آن انجام می‌شود، علاقه‌مندیم که توزیع تابعی از یک متغیر تصادفی را به دست آوریم. مانند آشکارساز دیودی. یا مثلاً مایلیم بدانیم که توزیع توان مصرفی مقاومتی که ولتاژ دو سر آن تصادفی است، چیست. یا اگر فاز سیگنال تصادفی باشد و $\mathbf{y} = A\cos(\omega t + \varphi)$، توزیع $\mathbf{y}$ چیست و امثالهم...

اگر $g(x)$ یک تابع حقیقی باشد، می‌توانیم به هر $x_i = \mathbf{x}(\omega_i)$، $y_i = g(x_i)$ را نسبت دهیم:



پس با ترکیب دو تابع مواجهیم که یک متغیر تصادفی جدید به ما می‌دهد:

$$\mathbf{y}(\omega) = g(\mathbf{x}(\omega))$$

متغیر تصادفی $\mathbf{y}$ را تابعی از متغیر تصادفی $\mathbf{x}$ گوییم.

حال ببینیم با دانستن توزیع $\mathbf{x}$ چگونه می‌توانیم توزیع $\mathbf{y}$ را به دست آوریم.

در حالت گسسته مسئله بسیار ساده است:

$$P_{\mathbf{y}}(y) = P\{\mathbf{y} = y\} = P\{g(\mathbf{x}) = y\} = \sum_{i:g(x_i)=y} P(x_i)$$

$$F_{\mathbf{y}}(y) = P\{\mathbf{y} \le y\} = P\{g(\mathbf{x}) \le y\} = \sum_{i:g(x_i)\le y} P(x_i)$$

مثلاً اگر متغیر تصادفی **x** مقادیر زیر را اختیار کند:

x_i	0	1	-1
P_i	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

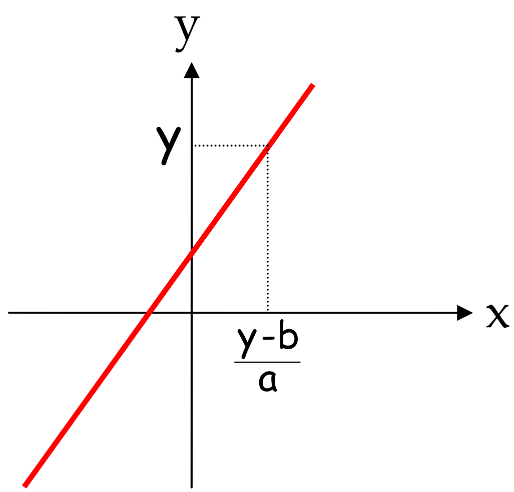
برای $\mathbf{y} = \mathbf{x}^2$ داریم:

$$P_{\mathbf{y}}(1) = P\{\mathbf{y} = 1\} = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}$$

و به سادگی می‌توانیم بقیهٔ احتمالها را نیز حساب کنیم.

اما برای حالتهای غیرگسسته ابتدا به محاسبهٔ تابع CDF می‌پردازیم که ساده‌تر است.

مثال ۱: $g(x) = ax + b$:



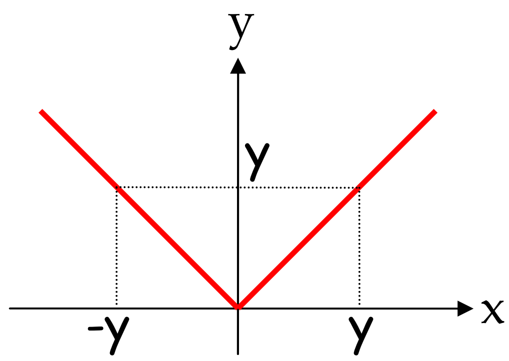
$$\mathbf{y} = a\mathbf{x} + b$$

$$F_{\mathbf{y}}(y) = P\{\mathbf{y} \le y\} = P\{a\mathbf{x} + b \le y\} = \begin{cases} P\{\mathbf{x} \le \frac{y-b}{a}\} = \overbrace{F_{\mathbf{x}}(\frac{y-b}{a})}^{\text{یعنی شیفت و scale}} & a > 0 \\ P\{\mathbf{x} \ge \frac{y-b}{a}\} = 1 - F_{\mathbf{x}}(\frac{y-b}{a}) & a < 0 \end{cases}$$

از اینجا داریم:

$$f_{\mathbf{y}}(y) = \begin{cases} \frac{1}{a} f_{\mathbf{x}}(\frac{y-b}{a}) & a > 0 \\ -\frac{1}{a} f_{\mathbf{x}}(\frac{y-b}{a}) & a < 0 \end{cases} = \frac{1}{|a|} f_{\mathbf{x}}(\frac{y-b}{a})$$

مثال ۲ (**یکسوساز تمام موج**): $g(x)=|x|$:



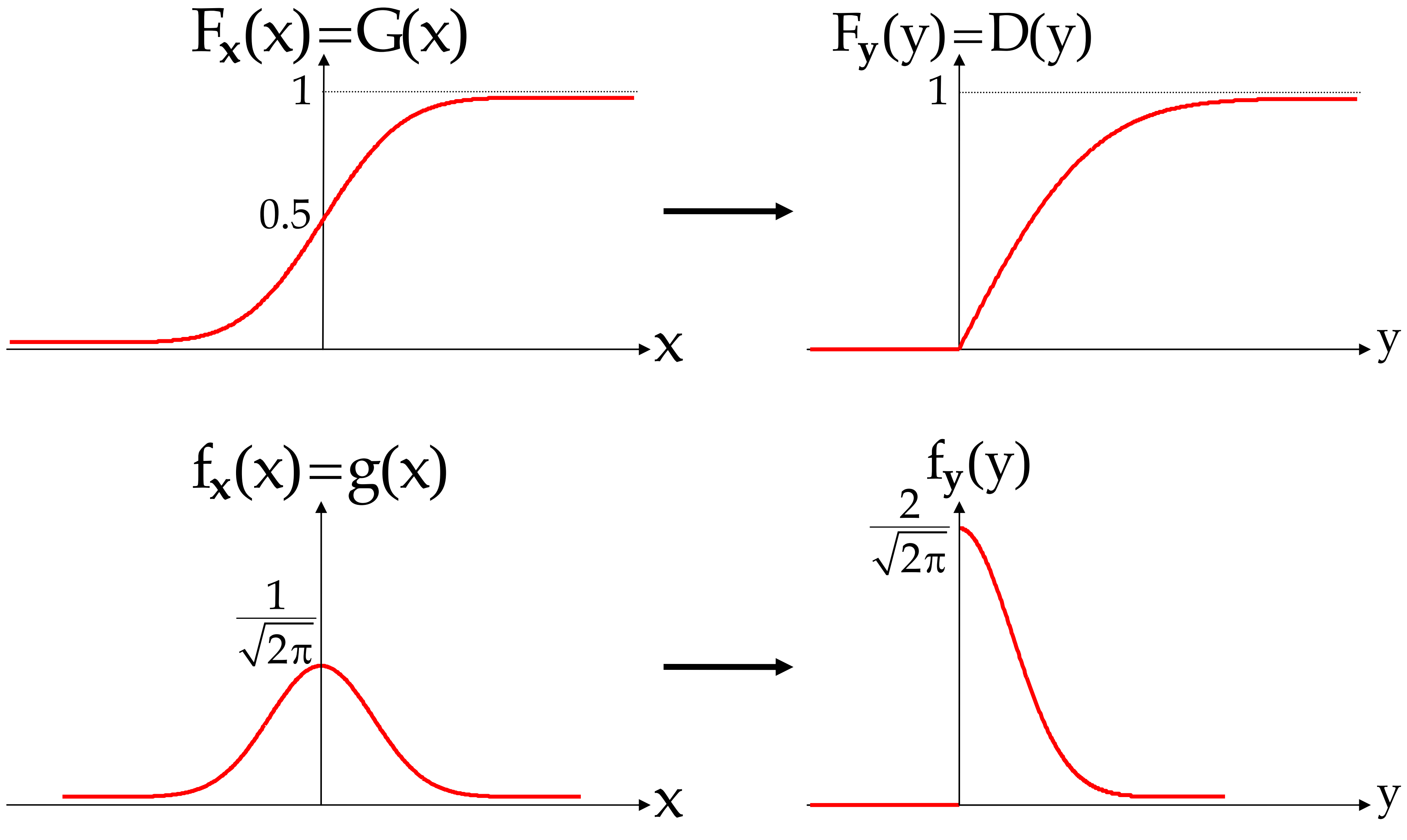
$$\mathbf{y}=|\mathbf{x}|$$

$$F_{\mathbf{y}}(y)=P\{\mathbf{y}\le y\}=P\{|\mathbf{x}|\le y\}=\begin{cases}0 & y<0\\ P\{-y\le \mathbf{x}\le y\}=F_{\mathbf{x}}(y)-F_{\mathbf{x}}(-y) & y\ge 0\end{cases}$$

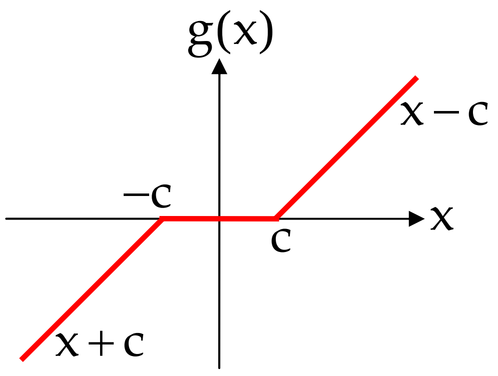
با مشتق‌گیری داریم:

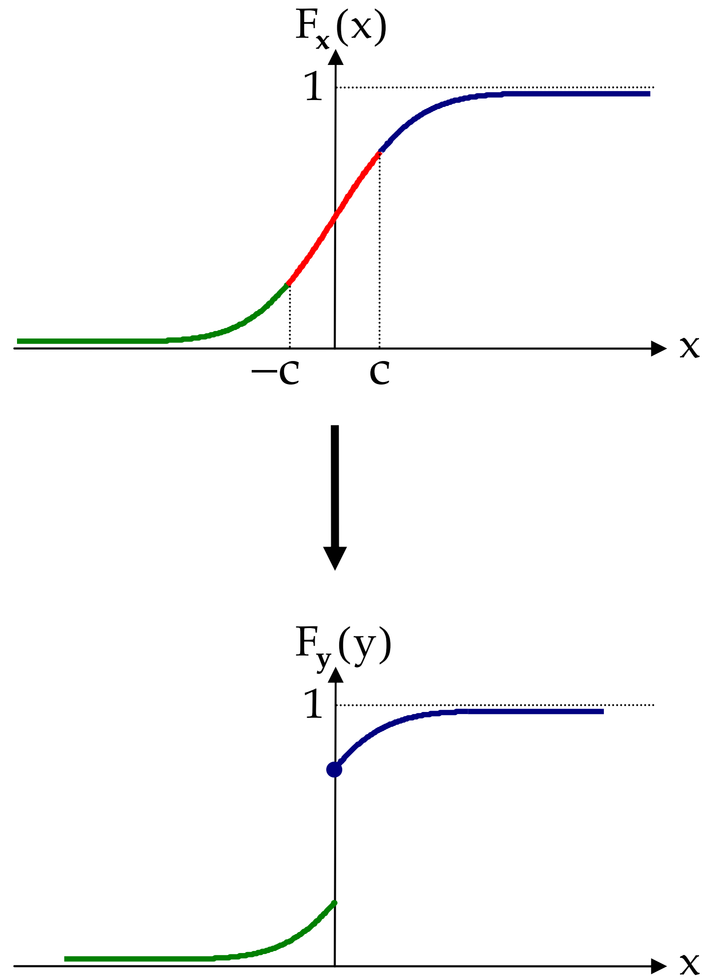
$$f_{\mathbf{y}}(y)=\begin{cases}f_{\mathbf{x}}(y)+f_{\mathbf{x}}(-y) & y\ge 0\\ 0 & y<0\end{cases}$$

مثلاً اگر $\mathbf{x} \sim N(0,1)$ باشد، خواهیم داشت:



مثال ۳: $g(x)=\begin{cases} x+c & x<-c \\ 0 & -c\le x\le c \\ x-c & x>c \end{cases}$:



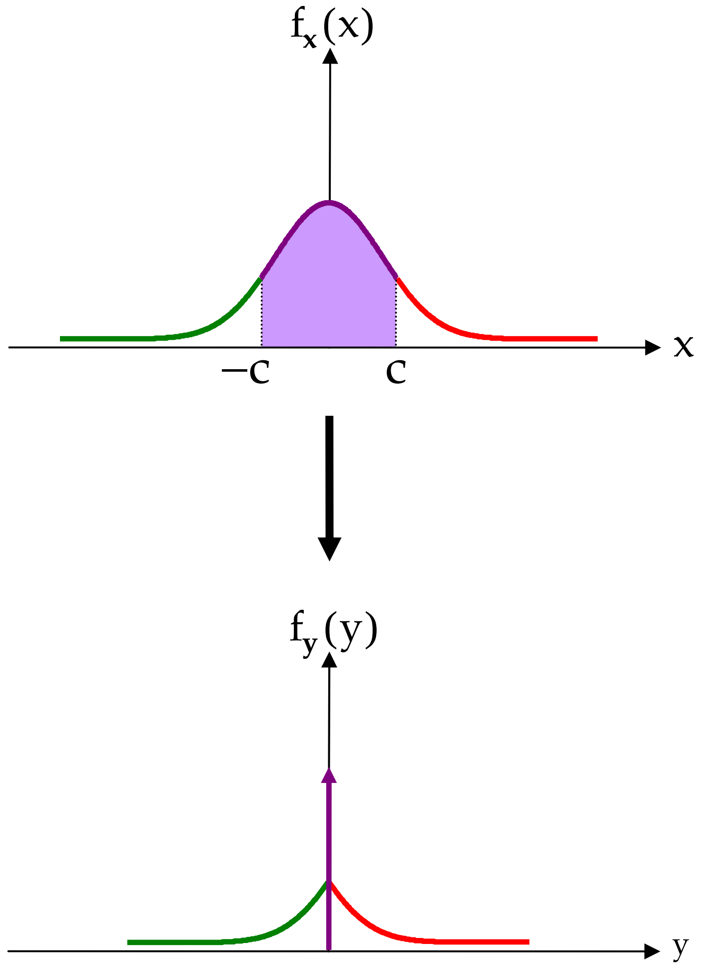


$$\mathbf{y}=g(\mathbf{x})$$

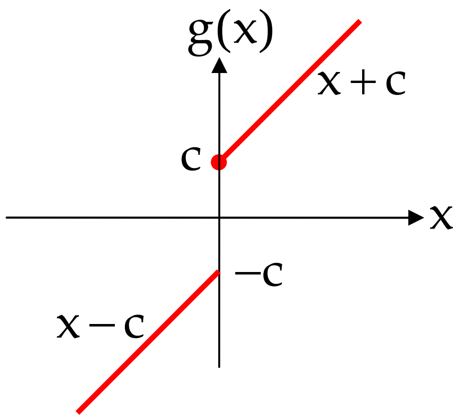
$$F_{\mathbf{y}}(y)=P\{\mathbf{y}\le y\}=\begin{cases} P\{\mathbf{x}+c\le y\} & y<0 \\ P\{\mathbf{x}-c\le y\} & y\ge 0 \end{cases}=\begin{cases} F_{\mathbf{x}}(y-c) & y<0 \\ F_{\mathbf{x}}(y+c) & y\ge 0 \end{cases}$$

$$P\{\mathbf{y}=0\}=F_{\mathbf{x}}(c)-F_{\mathbf{x}}(-c)$$

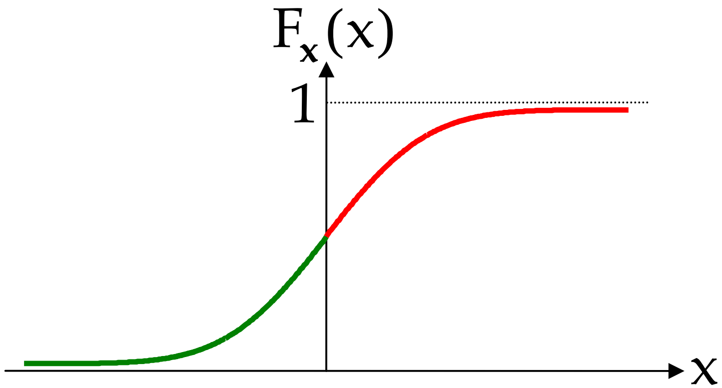
$$
f_{\mathbf{y}}(y)=\begin{cases} f_{\mathbf{x}}(y-c) & y<0 \\ \delta(y)\int_{-c}^{+c} f_{\mathbf{x}}(u)du & y=0 \\ f_{\mathbf{x}}(y+c) & y>0 \end{cases}
$$

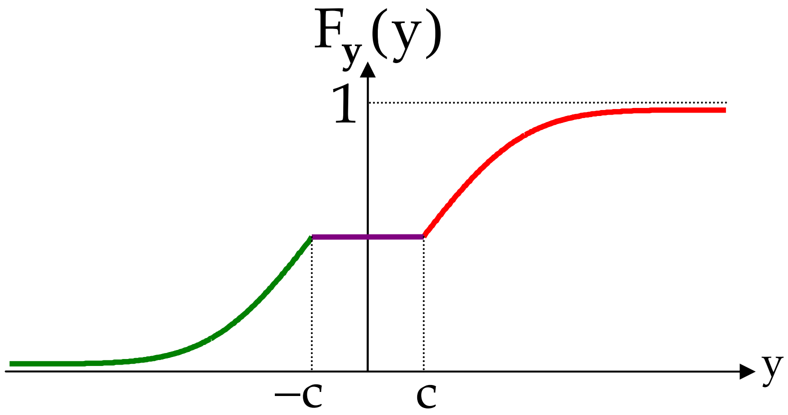


مثال ۴: $g(x)=\begin{cases} x+c & x\ge 0 \\ x-c & x<0 \end{cases}$

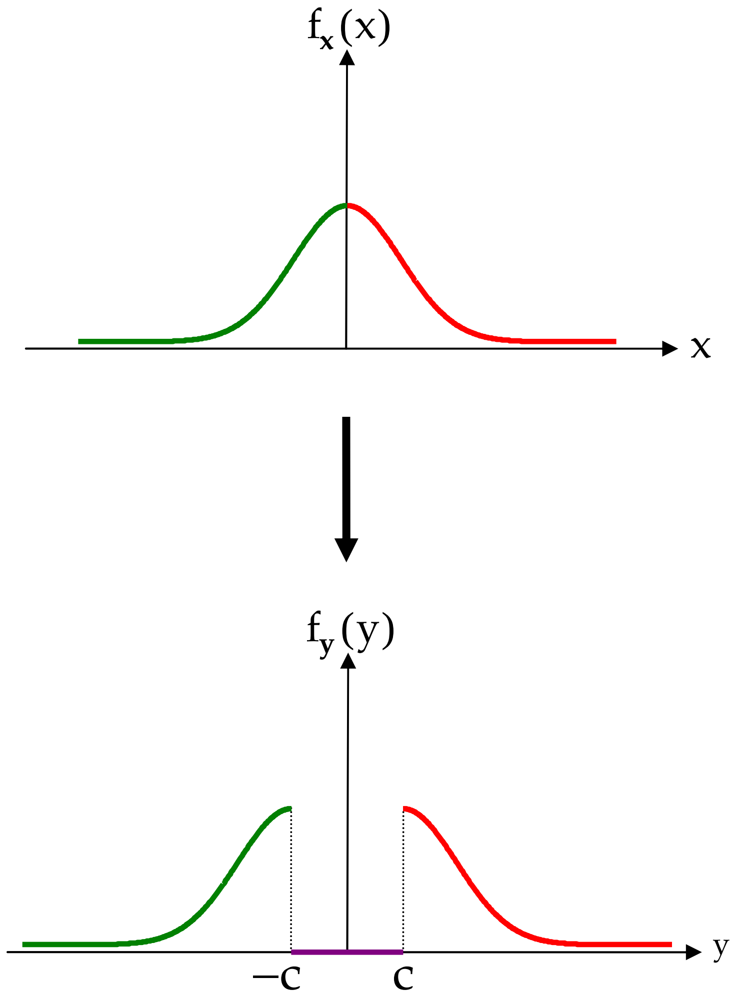


$$F_{\mathbf{y}}(y)=P\{\mathbf{y}\le y\}=\begin{cases} P\{\mathbf{x}+c\le y\} & y\ge c \\ P\{\mathbf{x}\le 0\} & -c<y<c \\ P\{\mathbf{x}-c\le y\} & y<-c \end{cases}=\begin{cases} F_{\mathbf{x}}(y-c) & y\ge c \\ F_{\mathbf{x}}(0) & -c<y<c \\ F_{\mathbf{x}}(y+c) & y<-c \end{cases}$$





$$f_y(y)=\begin{cases} f_x(y-c) & y \geq c \\ 0 & -c<y<c \\ f_x(y+c) & y<-c \end{cases}$$



به یاد داشته باشید که بی‌تغییر ماندن $g(x)$ در یک محدوده سبب پرش در F_y می‌شود و به عکس پرش در $g(x)$ موجب بی‌تغییر ماندن F_y در یک محدوده می‌شود.

می‌توانیم مستقیماً f_y را از روی f_x به دست آوریم.

قضیه: برای y داده شده، اگر معادلهٔ $g(x)=y$ دارای ریشه‌های x_1، x_2 و … باشد، یعنی: $y=g(x_1)=g(x_2)=\cdots$، داریم:

$$f_y(y)=\sum_i \frac{f_x(x_i)}{|g'(x_i)|}$$

که:

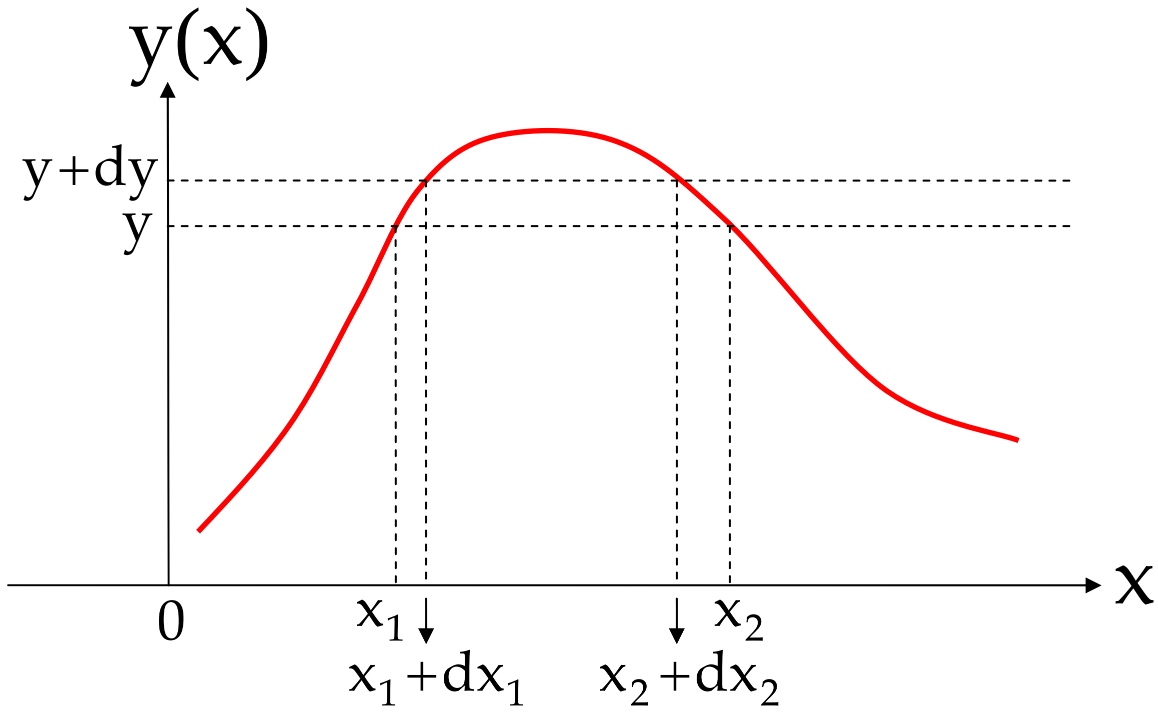
$$f_x(x_i)=f_x(x)\big|_{x=x_i(y)}$$

$$g'(x_i)=\frac{d}{dx}g(x)\bigg|_{x=x_i(y)}$$

(مشروط بر اینکه برای y داده شده، تعداد نقاط x_i قابل شمارش باشد و $F(x)$ در نقاط x_i مشتق‌پذیر باشد.)

ما در اینجا قضیه را برای وقتی که دو ریشه موجود باشد، اثبات می‌کنیم که به حالت کلی نیز قابل تعمیم است.

(در مثال خواهیم دید که در محلهایی که $g'=0$ می‌شود، f_y به سمت بی‌نهایت می‌رود، ولی جای نگرانی نیست.)



$$\rightarrow dx_1>0\ ,\ dx_2<0$$

$$f_{\mathbf{y}}(y)dy=P\{y<\mathbf{y}<y+dy\}=P\{x_1<\mathbf{x}<x_1+dx_1\}+P\{x_2+dx_2<\mathbf{x}<x_2\}=f_{\mathbf{x}}(x_1)dx_1+f_{\mathbf{x}}(x_2)\left|dx_2\right|$$

$$\Rightarrow f_{\mathbf{y}}(y)dy=f_{\mathbf{x}}(x_1)\frac{dy}{g'(x_1)}+f_{\mathbf{x}}(x_2)\frac{dy}{\left|g'(x_2)\right|}$$

زیرا:

$$y=g(x)\Big|_{x=x_1,x_2}\Rightarrow dy=g'(x)dx\Big|_{x=x_1,x_2}$$

$$\Rightarrow f_{\mathbf{y}}(y)=\frac{f_{\mathbf{x}}(x_1)}{g'(x_1)}+\frac{f_{\mathbf{x}}(x_2)}{\left|g'(x_2)\right|}$$

و در حالت کلی:

$$f_{\mathbf{y}}(y)=\sum_{i}\frac{f_{\mathbf{x}}(x_i(y))}{|g'(x_i(y))|}$$

روش دیگر:

$$f_{\mathbf{y}}(y)dy=f_{\mathbf{x}}(x_1(y))dx_1(y)+f_{\mathbf{x}}(x_2(y))|dx_2(y)|$$
$$\Rightarrow f_{\mathbf{y}}(y)=f_{\mathbf{x}}(x_1(y))\frac{dx_1(y)}{dy}+f_{\mathbf{x}}(x_2(y))\left|\frac{dx_2(y)}{dy}\right|$$

و در حالت کلی:

$$f_{\mathbf{y}}(y)=\sum_{i}f_{\mathbf{x}}(x_i(y))\left|\frac{dx_i(y)}{dy}\right|$$

ضمناً اگر برای y داده شده، $g(x)=y$ ریشه‌ای نداشته باشد، داریم: $f_{\mathbf{y}}(y)=0$.

مثال: $a > 0$, $\mathbf{y} = a\mathbf{x}^2$؛ یعنی $g(x) = ax^2$ (آشکارساز مربعی یا توان مقاومت $P = \frac{V^2}{R}$).

برای $y > 0$ داده شده، معادلهٔ $y = ax^2$ دو ریشه دارد:

$$x_1 = \sqrt{\frac{y}{a}} \; , \; x_2 = -\sqrt{\frac{y}{a}}$$

$$g'(x) = \frac{d}{dx} g(x) = 2ax \Rightarrow \begin{cases} g'(x_1(y)) = 2a\sqrt{\frac{y}{a}} = 2\sqrt{ay} \\ g'(x_2(y)) = -2a\sqrt{\frac{y}{a}} = -2\sqrt{ay} \end{cases}$$

$$f_{\mathbf{y}}(y) = \frac{f_{\mathbf{x}}(x_1)}{g'(x_1)} + \frac{f_{\mathbf{x}}(x_2)}{|g'(x_2)|} = \frac{f_{\mathbf{x}}(\sqrt{\frac{y}{a}})}{2\sqrt{ay}} + \frac{f_{\mathbf{x}}(-\sqrt{\frac{y}{a}})}{2\sqrt{ay}}$$

اگر $y < 0$ باشد، آنگاه $f_{\mathbf{y}}(y) = 0$.

یا از روش دیگر داریم:

$$f_{\mathbf{y}}(y) = f_{\mathbf{x}}(x_1(y))\frac{dx_1(y)}{dy} + f_{\mathbf{x}}(x_2(y))\left|\frac{dx_2(y)}{dy}\right|$$

$$x_1(y) = \sqrt{\frac{y}{a}} \;\Rightarrow\; \frac{d}{dy}x_1(y) = \frac{1}{2\sqrt{ay}} \quad , \quad x_2(y) = -\sqrt{\frac{y}{a}} \Rightarrow \frac{d}{dy}x_2(y) = -\frac{1}{2\sqrt{ay}}$$

$$f_{\mathbf{y}}(y) = \frac{f_{\mathbf{x}}(\sqrt{\frac{y}{a}})}{2\sqrt{ay}} + \frac{f_{\mathbf{x}}(-\sqrt{\frac{y}{a}})}{2\sqrt{ay}} : y > 0$$

یا از راه CDF نیز داریم:

$$F_{\mathbf{y}}(y) = P\{\mathbf{y} \le y\} = P\{-\sqrt{\frac{y}{a}} \le \mathbf{x} \le \sqrt{\frac{y}{a}}\} = F_{\mathbf{x}}(\sqrt{\frac{y}{a}}) - F_{\mathbf{x}}(-\sqrt{\frac{y}{a}}) : y > 0$$

$$\Rightarrow f_{\mathbf{y}}(y) = \frac{f_{\mathbf{x}}(\sqrt{\frac{y}{a}})}{2\sqrt{ay}} + \frac{f_{\mathbf{x}}(-\sqrt{\frac{y}{a}})}{2\sqrt{ay}} : y > 0$$

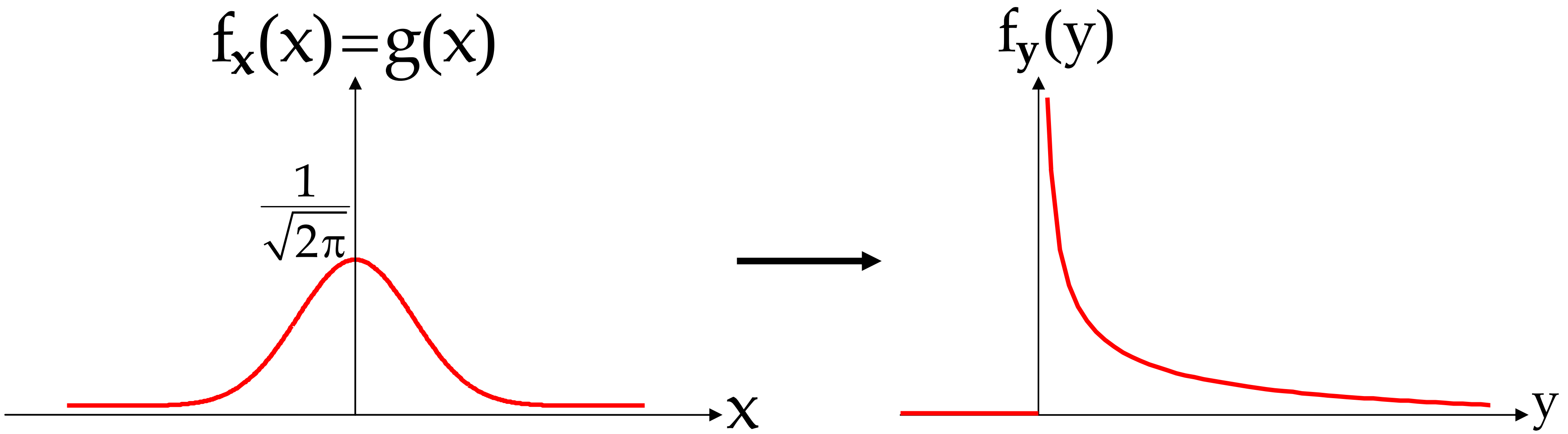
اگر $f_x(x)$ زوج باشد، خواهیم داشت:

$$f_y(y)=\frac{f_x(\sqrt{\frac{y}{a}})}{\sqrt{ay}} : y>0$$

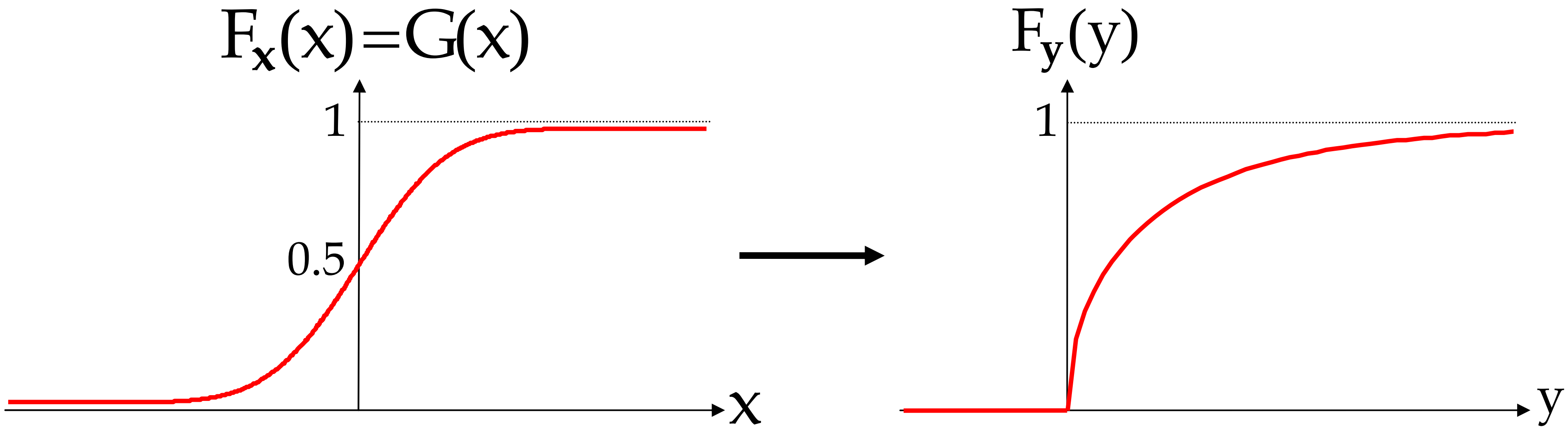
مثلاً اگر $\mathbf{x}\sim N(0,1)$ باشد و فرض کنیم: $a>0$, $\mathbf{y}=a\mathbf{x}^2$، داریم:

$$f_y(y)=\frac{f_x(\sqrt{\frac{y}{a}})}{\sqrt{ay}}=\frac{1}{\sqrt{2\pi ay}}e^{-\frac{y}{2a}}u(y)$$

که برای $a=1$، همان توزیع χ^2 با یک درجه آزادی است.



$$F_y(y) = 2F_x(\sqrt{\frac{y}{a}}) - 1 = 2G(\sqrt{\frac{y}{a}}) - 1 = D(\sqrt{\frac{y}{a}})$$



توجه کنید اگر چه در $y=0$، $f_y(y)$ به سمت بی‌نهایت می‌رود، اما در هر محدودهٔ dy، $f_y(y)dy$ احتمال است (محدود و کمتر از یک).

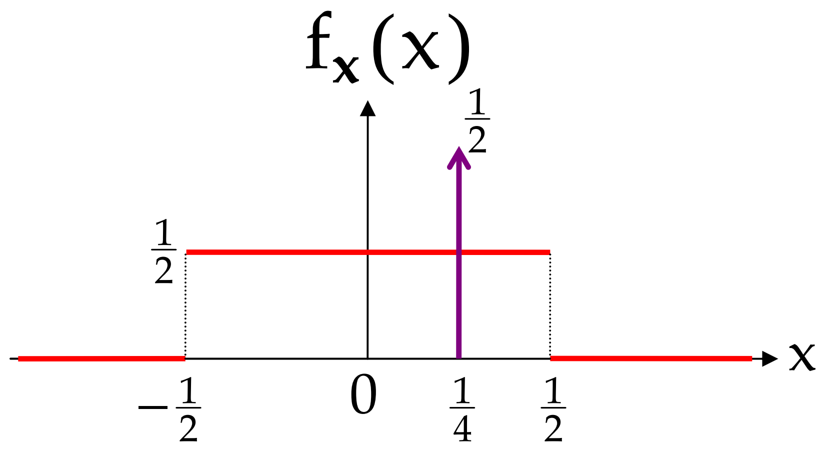
مثال: اگر $a>0$, $\mathbf{y}=a\mathbf{x}^2$ و $f_x(x)=\begin{cases}\frac{1}{2}+\frac{1}{2}\delta(x-\frac{1}{4}) & -\frac{1}{2}<x<\frac{1}{2}\\ 0 & \text{otherwise}\end{cases}$، داریم:

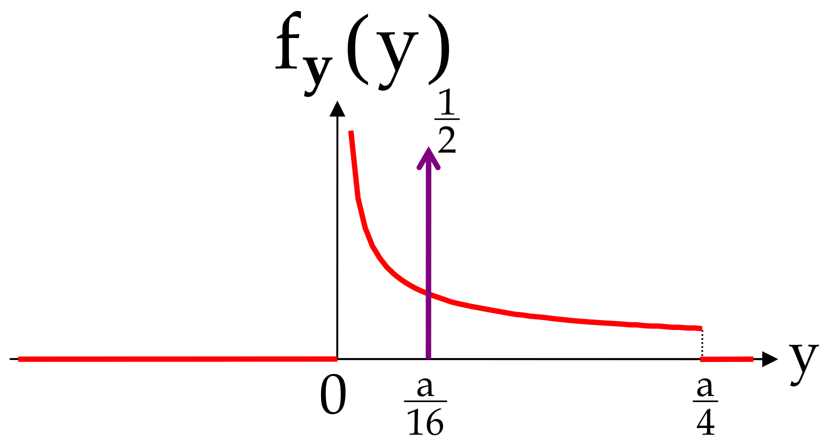
$$y>0\ ,\ \sqrt{\frac{y}{a}}<\frac{1}{2}\Rightarrow 0<y<\frac{a}{4}$$

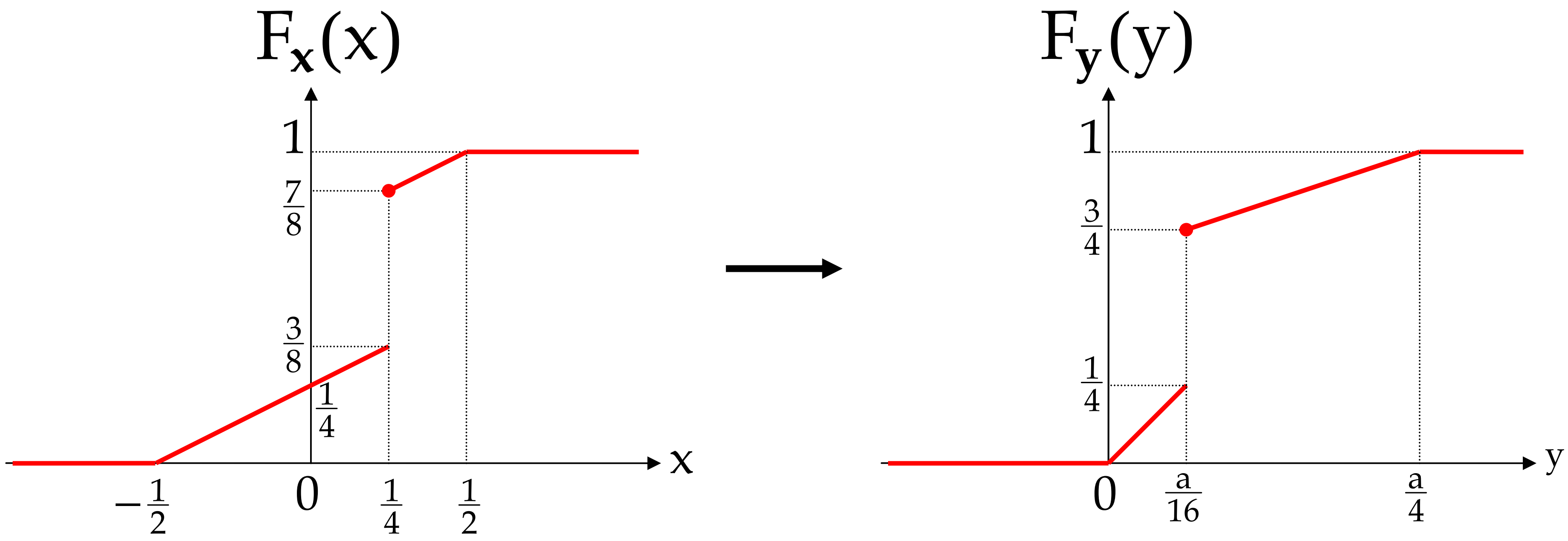
$$f_y(y)=\frac{\frac{1}{2}}{\sqrt{ay}}+\underbrace{\frac{1}{2}\delta(y-\frac{a}{16})}:0<y<\frac{a}{4}$$

x به احتمال $\frac{1}{2}$ برابر $\frac{1}{4}$ است،

پس **y** به احتمال $\frac{1}{2}$ برابر $\frac{a}{16}$ است.





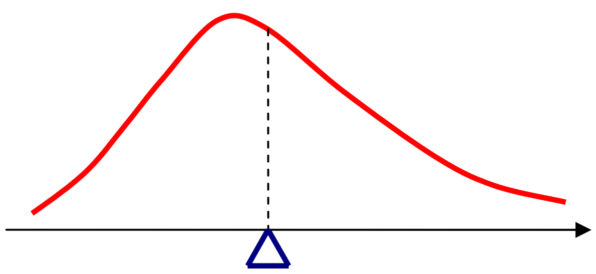


$F_{\mathbf{x}}(x)$
1
$\frac{7}{8}$
$\frac{3}{8}$
$\frac{1}{4}$
x
$-\frac{1}{2}$
0
$\frac{1}{4}$
$\frac{1}{2}$
$F_{\mathbf{y}}(y)$
1
$\frac{3}{4}$
$\frac{1}{4}$
y
0
$\frac{a}{16}$
$\frac{a}{4}$

میانگین و واریانس یک متغیر تصادفی:

آنچه متغیر تصادفی را به طور کامل مشخص می‌کرد، pdf یا CDF آن بود. دو پارامتر مهم توصیف کنندهٔ یک متغیر تصادفی **میانگین** و **واریانس** آن هستند.

میانگین (Mean) یا امید ریاضی (Expectation):



کمیت زیر را طبق تعریف، میانگین توزیع **x** یا میانگین **x** یا امید ریاضی **x** گویند:

$$\eta = E(\mathbf{x}) = \int_{-\infty}^{+\infty} x f(x) dx$$

(در تعریف امید ریاضی فرض می‌شود که $|x|f(x)$ انتگرال‌پذیر باشد، مثلاً وقتی که **x** از دو طرف محدود باشد.)

برای حالت گسسته داریم:

$$E(\mathbf{x}) = \sum_i x_i P_{\mathbf{x}}(x_i) \rightarrow \text{همان تابع احتمال است } P_{\mathbf{x}}(x)$$

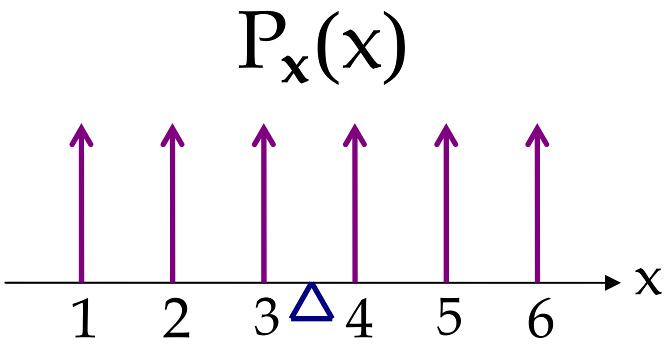
امید ریاضی را با η یا $\eta_{\mathbf{x}}$ (یا μ یا $\mu_{\mathbf{x}}$) نیز نمایش می‌دهیم.

مثال ۱: اگر برای هر ω، $\mathbf{x}(\omega) = c$ باشد، داریم:

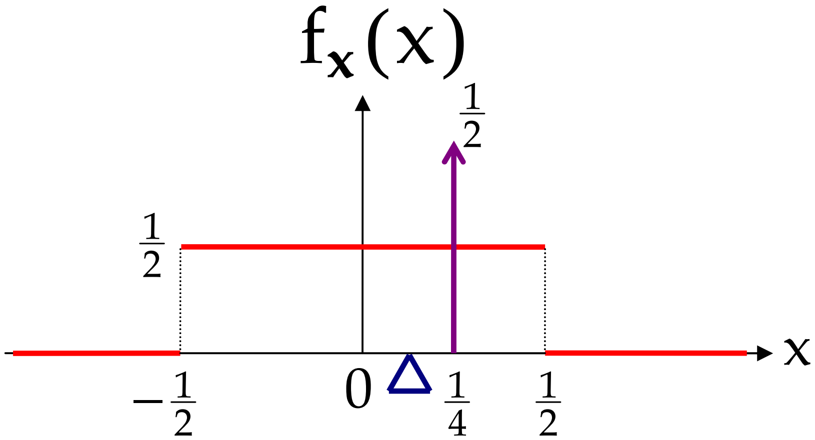
$$E(\mathbf{x}) = c\,P\{\mathbf{x} = c\} = c$$

مثال ۲: در پرتاب تاس اگر $\mathbf{x}(f_i) = i$ تعریف شود، داریم:

$$E(\mathbf{x}) = \frac{1}{6}(1+2+3+4+5+6) = 3.5$$



مثال ۳: برای این f_x داریم:



$$E(\mathbf{x}) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \frac{1}{2} x dx + \frac{1}{2} \times \frac{1}{4} = 0 + \frac{1}{8} = \frac{1}{8}$$

تعبیر تجربی امید ریاضی:

اگر $\mathbf{x}$ متغیر تصادفی با مقادیر ممکنه $x_i : i=1,2,\ldots,k$ باشد و nبار آزمایش را انجام دهیم و هر x_i، n_iمرتبه مشاهده شود، داریم:

$$\bar{x} = \frac{n_1x_1 + n_2x_2 + \cdots + n_kx_k}{\underbrace{\sum_{i=1}^{k} n_i}_{n}} = \sum_{i=1}^{k} \frac{n_i}{n} x_i \simeq \sum_{i=1}^{k} p_i x_i = E(\mathbf{x})$$

: متوسط حسابی نتایج حاصله

خواص امید ریاضی:

۱) امید توزیع متقارن:

اگر pdf متغیر تصادفی $\mathbf{x}$ حول نقطهٔ a متقارن باشد و متغیر تصادفی $\mathbf{x}$ دارای میانگین η باشد، آنگاه: $\eta = a$. یعنی:

$\forall x: f_{\mathbf{x}}(a+x) = f_{\mathbf{x}}(a-x) \Rightarrow E(\mathbf{x}) = a$ (در صورت وجود امید ریاضی)

درک شهودی این ویژگی با تعبیر مرکز ثقل کاملاً روشن است. برای اثبات ریاضی داریم:

$$E(\mathbf{x})=\int_{-\infty}^{+\infty} xf(x)dx=a+\int_{-\infty}^{+\infty}(x-a)f(x)dx=a+\int_{-\infty}^{a}(x-a)f(x)dx+\int_{a}^{+\infty}(x-a)f(x)dx$$

$$\to y=x-a\ ,\ z=a-x \Rightarrow E(\mathbf{x})=a-\int_{0}^{+\infty} z\underbrace{f(a-z)}_{f(a+z)}dz+\int_{0}^{+\infty} yf(a+y)dy=a$$

حالت خاص:

$f(x) = f(-x) \Rightarrow E(\mathbf{x}) = 0$

(ممکن است pdf حول هیچ نقطه‌ای متقارن نباشد. چنین pdfای را **چاوُله (Skewed)** گویند.)

۲) قضیۀ اساسی امید ریاضی:

اگر $\mathbf{y}=g(\mathbf{x})$ باشد، داریم:

$$E(\mathbf{y})=\int_{-\infty}^{+\infty} y f_{\mathbf{y}}(y)dy=\int_{-\infty}^{+\infty} g(x)f_{\mathbf{x}}(x)dx$$

(با فرض اینکه $|g(x)|f_{\mathbf{x}}(x)$ انتگرال‌پذیر باشد.)

یعنی لازم نیست که حتماً ابتدا $f_{\mathbf{y}}$ را حساب کنید تا بتوانید $E(\mathbf{y})$ را به دست آورید.

اثبات در کتاب، ص۱۲۴.

در حالت گسسته قضیه فوق به صورت زیر در می‌آید:

$$E(\mathbf{y})=\sum_{i} g(x_i)P_{\mathbf{x}}(x_i)$$

نتیجه: خطی بودن امید ریاضی:

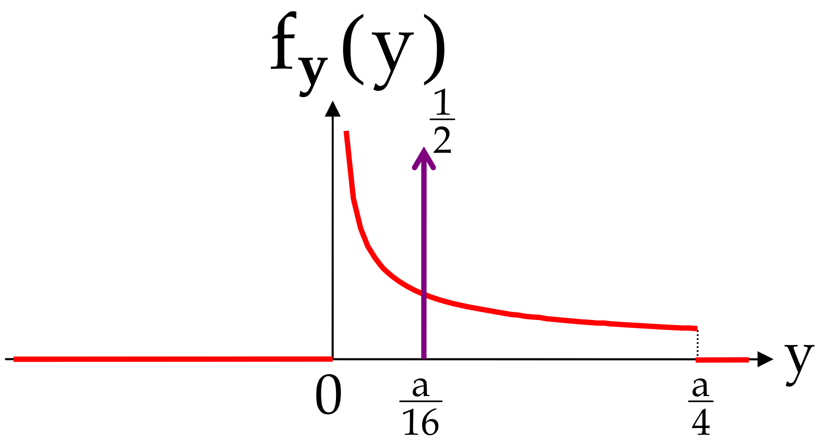
$$E\left(\sum_{i=1}^{n} a_i g_i(x)\right) = \int_{-\infty}^{+\infty} (\sum_{i=1}^{n} a_i g_i(x)) f_{\mathbf{x}}(x)dx = \sum_{i=1}^{n} a_i (\int_{-\infty}^{+\infty} g_i(x) f_{\mathbf{x}}(x)dx) = \sum_{i=1}^{n} a_i E(g_i(\mathbf{x}))$$

از جمله اینکه داریم:

$$E(a\mathbf{x}+b) = aE(\mathbf{x}) + b$$

مثال: $a>0$, $\mathbf{y}=a\mathbf{x}^2$ و نمودار f_x به صورت مقابل است. $E(\mathbf{y})$ را حساب کنید.

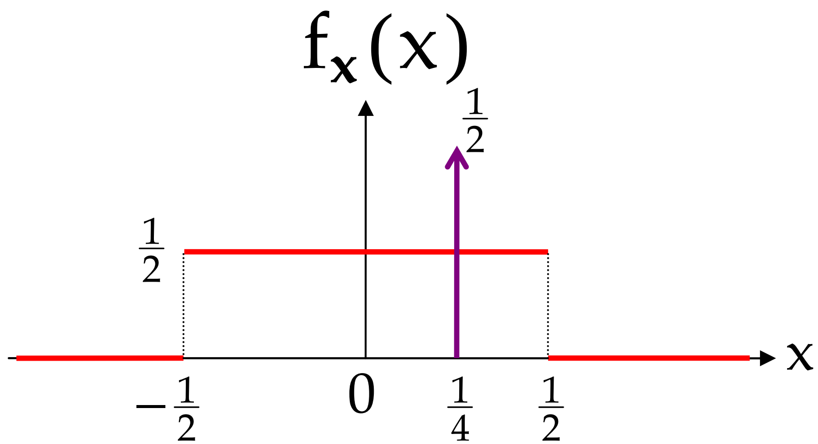
با استفاده از f_y داریم:



$$f_y(y)=\frac{1}{2\sqrt{ay}}+\frac{1}{2}\delta(y-\frac{a}{16}):0<y<\frac{a}{4}$$

$$E(\mathbf{y})=\int_0^{\frac{a}{4}}y\frac{dy}{2\sqrt{ay}}+\frac{1}{2}\times\frac{a}{16}=\frac{a}{24}+\frac{a}{32}=\frac{7a}{96}$$

بدون داشتن f_y نیز می‌توانیم $E(\mathbf{y})$ را محاسبه کنیم:



$$E(Y)=\int_{-\frac{1}{2}}^{+\frac{1}{2}}ax^2(\frac{1}{2})dx+a(\frac{1}{4})^2(\frac{1}{2})=\frac{a}{24}+\frac{a}{32}=\frac{7a}{96}$$

۳) اگر ثابت a وجود داشته باشد، به طوری‌که: $P\{\mathbf{x} \geq a\} = 1$، آنگاه: $E(\mathbf{x}) \geq a$ و حالت تساوی $(E(\mathbf{x}) = a)$ وقتی برقرار خواهد بود که: $P\{\mathbf{x} = a\} = 1$ ($x = a$ با احتمال یک، $x = a$ تقریباً مطمئن).

اگر ثابت b وجود داشته باشد، به طوری‌که $P\{\mathbf{x} \leq b\} = 1$، آنگاه: $E(\mathbf{x}) \leq b$ و حالت تساوی $(E(\mathbf{x}) = b)$ وقتی برقرار خواهد بود که: $P\{\mathbf{x} = b\} = 1$.

اثبات در کتاب DeGroot، ص ۱۸۸.

$$E(\mathbf{x}) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{a}^{+\infty} x f(x) dx \geq \int_{a}^{+\infty} a f(x) dx = a P\{\mathbf{x} \geq a\} = a$$

۴) نامساوی مارکف (Markoff's Inequality):

اگر $\mathbf{x}$ یک متغیر تصادفی مثبت باشد (یعنی: $f_{\mathbf{x}}(x) = 0 : x < 0$ یا $P\{\mathbf{x} < 0\} = 0$) و $\alpha > 0$ یک ثابت دلخواه باشد، داریم:

$$P\{\mathbf{x} \geq \alpha\} \leq \frac{E(\mathbf{x})}{\alpha}$$

مثلاً اگر $E(\mathbf{x}) = 1$ باشد، توزیع x هر چه باشد، حتماً داریم: $P\{\mathbf{x} \geq 100\} \leq 0.01$.

نامساوی مارکف معمولاً برای αهایی که نسبت به $E(\mathbf{x})$ بزرگ باشند، استفاده می‌شود.

(خاصیت قبل را می‌توان حالت خاصی از این نامساوی دانست.)

اثبات:

$$E(\mathbf{x})=\int_0^{+\infty} x f_{\mathbf{x}}(x)dx \geq \int_\alpha^{+\infty} x f_{\mathbf{x}}(x)dx \geq \int_\alpha^{+\infty} \alpha f_{\mathbf{x}}(x)dx = \alpha P\{\mathbf{x}\geq\alpha\}$$

(تساوی وقتی برقرار خواهد بود که x بزرگتر از α نشود و زیر α تنها $x=0$ می‌تواند دارای احتمال باشد.)

<u>نتیجه: نامساوی Bienayme:</u>

برای هر عدد حقیقی a و برای هر $\varepsilon>0$ و $n>0$ داریم:

$$P\{|\mathbf{x}-a|\geq\varepsilon\}\leq\frac{E(|\mathbf{x}-a|^n)}{\varepsilon^n}$$

زیرا:

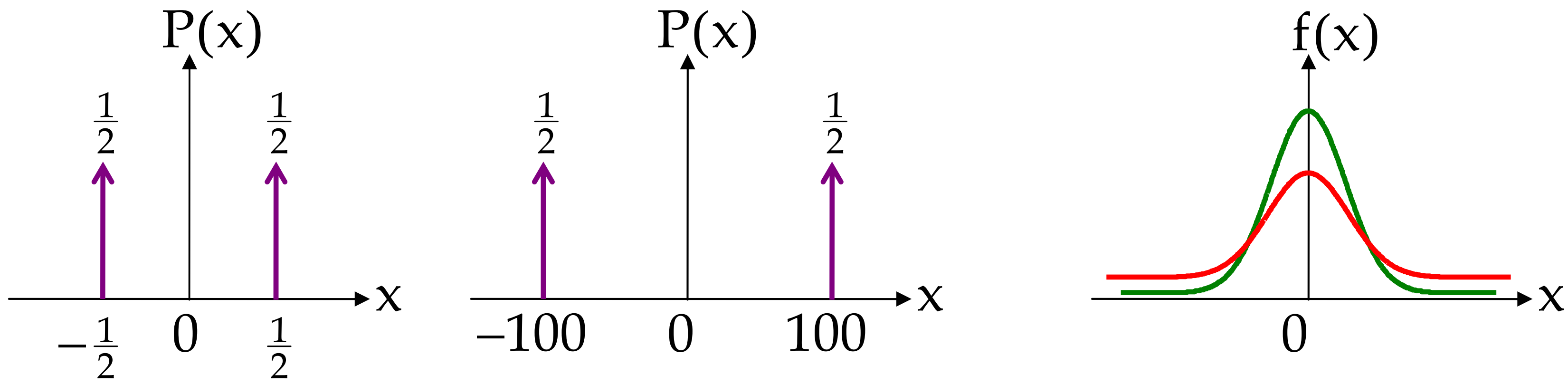
$$P\{|\mathbf{x}-a|\geq\varepsilon\}=P\{|\mathbf{x}-a|^n\geq\varepsilon^n\}\leq\frac{E(|\mathbf{x}-a|^n)}{\varepsilon^n}$$

(طبق قضیه مارکف)

واریانس:

میانگین نشان می‌داد که مقادیر **x** حول و حوش چه مقداری هستند. کمیت دیگری لازم داریم که میزان پراکندگی مقادیر **x** حول مقدار میانگین را به ما نشان دهد. واریانس بیانگر میزان پراکندگی احتمال حول مقدار متوسط (میانگین) است.

میانگین‌ها یکسان، ولی پراکندگی‌ها متفاوت:



تفاوت متغیر تصادفی با میانگین برابر است با: $|\mathbf{x}-\mathrm{E}(\mathbf{x})|$.

طبق تعریف، واریانس توزیع **x** یا واریانس متغیر تصادفی **x** به صورت زیر است:

$$\mathrm{var}(\mathbf{x})=\mathrm{E}((\mathbf{x}-\mathrm{E}(\mathbf{x}))^2)$$ (با فرض وجود این امید ریاضی)

(زیرا کار با آن ساده‌تر از $\mathrm{E}(|\mathbf{x}-\mathrm{E}(\mathbf{x})|)$ است.)

واریانس کمیتی نامنفی است، چون امید ریاضی یک متغیر تصادفی همواره نامنفی است. لذا متداول است که آن را با $\sigma_{\mathbf{x}}^2$ یا σ^2 نشان می‌دهند:

$$\sigma^2 = \mathrm{E}((\mathbf{x} - \eta)^2)$$

σ را انحراف معیار (Standard Deviation) گویند.

توجه دارید که $\mathrm{var}(\mathbf{x})$ دارای دیمانسیون مربع متغیر تصادفی است. σ از جنس خود متغیر تصادفی است.

از تعریف امید ریاضی برای حالت گسسته نتیجه می‌شود که:

$$\sigma^2 = \sum_{\mathrm{i}} (\mathrm{x_i} - \eta)^2 \mathrm{P}(\mathrm{x_i})$$

خواص واریانس:

۱) $\sigma^2 = E(\mathbf{x}^2) - E^2(\mathbf{x})$؛

$E(\mathbf{x}^2)$ را Mean Square گویند و داریم: $\text{rms (root mean square)} = \sqrt{E(\mathbf{x}^2)}$.

اثبات:

$$\sigma^2 = E((\mathbf{x}-\eta)^2) = E(\mathbf{x}^2 - 2\eta\mathbf{x} + \eta^2) = E(\mathbf{x}^2) - 2\eta E(\mathbf{x}) + \eta^2 = E(\mathbf{x}^2) - \eta^2$$

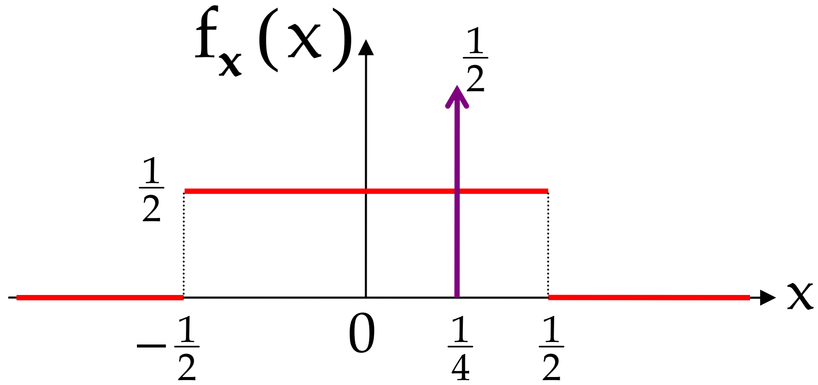
نتیجهٔ دیگر:

$$E(\mathbf{x}^2) = \sigma^2 + \eta^2$$
$$\Rightarrow E(\mathbf{x}^2) \geq E^2(\mathbf{x})$$

مثال: برای توزیع زیر قبلاً $E(a\mathbf{x}^2)$ را برابر با $\frac{7a}{96}$ به دست آورده بودیم. پس داریم:

$$E(\mathbf{x})=\frac{1}{8} \quad , \quad E(\mathbf{x}^2)=\frac{7}{96}$$

$$\Rightarrow \sigma^2=\frac{7}{96}-\frac{1}{64}=\frac{11}{192} \Rightarrow \sigma = 0.239$$



۲) اگر $\mathbf{y}=a\mathbf{x}+b$ باشد، داریم:

$$\sigma_{\mathbf{y}}^2=a^2\sigma_{\mathbf{x}}^2$$

اثبات:

$$\sigma_{\mathbf{y}}^2=E((\mathbf{y}-\eta_{\mathbf{y}})^2)=E((a\mathbf{x}+b-a\eta_{\mathbf{x}}-b)^2)=E(a^2(\mathbf{x}-\eta_{\mathbf{x}})^2)=a^2\sigma_{\mathbf{x}}^2$$

۳) $\operatorname{var}(\mathbf{x})=0$ اگر و تنها اگر عددی مانند η وجود داشته باشد که $P\{\mathbf{x}=\eta\}=1$ باشد (η همان $E(\mathbf{x})$ خواهد بود).

اثبات در کتاب DeGroot، ص ۱۹۵.

۴) نامساوی چبیشف (Tchebycheff’s Inequality):

$$P\{\eta-\varepsilon<\mathbf{x}<\eta+\varepsilon\}\geq 1-\frac{\sigma^2}{\varepsilon^2}$$

یا:

$$P\{\eta-k\sigma<\mathbf{x}<\eta+k\sigma\}\geq 1-\frac{1}{k^2}$$

(برای $\varepsilon \gg \sigma$ به خاصیت قبل می‌رسیم.)

اثبات: از نامساوی Bienayme داریم:

$$P\{|\mathbf{x}-a|\geq\varepsilon\}\leq\frac{E(|\mathbf{x}-a|^n)}{\varepsilon^n}$$

اگر بگیریم: $a=\eta$ و $n=2$ و $\varepsilon=k\sigma$، نتیجه می‌شود:

$$P\{|\mathbf{x}-\eta|\geq k\sigma\}\leq\frac{\sigma^2}{\varepsilon^2}\Rightarrow P\{\eta-k\sigma<\mathbf{x}<\eta+k\sigma\}\geq 1-\frac{1}{k^2}$$

که برای هر توزیع **x** صادق است (و می‌توان نشان داد که حد بهتری که برای هر توزیع صادق باشد، وجود ندارد).

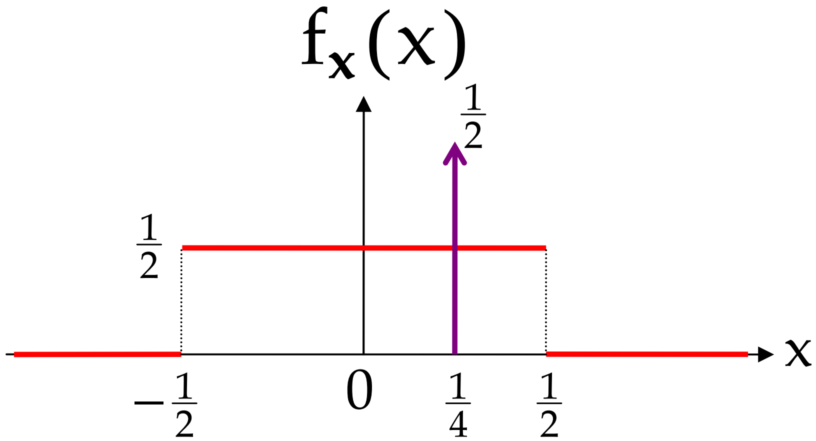
محاسبهٔ تقریبی $E(g(\mathbf{x}))$:

$$g(x) \simeq g(\eta_x) + g'(\eta_x)(x-\eta_x) + \frac{g''(\eta_x)}{2}(x-\eta_x)^2$$

$$\Rightarrow E(g(\mathbf{x})) \simeq g(\eta_x) + g''(\eta_x)\frac{\sigma_x^2}{2}$$

(تقریب بالا، تقریب $g(x)$ به صورت درجهٔ دو است. تقریب بیشتر: $E(g(\mathbf{x})) \simeq g(\eta_x)$ که تقریب $g(x)$ به صورت خطی است.)

مثال: در مثالی که داشتیم:



$$g(x) = ax^2 \Rightarrow g''(x) = 2a$$

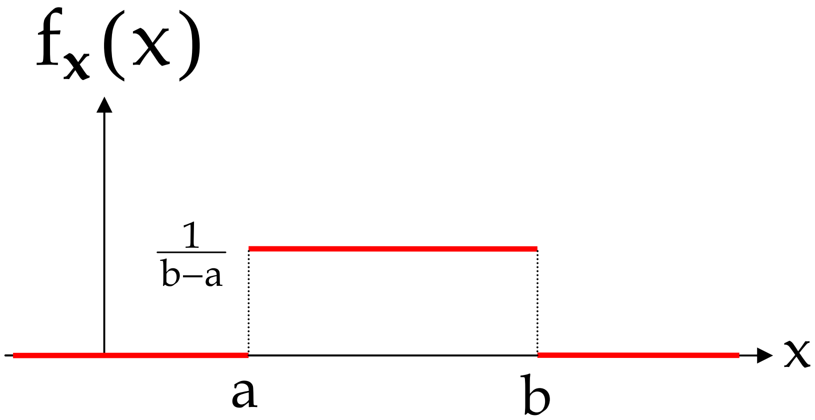
$$\eta_{\mathbf{x}} = \frac{1}{8} \quad , \quad \sigma_{\mathbf{x}}^2 = \frac{11}{192}$$

$$\Rightarrow E(g(\mathbf{x})) \simeq \frac{a}{64} + \frac{2a}{2} \times \frac{11}{192} = \frac{a}{64} + \frac{11}{192}a = \frac{14a}{192} = \frac{7a}{96}$$

که دقیقاً برابر مقدار واقعی است که قبلاً حساب کردیم (اگر تابع $g(x)$ از درجهٔ بالاتر بود، اختلاف ظاهر می‌شد).

میانگین و واریانس برخی توزیع‌های خاص:

۱) توزیع یکنواخت:

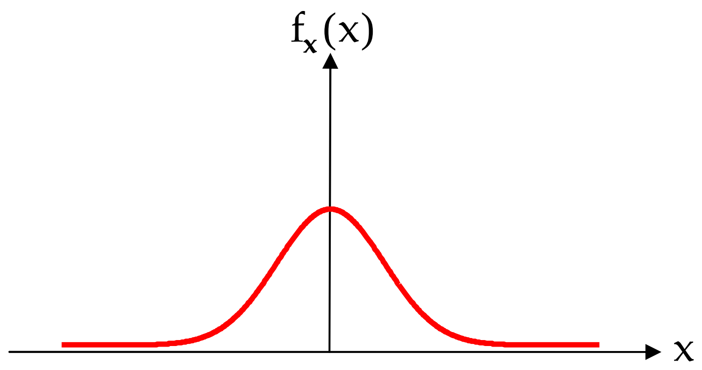


$$E(\mathbf{x})=\int_{-\infty}^{+\infty} x f_{\mathbf{x}}(x)dx=\int_{a}^{b} x\frac{1}{b-a}dx=\frac{a+b}{2}$$

چون این توزیع حول $\frac{a+b}{2}$ متقارن است از قبل نیز می‌توانستیم این را بگوییم.

$$\mathrm{var}(\mathbf{x})=\int_{-\infty}^{+\infty}(x-\eta_{\mathbf{x}})^2 f_{\mathbf{x}}(x)dx=\int_{a}^{b}(x-\frac{a+b}{2})^2\frac{1}{b-a}dx=\frac{(b-a)^2}{12}$$

۲) توزیع نرمال:



ابتدا نرمال استاندارد را در نظر می‌گیریم:

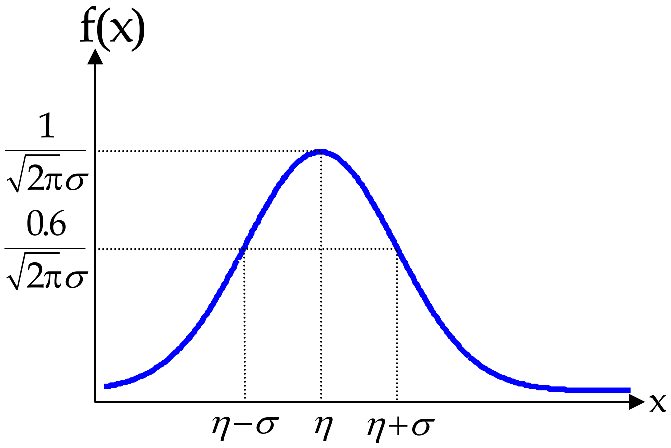
$$\mathbf{z} \sim N(0,1) \rightarrow f_{\mathbf{z}}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

این توزیع زوج است (حول صفر متقارن است) و داریم: $f_{\mathbf{z}}(z) = f_{\mathbf{z}}(-z)$، پس:

$$\eta_{\mathbf{z}} = E(\mathbf{z}) = 0$$

$$\text{var}(\mathbf{z}) = E((\mathbf{z}-0)^2) = E(\mathbf{z}^2) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} z^2 e^{-\frac{z^2}{2}} dz = 1$$

حال توزیع نرمال را در حالت کلی در نظر می‌گیریم. اگر $\mathbf{x} = a\mathbf{z} + b$ باشد، می‌دانیم که:



$$f_{\mathbf{x}}(x) = \frac{1}{|a|} f_{\mathbf{z}}\left(\frac{x-b}{a}\right)$$

با فرض $a > 0$ خواهیم داشت:

$$f_{\mathbf{x}}(x) = \frac{1}{a\sqrt{2\pi}} e^{-\frac{(x-b)^2}{2a^2}}$$

ولی می‌دانیم که:

$$\begin{cases} \eta_{\mathbf{x}} = a\eta_{\mathbf{z}} + b = b \\ \sigma_{\mathbf{x}} = a\sigma_{\mathbf{z}} = a \end{cases}$$

یعنی: $\mathbf{x} \sim N(b,a)$؛ به همین خاطر بود که از ابتدا نتاسیون $N(\eta,\sigma)$ را به کار می‌بردیم.

مثال: نویز حرارتی دارای توزیع گوسی است.

$$\mathbf{x} \sim N(0,\sigma)$$

$$\eta_{\mathbf{x}} = E(\mathbf{x}) = 0$$

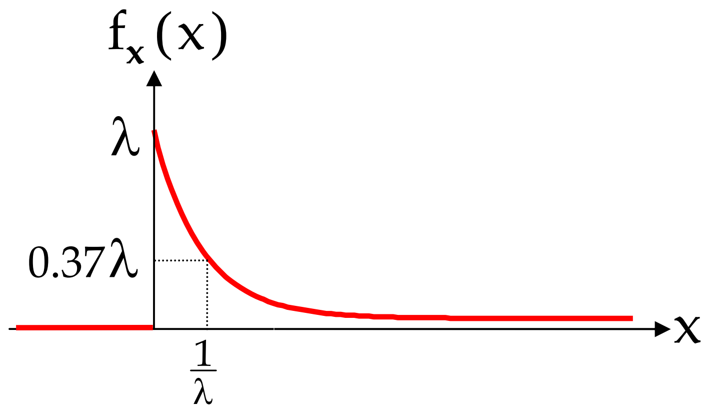
$$\sigma_{\mathbf{x}}^2 = E((\mathbf{x}-\eta_{\mathbf{x}})^2) = E(\frac{\mathbf{x}^2}{1\Omega}) = \text{قدرت نویز} = \int_{-\infty}^{+\infty} x^2 f_{\mathbf{x}}(x)dx$$

(هر چه قدرت نویز بیشتر باشد، دامنه‌های بزرگتری می‌تواند داشته باشد.)

برقراری نامساوی چبیشف:

$$P\{|\mathbf{x}-\eta| < k\sigma\} \geq 1-\frac{1}{k^2} \rightarrow P\{-2\sigma < \mathbf{x} < 2\sigma\} = 0.954 > 0.75 \ ☑$$

۳) توزیع نمایی:



$$f_{\mathbf{x}}(x)=\lambda e^{-\lambda x}u(x)$$

$$E(\mathbf{x})=\int_0^{+\infty}x\lambda e^{-\lambda x}dx=\frac{1}{\lambda}$$

$$E(\mathbf{x}^2)=\int_0^{+\infty}x^2\lambda e^{-\lambda x}dx=\frac{2}{\lambda^2}\Rightarrow \mathrm{var}(\mathbf{x})=\frac{2}{\lambda^2}-\frac{1}{\lambda}=\frac{1}{\lambda^2}$$

$$\Rightarrow \eta=\sigma=\frac{1}{\lambda}$$

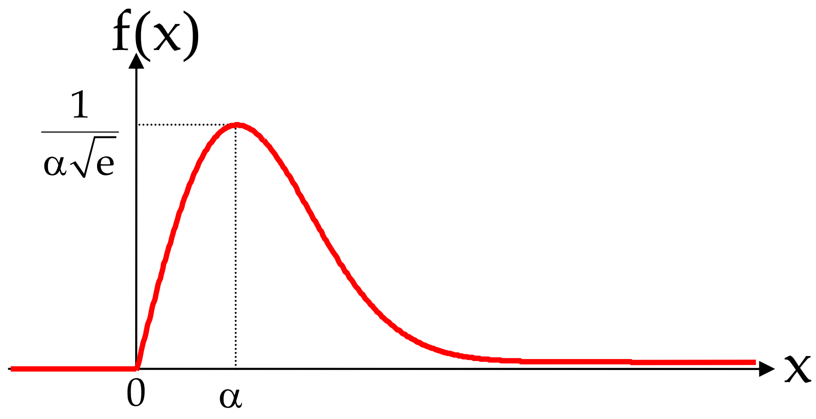
برقراری نامساوی مارکوف: برای $\alpha>0$ داریم:

$$P\{\mathbf{x}\ge\alpha\}=1-F_{\mathbf{x}}(\alpha)=e^{-\lambda\alpha}$$

$$P\{\mathbf{x}\ge\alpha\}\le\frac{E(\mathbf{x})}{\alpha}\Leftrightarrow e^{-\lambda\alpha}\le\frac{1}{\lambda\alpha}\ \boxtimes$$

مثلاً داریم: $e^{-2}<\frac{1}{2}$.

۴) توزیع رایلی:

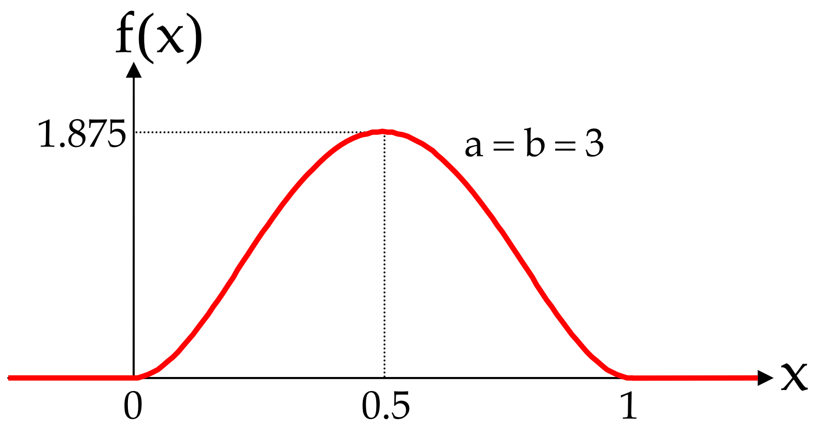


$$f_{\mathbf{x}}(x) = \frac{x}{\alpha} e^{-\frac{x^2}{2\alpha^2}} \; : x \geq 0$$

می‌توان نشان داد (در تمرین نشان می‌دهید) که:

$$\left.\begin{aligned} E(\mathbf{x}) &= \alpha\sqrt{\frac{\pi}{2}} \\ E(\mathbf{x}^2) &= 2\alpha^2 \end{aligned}\right\} \Rightarrow \operatorname{var}(\mathbf{x}) = (2 - \frac{\pi}{2})\alpha^2$$

۵) توزیع بتا:

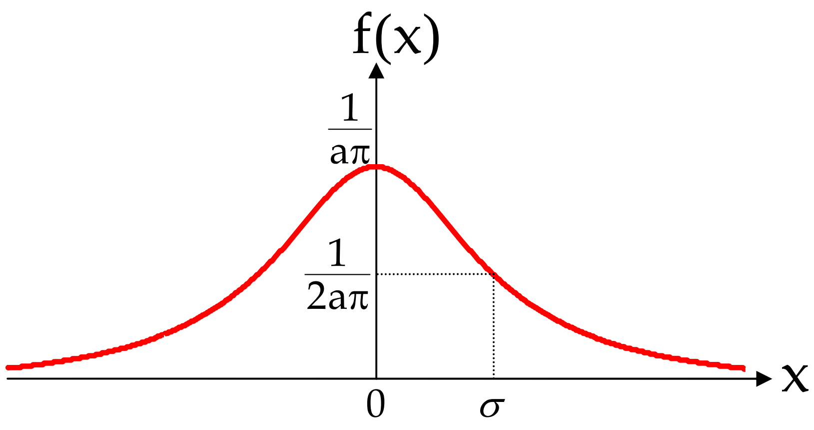


$$f(x)=\begin{cases}\dfrac{1}{\beta(a,b)}x^{a-1}(1-x)^{b-1} & 0\le x\le 1\\ 0 & \text{otherwise}\end{cases}$$

می‌توان نشان داد که:

$$\eta=\frac{a}{a+b}\quad,\quad \sigma^2=\frac{ab}{(a+b)^2(a+b+1)}$$

۶) توزیع کوشی:

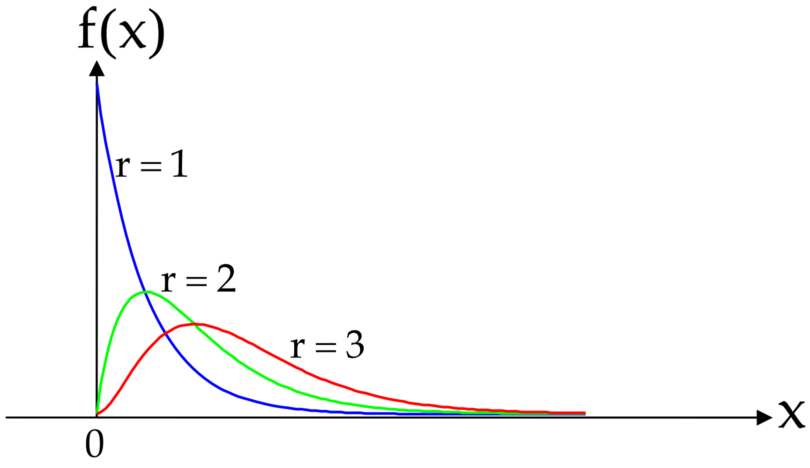


$$f(x)=\frac{\frac{a}{\pi}}{x^2+a^2}$$

برای این توزیع، امید ریاضی و واریانس وجود ندارد.

۷) توزیع گاما:

$$f(x) = A\,x^{r-1}\,e^{-\lambda x}\,u(x) : r > 0, \lambda > 0$$



می‌توان نشان داد که:

$$E(\mathbf{x}) = \frac{r}{\lambda} \quad , \quad \sigma_{\mathbf{x}}^2 = \frac{r}{\lambda^2}$$

۸) توزیع پواسن:

$$P\{\mathbf{x}=k\}=e^{-a}\frac{a^k}{k!}:k=0,1,2,\ldots$$

$$E(\mathbf{x})=\sum_{k=0}^{+\infty}k\,P\{\mathbf{x}=k\}=\sum_{k=0}^{+\infty}k\,e^{-a}\frac{a^k}{k!}=e^{-a}\,a\underbrace{\sum_{k=1}^{+\infty}\frac{a^{k-1}}{(k-1)!}}_{e^a}=a$$

به همین خاطر بود که از پیش به جای a، نام η را به کار برده بودیم. پس وقتی می‌گوییم:

$$P\{\tau \text{ نقطه در فاصلهٔ } k\}=e^{-\lambda\tau}\frac{(\lambda\tau)^k}{k!}$$

$\eta=\lambda\tau$، میانگین تعداد نقاط در فاصلهٔ زمانی τ بوده و λ، میانگین (امید ریاضی) تعداد نقاط در واحد زمان است.

$$E(\mathbf{x}^2)=\sum_{k=0}^{+\infty}k^2e^{-a}\frac{a^k}{k!}=e^{-a}\sum_{k=0}^{+\infty}k^2\frac{a^k}{k!}=e^{-a}\left(\sum_{k=0}^{+\infty}k\frac{a^k}{k!}+\sum_{k=0}^{+\infty}k(k-1)\frac{a^k}{k!}\right)$$

$$=e^{-a}(ae^a+a^2\underbrace{\sum_{k=2}^{+\infty}\frac{a^{k-2}}{(k-2)!}}_{e^a})=a^2+a$$

$$\Rightarrow\sigma^2=E(\mathbf{x}^2)-E^2(\mathbf{x})=a^2+a-a^2=a$$

$\rightarrow\eta=\sigma^2=a$ (پارامتر توزیع پواسن)

روش دیگر برای محاسبهٔ $E(\mathbf{x})$ و $E(\mathbf{x}^2)$:

$$\sum_{k=0}^{+\infty}\frac{a^k}{k!}=e^a\xrightarrow{\text{مشتق نسبت به } a}\sum_{k=0}^{+\infty}k\frac{a^{k-1}}{k!}=e^a\xrightarrow{\text{ضرب در } a}\sum_{k=0}^{+\infty}k\frac{a^k}{k!}=ae^a\Rightarrow E(\mathbf{x})=a$$

$$\sum_{k=0}^{+\infty}k\frac{a^{k-1}}{k!}=e^a\xrightarrow{\text{مشتق نسبت به } a}\sum_{k=0}^{+\infty}k(k-1)\frac{a^{k-2}}{k!}=e^a\xrightarrow{\text{ضرب در } a^2}\sum_{k=0}^{+\infty}k(k-1)\frac{a^k}{k!}=a^2e^a$$

$$\Rightarrow\sum_{k=0}^{+\infty}k^2\frac{a^k}{k!}=a^2e^a+ae^a\Rightarrow E(\mathbf{x}^2)=a^2+a$$

۹) توزیع برنولی:

$$P_{\mathbf{x}}(k) = P\{\mathbf{x} = k\} = p^k q^{1-k} : k = 0,1$$

$$E(\mathbf{x}) = 0 \times P\{\mathbf{x} = 0\} + 1 \times P\{\mathbf{x} = 1\} = p$$
$$E(\mathbf{x}^2) = 0^2 \times P\{\mathbf{x} = 0\} + 1^2 \times P\{\mathbf{x} = 1\} = p$$
$$\mathrm{var}(\mathbf{x}) = E(\mathbf{x}^2) - E^2(\mathbf{x}) = p - p^2 = pq$$

۱۰) توزیع دوجمله‌ای:

$$P_{\mathbf{x}}(k) = P\{\mathbf{x} = k\} = \binom{n}{k} p^k q^{n-k} : k = 0, 1, \ldots, n$$

می‌توان نشان داد که:

$$\eta = E(\mathbf{x}) = np$$
$$\sigma^2 = \mathrm{var}(\mathbf{x}) = npq$$

(در واقع متغیر تصادفی دوجمله‌ای را می‌توان مجموع nتا متغیر تصادفی برنولی دانست و لذا η و σ^2 این چنین می‌شوند. می‌توان این تساوی‌ها را مستقیماً نیز نشان داد. یک راه دیگر استفاده از رابطهٔ $i\binom{n}{i} = n\binom{n-1}{i-1}$ است (در کتاب راس، ص ۱۵۲ و ۱۵۳). راه دیگر با استفاده از تابع مشخصه را بعداً خواهیم دید.)

۱۱) توزیع فوق‌هندسی:

$$P_{\mathbf{x}}(k)=P\{\mathbf{x}=k\}=\frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}}:\max(0,n-N+K)\le k\le\min(n,K)\quad,\quad p=\frac{K}{N}$$

می‌توان نشان داد که:

$$E(\mathbf{x})=n\frac{K}{N}=np$$

$$\mathrm{var}(\mathbf{x})=\underbrace{\frac{nK(N-K)}{N^2}}_{npq}\times\frac{N-n}{N-1}$$

۱۲) توزیع دوجمله‌ای منفی:

$$P_{\mathbf{x}}(k) = P\{\mathbf{x} = k\} = \binom{r+k-1}{k} p^r q^k : k = 0,1,2,\ldots$$

می‌توان نشان داد که:

$$E(\mathbf{x}) = \frac{rq}{p}$$

$$\mathrm{var}(\mathbf{x}) = \frac{rq}{p^2}$$

گشتاورهای یک متغیر تصادفی (Moments of a RV):

میانگین و واریانس اعدادی بودند که تا حدودی توصیفی از متغیر تصادفی را به دست می‌دادند. در حالت کلی‌تر گشتاور (گشتاور ابتدایی (Initial)) مرتبهٔ nام متغیر تصادفی $\mathbf{x}$ به صورت زیر تعریف می‌شود:

$$m_n = E(\mathbf{x}^n) = \int_{-\infty}^{+\infty} x^n f(x)dx$$

و گشتاور مرکزی مرتبهٔ nام متغیر تصادفی $\mathbf{x}$ به صورت زیر تعریف می‌شود:

$$\mu_n = E((\mathbf{x}-\eta)^n) = \int_{-\infty}^{+\infty} (x-\eta)^n f(x)dx$$

روشن است که:

$$\eta = m_1 \ , \ \sigma^2 = \mu_2 \ , \ \mu_1 = 0 \ , \ \mu_0 = m_0 = 1$$

همچنین $E(|\mathbf{x}|^n)$ را گشتاور مطلق مرتبهٔ nام و $E(\mathbf{x}-a)$ را گشتاور $\mathbf{x}$ حول نقطهٔ a (گشتاور تعمیم‌یافته) گویند. اگر $f(x)$ زوج باشد، $\eta = 0$ و لذا: $m_n = \mu_n$ و همچنین داریم: $\mu_{2n+1} = 0$. اصولاً اگر $f(x)$ حول نقطه‌ای متقارن باشد، داریم: $\mu_{2n+1} = 0$.

مثال: گشتاور مرتبهٔ nام توزیع گاما:

$$f(x) = \frac{\lambda^r x^{r-1} e^{-\lambda x}}{\Gamma(r)} u(x)$$

$$m_n = E(\mathbf{x}^n) = \int_0^{+\infty} \frac{\lambda^r x^{n+r-1} e^{-\lambda x}}{\Gamma(r)} dx = \frac{\lambda^r}{\Gamma(r)} \int_0^{+\infty} x^{n+r-1} e^{-\lambda x} dx = \frac{\lambda^r}{\Gamma(r)} \frac{\Gamma(n+r)}{\lambda^{n+r}} = \frac{\Gamma(n+r)}{\Gamma(r)\lambda^n}$$

$$\Rightarrow \eta = E(\mathbf{x}) = \frac{r}{\lambda}$$

$$\Rightarrow E(\mathbf{x}^2) = \frac{r(r+1)}{\lambda^2} \Rightarrow \sigma^2 = \frac{r}{\lambda^2}$$

تابع مولد گشتاور و تابع مشخصه:

در کاربردهای مختلفی از جمله محاسبهٔ گشتاورهای یک متغیر تصادفی، تابع مولد گشتاور (یا تابع مشخصه) کمک می‌کند.

طبق تعریف، تابع مولد گشتاور (Moment Generating Function) متغیر تصادفی **x** عبارت است از:

$$\Phi_{\mathbf{x}}(s) = E(e^{s\mathbf{x}}) = \int_{-\infty}^{+\infty} e^{sx} f_{\mathbf{x}}(x)dx$$

(البته برای sهایی که این انتگرال وجود داشته باشد.)

یعنی تابع مولد گشتاور، همان تبدیل لاپلاس تابع چگالی است (با تبدیل s به s-).

در حالت گسسته، تابع مولد گشتاور به مجموع زیر تبدیل می‌شود:

$$\Phi_{\mathbf{x}}(s) = E(e^{s\mathbf{x}}) = \sum_{i} e^{sx_i} P_{\mathbf{x}}(x_i)$$

خواص تابع مولد گشتاور:

۱) **قضیهٔ گشتاور:** اصولاً علت اینکه به این تابع، مولد گشتاور می‌گویند، این است که:

$$m_n = E(\mathbf{x}^n) = \Phi^{(n)}(0)$$

یعنی مشتق nام تابع در نقطهٔ صفر، گشتاور nام را می‌دهد.

اثبات:

$$\frac{d^n\Phi(s)}{ds^n} = \int_{-\infty}^{+\infty} x^n e^{sx} f_{\mathbf{x}}(x)dx$$

$$\Rightarrow \left.\frac{d^n\Phi(s)}{ds^n}\right|_{s=0} = \int_{-\infty}^{+\infty} x^n f_{\mathbf{x}}(x)dx = m_n$$

برای $n=0$ نتیجه می‌شود که: $\Phi(0) = m_0 = 1$، زیرا:

$$\Phi(0) = \int_{-\infty}^{+\infty} f_{\mathbf{x}}(x)dx = 1$$

۲) اگر $\mathbf{y} = a\mathbf{x} + b$ باشد، داریم:

$$\Phi_{\mathbf{y}}(s) = e^{bs}\Phi_{\mathbf{x}}(as)$$

زیرا:

$$\Phi_{\mathbf{y}}(s) = E(e^{(a\mathbf{x}+b)s}) = e^{bs}\,E(e^{as\mathbf{x}}) = e^{bs}\Phi_{\mathbf{x}}(as)$$

۳) اگر $\mathbf{y} = g(\mathbf{x})$، بدون نیاز به محاسبهٔ $f_{\mathbf{y}}$ می‌توان $\Phi_{\mathbf{y}}(s)$ را محاسبه کرد:

$$\Phi_{\mathbf{y}}(s) = E(e^{sy}) = \int_{-\infty}^{+\infty} e^{sy} f_{\mathbf{y}}(y)dy = \int_{-\infty}^{+\infty} e^{s\,g(x)} f_{\mathbf{x}}(x)dx$$

طبق قضیهٔ اساسی امید ریاضی

مثال: توزیع نمایی:

$$f(x)=\lambda e^{-\lambda x}u(x)$$

$$\Phi(s)=\int_{0}^{+\infty}e^{sx}\lambda e^{-\lambda x}dx=\lambda\int_{0}^{+\infty}e^{-(\lambda-s)x}dx=\frac{\lambda}{\lambda-s}:\mathrm{Re}(s)<\lambda$$

از جداول تبدیل لاپلاس داریم:

$$e^{-\lambda t}u(t)\xleftrightarrow{\text{Laplace}}\frac{1}{s+\lambda}:\mathrm{Re}(s)>-\lambda$$

پس با تبدیل s به s- خواهیم داشت:

$$\mathrm{Re}(-s)>-\lambda\Rightarrow \mathrm{Re}(s)<\lambda$$

$$\Rightarrow\Phi(s)=\frac{\lambda}{\lambda-s}:\mathrm{Re}(s)<\lambda$$

در خیلی از موارد محاسبهٔ گشتاور از این راه ساده‌تر است.

مثال: توزیع دوجمله‌ای:

$$P(k)=\binom{n}{k}p^k q^{n-k} : k=0,1,\dots,n$$

محاسبهٔ مستقیم η و σ^2 نسبتاً مشکل است.

$$\Phi(s)=\sum_i e^{sx_i}P(x_i)=\sum_{k=0}^{n}e^{sk}\binom{n}{k}p^k q^{n-k}=\sum_{k=0}^{n}\binom{n}{k}(pe^s)^k q^{n-k}=(pe^s+q)^n$$

$$\eta=\left.\frac{d\Phi(s)}{ds}\right|_{s=0}=n\,(pe^s+q)^{n-1}\,pe^s\Big|_{s=0}=n\,(p+q)^{n-1}\,p=np$$

$$E(\mathbf{x}^2)=\left.\frac{d^2\Phi(s)}{ds^2}\right|_{s=0}=npq+n^2p^2$$

$$\Rightarrow\sigma^2=npq$$

تابع مشخصه:

طبق تعریف، تابع مشخصهٔ متغیر تصادفی **x** عبارت است از:

$$\Phi_{\mathbf{x}}(j\omega)=E(e^{j\omega\mathbf{x}})=\int_{-\infty}^{+\infty}e^{j\omega x}f_{\mathbf{x}}(x)dx$$

یعنی همان $\Phi_{\mathbf{x}}(s)$ که به جای s، $j\omega$ قرار گرفته است.

پس همان خواص $\Phi(s)$ را دارد و گاهی اوقات نیز آن را با $\Phi(\omega)$ نشان می‌دهیم.

تابع مشخصه در واقع تبدیل فوریهٔ pdf است (با تبدیل ω به $-\omega$).

با توجه به رابطهٔ عکس تبدیل فوریه داریم:

$$f_{\mathbf{x}}(x)=\frac{1}{2\pi}\int_{-\infty}^{+\infty}e^{-j\omega x}\Phi_{\mathbf{x}}(j\omega)d\omega$$

پس می‌توان f را از روی Φ و Φ را از روی f به دست آورد (تابع مشخصه یگانه است و با داشتن آن، متغیر تصادفی به طور کامل مشخص شده است).

به راحتی ملاحظه می‌شود که:

$$m_n = \left.\frac{d^n\Phi(j\omega)}{d(j\omega)^n}\right|_{j\omega=0} = \frac{1}{j^n}\left.\frac{d^n\Phi(j\omega)}{d\omega^n}\right|_{\omega=0}$$

(اگر تمام گشتاورها را داشته باشیم، تمام مشتقات Φ را در نقطهٔ صفر داریم. پس با بسط مک‌لورن Φ، $\Phi(\omega)$ را برای هر ω داریم و اگر Φ معلوم باشد، f نیز معلوم است.)

بسط مک‌لورن Φ:

$$\Phi(\omega) = 1 + j\omega m_1 + \cdots + \frac{(j\omega)^i}{i!}m_i + \cdots = \sum_{n=0}^{+\infty}\frac{(j\omega)^n}{n!}m_n$$

مثال: برای توزیع نمایی با پارامتر λ داریم:

$$\Phi_{\mathbf{x}}(\omega) = \frac{\lambda}{\lambda - j\omega} \quad (e^{-\lambda t}u(t) \xleftrightarrow{\mathfrak{I}} \frac{1}{\lambda + j\omega} : \mathrm{Re}(\lambda) > 0)$$

مثال: برای $\mathbf{x} \sim N(\eta, \sigma)$، $\Phi_{\mathbf{x}}$ را پیدا کنید.

$$\mathbf{z} = \frac{\mathbf{x} - \eta}{\sigma} \sim N(0,1)$$

$$\Phi_{\mathbf{z}}(j\omega) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{j\omega z} e^{-\frac{z^2}{2}} dz \qquad \to j\omega z - \frac{z^2}{2} = -\frac{1}{2}(z - j\omega)^2 - \frac{\omega^2}{2}$$

$$= e^{-\frac{\omega^2}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-j\omega)^2}{2}} dz = e^{-\frac{\omega^2}{2}}$$

اصولاً به یاد داشته باشید که:

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \xleftrightarrow{\mathfrak{I}} e^{-\frac{\omega^2}{2}}$$

پس اکنون برای $\mathbf{x} = \sigma \mathbf{z} + \eta$ داریم:

$$\Phi_{\mathbf{x}}(j\omega) = e^{j\omega\eta} \Phi_{\mathbf{z}}(j\omega\sigma) = e^{j\omega\eta - \frac{\sigma^2\omega^2}{2}}$$

بهتر است که این رابطه را به خاطر بسپارید.

مثال: برای $\mathbf{x} \sim N(\eta,\sigma)$، μ_nها را پیدا کنید.

$$\mathbf{x} \sim N(\eta,\sigma) \Rightarrow \mathbf{z} = \frac{\mathbf{x}-\eta}{\sigma} \sim N(0,1)$$

$$\mu_n(\mathbf{x}) = E((\mathbf{x}-\eta)^n) = \sigma^n E((\frac{\mathbf{x}-\eta}{\sigma})^n) = \sigma^n m_n(\mathbf{z})$$

$$(\frac{1}{j})^n = (-j)^n$$

$$m_n(\mathbf{z}) = \frac{1}{j^n}\frac{d^n}{d\omega^n}(e^{-\frac{\omega^2}{2}})\Big|_{\omega=0} = j^n(-1)^n\frac{d^n}{d\omega^n}(e^{-\frac{\omega^2}{2}})\Big|_{\omega=0}$$

چندجمله‌ای هرمیت:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}}\frac{d^n}{dx^n}(e^{-\frac{x^2}{2}})$$

$$H_0(x) = 1 \text{ , } H_1(x) = x \text{ , } H_2(x) = x^2 - 1 \text{ , } H_3(x) = x^3 - 3x \text{ , } H_4(x) = x^4 - 6x^2 + 3 \text{ , } \ldots$$

چندجمله‌ای هرمیت دارای ویژگی‌های زیر است:

$$H_{k+1}(x) = xH_k(x) - kH_{k-1}(x)$$

$$H_n(x) = x^n - \binom{n}{2}x^{n-2} + 1\times 3\binom{n}{4}x^{n-4} \mp \cdots$$

پس:

$$m_n(\mathbf{z}) = j^n H_n(0)$$

از طرفی داریم:

$$H_n(0) = \begin{cases} (-1)^{\frac{n}{2}} \times 1\times 3\times\cdots\times(n-1) & n \text{ زوج} \\ 0 & n \text{ فرد} \end{cases}$$

پس:

$$m_n(\mathbf{z}) = \begin{cases} 1\times 3\times\cdots\times(n-1) & \text{زوج } n \\ 0 & \text{فرد } n \end{cases}$$

و لذا:

$$\mu_n(\mathbf{x}) = \begin{cases} 1\times 3\times\cdots\times(n-1)\times\sigma^n & \text{زوج } n \\ 0 & \text{فرد } n \end{cases}$$

مثلاً داریم: $\mu_4 = 4\sigma^4$.

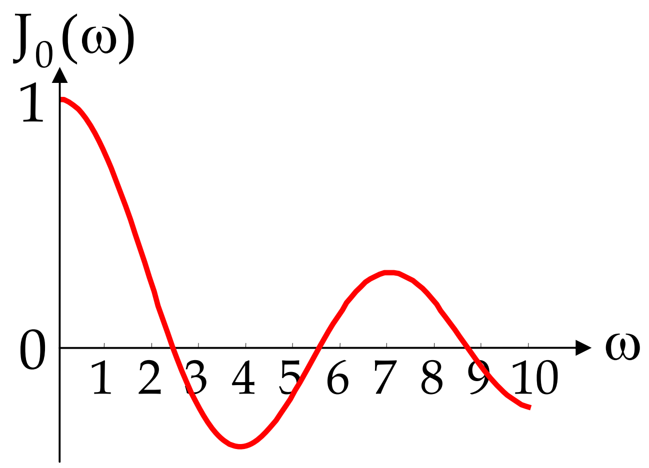
در اینجا حساب کردن گشتاور از این راه چندان ساده‌تر نبود، ولی در خیلی موارد ساده‌تر می‌شود (راه مستقیم در کتاب، ص ۱۵۱).

مثال: اگر $\mathbf{x} \sim a\sin(\Omega t+\theta)$ و $\theta \sim u(0,2\pi)$، $\Phi_{\mathbf{x}}(j\omega)$ را حساب کنید.

با توجه به خاصیت ۳ داریم:

$$\Phi_{\mathbf{x}}(j\omega)=\int_0^{2\pi} e^{j\omega a\sin(\Omega t+\theta)} \overbrace{f_\theta(\theta)}^{\frac{1}{2\pi}} d\theta \qquad \to u=\Omega t+\theta$$

$$=\frac{1}{2\pi}\int_{\Omega t}^{\Omega t+2\pi} e^{j\omega a\sin u} du = J_0(\omega a)$$: تابع بسل نوع اول مرتبهٔ صفر



تابع بسل نوع اول مرتبهٔ n:

$$J_n(x)=\frac{1}{2\pi}\int_c^{c+2\pi} e^{-jnu+jx\sin u} du$$

فصل ۵: دو متغیر تصادفی

Chapter 5

۱. pmf مشترک

۲. CDF مشترک

۳. pdf مشترک

۴. استقلال دو متغیر تصادفی

۵. امید ریاضی و همبستگی

۶. گشتاور مشترک

۷. توابعی از دو متغیر تصادفی

۸. متغیرهای تصادفی مشترکاً نرمال

دو متغیر تصادفی:

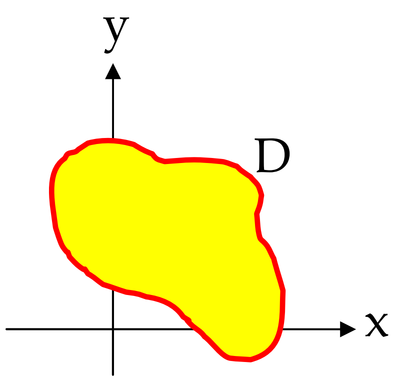
متغیر تصادفی $\mathbf{x}$ تابعی بود که به نقاط ω_i، x_iها را نسبت می‌داد:

$$x_1 = \mathbf{x}(\omega_1), x_2 = \mathbf{x}(\omega_2), \ldots, x_N = \mathbf{x}(\omega_N)$$

به همین ترتیب اگر متغیر تصادفی دیگری مانند $\mathbf{y}$ را روی همین فضای نمونه در نظر بگیریم، به نقاط ω_i، اعداد دیگری (y_iها) را نسبت می‌دهد:

$$y_1 = y(\omega_1), y_2 = \mathbf{y}(\omega_2), \ldots, y_N = \mathbf{y}(\omega_N)$$

در رابطه با $\mathbf{x}$ به دنبال احتمال واقعه‌هایی مثل $\{\mathbf{x} = x_1\}$ یا $\{\mathbf{x} \le x_0\}$ و امثالهم بودیم و در مورد $\mathbf{y}$ نیز به همین ترتیب.



حال می‌توان با در نظر گرفتن زوج مرتب‌های (x_i, y_i)، احتمال واقعه‌هایی مثل $\{\mathbf{x} = x_1, \mathbf{y} = y_1\}$ یا $\{\mathbf{x} \le x_0, \mathbf{y} \le y_0\}$ و به طور کلی $\{(\mathbf{x},\mathbf{y})\} \in D$ (یعنی $\{\omega : (\mathbf{x}(\omega), \mathbf{y}(\omega))\} \in D$) را مطرح کرد.

در حالت کلی با صرف داشتن توزیع $\mathbf{x}$ و توزیع $\mathbf{y}$، احتمال این واقعه‌ها (دو بعدی) مشخص نمی‌شود (مگر آنکه $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند که بعداً خواهیم دید) و برای مشخص کردن احتمال این واقعه‌ها به ابزارهای جدیدی نیاز داریم.

تابع احتمال مشترک (Joint Probability Function or Joint pmf):

اگر **x** و **y** متغیرهای تصادفی گسسته باشند، آنگاه:

$$P_{\mathbf{xy}}(x,y) = P\{\mathbf{x}=x, \mathbf{y}=y\}$$

اگر مقادیر ممکنه برای **x**، x_1، x_2، x_3 و ... و برای **y**، y_1، y_2، y_3 و ... باشند، داریم:

$$P_{\mathbf{xy}}(x,y) = \begin{cases} p_{ij} & x=x_i\,, y=y_i \\ 0 & \text{otherwise} \end{cases}$$

پس خواهیم داشت:

$$\sum_i \sum_j P_{\mathbf{xy}}(x_i, y_j) = \sum_i \sum_j p_{ij} = 1$$

از طرف دیگر چون می‌دانیم که:

$$\{\mathbf{x}=x_{\mathbf{i}}\}=\bigcup_{j}\{\mathbf{x}=x_{\mathbf{i}},\mathbf{y}=y_j\}$$

پس:

$$P\{\mathbf{x}=x_{\mathbf{i}}\}=\sum_{j}P\{\mathbf{x}=x_{\mathbf{i}},\mathbf{y}=y_j\}$$

در نتیجه توابع احتمال حاشیه‌ای (Marginal pmf) زیر را داریم:

$$\begin{cases} P_{\mathbf{x}}(x_i)=\sum_{j}P_{\mathbf{xy}}(x_i,y_j) & :\mathbf{x}\ \text{حاشیه‌ای pmf} \\ P_{\mathbf{y}}(y_j)=\sum_{i}P_{\mathbf{xy}}(x_i,y_j) & :\mathbf{y}\ \text{حاشیه‌ای pmf} \end{cases}$$

و برای هر زیر مجموعهٔ D از $\mathbb{R}^2$ داریم:

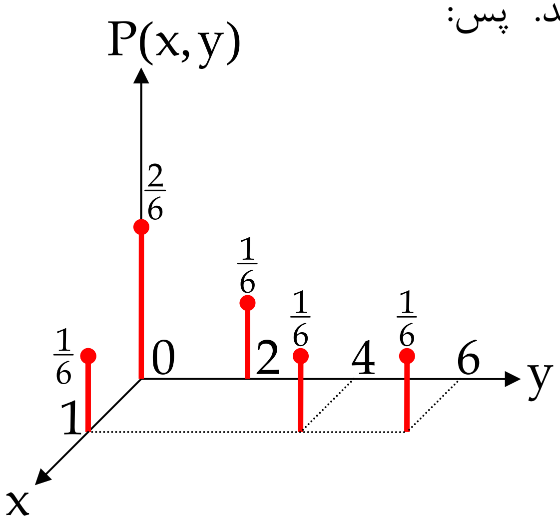
$$P\{(\mathbf{x},\mathbf{y})\in D\}=\sum_{(x_i,y_j)\in D}P(x_i,y_j)$$

مثال: تاسی را می‌اندازیم و متغیرهای تصادفی **x** و **y** را به صورت زیر تعریف می‌کنیم:

$$\mathbf{x}(f_i)=\begin{cases}0 & i<4\\ 1 & i\geq 4\end{cases} \qquad \text{و} \qquad \mathbf{y}(f_i)=\begin{cases}i & \text{زوج } i\\ 0 & \text{فرد } i\end{cases}$$

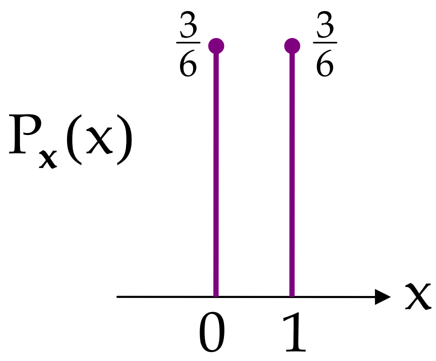
x می‌تواند دو مقدار 0 و 1 را داشته باشد. **y** نیز می‌تواند چهار مقدار 0، 2، 4 و 6 را اختیار کند. پس:

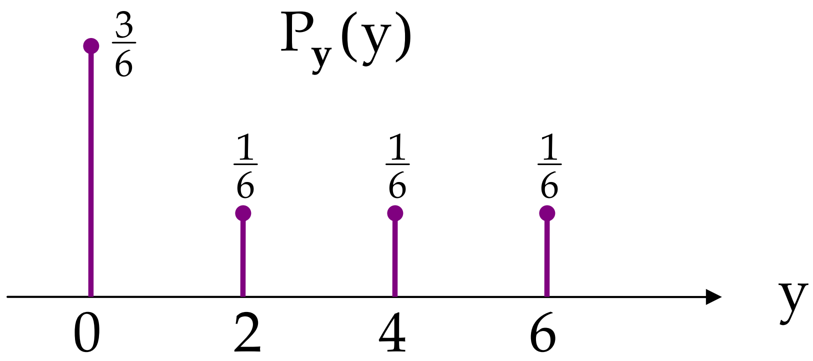
$$1\rightarrow(0,0) \qquad 4\rightarrow(1,4)$$
$$2\rightarrow(0,2) \qquad 5\rightarrow(1,0)$$
$$3\rightarrow(0,0) \qquad 6\rightarrow(1,6)$$



(از ۸تا p_{ij}، ۵تا غیر صفر هستند.)

توابع احتمال **x** و **y** به صورت زیر هستند:





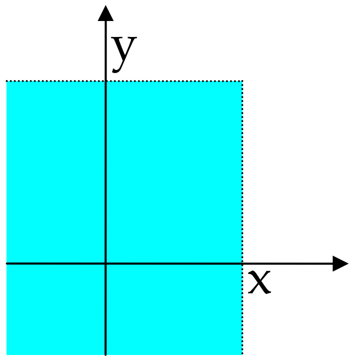
در حالت پیوسته، $P\{\mathbf{x} = x_i , \mathbf{y} = y_j\}$ صفر است و لذا تابع احتمال، مشخص‌کنندهٔ احتمال واقعه‌ها نیست.
لذا CDF مشترک و pdf مشترک را معرفی می‌کنیم.

تابع توزیع انباشتهٔ مشترک (Joint CDF):

طبق تعریف:

$$F_{\mathbf{xy}}(x, y) = P\{\mathbf{x} \le x, \mathbf{y} \le y\}$$

یعنی ناحیهٔ D مطابق شکل مقابل است:



در حالت گسسته:

$$F_{\mathbf{xy}}(x, y) = \sum_{x_i \le x} \sum_{y_j \le y} P_{\mathbf{xy}}(x_i , y_j)$$

مثلاً در مثال قبل داریم:

$$F_{\mathbf{xy}}(0.2, 2) = P(0,0) + P(0,2) = \frac{2}{6} + \frac{1}{6} = \frac{3}{6}$$

برخی خواص CDF مشترک:

1) $F_{\mathbf{xy}}(-\infty,-\infty)=0 \quad , \quad F_{\mathbf{xy}}(-\infty,y)=0 \quad , \quad F_{\mathbf{xy}}(x,-\infty)=0 \quad , \quad F_{\mathbf{xy}}(+\infty,+\infty)=1$

2) $\begin{Bmatrix} F_{\mathbf{xy}}(x,+\infty)=P\{\mathbf{x}\le x\}=F_{\mathbf{x}}(x) \\ F_{\mathbf{xy}}(+\infty,y)=P\{\mathbf{y}\le y\}=F_{\mathbf{y}}(y) \end{Bmatrix}$: (Marginal CDF) توابع CDF حاشیه‌ای

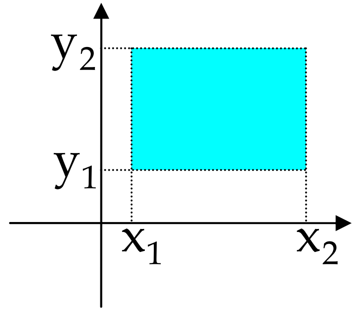
پس برای $y_1 < y_2$ داریم:

$$P\{\mathbf{x}\le x, y_1<\mathbf{y}\le y_2\}=F_{\mathbf{xy}}(x,y_2)-F_{\mathbf{xy}}(x,y_1)$$
$$\Rightarrow F_{\mathbf{xy}}(x,y_2)-F_{\mathbf{xy}}(x,y_1)\ge 0 : \forall x, y_1<y_2$$

و برای $x_1 < x_2$ نیز داریم:

$$P\{x_1<\mathbf{x}\le x_2, \mathbf{y}\le y\}=F_{\mathbf{xy}}(x_2,y)-F_{\mathbf{xy}}(x_1,y)$$
$$\Rightarrow F_{\mathbf{xy}}(x_2,y)-F_{\mathbf{xy}}(x_1,y)\ge 0 : \forall y, x_1<x_2$$

و سرانجام خواهیم داشت:



$$
\begin{aligned}
&P\{x_1 < \mathbf{x} \le x_2, y_1 < \mathbf{y} \le y_2\} \\
&= P\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y_2\} - P\{x_1 < \mathbf{x} \le x_2, \mathbf{y} \le y_1\} \\
&= (P\{\mathbf{x} \le x_2, \mathbf{y} \le y_2\} - P\{\mathbf{x} \le x_1, \mathbf{y} \le y_2\}) - (P\{\mathbf{x} \le x_2, \mathbf{y} \le y_1\} - P\{\mathbf{x} \le x_1, \mathbf{y} \le y_1\}) \\
&= F_{\mathbf{xy}}(x_2, y_2) - F_{\mathbf{xy}}(x_1, y_2) - F_{\mathbf{xy}}(x_2, y_1) + F_{\mathbf{xy}}(x_1, y_1)
\end{aligned}
$$

پس از جمله خواص تابع F این است که مجموع فوق برای هر x_1، x_2، y_1 و y_2 که $x_2 > x_1$ و $y_2 > y_1$، نامنفی است.
یعنی اگر $x_1 < x_2$ و $y_1 < y_2$، داریم:

$$F_{\mathbf{xy}}(x_2, y_2) - F_{\mathbf{xy}}(x_1, y_2) - F_{\mathbf{xy}}(x_2, y_1) + F_{\mathbf{xy}}(x_1, y_1) \ge 0$$

توجه کنید که اگر $\mathbf{x}$ و $\mathbf{y}$ پیوسته باشند، آنگاه: $P\{\mathbf{x} = x, \mathbf{y} = y\} = 0$. یعنی یک نقطهٔ تنها احتمالی ندارد. همچنین داریم: $P\{\mathbf{x} = x\} = 0$. یعنی یک خط هم احتمالی ندارد و حتماً باید ناحیهٔ D سطحی غیرصفر باشد.

تابع چگالی احتمال مشترک (Joint pdf):

طبق تعریف:

$$f_{\mathbf{xy}}(x,y)=\frac{\partial^2 F_{\mathbf{xy}}(x,y)}{\partial x \partial y}$$

از تعریف می‌توان نشان داد که:

$$f_{\mathbf{xy}}(x,y)dxdy=P\{x<\mathbf{x}\le x+dx, y<\mathbf{y}\le y+dy\}$$

لذا:

$$P\{(\mathbf{x},\mathbf{y})\in D\}=\iint_D f_{\mathbf{xy}}(x,y)dxdy$$

یعنی حجم زیر تابع $f_{\mathbf{xy}}(x,y)$ در هر ناحیه، بیانگر احتمال آن ناحیه است.

اگر D را ناحیهٔ $\mathbf{x}\le x_0$ و $\mathbf{y}\le y_0$ در نظر بگیریم و با توجه به تعریف CDF داریم:

$$F_{\mathbf{xy}}(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y} f(u,v)dudv=P\{\mathbf{x}\le x,\mathbf{y}\le y\}$$

به عنوان حالت خاص نیز خواهیم داشت:

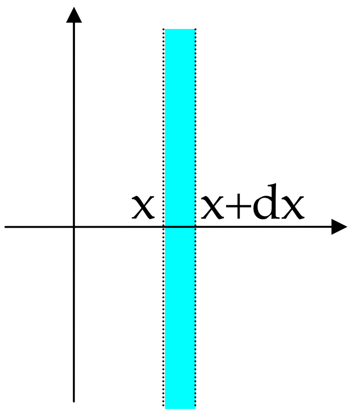
$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = F(+\infty,+\infty) = 1$$

می‌دانیم که: $F_{\mathbf{x}}(x) = F_{\mathbf{xy}}(x,+\infty)$؛ لذا: $f_{\mathbf{x}}(x) = \dfrac{d}{dx}\int_{-\infty}^{x}\int_{-\infty}^{+\infty} f_{\mathbf{xy}}(u,v)\,du\,dv$ و در نتیجه:

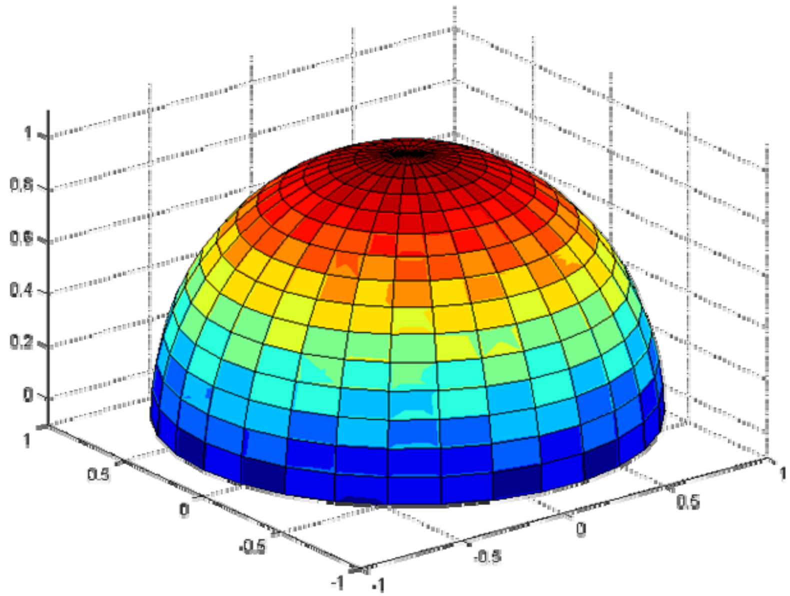
$$f_{\mathbf{x}}(x) = \int_{-\infty}^{+\infty} f_{\mathbf{xy}}(x,y)dy$$ تابع چگالی حاشیه‌ای $\mathbf{x}$:

و مشابهاً داریم:

$$f_{\mathbf{y}}(y) = \int_{-\infty}^{+\infty} f_{\mathbf{xy}}(x,y)dx$$ تابع چگالی حاشیه‌ای $\mathbf{y}$:

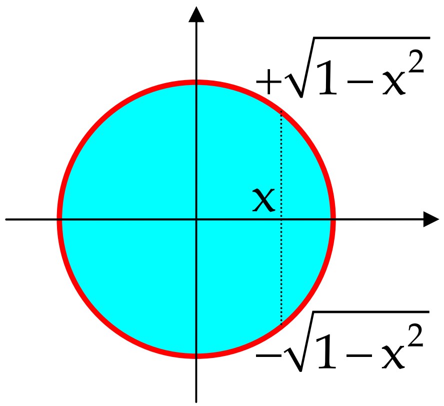


مثال: اگر $f_{xy}(x,y)=\begin{cases} c(1-x^2-y^2) & 0<x^2+y^2<1 \\ 0 & x^2+y^2>1 \end{cases}$ باشد، مقدار c، تابع $f_x(x)$ و مقدار $P\{x^2+y^2<\frac{1}{2}\}$ را محاسبه نمایید.



$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\,dx\,dy=1$$

$$\Rightarrow \iint\limits_{\text{روی دایرهٔ واحد}} c(1-x^2-y^2)\,dx\,dy=1 \Rightarrow c\int_0^{2\pi}\int_0^1 (1-r^2)r\,dr\,d\theta=1 \Rightarrow c=\frac{2}{\pi}$$



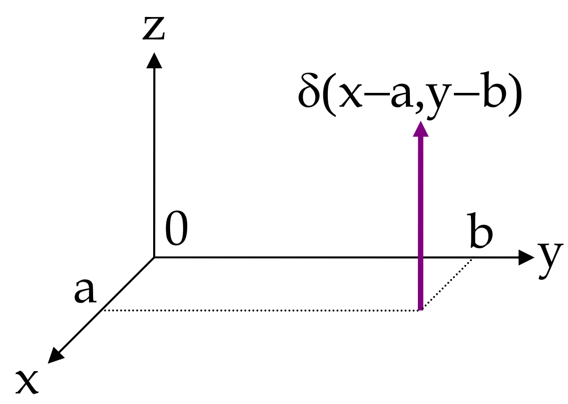
$$f_{\mathbf{x}}(x)=\int_{-\infty}^{+\infty} f_{\mathbf{xy}}(x,y)dy=\begin{cases}\displaystyle\int_{-\sqrt{1-x^2}}^{+\sqrt{1-x^2}}\frac{2}{\pi}(1-x^2-y^2)dy=\frac{8}{3\pi}(1-x^2)^{\frac{3}{2}} & -1<x<1\\ 0 & |x|>1\end{cases}$$

$$P\{x^2+y^2<\frac{1}{2}\}=\underset{\text{روی دایره به شعاع } \frac{\sqrt{2}}{2}}{\iint}\frac{2}{\pi}(1-x^2-y^2)\,dx\,dy=\int_0^{\frac{\sqrt{2}}{2}}\int_0^{2\pi}\frac{2}{\pi}(1-r^2)r\,dr\,d\theta=\frac{3}{4}$$

اگر $\mathbf{x}$ و $\mathbf{y}$ هر دو گسسته باشند، pdf یک سری δ خواهد بود (δی دوبُعدی) و نقاط تنها احتمال دارند (جرم نقطه‌ای).

داریم:

$$\delta(x-a, y-b) = \delta(x-a)\,\delta(y-b)$$



اگر $(a,b) \in D$ باشد، داریم:

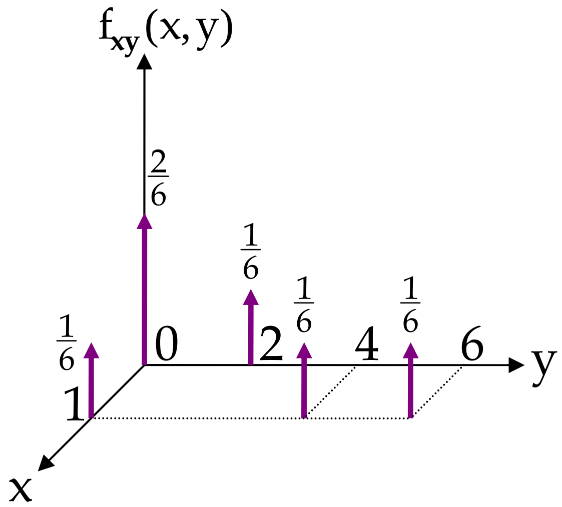
$$\iint_D g(x,y)\,\delta(x-a, y-b)\,dx\,dy = g(a,b)$$

از جمله اگر $g(x,y)=1$ باشد، انتگرال فوق برابر یک می‌شود (زیرا حجم زیر δ یک است).

برای $\mathbf{x}$ و $\mathbf{y}$ گسسته (جرم نقطه‌ای) داریم:

تابع چگالی : $f_{\mathbf{xy}}(x,y)=\sum_i\sum_j p_{ij}\,\delta(x-x_i\,,y-y_j)$

مثلاً در مثال تاس که قبلاً داشتیم:



اگر از دو متغیر تصادفی، یکی پیوسته و دیگری گسسته باشد، احتمال فقط در یک سری صفحات خواهیم داشت (جرم خطی).

مثلاً اگر متغیر تصادفی **x** از نوع گسسته بوده و دارای توزیع زیر باشد:

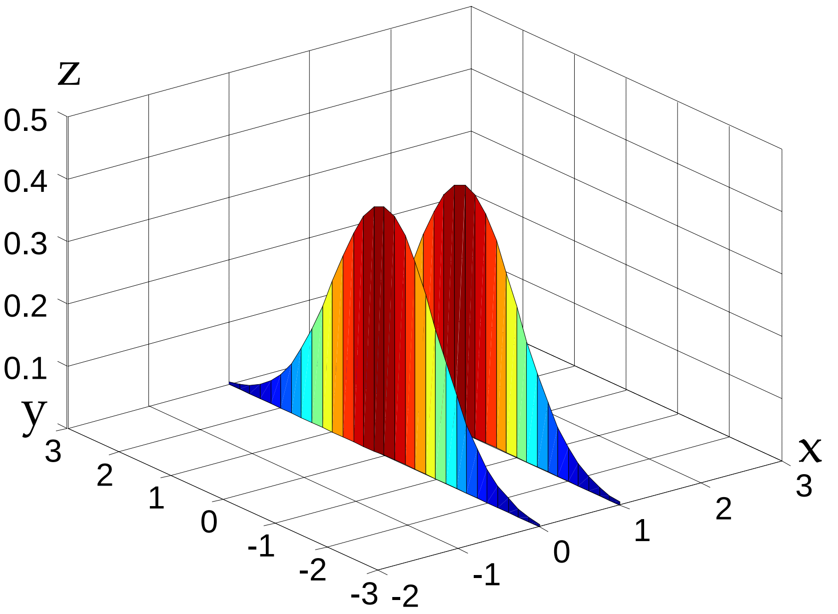
$$P\{\mathbf{x}=0\}=P\{\mathbf{x}=1\}=\frac{1}{2}$$

ولی متغیر تصادفی **y** از نوع پیوسته و دارای توزیع نرمال زیر باشد:

$$f_{\mathbf{y}}(y)=\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$$

تابع چگالی مشترک آنها برابر خواهد بود با:

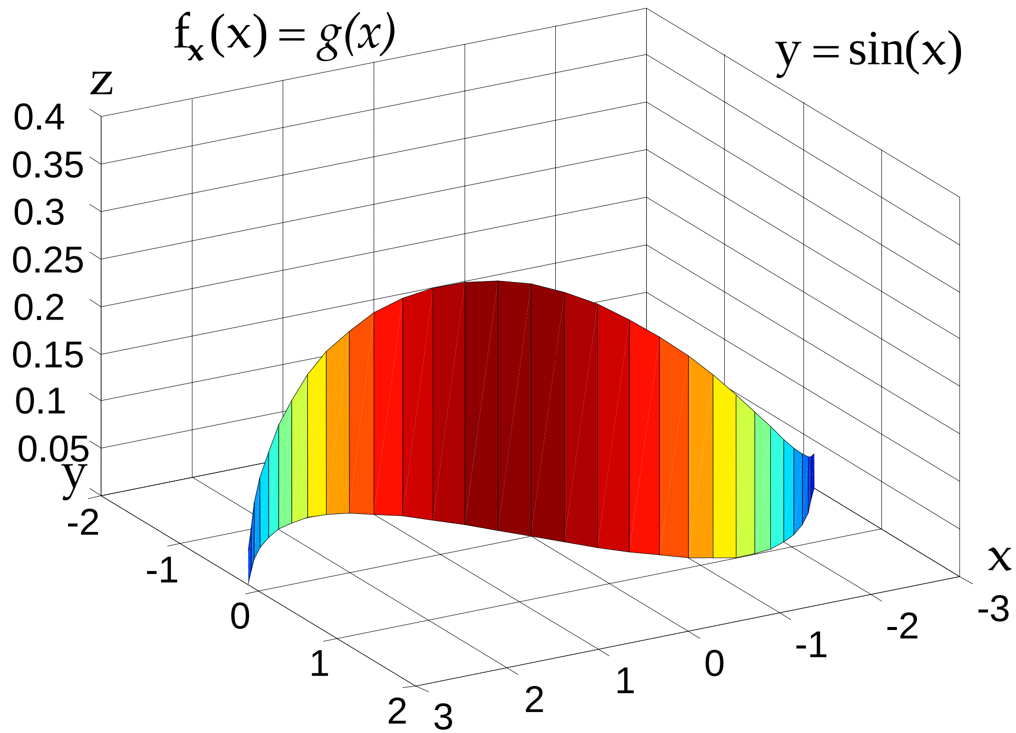
$$f_{\mathbf{xy}}(x,y)=\sum_{i=0}^{1}\frac{1}{2}\delta(x-i)\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}$$



ممکن است $\mathbf{x}$ و $\mathbf{y}$ هر دو پیوسته باشند، ولی جرم خطی داشته باشیم.

اگر $\mathbf{y} = g(\mathbf{x})$ باشد، کلیهٔ نقاط به غیر از $(x,g(x))$ احتمال صفر دارند، یعنی:

$$f_{\mathbf{xy}}(x,y) = f_{\mathbf{x}}(x)\delta(y - g(x))$$



کلاً اگر زوج (x,y)های ممکنه روی یک منحنی در صفحهٔ xy قرار داشته باشند، این حالت پیش می‌آید.

متغیرهای تصادفی مستقل:

دو متغیر تصادفی $\mathbf{x}$ و $\mathbf{y}$ را مستقل گویند، هرگاه برای هر x و y، واقعه‌های $\{\mathbf{x} \le x\}$ و $\{\mathbf{y} \le y\}$ مستقل باشند، یعنی:

$$P\{\mathbf{x} \le x, \mathbf{y} \le y\} = P\{\mathbf{x} \le x\}P\{\mathbf{y} \le y\}$$

یا به عبارت دیگر:

$$F_{\mathbf{xy}}(x,y) = F_{\mathbf{x}}(x)F_{\mathbf{y}}(y)$$

یا معادلاً (اگر از دو طرف $\dfrac{\partial^2}{\partial x \partial y}$ بگیریم):

$$f_{\mathbf{xy}}(x,y) = f_{\mathbf{x}}(x)f_{\mathbf{y}}(y)$$

در حالت گسسته نیز داریم:

$$\underbrace{P\{\mathbf{x} = x_i, \mathbf{y} = y_j\}}_{\downarrow \atop P_{\mathbf{xy}}(x_i,y_j)} = \underbrace{P\{\mathbf{x} = x_i\}}_{\downarrow \atop P_{\mathbf{x}}(x_i)}\underbrace{P\{\mathbf{y} = y_j\}}_{\downarrow \atop P_{\mathbf{y}}(y_j)}$$

(اگر هر یک از شروط فوق برقرار باشد، واقعه‌های $\{\mathbf{x} \in D\}$ و $\{\mathbf{y} \in D'\}$ برای هر D و D' مستقل خواهند بود.)

مثال: در مثالی که داشتیم، یعنی: $f_{xy}(x,y)=\begin{cases}\frac{2}{\pi}(1-x^2-y^2) & 0<x^2+y^2<1\\ 0 & x^2+y^2>1\end{cases}$، تابع $f_{xy}(x,y)$ قابل نوشتن به صورت تابعی از x ضربدر تابعی از y نیست. در نتیجه **x** و **y** مستقل نیستند.

(ضمناً دیدیم که:

$$f_x(x)=\frac{8}{3\pi}(1-x^2)^{\frac{3}{2}} : -1<x<1$$

و f_y نیز به همین شکل است و حاصلضرب این دو f_{xy} نمی‌شود.)

مثال: $f_{xy}(x,y)=\dfrac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$؛

$$f_{xy}(x,y)=\left(\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{x^2}{2\sigma^2}}\right)\left(\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{y^2}{2\sigma^2}}\right)=f_x(x)f_y(y)$$

پس $\mathbf{x}$ و $\mathbf{y}$ مستقل بوده و هر دو $N(0,\sigma)$ هستند.

تعریف: گوییم f_{xy} دارای تقارن دایروی (Circular Symmetry) است، هرگاه $f_{xy}(x,y)$ صرفاً تابعی از فاصلهٔ (x,y) از مبدأ باشد. یعنی:

$$f_{xy}(x,y)=g(r):r=\sqrt{x^2+y^2}$$

در هر دو مثال گذشته، f_{xy} دارای تقارن دایروی بود.

قضیهٔ بسیار جالبی داریم که اهمیت توزیع نرمال را نشان می‌دهد. اگر بخواهیم f_{xy} تقارن دایروی داشته باشد و در عین حال $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، $\mathbf{x}$ و $\mathbf{y}$ باید هر دو $N(0,\sigma)$ باشند.

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل بوده و $f_{\mathbf{xy}}$ دارای تقارن دایروی باشد، $\mathbf{x}$ و $\mathbf{y}$ هر دو نرمال با میانگین صفر و واریانس یکسان، یعنی $N(0,\sigma)$ خواهند بود (اثبات در کتاب، ص۱۴۳).

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، آنگاه $\mathbf{z} = g(\mathbf{x})$ و $\mathbf{w} = h(\mathbf{y})$ نیز مستقل خواهند بود.

اثبات در حالت خاصی که g و h صعودی باشند (اثبات کلی در کتاب، ص۱۴۳):

$$\begin{aligned}
P\{g(\mathbf{x}) \le z, h(\mathbf{y}) \le w\} &= P\{\mathbf{x} \le g^{-1}(z), \mathbf{y} \le h^{-1}(w)\} \\
&= P\{\mathbf{x} \le g^{-1}(z)\} P\{\mathbf{y} \le h^{-1}(w)\} \quad \text{(به دلیل استقلال } \mathbf{x} \text{ و } \mathbf{y}) \\
&= P\{g(\mathbf{x}) \le z\} P\{h(\mathbf{y}) \le w\}
\end{aligned}$$

امید ریاضی و همبستگی:

اگر $\mathbf{z}$ تابعی از $\mathbf{x}$ و $\mathbf{y}$ باشد، یعنی: $\mathbf{z} = g(\mathbf{x}, \mathbf{y})$، امید ریاضی آن برابر است با:

$$E(\mathbf{z}) = \int_{-\infty}^{+\infty} z f_{\mathbf{z}}(z) dz$$

ولی همان طور که در مورد تابع یک متغیر تصادفی گفتیم، در اینجا نیز بدون محاسبهٔ $f_{\mathbf{z}}$ می‌توان $E(\mathbf{z})$ را از روی $f_{\mathbf{xy}}$ به دست آورد.
پس با تعمیم قضیهٔ اساسی امید ریاضی داریم:

قضیه:

$$E(g(\mathbf{x}, \mathbf{y})) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f_{\mathbf{xy}}(x, y) dx dy$$

(اثبات در کتاب، صفحهٔ ۱۴۴)

(توجه کنید که این تعریف با تعریف یک‌بُعدی امید ریاضی سازگار است و اگر g صرفاً تابعی از مثلاً $\mathbf{x}$ باشد، به همان بر می‌گردد.)

گاهی برای بیان اینکه امید ریاضی با انتگرال‌گیری روی کدام چگالی به دست آمده، برای آن زیرنویس می‌گذارند , مثلاً برای امید ریاضی بالا می‌نویسند: $E_{\mathbf{xy}}(g(\mathbf{x}, \mathbf{y}))$.

نتیجه : خطی بودن امید ریاضی (در حالت دو بُعدی):

$$E\left[\sum_{i=1}^{n} a_i g_i(\mathbf{x},\mathbf{y})\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\sum_{i=1}^{n} a_i g_i(x,y)\right) f_{\mathbf{xy}}(x,y)\,dx\,dy$$

$$= \sum_{i=1}^{n} a_i \left(\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g_i(x,y) f_{\mathbf{xy}}(x,y)\,dx\,dy\right)$$

$$= \sum_{i=1}^{n} a_i E(g_i(\mathbf{x},\mathbf{y}))$$

یعنی جای مجموع و امید ریاضی را می‌توان عوض کرد.

به عنوان حالت خاص داریم:

$$E(a\mathbf{x} + b\mathbf{y}) = aE(\mathbf{x}) + bE(\mathbf{y})$$

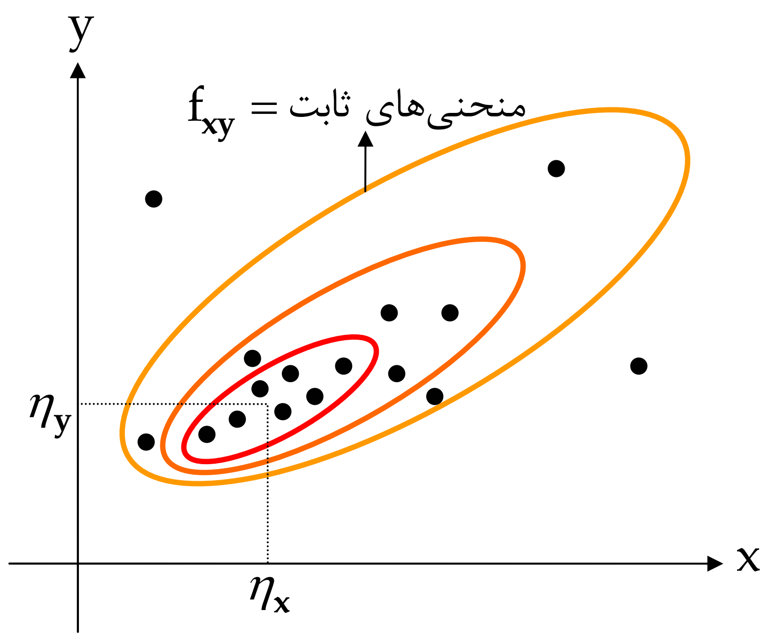
حال توجه کنید که:

$$E(\mathbf{x}) = \eta_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f_{xy}(x,y)\,dx\,dy = \int_{-\infty}^{+\infty} x\, f_x(x)\,dx$$

$$\sigma_x^2 = E((\mathbf{x}-\eta_x)^2) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\eta_x)^2 f_{xy}(x,y)\,dx\,dy = \int_{-\infty}^{+\infty} (x-\eta_x)^2 f_x(x)\,dx$$

و به همین ترتیب برای $\mathbf{y}$ نیز مقادیر η_y و σ_y به همین صورت محاسبه می‌شوند.

نقطهٔ (η_x,η_y)، مرکز ثقل f_{xy} است و σ_x، پراکندگی نقاط تابع توزیع را حول خط $x=\eta_x$ را نشان می‌دهد. σ_y نیز پراکندگی نقاط حول خط $y=\eta_y$ را نشان می‌دهد.



پارامتر پنجم، پارامتری است که میزان بستگی $\mathbf{x}$ و $\mathbf{y}$ را به هم نشان می‌دهد.

طبق تعریف، کُواریانس $\mathbf{x}$ و $\mathbf{y}$ برابر است با:

$$\operatorname{cov}(\mathbf{x},\mathbf{y})=\mu_{\mathbf{xy}}=\mathrm{E}\left[(\mathbf{x}-\eta_{\mathrm{x}})(\mathbf{y}-\eta_{\mathrm{y}})\right]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(\mathrm{x}-\eta_{\mathrm{x}})(\mathrm{y}-\eta_{\mathrm{y}})\mathrm{f}_{\mathbf{xy}}(\mathrm{x},\mathrm{y})\,\mathrm{dx}\,\mathrm{dy}$$

توجه کنید که بر خلاف واریانس که کمیتی نامنفی بود، کواریانس می‌تواند مثبت یا منفی باشد.

(بعضی کتابها کواریانس را با $\sigma_{\mathbf{xy}}$ نشان می‌دهند.)

خواص کواریانس:

۱)

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = \mathrm{E}\left[(\mathbf{x}-\eta_{\mathbf{x}})(\mathbf{y}-\eta_{\mathbf{y}})\right] = \mathrm{E}(\mathbf{xy}) - \eta_{\mathbf{x}}\mathrm{E}(\mathbf{y}) - \eta_{\mathbf{y}}\mathrm{E}(\mathbf{x}) + \eta_{\mathbf{x}}\eta_{\mathbf{y}}$$
$$\Rightarrow \operatorname{cov}(\mathbf{x},\mathbf{y}) = \mathrm{E}(\mathbf{xy}) - \mathrm{E}(\mathbf{x})\mathrm{E}(\mathbf{y})$$

۲)

$$\operatorname{cov}(\mathbf{x},\mathbf{x}) = \mathrm{E}((\mathbf{x}-\eta_{\mathbf{x}})^2)$$
$$\Rightarrow \operatorname{cov}(\mathbf{x},\mathbf{x}) = \operatorname{var}(\mathbf{x})$$

۳) اگر $\mathbf{z}=\mathbf{x}+\mathbf{y}$ باشد، می‌دانیم که: $\eta_{\mathbf{z}}=\eta_{\mathbf{x}}+\eta_{\mathbf{y}}$؛ از طرف دیگر داریم:

$$\operatorname{var}(\mathbf{z})=\mathrm{E}((\mathbf{z}-\eta_{\mathbf{z}})^2)=\mathrm{E}(\left[(\mathbf{x}-\eta_{\mathbf{x}})+(\mathbf{y}-\eta_{\mathbf{y}})\right]^2)=\mathrm{E}((\mathbf{x}-\eta_{\mathbf{x}})^2)+\mathrm{E}((\mathbf{y}-\eta_{\mathbf{y}})^2)+2\mathrm{E}((\mathbf{x}-\eta_{\mathbf{x}})(\mathbf{y}-\eta_{\mathbf{y}}))$$

پس:

$$\operatorname{var}(\mathbf{x}+\mathbf{y}) = \operatorname{var}(\mathbf{x}) + \operatorname{var}(\mathbf{y}) + 2\operatorname{cov}(\mathbf{x},\mathbf{y})$$

یا:

$$\sigma_{\mathbf{x}+\mathbf{y}}^2 = \sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 + 2\mu_{\mathbf{xy}}$$

در حالت کلی‌تر اگر $\mathbf{z} = a\mathbf{x} + b\mathbf{y} + c$ باشد، داریم:

$$\mathrm{var}(\mathbf{z}) = a^2\,\mathrm{var}(\mathbf{x}) + b^2\,\mathrm{var}(\mathbf{y}) + 2ab\,\mathrm{cov}(\mathbf{x},\mathbf{y})$$

تعریف: ضریب همبستگی (Correlation Coefficient) $\mathbf{x}$ و $\mathbf{y}$ به صورت زیر تعریف می‌شود:

$$r_{\mathbf{xy}} = \frac{\mathrm{cov}(\mathbf{x},\mathbf{y})}{\sqrt{\mathrm{var}(\mathbf{x})\,\mathrm{var}(\mathbf{y})}} = \frac{\mu_{\mathbf{xy}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$$

می‌توان نشان داد که:

$$r_{\mathbf{xy}} = \mathrm{cov}\left(\frac{\mathbf{x}-\eta_{\mathbf{x}}}{\sigma_{\mathbf{x}}},\frac{\mathbf{y}-\eta_{\mathbf{y}}}{\sigma_{\mathbf{y}}}\right)$$

یعنی $r_{\mathbf{xy}}$ کواریانس نرمالیزه است. بعداً ثابت خواهیم کرد که: $-1 \le r_{\mathbf{xy}} \le 1$.

گاهی اگر **x** بالاتر از متوسطش برود، معمولاً **y** هم بالاتر از متوسطش خواهد شد. در این صورت: $r > 0$؛

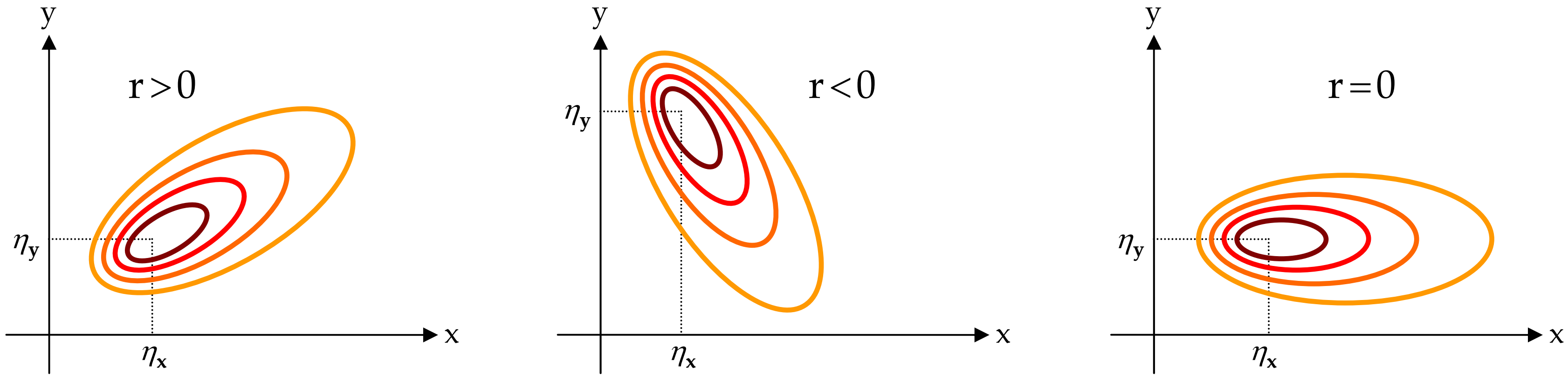
گاهی نیز به عکس اگر **x** بالاتر از متوسطش برود، **y** پایین‌تر از متوسطش خواهد شد. در این صورت: $r < 0$؛

گاهی هم **x** و **y** تأثیری روی هم ندارند. در این صورت: $r = 0$.

مثال: الف) میزان خوردن نوشابه و میزان درجه حرارت؛ علی‌الاصول با افزایش درجه حرارت، میزان خوردن نوشابه نیز افزایش می‌یابد، یعنی: $r > 0$؛

ب) میزان خوردن نوشابه و میزان رطوبت هوا؛ علی‌الاصول با افزایش حرارت، رطوبت کاهش می‌یابد، یعنی: $r < 0$؛

ج) میزان خوردن نوشابه و میزان کتابهای فروخته شده؛ علی‌الاصول ربطی به هم ندارند، یعنی: $r = 0$.



اگر $r = 0$ باشد، محور تقارن موازی یکی از محورها خواهد بود (توجه کنید که ممکن است اصولاً محور تقارنی نداشته باشیم، ولی محور ثقل در حالت کلی داریم).

تعریف: متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ را ناهمبسته (Uncorrelated) گویند، هرگاه $r_{\mathbf{xy}} = 0$ باشد، یعنی:

$$\operatorname{cov}(\mathbf{x},\mathbf{y}) = 0$$

یا معادلاً:

$$R_{\mathbf{xy}} = E(\mathbf{xy}) = E(\mathbf{x})E(\mathbf{y})$$

(در حالی که در حالت کلی داشتیم: $R_{\mathbf{xy}} = E(\mathbf{xy}) = E(\mathbf{x})E(\mathbf{y}) + \operatorname{cov}(\mathbf{x},\mathbf{y})$)

تعریف: متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ را متعامد (Orthogonal) گویند، هرگاه:

$$R_{\mathbf{xy}} = E(\mathbf{xy}) = 0$$

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ ناهمبسته باشند، داریم:

$$\sigma^2_{\mathbf{x+y}} = \sigma^2_{\mathbf{x}} + \sigma^2_{\mathbf{y}}$$

یعنی واریانس مجموع برابر مجموع واریانس‌ها است (اصل سوپرپوزیشن قدرت: سیگنال + نویز).

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ متعامد باشند، داریم:

$$E((\mathbf{x}+\mathbf{y})^2) = E(\mathbf{x}^2) + E(\mathbf{y}^2)$$

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ ناهمبسته باشند، $\mathbf{x}-\eta_{\mathbf{x}}$ و $\mathbf{y}-\eta_{\mathbf{y}}$ متعامدند و برعکس، زیرا:

$$E((\mathbf{x}-\eta_{\mathbf{x}})(\mathbf{y}-\eta_{\mathbf{y}}))=\mu_{\mathbf{xy}}=0$$

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، ناهمبسته خواهند بود (ولی عکس آن لزوماً درست نیست)، زیرا:

$$\begin{aligned}E(\mathbf{xy})&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_{\mathbf{xy}}(x,y)\,dx\,dy\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf_{\mathbf{x}}(x)f_{\mathbf{y}}(y)\,dx\,dy\\&=\left(\int_{-\infty}^{+\infty}xf_{\mathbf{x}}(x)dx\right)\left(\int_{-\infty}^{+\infty}yf_{\mathbf{y}}(y)dy\right)\\&=E(\mathbf{x})E(\mathbf{y})\quad\Rightarrow\quad \mathbf{x}\text{ و }\mathbf{y}\text{ ناهمبسته‌اند}\end{aligned}$$

مثال نقض برای عکس مسأله:

$$f_{\mathbf{x}}(x)=\begin{cases}\frac{1}{2} & -1<x<1\\ 0 & \text{otherwise}\end{cases} \quad , \quad \mathbf{y}=\mathbf{x}^2$$

در اینجا $\mathbf{x}$ و $\mathbf{y}$ مستقل نیستند و داریم:

$$E(\mathbf{x})=0$$

$$E(\mathbf{y})=E(\mathbf{x}^2)=\int_{-1}^{1}\frac{1}{2}x^2dx=\frac{1}{3}$$

$$E(\mathbf{xy})=E(\mathbf{x}^3)=0$$

$$\operatorname{cov}(\mathbf{x},\mathbf{y})=E(\mathbf{xy})-E(\mathbf{x})E(\mathbf{y})=0-0=0$$

یعنی $\mathbf{x}$ و $\mathbf{y}$ ناهمبسته هستند، در حالی که مستقل نیستند.

مثال دیگر: اگر $\mathbf{x}=\cos\varphi$، $\mathbf{y}=\sin\varphi$ و $\varphi\sim u(0,2\pi)$ باشد، در نتیجه $\mathbf{y}=\sqrt{1-\mathbf{x}^2}$ بوده و $\mathbf{x}$ و $\mathbf{y}$ مستقل نخواهند بود، ولی داریم:

$$\operatorname{cov}(\mathbf{x},\mathbf{y})=E(\cos\varphi\sin\varphi)-\underbrace{E(\cos\varphi)}_{0}\underbrace{E(\sin\varphi)}_{0}=\frac{1}{2}E(\sin2\varphi)=0$$

در واقع استقلال بسیار قویتر از ناهمبستگی است. ناهمبستگی فقط روی متوسط‌ها کار می‌کند، ولی استقلال باید در هر نقطه برقرار باشد. ضمناً توجه کنید که اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، $g(\mathbf{x})$ و $h(\mathbf{y})$ نیز ناهمبسته خواهند بود (زیرا $g(\mathbf{x})$ و $h(\mathbf{y})$ نیز مستقلند). در حالی که این حرف را برای ناهمبستگی نمی‌توانیم بزنیم.

نامساوی شوارتز:

$$E^2(\mathbf{xy}) \le E(\mathbf{x}^2)E(\mathbf{y}^2)$$

اثبات: برای هر مقدار ثابت α داریم:

$$E(\alpha\mathbf{x}-\mathbf{y})^2 \ge 0 \Rightarrow \alpha^2E(\mathbf{x}^2)-2\alpha E(\mathbf{xy})+E(\mathbf{y}^2) \ge 0$$

اکنون اگر بگیریم: $\alpha = \dfrac{E(\mathbf{xy})}{E(\mathbf{x}^2)}$، برای این مقدار α داریم:

$$\frac{E^2(\mathbf{xy})}{E(\mathbf{x}^2)}-\frac{2E^2(\mathbf{xy})}{E(\mathbf{x}^2)}+E(\mathbf{y}^2)\ge 0 \Rightarrow E(\mathbf{y}^2) \ge \frac{E^2(\mathbf{xy})}{E(\mathbf{x}^2)} \Rightarrow E^2(\mathbf{xy}) \le E(\mathbf{x}^2)E(\mathbf{y}^2)$$

تساوی را در نامساوی شوارتز وقتی داریم که: $\mathbf{y}=c\mathbf{x}$ باشد، زیرا به ازای $\mathbf{y}=c\mathbf{x}$ داریم:

$$\left.\begin{aligned} E^2(\mathbf{xy}) &= \left[E(c\mathbf{x}^2)\right]^2 = c^2E^2(\mathbf{x}^2) \\ E(\mathbf{x}^2)E(\mathbf{y}^2) &= E(\mathbf{x}^2)E(c^2\mathbf{x}^2) = c^2E^2(\mathbf{x}^2) \end{aligned}\right\} \Rightarrow E^2(\mathbf{xy}) = E(\mathbf{x}^2)E(\mathbf{y}^2)$$

ضمناً $\alpha = \dfrac{E(\mathbf{xy})}{E(\mathbf{x}^2)} = c$ خواهد بود.

به عکس در صورت تساوی داریم:

$$E^2(\mathbf{xy}) = E(\mathbf{x}^2)E(\mathbf{y}^2) \Rightarrow E\left(\left(\frac{E(\mathbf{xy})}{E(\mathbf{x}^2)}\mathbf{x}-\mathbf{y}\right)^2\right)=0 \Rightarrow \mathbf{y}=c\mathbf{x} \text{ in Probability },\ c=\frac{E(\mathbf{xy})}{E(\mathbf{x}^2)}$$

اگر در نامساوی شوارتز به جای $\mathbf{x}$، $\mathbf{x}-\eta_x$ و به جای $\mathbf{y}$، $\mathbf{y}-\eta_y$ قرار دهیم، خواهیم داشت:

$$\left[\operatorname{cov}(\mathbf{x},\mathbf{y})\right]^2 \le \operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{y})$$

یعنی کواریانس گر چه می‌تواند مثبت و منفی شود، اما تغییرات آن از دو سو حدی برابر با $\sqrt{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{y})}$ دارد.

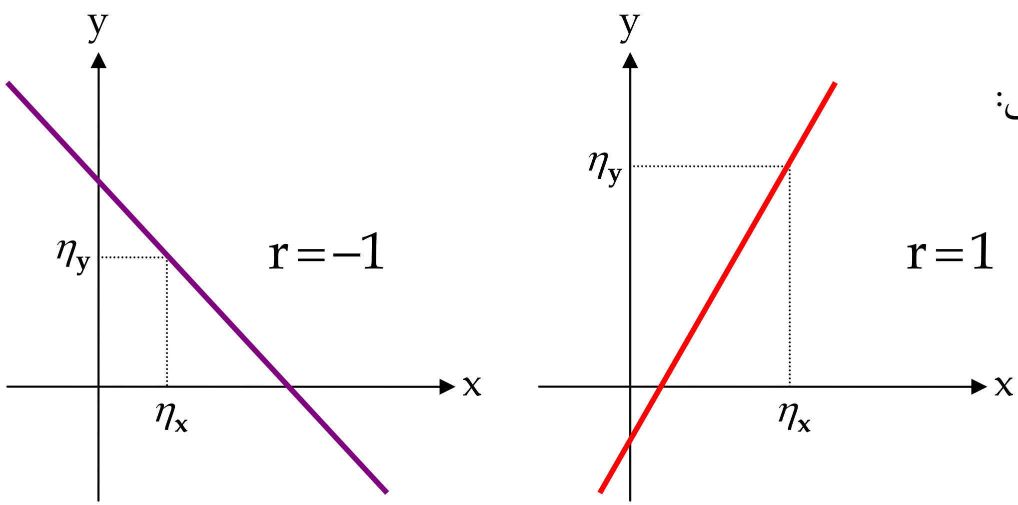
پس در مورد $r_{xy}=\dfrac{\operatorname{cov}(\mathbf{x},\mathbf{y})}{\sqrt{\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{y})}}$ داریم: $r_{xy}^2\le 1$، یعنی: $-1\le r_{xy}\le 1$.

تساوی وقتی برقرار خواهد بود که: $\mathbf{y}-\eta_y=c(\mathbf{x}-\eta_x)$ باشد. در این صورت داریم: $\left[\operatorname{cov}(\mathbf{x},\mathbf{y})\right]^2=\operatorname{var}(\mathbf{x})\operatorname{var}(\mathbf{y})$،

یعنی: $r_{xy}^2=1$. اگر c مثبت باشد، $r_{xy}=1$ و اگر c منفی باشد، $r_{xy}=-1$ خواهد بود، زیرا:

$$c=\frac{\mu_{xy}}{\sigma_x^2}=\frac{r\,\sigma_x\,\sigma_y}{\sigma_x^2}=\pm\frac{\sigma_y}{\sigma_x}$$

جرم خطی:



از جمله کاربردهای ضریب همبستگی، رگرسیون خطی است. کتاب، صفحات ۱۴۹ و ۱۵۰ را ملاحظه نمایید (دربارهٔ رگرسیون خطی، بعداً در فصل ۶ صحبت خواهیم کرد. این مبحث در فصل ۱۱ کتاب به طور کامل آمده است).

گشتاور مشترک:

طبق تعریف، گشتاور مشترک متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ برابر است با:

$$m_{kr} = E(\mathbf{x}^k \mathbf{y}^r) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^k\, y^r\, f_{\mathbf{xy}}(x,y)\,dx\,dy$$

گشتاور مرکزی مشترک:

طبق تعریف، گشتاور مرکزی مشترک متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ برابر است با:

$$\mu_{\mathbf{xy}} = E\left[(\mathbf{x}-\eta_{\mathbf{x}})^k(\mathbf{y}-\eta_{\mathbf{y}})^r\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\eta_{\mathbf{x}})^k\,(y-\eta_{\mathbf{y}})^r\, f_{\mathbf{xy}}(x,y)\,dx\,dy$$

که $k+r$ را مرتبهٔ گشتاور گویند و داریم:

$$m_{11} = R_{\mathbf{xy}} \quad , \quad \mu_{11} = \operatorname{cov}(\mathbf{x},\mathbf{y}) = \mu_{\mathbf{xy}} \quad , \quad \mu_{20} = \sigma_{\mathbf{x}}^2 \quad , \quad \ldots$$

تابع مولد گشتاور مشترک:

$$\Phi_{\mathbf{xy}}(s_1,s_2)=E(e^{s_1\mathbf{x}+s_2\mathbf{y}})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{s_1x+s_2y}f_{\mathbf{xy}}(x,y)\,dx\,dy$$

$$\left.\frac{\partial^k\partial^r\Phi(s_1,s_2)}{\partial s_1^k\partial s_2^r}\right|_{s_1,s_2=0}=m_{kr}$$

تابع مشخصهٔ مشترک:

$$\Phi_{\mathbf{xy}}(j\omega_1,j\omega_2)=E(e^{j(\omega_1\mathbf{x}+\omega_2\mathbf{y})})=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{j(\omega_1x+\omega_2y)}f_{\mathbf{xy}}(x,y)\,dx\,dy$$

$$f_{\mathbf{xy}}(x,y)=\frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-j(\omega_1x+\omega_2y)}\Phi_{\mathbf{xy}}(j\omega_1,j\omega_2)\,d\omega_1\,d\omega_2$$

خواص تابع مشخصهٔ مشترک: (توجه داشته باشید که خواص تابع مولد گشتاور و تابع مشخصه یکی هستند)

۱)

$$\frac{1}{j^{k+r}} \frac{\partial^k \partial^r \Phi_{\mathbf{xy}}(j\omega_1, j\omega_2)}{\partial\omega_1^k \partial\omega_2^r}\bigg|_{\omega_1, \omega_2 = 0} = m_{kr}$$

۲) توابع مشخصهٔ حاشیه‌ای:

$$\Phi_{\mathbf{x}}(j\omega) = \Phi_{\mathbf{xy}}(j\omega, 0)$$
$$\Phi_{\mathbf{y}}(j\omega) = \Phi_{\mathbf{xy}}(0, j\omega)$$

۳) اگر $\mathbf{z} = a\mathbf{x} + b\mathbf{y}$ باشد، آنگاه:

$$\Phi_{\mathbf{z}}(j\omega) = E(e^{j\omega\mathbf{z}}) = E(e^{j(a\omega\mathbf{x} + b\omega\mathbf{y})}) \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega) = \Phi_{\mathbf{xy}}(a\omega, b\omega)$$

اگر $\mathbf{z} = a\mathbf{x} + b\mathbf{y} + c$ باشد، $\Phi_{\mathbf{xy}}(a\omega, b\omega)$ در $e^{j\omega c}$ نیز ضرب می‌شود:

$$\Phi_{\mathbf{z}}(j\omega) = e^{j\omega c}\, \Phi_{\mathbf{xy}}(a\omega, b\omega)$$

۴) متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ مستقلند، اگر و تنها اگر:

$$\Phi_{\mathbf{xy}}(j\omega_1, j\omega_2) = \Phi_{\mathbf{x}}(j\omega_1)\Phi_{\mathbf{y}}(j\omega_2)$$

زیرا اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، $e^{j\omega\mathbf{x}}$ و $e^{j\omega\mathbf{y}}$ نیز مستقل بوده و لذا ناهمبسته نیز خواهند بود. پس خواهیم داشت:

$$\Phi_{\mathbf{xy}}(j\omega_1, j\omega_2) = E(e^{j(\omega_1\mathbf{x}+\omega_2\mathbf{y})}) = E(e^{j\omega_1\mathbf{x}}\, e^{j\omega_2\mathbf{y}}) = E(e^{j\omega_1\mathbf{x}})E(e^{j\omega_2\mathbf{y}}) = \Phi_{\mathbf{x}}(j\omega_1)\Phi_{\mathbf{y}}(j\omega_2)$$

از طرف دیگر اگر تساوی $\Phi_{\mathbf{xy}}(j\omega_1, j\omega_2) = \Phi_{\mathbf{x}}(j\omega_1)\Phi_{\mathbf{y}}(j\omega_2)$ برقرار باشد، داریم:

$$\begin{aligned}
f_{\mathbf{xy}}(x,y) &= \frac{1}{(2\pi)^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\Phi_{\mathbf{xy}}(j\omega_1, j\omega_2)e^{-j(\omega_1 x+\omega_2 y)}d\omega_1 d\omega_2 \\
&= \left(\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Phi_{\mathbf{x}}(j\omega_1)e^{-j\omega_1 x}d\omega_1\right)\left(\frac{1}{2\pi}\int_{-\infty}^{+\infty}\Phi_{\mathbf{y}}(j\omega_2)e^{-j\omega_2 y}d\omega_2\right) \\
&= f_{\mathbf{x}}(x)\, f_{\mathbf{y}}(y)
\end{aligned}$$

۵) اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{z=x+y}$، آنگاه:

$$\Phi_{\mathbf{z}}(j\omega) = \Phi_{\mathbf{xy}}(j\omega, j\omega) = \Phi_{\mathbf{x}}(j\omega)\Phi_{\mathbf{y}}(j\omega)$$

مثال ۱: اگر $\mathbf{x}\sim N(\eta_x,\sigma_x)$ و $\mathbf{y}\sim N(\eta_y,\sigma_y)$ بوده و $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{z=x+y}$، داریم:

$$\Phi_z(j\omega)=\Phi_x(j\omega)\Phi_y(j\omega)=e^{j\omega\eta_x-\frac{\sigma_x^2\omega^2}{2}}e^{j\omega\eta_y-\frac{\sigma_y^2\omega^2}{2}}=e^{j\omega(\eta_x+\eta_y)-\frac{(\sigma_x^2+\sigma_y^2)\omega^2}{2}}$$

یعنی $\mathbf{z}$ هم نرمال است و داریم:

$$\eta_z=\eta_x+\eta_y$$

$$\sigma_z^2=\sigma_x^2+\sigma_y^2$$

قضیه‌ای هست که می‌گوید امکان ندارد مجموع دو متغیر تصادفی مستقل غیرنرمال، نرمال شود. البته اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل نباشند، این امکان وجود دارد (مثال در فصل ۷، کتاب فرآیند Papoulis).

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{x+y}$ نرمال باشد، $\mathbf{x}$ و $\mathbf{y}$ نرمال خواهند بود.

(اثبات در کتاب Lukacs)

مثال ۲: اگر $\mathbf{x}\sim \mathrm{Binomial}(n,p)$ و $\mathbf{y}\sim \mathrm{Binomial}(m,p)$ بوده و $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{z}=\mathbf{x}+\mathbf{y}$، توزیع $\mathbf{z}$ را بیابید.

$$\Phi_{\mathbf{x}}(j\omega)=(pe^{j\omega}+q)^{n}\ ,\ \Phi_{\mathbf{y}}(j\omega)=(pe^{j\omega}+q)^{m}\ ,\ q=1-p$$

$$\Phi_{\mathbf{z}}(j\omega)=\Phi_{\mathbf{x}}(j\omega)\Phi_{\mathbf{y}}(j\omega)=(pe^{j\omega}+q)^{n+m}$$

یعنی:

$$\mathbf{z}\sim \mathrm{Binomial}(n+m,p)$$

البته این امر که کانولوشن یک توزیع در خودش، توزیعی از نوع خودش شود، در معدود توابعی صادق است.

توابعی از دو متغیر تصادفی:

اگر تابعی از دو متغیر تصادفی داشته باشیم، یعنی:

$$\mathbf{z}=g(\mathbf{x},\mathbf{y})$$

$\mathbf{z}$ خود یک متغیر تصادفی است که برای هر نقطه از فضای نمونه مثل ω داریم:

$$\mathbf{z}(\omega)=g(\mathbf{x}(\omega),\mathbf{y}(\omega))$$

برای به دست آوردن توزیع $\mathbf{z}$، می‌نویسیم:

$$F_{\mathbf{z}}(z)=P\{\mathbf{z}\le z\}=P\{g(\mathbf{x},\mathbf{y})\le z\}$$

اگر ناحیه‌ای در صفحهٔ $x\text{-}y$ را که برای آن $g(x,y)\le z$ می‌شود، $D(z)$ بنامیم، خواهیم داشت:

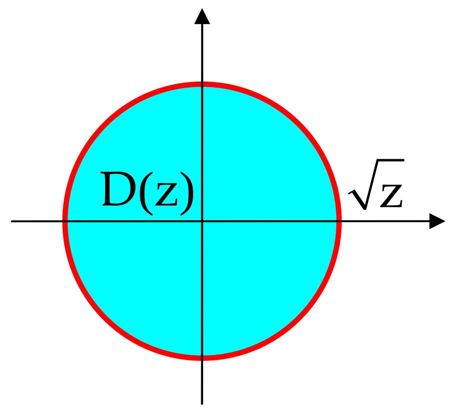
$$F_{\mathbf{z}}(z)=P\{(\mathbf{x},\mathbf{y})\in D(z)\}=\iint_{D(z)} f_{\mathbf{xy}}(x,y)\,dx\,dy$$

مثال ۱: اگر $\mathbf{z}=\mathbf{x}^2+\mathbf{y}^2$ و $f_{\mathbf{xy}}(x,y)=\dfrac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$ (یعنی **x** و **y** هر یک $N(0,\sigma)$ و مستقل از هم) باشند، داریم:

برای $z<0$:

$$F_{\mathbf{z}}(z)=P\{\mathbf{z}\le z\}=0$$

برای $z>0$:



$$F_{\mathbf{z}}(z)=P\{\mathbf{x}^2+\mathbf{y}^2\le z\}=\iint\limits_{D(z)} f_{\mathbf{xy}}(x,y)dxdy=\frac{1}{2\pi\sigma^2}\iint\limits_{D(z)} e^{-\frac{x^2+y^2}{2\sigma^2}}dxdy$$

$$=\frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_0^{\sqrt{z}} e^{-\frac{r^2}{2\sigma^2}}r\,dr\,d\theta=1-e^{-\frac{z}{2\sigma^2}}$$

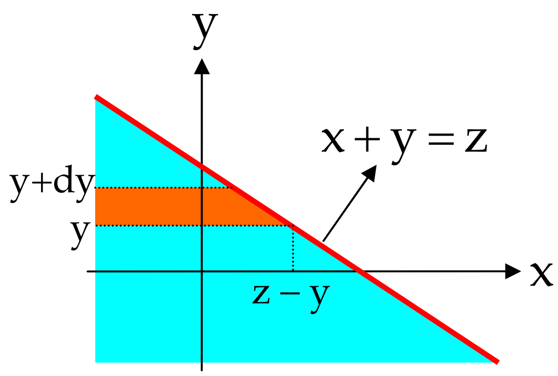
پس:

$$F_{\mathbf{z}}(z)=(1-e^{-\frac{z}{2\sigma^2}})u(z)$$

$$f_{\mathbf{z}}(z)=\frac{dF_{\mathbf{z}}(z)}{dz}=\frac{1}{2\sigma^2}e^{-\frac{z}{2\sigma^2}}u(z)$$ توزیع نمایی :

(برای $\sigma=1$، χ^2 با ۲ درجه آزادی است. اگر $\mathbf{z}=\mathbf{x}_1^2+\mathbf{x}_2^2+\cdots+\mathbf{x}_n^2$ بود، χ^2 با n درجه آزادی می‌شد.)

مثال ۲: $\mathbf{z}=\mathbf{x}+\mathbf{y}$ که $\mathbf{x}$ و $\mathbf{y}$ متغیرهای تصادفی دلخواه هستند. داریم:



$$F_{\mathbf{z}}(z)=P\{\mathbf{x}+\mathbf{y}\le z\}=\iint\limits_{x+y\le z} f_{\mathbf{xy}}(x,y)dxdy=\int_{-\infty}^{+\infty}\int_{-\infty}^{z-y} f_{\mathbf{xy}}(x,y)dxdy$$

$$f_{\mathbf{z}}(z)=\frac{dF_{\mathbf{z}}(z)}{dz}=\int_{-\infty}^{+\infty} f_{\mathbf{xy}}(z-y,y)dy$$

برای یافتن $f(z)$، با مشتق گرفتن از انتگرالهایی که حدود آن توابعی از z هستند روبرو می‌شویم.

یادآوری از حساب دیفرانسیل و انتگرال:

قاعدهٔ لایبنیتز:

$$\frac{\partial}{\partial z}\int_{a(z)}^{b(z)} g(x,z)dx=\frac{db}{dz}g(b,z)-\frac{da}{dz}g(a,z)+\int_{a(z)}^{b(z)}\frac{\partial}{\partial z}g(x,z)dx$$

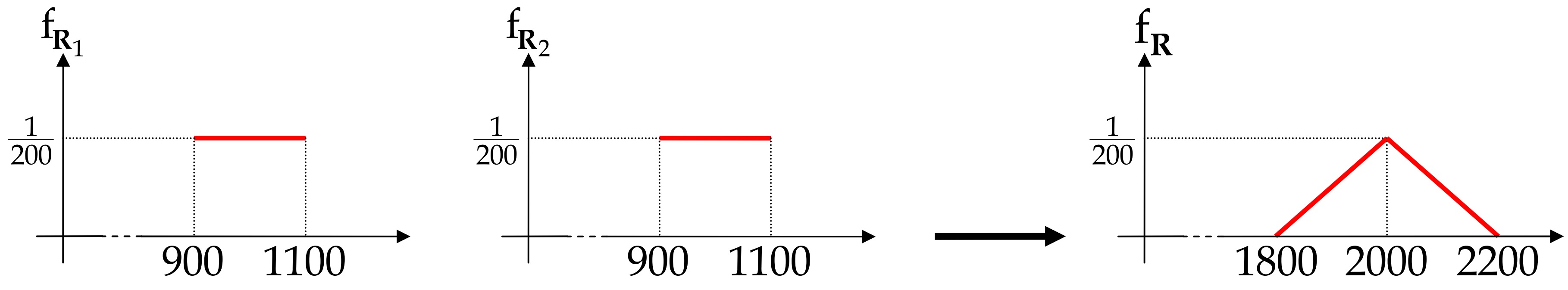
حال اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{z=x+y}$ باشد، داریم:

$$f_{\mathbf{z}}(z)=\int_{-\infty}^{+\infty} f_{\mathbf{x}}(z-y)f_{\mathbf{y}}(y)dy=\int_{-\infty}^{+\infty} f_{\mathbf{y}}(z-x)f_{\mathbf{x}}(x)dx$$

یعنی کانولوشن $f_{\mathbf{x}}$ و $f_{\mathbf{y}}$. قبلاً هم دیده بودیم که اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند و $\mathbf{z=x+y}$ باشد، داریم:

$$\Phi_{\mathbf{z}}(j\omega)=\Phi_{\mathbf{x}}(j\omega)\Phi_{\mathbf{y}}(j\omega)\rightarrow f_{\mathbf{z}}=f_{\mathbf{x}}*f_{\mathbf{y}}$$

مثال: اگر $\mathbf{R}_1\sim u(900,1100)$ و $\mathbf{R}_2\sim u(900,1100)$ و مقادیر $\mathbf{R}_1$ و $\mathbf{R}_2$ مستقل باشند، تابع چگالی $\mathbf{R}=\mathbf{R}_1+\mathbf{R}_2$ را بیابید.



توزیع مثلثی (سیمسون Simpson)

به روش دیگری نیز می‌توانیم مستقیماً $f_{\mathbf{z}}$ را به دست آوریم:

$$\mathbf{z}=g(\mathbf{x},\mathbf{y})$$

$$f_{\mathbf{z}}(z)dz=P\{z\le\mathbf{z}\le z+dz\}=P\{z\le g(\mathbf{x},\mathbf{y})\le z+dz\}$$

اگر ناحیهٔ دیفرانسیلی از صفحهٔ $x\text{-}y$ را که در آن $z\le g(x,y)\le z+dz$ است، $\Delta D(z)$ بنامیم، داریم:

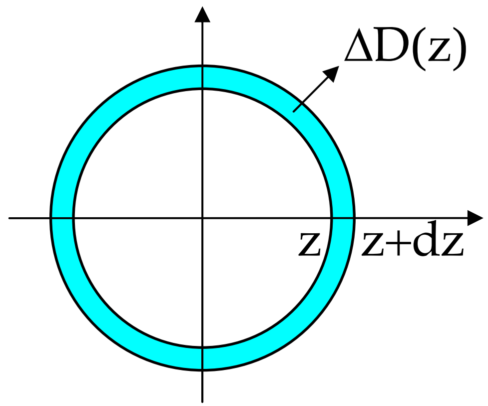
$$f_{\mathbf{z}}(z)dz=P\{(\mathbf{x},\mathbf{y})\in\Delta D(z)\}=\iint_{\Delta D(z)} f_{\mathbf{xy}}(x,y)dxdy$$

مثال: اگر $\mathbf{z}=\sqrt{\mathbf{x}^2+\mathbf{y}^2}$ و $\mathbf{x}\sim N(0,\sigma)$ و $\mathbf{y}\sim N(0,\sigma)$ و $\mathbf{x}$ و $\mathbf{y}$ مستقل از هم باشند، داریم:

برای $z<0$:

$$f_{\mathbf{z}}(z)dz=P\{z\le\mathbf{z}\le z+dz\}=0$$

برای $z>0$:



$$f_{\mathbf{z}}(z)dz=P\{z\le\sqrt{\mathbf{x}^2+\mathbf{y}^2}\le z+dz\}=\iint_{\Delta D(z)}\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}dxdy$$

$$=\frac{1}{2\pi\sigma^2}\int_0^{2\pi}\int_z^{z+dz}e^{-\frac{r^2}{2\sigma^2}}r\,dr\,d\theta=\left(\frac{1}{2\pi}\int_0^{2\pi}d\theta\right)\underbrace{\left(\frac{1}{\sigma^2}\int_z^{z+dz}re^{-\frac{r^2}{2\sigma^2}}dr\right)}$$

$$=\frac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}dz$$

این عبارت برابر است با dz ضربدر مقدار تابع در نقطهٔ z، زیرا: $\int_z^{z+dz}g(x)dx=g(z)dz$

یعنی:

$$f_{\mathbf{z}}(z)=\frac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}u(z)$$: توزیع رایلی

حال اگر دو تابع از $\mathbf{x}$ و $\mathbf{y}$ داشته باشیم، یعنی:

$$\begin{cases} \mathbf{z}=g(\mathbf{x},\mathbf{y}) \\ \mathbf{w}=h(\mathbf{x},\mathbf{y}) \end{cases}$$

قبلاً فهمیدیم که چگونه تابع توزیع (یا چگالی) هر یک را با داشتن $f_{\mathbf{xy}}$ به دست آوریم. اما $f_{\mathbf{zw}}$ چه خواهد بود؟

با تعمیم قضیه‌ای که در حالت تک متغیر تصادفی داشتیم، می‌توان نشان داد که:

قضیه: اگر دستگاه معادلات $\begin{cases} z=g(x,y) \\ w=h(x,y) \end{cases}$ دارای ریشه‌های (x_1,y_1)، (x_2,y_2) و ... باشد، آنگاه:

$$f_{\mathbf{zw}}(z,w)=\sum_i \frac{f_{\mathbf{xy}}(x_i,y_i)}{|J(x_i,y_i)|}$$

که:

$$J(x,y)=\det\begin{bmatrix} \dfrac{\partial g(x,y)}{\partial x} & \dfrac{\partial g(x,y)}{\partial y} \\ \dfrac{\partial h(x,y)}{\partial x} & \dfrac{\partial h(x,y)}{\partial y} \end{bmatrix}$$

اثبات در صفحهٔ ۱۵۷ کتاب.

یا معادلاً می‌توان نوشت:

$$f_{\mathbf{zw}}(z,w)=\sum_i f_{\mathbf{xy}}(x_i(z,w),y_i(z,w))|\det\begin{bmatrix} \dfrac{\partial x_i}{\partial z} & \dfrac{\partial x_i}{\partial w} \\ \dfrac{\partial y_i}{\partial z} & \dfrac{\partial y_i}{\partial w} \end{bmatrix}|$$

روشن است که از این روش می‌توان $f_{\mathbf{zw}}$ و از آنجا $f_{\mathbf{z}}$ یا $f_{\mathbf{w}}$ را به دست آورد (به جای روشهای گذشته).

مثال ۱: $\mathbf{z}=\sqrt{\mathbf{x}^2+\mathbf{y}^2}$ و $\mathbf{w}=\dfrac{\mathbf{x}}{\mathbf{y}}$؛

ابتدا دستگاه $\begin{cases}\mathbf{z}=\sqrt{\mathbf{x}^2+\mathbf{y}^2}\\ \mathbf{w}=\dfrac{\mathbf{x}}{\mathbf{y}}\end{cases}$ را حل می‌کنیم. برای $z>0$، این دستگاه دو ریشه دارد:

$$\begin{cases}x_1=\dfrac{zw}{\sqrt{1+w^2}}\\ y_1=\dfrac{z}{\sqrt{1+w^2}}\end{cases},\quad \begin{cases}x_2=\dfrac{-zw}{\sqrt{1+w^2}}\\ y_2=\dfrac{-z}{\sqrt{1+w^2}}\end{cases}$$

$$J(x,y)=\det\begin{bmatrix}\dfrac{\partial\sqrt{x^2+y^2}}{\partial x} & \dfrac{\partial\sqrt{x^2+y^2}}{\partial y}\\ \dfrac{\partial(\frac{x}{y})}{\partial x} & \dfrac{\partial(\frac{x}{y})}{\partial y}\end{bmatrix}=\det\begin{bmatrix}\dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}}\\ \dfrac{1}{y} & \dfrac{-x}{y^2}\end{bmatrix}=-\dfrac{\dfrac{x^2}{y^2}+1}{\sqrt{x^2+y^2}}$$

$$\Rightarrow J(x_1,y_1)=J(x_2,y_2)=-\dfrac{1+w^2}{z}$$

$$f_{zw}(z,w)=\frac{f_{xy}(x_1,y_1)}{|J(x_1,y_1)|}+\frac{f_{xy}(x_2,y_2)}{|J(x_2,y_2)|}:z>0$$

$$=\frac{z}{1+w^2}\left[f_{xy}\left(\frac{zw}{\sqrt{1+w^2}},\frac{z}{\sqrt{1+w^2}}\right)+f_{xy}\left(\frac{-zw}{\sqrt{1+w^2}},\frac{-z}{\sqrt{1+w^2}}\right)\right]u(z)$$

مثلاً اگر $f_{xy}(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$ (یعنی $\mathbf{x}$ و $\mathbf{y}$ هر دو $N(0,\sigma)$ و مستقل از هم باشند)، داریم:

$$f_{zw}(z,w)=\frac{2z}{1+w^2}\frac{1}{2\pi\sigma^2}e^{-\frac{z^2}{2\sigma^2}}u(z)$$

یعنی:

توزیع کوشی : $f_w(w)=\frac{\frac{1}{\pi}}{1+w^2}$ و توزیع رایلی : $f_z(z)=\frac{z}{\sigma^2}e^{-\frac{z^2}{2\sigma^2}}u(z)$

پس $\mathbf{z}$ و $\mathbf{w}$ مستقل هستند (بنابراین اگر $\mathbf{x},\mathbf{y}\sim N(0,\sigma)$ و مستقل باشند، $\sqrt{\mathbf{x}^2+\mathbf{y}^2}$ و $\frac{\mathbf{x}}{\mathbf{y}}$ نیز مستقلند).

یا از روش معادل داریم:

$$f_{zw}(z,w)=\sum_i f_{xy}(x_i,y_i)|\det\begin{bmatrix}\dfrac{\partial x_i}{\partial z} & \dfrac{\partial x_i}{\partial w}\\ \dfrac{\partial y_i}{\partial z} & \dfrac{\partial y_i}{\partial w}\end{bmatrix}|$$

در اینجا:

$$\det\begin{bmatrix}\dfrac{\partial x_1}{\partial z} & \dfrac{\partial x_1}{\partial w}\\ \dfrac{\partial y_1}{\partial z} & \dfrac{\partial y_1}{\partial w}\end{bmatrix}=\det\begin{bmatrix}\dfrac{w}{\sqrt{1+w^2}} & \dfrac{z}{(1+w^2)^{\frac{3}{2}}}\\ \dfrac{1}{\sqrt{1+w^2}} & \dfrac{-zw}{(1+w^2)^{\frac{3}{2}}}\end{bmatrix}=\frac{-z}{1+w^2}=\det\begin{bmatrix}\dfrac{\partial x_2}{\partial z} & \dfrac{\partial x_2}{\partial w}\\ \dfrac{\partial y_2}{\partial z} & \dfrac{\partial y_2}{\partial w}\end{bmatrix}$$

لذا:

$$f_{zw}(z,w)=\frac{z}{1+w^2}\left[f_{xy}\left(\frac{zw}{\sqrt{1+w^2}},\frac{z}{\sqrt{1+w^2}}\right)+f_{xy}\left(\frac{-zw}{\sqrt{1+w^2}},\frac{-z}{\sqrt{1+w^2}}\right)\right]u(z)$$

مثال ۲: تبدیل خطی $\begin{cases}\mathbf{z}=a_{11}\mathbf{x}+a_{12}\mathbf{y}\\ \mathbf{w}=a_{21}\mathbf{x}+a_{22}\mathbf{y}\end{cases}$ یا $\begin{bmatrix}\mathbf{z}\\ \mathbf{w}\end{bmatrix}=A\begin{bmatrix}\mathbf{x}\\ \mathbf{y}\end{bmatrix}$ که $A=\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}$؛

ابتدا دستگاه $\begin{bmatrix}z\\ w\end{bmatrix}=A\begin{bmatrix}x\\ y\end{bmatrix}$ را حل می‌کنیم. با فرض $\det(A)\neq 0$، این دستگاه یک ریشه دارد:

$$\begin{bmatrix}x\\ y\end{bmatrix}=B\begin{bmatrix}z\\ w\end{bmatrix} \quad\rightarrow\quad B=A^{-1} \quad\rightarrow\quad b_{11}=\frac{a_{22}}{a_{11}a_{22}-a_{12}a_{21}},\cdots$$

یعنی:

$$\begin{cases}x=b_{11}z+b_{12}w\\ y=b_{21}z+b_{22}w\end{cases}$$

$$J(x,y)=\det\begin{bmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}=\det(A)=a_{11}a_{22}-a_{12}a_{21}=\frac{1}{\det(B)}$$

$$f_{\mathbf{zw}}(z,w)=\frac{f_{\mathbf{xy}}(b_{11}z+b_{12}w,b_{21}z+b_{22}w)}{\left|\det(A)\right|}$$

(در روش معادل هم $\left|\det(B)\right|$ ضرب می‌شود.)

مثلاً اگر $\begin{cases}\mathbf{z}=\mathbf{x}+\mathbf{y}\\ \mathbf{w}=\mathbf{x}-\mathbf{y}\end{cases}$، داریم:

$$\det(A)=\det\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}=-2$$

و

یک ریشه : $\begin{cases}x_1=\dfrac{z+w}{2}\\ y_1=\dfrac{z-w}{2}\end{cases}$

$$\Rightarrow f_{\mathbf{zw}}(z,w)=\frac{1}{2}f_{\mathbf{xy}}(\frac{z+w}{2},\frac{z-w}{2})$$

برای مثال اگر $\mathbf{x},\mathbf{y}\sim N(0,\sigma)$ و مستقل از هم باشند، خواهیم داشت:

$$f_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}$$

$$f_{\mathbf{zw}}(z,w)=\frac{1}{2}\times\frac{1}{2\pi\sigma^2}e^{-\frac{(z+w)^2+(z-w)^2}{8\sigma^2}}=\frac{1}{4\pi\sigma^2}e^{-\frac{z^2+w^2}{4\sigma^2}}=\left(\frac{1}{\sqrt{2\pi}\sqrt{2}\,\sigma}e^{-\frac{z^2}{2(\sqrt{2}\,\sigma)^2}}\right)\left(\frac{1}{\sqrt{2\pi}\sqrt{2}\,\sigma}e^{-\frac{w^2}{2(\sqrt{2}\,\sigma)^2}}\right)$$

یعنی $\mathbf{z}$ و $\mathbf{w}$ هر یک $N(0,\sqrt{2}\,\sigma)$ بوده و مستقل از هم هستند. یعنی برای توزیع نرمال، اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل بوده چگالی احتمال یکسان $N(0,\sigma)$ داشته باشند، $\mathbf{x}+\mathbf{y}$ و $\mathbf{x}-\mathbf{y}$ نیز مستقل و نرمال هستند.

کلاً با به دست آوردن $f_{\mathbf{zw}}$ می‌توانیم توابع چگالی حاشیه‌ای را نیز به دست آوریم.

گاهی وقتی که فقط با یک تابع $z=g(x,y)$ مواجهیم و بخواهیم از روی f_{xy}، f_z را به دست آوریم، به جای دو روشی که قبلاً گفتیم، به کار بردن متغیر کمکی w و استفاده از قضیهٔ فوق می‌تواند کمک کند که f_z را از این طریق به دست آوریم.

پس روش دیگر برای به دست آوردن توزیع یک تابع از دو متغیر تصادفی، **استفاده از متغیر تصادفی کمکی** است.

مثال: $z=xy$؛

روش اول:

$$F_z(z)=\iint_{D(z)} f_{xy}(x,y)dxdy$$

روش دوم:

$$f_z(z)dz=\iint_{\Delta D(z)} f_{xy}(x,y)dxdy$$

روش سوم: استفاده از متغیر تصادفی کمکی $\mathbf{w=x}$؛

حل دستگاه $\begin{cases} z=xy \\ w=x \end{cases}$ نتیجه می‌دهد:

$$\begin{cases} x_1=w \\ y_1=\dfrac{z}{w} \end{cases}$$

$$J(x,y)=\begin{vmatrix} y & x \\ 1 & 0 \end{vmatrix}=-x \quad \Rightarrow \quad J(x_1,y_1)=-w$$

$$\Rightarrow f_{\mathbf{zw}}(z,w)=\frac{1}{|w|}f_{\mathbf{xy}}(w,\frac{z}{w})$$

$$\Rightarrow f_{\mathbf{z}}(z)=\int_{-\infty}^{+\infty}f_{\mathbf{zw}}(z,w)dw=\int_{-\infty}^{+\infty}\frac{1}{|w|}f_{\mathbf{xy}}(w,\frac{z}{w})dw$$

متغیرهای تصادفی مشترکاً نرمال:

تعریف: $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال هستند، اگر $\mathbf{z}=a\mathbf{x}+b\mathbf{y}$ برای هر a و b نرمال باشد.

نتیجه: اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال باشند، $\mathbf{x}$ و $\mathbf{y}$ هر کدام نرمال هستند.

ولی عکس این مطلب لزوماً صحیح نیست (مثال نقض در فصل ۶ کتاب فرآیند Papoulis). مثلاً ممکن است $\mathbf{x}$ و $\mathbf{y}$ هر کدام نرمال باشند، ولی $\mathbf{x}+\mathbf{y}$ نرمال نباشد (مثال در فصل ۷ کتاب فرآیند Papoulis). توجه کنید که ممکن است $\mathbf{x}$ و $\mathbf{y}$ و $\mathbf{x}+\mathbf{y}$ هر سه نرمال باشند، ولی $\mathbf{x}$ و $\mathbf{y}$ باز هم مشترکاً نرمال نباشند (مثال در فصل ۷ کتاب فرآیند Papoulis).

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ هر کدام نرمال بوده و مستقل از هم باشند، مشترکاً نرمال خواهند بود (یعنی عکس نتیجهٔ فوق در حالت استقلال $\mathbf{x}$ و $\mathbf{y}$ صادق است). برای اثبات این قضیه، به سادگی می‌توان از طریق به دست آوردن تابع مشخصهٔ $\mathbf{z}=a\mathbf{x}+b\mathbf{y}$ نشان داد که $\mathbf{z}$ برای هر a و b نرمال خواهد بود. مانند کاری که در مثال ۱، صفحهٔ ۳۷ همین فصل برای $\mathbf{z}=\mathbf{x}+\mathbf{y}$ کردیم:

$$\Phi_{\mathbf{z}}(j\omega)=\Phi_{\mathbf{x}}(aj\omega)\Phi_{\mathbf{y}}(bj\omega)=e^{j\omega(a\eta_x+b\eta_y)-\frac{(a^2\sigma_x^2+b^2\sigma_y^2)\omega^2}{2}}$$

قضیه: $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال هستند، اگر و تنها اگر:

$$f_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-\eta_{\mathbf{x}})^2}{\sigma_{\mathbf{x}}^2}-2r\frac{(x-\eta_{\mathbf{x}})(y-\eta_{\mathbf{y}})}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}+\frac{(y-\eta_{\mathbf{y}})^2}{\sigma_{\mathbf{y}}^2}\right]\right\}$$

یا معادلاً:

$$\Phi_{\mathbf{xy}}(j\omega_1,j\omega_2)=e^{j(\omega_1\eta_{\mathbf{x}}+\omega_2\eta_{\mathbf{y}})}\,e^{-\frac{1}{2}(\sigma_{\mathbf{x}}^2\omega_1^2+2r\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}\omega_1\omega_2+\sigma_{\mathbf{y}}^2\omega_2^2)}$$

اثبات در صفحات ۱۶۲ و ۱۶۳ کتاب.

یعنی برای توزیع نرمال، ۵ پارامتر $\sigma_{\mathbf{x}}$، $\sigma_{\mathbf{y}}$، $\eta_{\mathbf{x}}$، $\eta_{\mathbf{y}}$ و r به طور کامل توزیع مشترک $\mathbf{x}$ و $\mathbf{y}$ را مشخص می‌کنند (اکنون اثبات راحت‌تر برای اینکه اگر $\mathbf{x}$ و $\mathbf{y}$ نرمال و مستقل باشند، مشترکاً نرمال خواهند بود، این است که $f_{\mathbf{xy}}=f_{\mathbf{x}}f_{\mathbf{y}}$، فرم $f_{\mathbf{xy}}$ گوسی را دارد).

تعریف معادل: $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال هستند، هرگاه $f_{\mathbf{xy}}$ به صورت $Ae^{-Q(x,y)}$ باشد که $Q(x,y)$ فرم درجهٔ دوم معین مثبت است، یعنی:

$$Q(x,y)=ax^2+bxy+cy^2+dx+ey+f$$

به طوری که برای هر $\mathbf{x}$ و $\mathbf{y}$ داشته باشیم: $Q(x,y)\geq 0$.

اثبات معادل بودن این تعریف با تعریف قبلی در کتاب، صفحهٔ ۱۶۳ است.

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال بوده و ناهمبسته نیز باشند $(r=0)$، آنگاه مستقل خواهند بود.

اثبات:

$$r=0\Rightarrow f_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}e^{-\frac{1}{2}\left[\frac{(x-\eta_{\mathbf{x}})^2}{\sigma_{\mathbf{x}}^2}+\frac{(y-\eta_{\mathbf{y}})^2}{\sigma_{\mathbf{y}}^2}\right]}=f_{\mathbf{x}}(x)f_{\mathbf{y}}(y)$$

می‌دانیم که در حالت کلی: استقلال $\Leftarrow$ ناهمبستگی، اما: ناهمبستگی $\not\Leftarrow$ استقلال؛

ولی در مورد متغیرهای تصادفی مشترکاً نرمال داریم: استقلال $\Leftrightarrow$ ناهمبستگی.

قضیه: اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال بوده و $\begin{cases} \mathbf{z}=a_{11}\mathbf{x}+a_{12}\mathbf{y} \\ \mathbf{w}=a_{21}\mathbf{x}+a_{22}\mathbf{y} \end{cases}$، آنگاه $\mathbf{z}$ و $\mathbf{w}$ نیز مشترکاً نرمال خواهند بود.

اثبات در کتاب، صفحهٔ ۱۶۳ (از طریق $f_{\mathbf{zw}}$ و کوادراتیک بودن یا می‌توان از طریق $\Phi_{\mathbf{zw}}$ اثبات کرد یا می‌گوییم هر ترکیب خطی از $\mathbf{z}$ و $\mathbf{w}$، یک ترکیب خطی از $\mathbf{x}$ و $\mathbf{y}$ خواهد بود، پس نرمال است).

فصل ۶: توزیع‌های شرطی

Chapter 6

۱. *pmf*، *CDF* و *pdf* شرطی

۲. قضیهٔ احتمال کل و قضیهٔ بیز

۳. شرط واقعه‌ای در ارتباط با همان متغیر تصادفی

۴. قابلیت اعتماد

۵. شرط واقعه‌ای در ارتباط با متغیر تصادفی دیگر

۶. امید ریاضی شرطی

۷. امید ریاضی شرطی تابعی از دو متغیر تصادفی

۸. تخمین *ls*

۹. تخمین *lls*

می‌دانیم که با فرض $P(M) \neq 0$، داریم:

$$P(A|M) = \frac{P(A \cap M)}{P(M)}$$

تابع احتمال شرطی (pmf شرطی):

اگر **x** یک متغیر تصادفی گسسته باشد که مقادیر x_i $(i = 0,1,2,\ldots)$ را اختیار می‌کند، تابع احتمال شرطی **x** به شرط واقعهٔ M، طبق تعریف برابر است با:

$$P_{\mathbf{x}}(x_i | M) = P\{\mathbf{x} = x_i | M\} = \frac{P\{\mathbf{x} = x_i, M\}}{P(M)} = \frac{P\{\omega : \mathbf{x}(\omega) = x_i, \omega \in M\}}{P(M)}$$

ثابت کرده بودیم که احتمال شرطی تمام خواص احتمال را دارد. لذا $P_{\mathbf{x}}(x|M)$ نیز تمام خواص تابع احتمال غیرشرطی را دارد، از جمله اینکه داریم:

$$1)\ \sum_i P_{\mathbf{x}}(x_i | M) = 1$$

$$2)\ \forall D \subset \mathbb{R} : P\{\mathbf{x} \in D | M\} = \sum_{x_i \in D} P_{\mathbf{x}}(x_i | M)$$

تابع توزیع انباشتهٔ شرطی (CDF شرطی):

طبق تعریف، تابع توزیع انباشتهٔ شرطی **x** به شرط واقعهٔ M برابر است با:

$$F_{\mathbf{x}}(x|M) = P\{\mathbf{x} \leq x|M\}$$

این تابع نیز تمام خواص CDF غیرشرطی را دارد، از جمله داریم:

1) $F(-\infty|M) = 0$

2) $F(+\infty|M) = 1$

3) $P(x_1 \leq \mathbf{x} \leq x_2|M) = F(x_2|M) - F(x_1|M)$

تابع چگالی شرطی (pdf شرطی):

طبق تعریف، تابع چگالی شرطی $\mathbf{x}$ به شرط واقعهٔ M برابر است با:

$$f_{\mathbf{x}}(x|M) = \frac{dF_{\mathbf{x}}(x|M)}{dx}$$

این تابع نیز تمام خواص pdf غیرشرطی را دارد، از جمله داریم:

1) $f_{\mathbf{x}}(x|M)dx = P\{x \le \mathbf{x} \le x + dx|M\}$

2) $\int_{-\infty}^{+\infty} f_{\mathbf{x}}(x|M)dx = F(+\infty|M) - F(-\infty|M) = 1$

یادآوری: اگر A_iها $(i = 0,1,2,\ldots,m)$ افرازی از Ω باشند، قضایای زیر را داشتیم:

۱. قضیهٔ احتمال کل:

$$P(B) = \sum_{i=1}^{m} P(B|A_i)P(A_i)$$

۲. قضیهٔ بِیز:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{P(B)} = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^{m} P(B|A_i)P(A_i)}$$

حال اگر **x** یک متغیر تصادفی گسسته باشد، در رابطه با pmf شرطی آن داریم:

$$P\{\mathbf{x}=x\}=\sum_{i=1}^{m}P\{\mathbf{x}=x\,|\,A_i\}P(A_i)$$

یعنی:

۱. قضیهٔ احتمال کل:

$$P_{\mathbf{x}}(x)=\sum_{i=1}^{m}P_{\mathbf{x}}(x\,|\,A_i)P(A_i)$$

همچنین داریم:

$$P\{A\,|\,\mathbf{x}=x_k\}=\frac{P\{\mathbf{x}=x_k\,|\,A\}P(A)}{P\{\mathbf{x}=x_k\}}$$

یعنی:

۲. قضیهٔ بِیز:

$$P(A\,|\,x_k)=\frac{P_{\mathbf{x}}(x_k\,|\,A)P(A)}{P_{\mathbf{x}}(x_k)}$$

و نیز اگر $\{\mathbf{x}=x_i\}$ها را به عنوان وقایع افرازکننده در نظر بگیریم، خواهیم داشت:

۱. قضیهٔ احتمال کل:

$$P(A)=\sum_i P(A\,|\,\mathbf{x}=x_i)P\{\mathbf{x}=x_i\}=\sum_i P(A\,|\,x_i)P_{\mathbf{x}}(x_i)$$

همچنین داریم:

$$P\{\mathbf{x}=x_k\,|\,A\}=\frac{P(A\,|\,\mathbf{x}=x_k)P\{\mathbf{x}=x_k\}}{P(A)}$$

پس:

۲. قضیهٔ بِیز:

$$P_{\mathbf{x}}(x_k\,|\,A)=\frac{P(A\,|\,x_k)P_{\mathbf{x}}(x_k)}{P(A)}=\frac{P(A\,|\,x_k)P_{\mathbf{x}}(x_k)}{\sum_i P(A\,|\,x_i)P_{\mathbf{x}}(x_i)}$$

در رابطه با CDF شرطی نیز داریم:

۱. قضیهٔ احتمال کل:

$$F_{\mathbf{x}}(x)=\sum_{i=1}^{m}F_{\mathbf{x}}(x\,|\,A_i)P(A_i)$$

۲. قضیهٔ بیز:

$$P(A\,|\,\mathbf{x}\le x)=\frac{F_{\mathbf{x}}(x\,|\,A)P(A)}{F_{\mathbf{x}}(x)}$$

در نهایت، در رابطه با pdf شرطی خواهیم داشت:

۱. قضیهٔ احتمال کل:

$$f_{\mathbf{x}}(x)=\sum_{i=1}^{m}f_{\mathbf{x}}(x\,|\,A_i)P(A_i)$$

۲. قضیهٔ بیز:

$$P(A\,|\,\underbrace{\mathbf{x}=x}_{\boxed{\text{احتمالش صفر است}}})=\lim_{\Delta x\to 0}P(A\,|\,x\le\mathbf{x}\le x+\Delta x)=\frac{f_{\mathbf{x}}(x\,|\,A)P(A)\Delta x}{f_{\mathbf{x}}(x)\Delta x}=\frac{f_{\mathbf{x}}(x\,|\,A)P(A)}{f_{\mathbf{x}}(x)}$$

و اگر $\{\mathbf{x}=x\}$ها را وقایع افرازکننده در نظر بگیریم، داریم:

۱. قضیهٔ احتمال کل:

$$P(A)=\int_{-\infty}^{+\infty}P(A|x)f_{\mathbf{x}}(x)dx$$

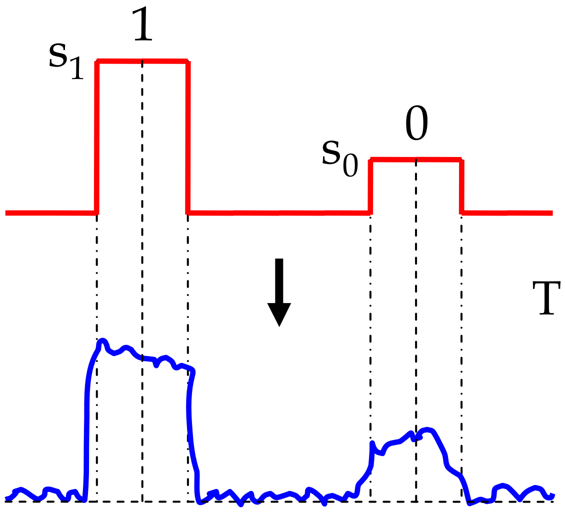
۲.قضیهٔ بیز:

$$f_{\mathbf{x}}(x|A)=\frac{P(A|x)f_{\mathbf{x}}(x)}{P(A)}=\frac{P(A|x)f_{\mathbf{x}}(x)}{\int_{-\infty}^{+\infty}P(A|x)f_{\mathbf{x}}(x)dx}$$

(برای اثبات رابطهٔ بالایی می‌توان از رابطهٔ پایینی (بیز) استفاده کرد و از طرفین، از $-\infty$ تا $+\infty$ انتگرال گرفت.)

مثالی از مخابرات دیجیتال:

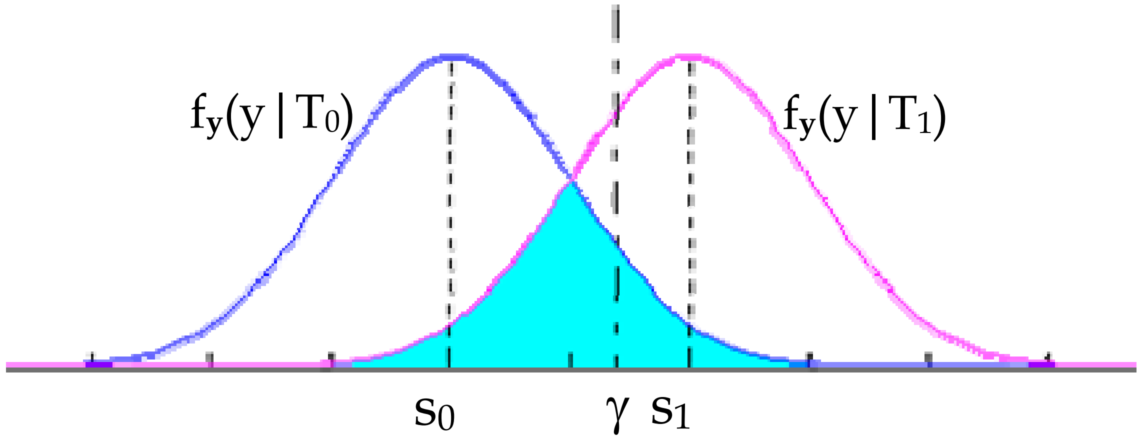
برای ارسال 0 و 1، ولتاژهای s_0 و s_1 را به کار می‌بریم، ولی به دلیل وجود نویز در کانال، گیرنده خود s_0 و s_1 را دریافت نمی‌کند.



فرض کنید نویز دارای توزیع گوسی $N(0,\sigma)$ است. می‌توان نشان داد برای اینکه گیرنده از روی این داده‌های نویزی بتواند به طور بهینه (کمترین احتمال خطا) تصمیم بگیرد که 0 یا 1 ارسال شده بوده، باید ولتاژ دریافتی را با سطح آستانه‌ای مقایسه کند. اگر از این سطح آستانه بزرگتر بود، 1 و اگر کوچکتر بود، 0 فرض می‌شود. مقدار سطح آستانه در صورتی که $P(T_0)=P(T_1)=\dfrac{1}{2}$ باشد، برابر $\gamma=\dfrac{s_0+s_1}{2}$ است (در حالت کلی داریم: $\gamma=\dfrac{s_0+s_1}{2}+\dfrac{\sigma^2}{s_0-s_1}\ln\dfrac{P(T_1)}{P(T_0)}$).

در هر بیت، احتمال خطا، یعنی P_{error}، چقدر است؟

$$
\begin{aligned}
P_{error} &= P(E \mid T_1)P(T_1) + P(E \mid T_0)P(T_0) \\
&= P(R_0 \mid T_1)P(T_1) + P(R_1 \mid T_0)P(T_0) \\
&= P\{\mathbf{y} < \gamma \mid T_1)P(T_1) + P\{\mathbf{y} > \gamma \mid T_0)P(T_0) \\
&= F_{\mathbf{y}}(\gamma \mid T_1)P(T_1) + \left(1 - F_{\mathbf{y}}(\gamma \mid T_0)\right)P(T_0) \\
&= P(T_1)\int_{-\infty}^{\gamma} f_{\mathbf{y}}(y \mid T_1)dy + P(T_0)\int_{\gamma}^{+\infty} f_{\mathbf{y}}(y \mid T_0)dy
\end{aligned}
$$



در واقع، به شرط T_1 یعنی: $\mathbf{y} = s_1 + \mathbf{n}$ (مثل $\mathbf{y} = \mathbf{x} + 2$)، پس:

$$f_{\mathbf{y}}(y \mid T_1) = f_{\mathbf{n}}(y - s_1)$$

و به همین ترتیب خواهیم داشت:

$$f_{\mathbf{y}}(y \mid T_0) = f_{\mathbf{n}}(y - s_0)$$

در نتیجه:

$$\begin{aligned} P_{error} &= \frac{1}{2}\int_{-\infty}^{\gamma} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-s_1)^2}{2\sigma^2}} dy + \frac{1}{2}\int_{\gamma}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-s_0)^2}{2\sigma^2}} dy \\ &= \frac{1}{2} G(\frac{\gamma - s_1}{\sigma}) + \frac{1}{2}\left[1 - G(\frac{\gamma - s_0}{\sigma})\right] \end{aligned}$$

ولی $\gamma = \dfrac{s_0 + s_1}{2}$، پس:

$$\begin{aligned} P_{error} &= \frac{1}{2} G(\frac{s_0 - s_1}{2\sigma}) + \frac{1}{2}\left[1 - G(\frac{s_1 - s_0}{2\sigma})\right] = \frac{1}{2}\left[1 - G(\frac{s_1 - s_0}{2\sigma})\right] + \frac{1}{2}\left[1 - G(\frac{s_1 - s_0}{2\sigma})\right] \\ &= 1 - G(\frac{s_1 - s_0}{2\sigma}) = Q(\frac{s_1 - s_0}{2\sigma}) \end{aligned}$$

هر چه $\frac{s_1 - s_0}{\sigma}$ بزرگتر باشد (یعنی نسبت سیگنال به نویز بیشتر باشد)، خطا کوچکتر خواهد بود. مثلاً اگر $s_1 = 2V$ و $s_0 = -2V$ و $\sigma = 1V$ باشد، داریم:

$$P_{error} = Q(2) = 0.0227$$

یعنی ۲٪ احتمال خطا وجود دارد.

شرط واقعه‌ای در ارتباط با همان متغیر تصادفی:

ممکن است واقعهٔ مشروط کننده، واقعه‌ای در ارتباط با همان متغیر تصادفی باشد. مثلاً اگر:

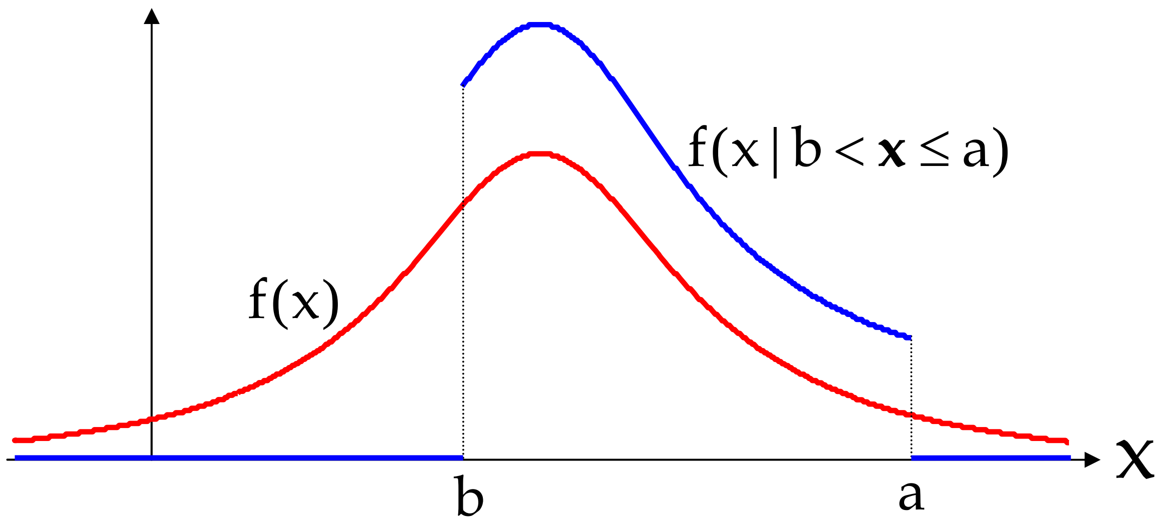
$$M = \{b < \mathbf{x} \le a\}$$

آنگاه داریم:

$$F_{\mathbf{x}}(x \mid b < \mathbf{x} \le a) = \frac{P\{\mathbf{x} \le x \mid b < \mathbf{x} \le a\}}{P\{b < \mathbf{x} \le a\}} = \begin{cases} 1 & x \ge a \\ 0 & x < b \\ \dfrac{P\{b < \mathbf{x} \le x\}}{P\{b < \mathbf{x} \le a\}} = \dfrac{F_{\mathbf{x}}(x) - F_{\mathbf{x}}(b)}{F_{\mathbf{x}}(a) - F_{\mathbf{x}}(b)} & b \le x < a \end{cases}$$

با مشتق‌گیری خواهیم داشت:

$$f_{\mathbf{x}}(x \mid b < \mathbf{x} \le a) = \begin{cases} 0 & x \ge a \\ 0 & x < b \\ \dfrac{f_{\mathbf{x}}(x)}{F_{\mathbf{x}}(a) - F_{\mathbf{x}}(b)} = \dfrac{f_{\mathbf{x}}(x)}{\int_b^a f_{\mathbf{x}}(x)dx} & b \le x < a \end{cases}$$



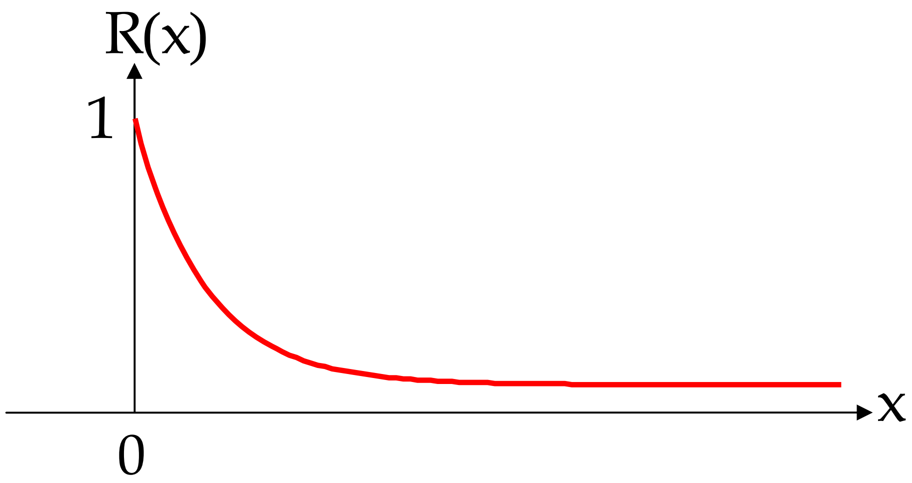
مثالی از کاربرد توزیع مشروط: قابلیت اعتماد

فرض کنید متغیر تصادفی $\mathbf{x}$ زمان خراب شدن سیستمی باشد که از $t=0$ شروع به کار کرده است (متغیر تصادفی $\mathbf{x}$ را عمر سیستم (Time to Failure) می‌گویند). اصطلاحاً $1-F(x)$ را قابلیت اعتماد سیستم می‌گویند:

$$R(x)=1-F(x)=P\{\mathbf{x}>x\}$$

یعنی احتمال اینکه سیستم لااقل تا زمان x کار کند.

پس: $R(0)=1$، $R(+\infty)=0$ و $R(x)$ همواره نزولی است.



امید ریاضی $\mathbf{x}$ را Mean Time to Failure (MTTF) گویند (عمر متوسط سیستم):

$$E(\mathbf{x})=\int_0^{+\infty} xf(x)dx=\int_0^{+\infty} R(x)dx$$

(چون: $\int_{-\infty}^{+\infty} xf(x)dx=\int_{-\infty}^{0} F(x)dx+\int_0^{+\infty}(1-F(x))dx$)

حال می‌خواهیم $F(x|\mathbf{x}\ge t)$ و $f(x|\mathbf{x}\ge t)$ را حساب کنیم.

$F(x|\mathbf{x}\ge t)$ یعنی احتمال اینکه سیستمی که تا لحظهٔ t کار می‌کرده است قبل از لحظهٔ x خراب شود:

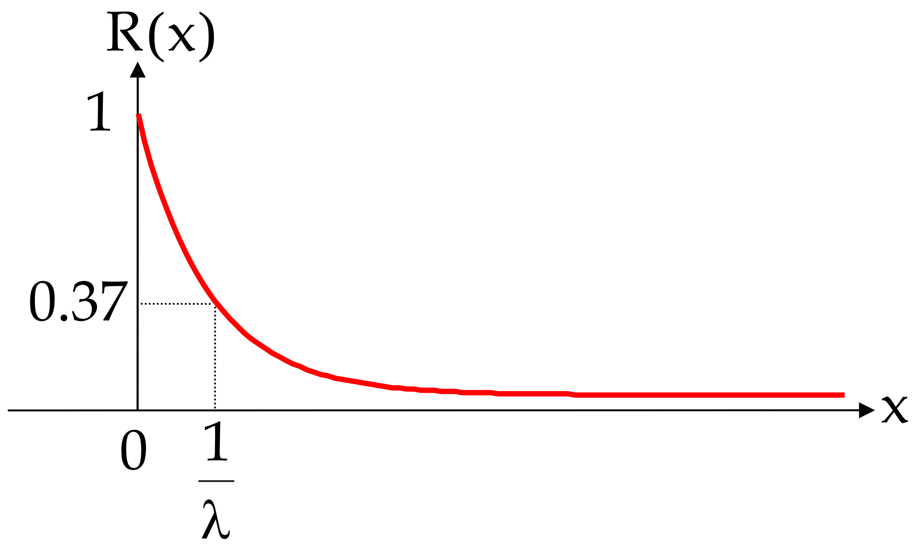
$$F(x|\mathbf{x}\ge t)=\frac{P\{\mathbf{x}\le x,\mathbf{x}\ge t\}}{P\{\mathbf{x}\ge t\}}=\begin{cases}0 & x<t\\ \dfrac{F(x)-F(t)}{1-F(t)} & x\ge t\end{cases}$$

با مشتق‌گیری، $f(x|\mathbf{x}\ge t)$ به دست می‌آید. تابع فوق در $x=t$ پیوسته است، پس f شامل δ نخواهد شد:

$$f(x|\mathbf{x}\ge t)=\begin{cases}0 & x<t\\ \dfrac{f(x)}{1-F(t)} & x\ge t\end{cases}$$

$f(x|\mathbf{x}\ge t)dx$ یعنی احتمال اینکه سیستمی که تا لحظهٔ t کار می‌کرده است بین لحظات x و $x+dx$ خراب شود.

مثال: اگر $f(x)=\lambda e^{-\lambda x}u(x)$ باشد، داریم:



$$F(x)=(1-e^{-\lambda x})u(x)\Rightarrow R(x)=e^{-\lambda x}\ :x\geq 0$$

$$E(\mathbf{x})=\frac{1}{\lambda}$$

$$f(x\,|\,\mathbf{x}\geq t)=\frac{f(x)}{1-F(t)}=\frac{\lambda e^{-\lambda x}}{e^{-\lambda t}}=\lambda e^{-\lambda(x-t)}=f(x-t):x\geq t$$

که همان خاصیت بی‌حافظه بودن توزیع نمایی را نشان می‌دهد.

در واقع وقتی $f(x)$ نمایی باشد، نرخ خرابی شرطی ثابت است.

نرخ خرابی شرطی:

طبق تعریف، نرخ خرابی شرطی متغیر تصادفی $\mathbf{x}$ برابر است با:

$$\beta(t) \triangleq f(t \mid \mathbf{x} \ge t)$$

$\beta(t)dt$ بیانگر احتمال این است که سیستمی که تا لحظهٔ t سالم بوده، در لحظهٔ t (یعنی بین t و $t+dt$) خراب شود.

$$\beta(t) = \frac{f(t)}{1-F(t)} = \frac{F'(t)}{1-F(t)} = \frac{-R'(t)}{R(t)}$$

اگر از طرفین انتگرال بگیریم و با توجه به اینکه: $R(0)=1$، داریم:

$$-\int_0^x \beta(t)dt = \ln R(x)$$

$$\Rightarrow R(x) = e^{-\int_0^x \beta(t)dt} \quad : x \ge 0$$

$$\Rightarrow F(x) = 1 - e^{-\int_0^x \beta(t)dt} \quad : x \ge 0$$

$$\Rightarrow f(x) = \beta(x) e^{-\int_0^x \beta(t)dt} \quad : x \ge 0$$

خواص $\beta(t)$:

$$1)\ \beta(t)=\frac{f(t)}{1-F(t)}\Rightarrow \beta(t)\geq 0$$

$$2)\ F(+\infty)=1\Rightarrow \int_0^{+\infty}\beta(t)dt\rightarrow +\infty$$

مثلاً اگر $\beta(t)=\lambda$، یعنی مقدار ثابت باشد، داریم:

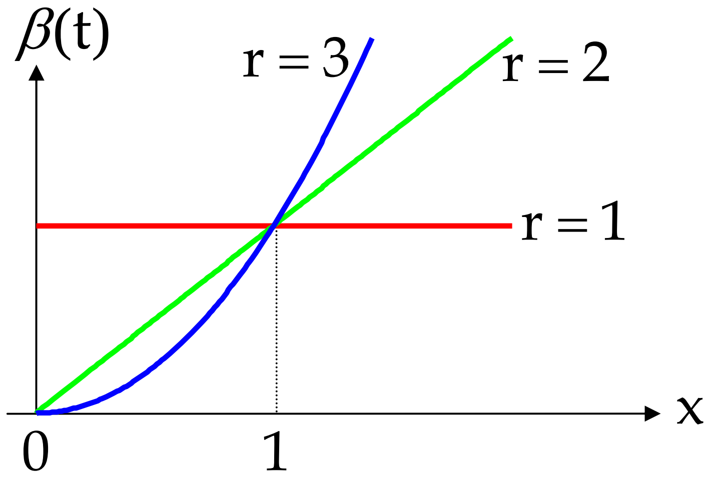
$$f(x)=\lambda e^{-\lambda x}u(x)$$

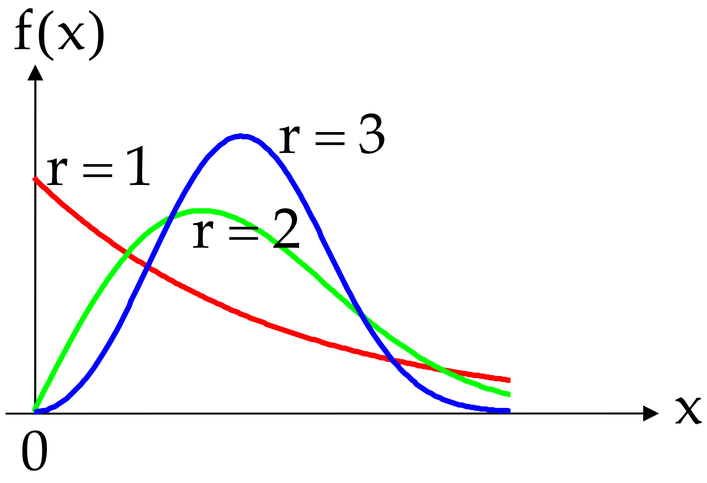
اگر $\beta(t)=\lambda t^{r-1}$ باشد، در بسیاری موارد تابع مناسبی است و داریم:

$$f(x)=\lambda x^{r-1}e^{-\lambda\frac{x^r}{r}}u(x)$$

این توزیع را <u>توزیع ویبول (Weibull)</u> می‌گویند که برای $r=1$ همان توزیع نمایی و برای $r=2$ همان توزیع رایلی می‌شود.

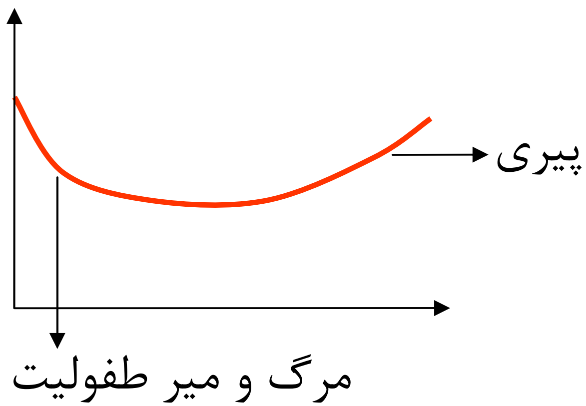
برای $r=1$، $\beta(t)$ ثابت است و برای rهای بزرگتر، فرسودگی را مدل می‌کند:





$$\mathrm{MTTF} = E(\mathbf{x}) = \left(\frac{r}{\lambda}\right)^{\frac{1}{r}} \Gamma\left(\frac{r+1}{r}\right)$$

ولی خیلی اوقات $\beta(t)$ به این شکل است:



شرط واقعه‌ای در ارتباط با متغیر تصادفی دیگر:

گاهی واقعهٔ مشروط کننده، واقعه‌ای در ارتباط با یک متغیر تصادفی دیگر است.

اگر $\mathbf{x}$ و $\mathbf{y}$ گسسته باشند، تابع احتمال $\mathbf{y}$ به شرط $x_1 < \mathbf{x} \le x_2$ برابر است با:

$$P_{\mathbf{y}}(y_j \mid x_1 < \mathbf{x} \le x_2) = \frac{P\{\mathbf{y} = y_j, x_1 < \mathbf{x} \le x_2\}}{P\{x_1 < \mathbf{x} \le x_2\}} = \frac{\sum_{x_1 < x_i \le x_2} P_{\mathbf{xy}}(x_i, y_j)}{\sum_{x_1 < x_i \le x_2} P_{\mathbf{x}}(x_i)}$$

و تابع احتمال $\mathbf{y}$ به شرط $\mathbf{x} = x_i$ نیز برابر است با:

$$P_{\mathbf{y}}(y_j \mid \mathbf{x} = x_i) = P_{\mathbf{y}}(y_j \mid x_i) = \frac{P\{\mathbf{y} = y_j, \mathbf{x} = x_i\}}{P\{\mathbf{x} = x_i\}} = \frac{P_{\mathbf{xy}}(x_i, y_j)}{P_{\mathbf{x}}(x_i)}$$

در مورد pdf نیز داریم:

$$f_{\mathbf{y}}(y \mid x_1 < \mathbf{x} \leq x_2) = \frac{\int_{x_1}^{x_2} f_{\mathbf{xy}}(x,y)dx}{\int_{x_1}^{x_2} f_{\mathbf{x}}(x)dx}$$

و همچنین:

$$f_{\mathbf{y}}(y \mid x) = \frac{f_{\mathbf{xy}}(x,y)}{f_{\mathbf{x}}(x)}$$

در حالت خاص اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، آنگاه برای هر x و y داریم: $f_{\mathbf{xy}}(x,y) = f_{\mathbf{x}}(x) f_{\mathbf{y}}(y)$، لذا:

$$f_{\mathbf{y}}(y \mid x) = f_{\mathbf{y}}(y)$$

و نیز:

$$f_{\mathbf{x}}(x \mid y) = f_{\mathbf{x}}(x)$$

می‌دانیم که:

$$f_{\mathbf{y}}(y) = \int_{-\infty}^{+\infty} f_{\mathbf{xy}}(x,y)dx$$

ولی داریم:

$$f_{xy}(x,y)=f_y(y|x)f_x(x)$$

پس:

$$f_y(y)=\int_{-\infty}^{+\infty}f_y(y|x)f_x(x)dx$$

که همان قضیهٔ احتمال کل است.

حال با تعویض نقش **x** و **y** داریم:

$$f_x(x|y)=\frac{f_{xy}(x,y)}{f_y(y)}$$

$$\Rightarrow f_x(x|y)=\frac{f_y(y|x)f_x(x)}{f_y(y)}=\frac{f_y(y|x)f_x(x)}{\int_{-\infty}^{+\infty}f_y(y|x)f_x(x)dx}$$

که همان قضیهٔ بیز است.

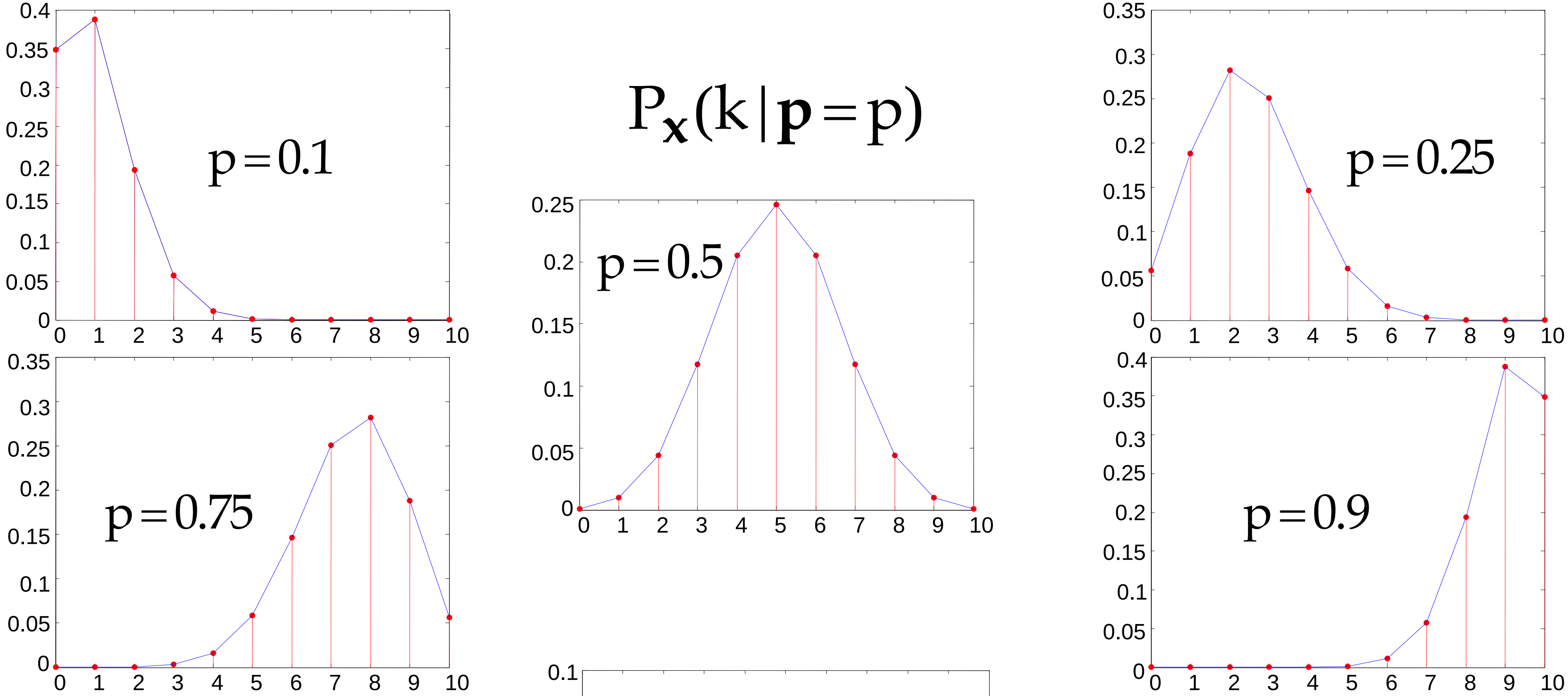
مثال ۱: $\mathbf{x} \sim \text{Binomial}(n,\mathbf{p})$، ولی احتمال موفقیت را نمی‌دانیم و $\mathbf{p}$ خود یک متغیر تصادفی است: $\mathbf{p} \sim u(0,1)$؛ مثلاً سکه‌ای که احتمال شیر آمدنش را نمی‌دانیم.

$$P_{\mathbf{x}}(k \mid \mathbf{p}=p) = P\{\mathbf{x}=k \mid \mathbf{p}=p\} = \binom{n}{k} p^k (1-p)^{n-k} : k=0,1,\ldots,n$$

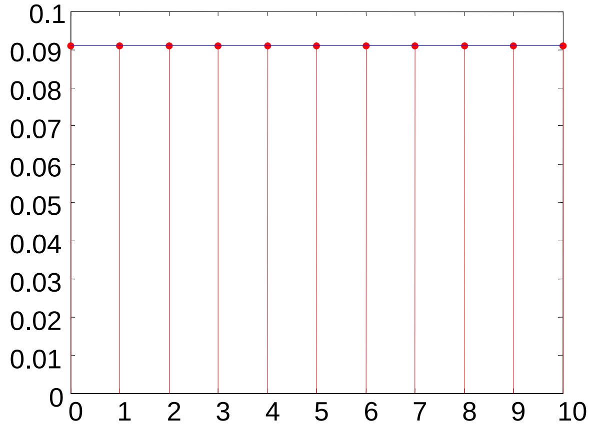
$$f_{\mathbf{p}}(p) = \begin{cases} 1 & 0 \le p < 1 \\ 0 & \text{otherwise} \end{cases}$$

توابع $P_{\mathbf{x}}(k) = P\{\mathbf{x}=k\}$ و $f_{\mathbf{p}}(p \mid \mathbf{x}=k)$ را بیابید.

$$\begin{aligned} P_{\mathbf{x}}(k) &= \int_{-\infty}^{+\infty} P_{\mathbf{x}}(k \mid \mathbf{p}=p) f_{\mathbf{p}}(p)\,dp : k=0,1,\ldots,n \\ &= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,dp \qquad \to \beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \\ &= \binom{n}{k} \beta(k+1, n-k+1) = \frac{n!}{k!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!} \\ &= \frac{1}{n+1} : k=0,1,\ldots,n \end{aligned}$$



$P_{\mathbf{x}}(k)$:



$$f_{\mathbf{p}}(p|\mathbf{x}=k)=\frac{P\{\mathbf{x}=k\,|\,\mathbf{p}=p\}f_{\mathbf{p}}(p)}{P\{\mathbf{x}=k\}}=\frac{P_{\mathbf{x}}(k|p)f_{\mathbf{p}}(p)}{P_{\mathbf{x}}(k)}=\frac{\binom{n}{k}p^{k}(1-p)^{n-k}}{\frac{1}{n+1}}:0\le p<1$$

$$=\frac{(n+1)!}{k!(n-k)!}p^{k}(1-p)^{n-k}:0\le p<1$$

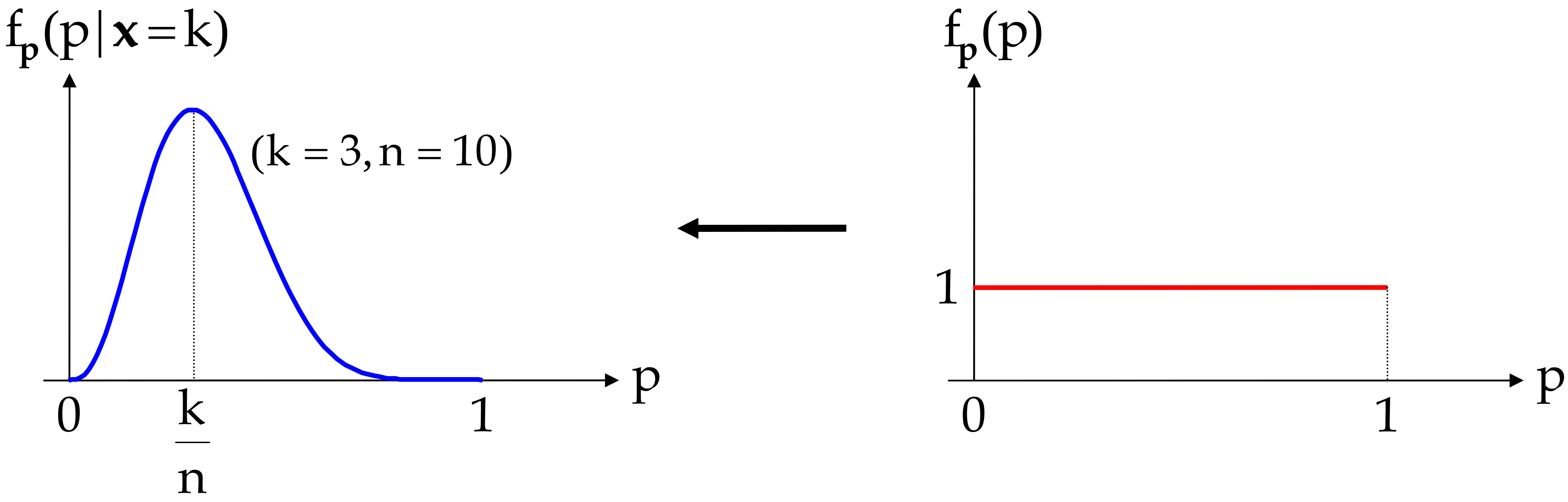
$$\Rightarrow(\mathbf{p}|\mathbf{x}=k)\sim \mathrm{Beta}(k+1,n-k+1)$$

***<u>یادآوری:</u>* توزیع بتا:**

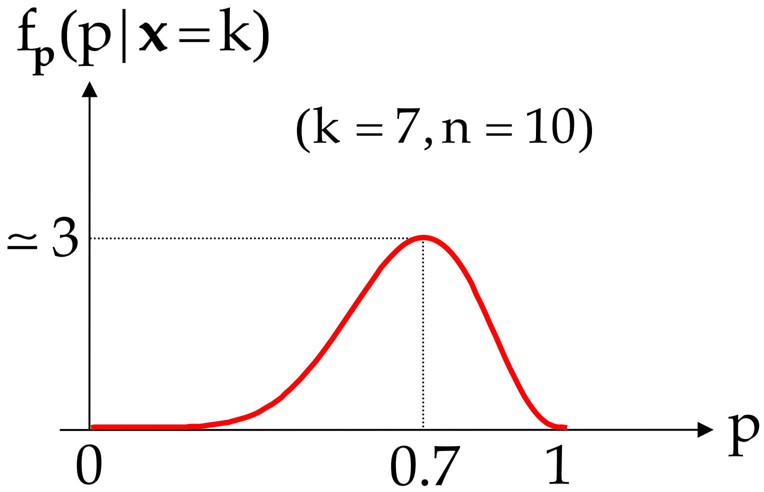
$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}:0<x<1$$

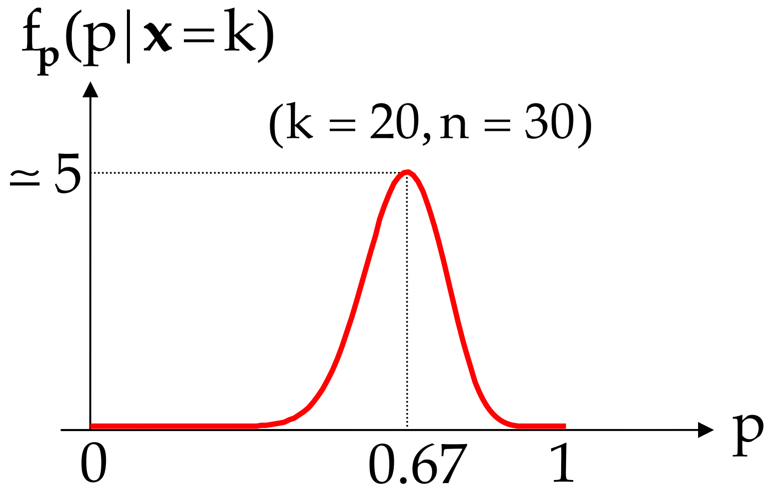
یعنی $\mathbf{p}$ که توزیع یکنواخت داشت، با مشاهدهٔ مقدار $\mathbf{x}$ (مثلاً تعداد شیرهای آمده در nبار پرتاب سکه) توزیع بتا پیدا می‌کند.

ماکزیمم توزیع بتا در $\frac{a-1}{a+b-2}$ یعنی $\frac{k}{n}$ اتفاق می‌افتد. هر چه n بزرگتر باشد، نمودار تیزتر می‌شود.



کتاب در فصل ۶ (Section 6.1) در مورد تخمین بیز نیز بحث کرده است که ما بعداً مفصل دربارهٔ آن صحبت خواهیم کرد.


$f_{\mathbf{p}}(p|\mathbf{x}=k)$
$(k=7, n=10)$
$\simeq 3$
p
0
0.7
1


$f_{\mathbf{p}}(p|\mathbf{x}=k)$
$(k=20, n=30)$
$\simeq 5$
p
0
0.67
1

مثال ۲: مقادیر $P\{\mathbf{x}>1|\mathbf{y}=y\}$، $E(\mathbf{x}|\mathbf{y}=y)$ و $var(\mathbf{x}|\mathbf{y}=y)$ را برای توزیع زیر پیدا کنید:

$$f_{\mathbf{xy}}(x,y)=\begin{cases}\dfrac{e^{-\frac{x}{y}}e^{-y}}{y} & x>0, y>0\\ 0 & \text{otherwise}\end{cases}$$

حل: ابتدا باید $f_{\mathbf{x}}(x|y)$ را به دست آوریم:

$$f_{\mathbf{x}}(x|y)=\frac{f_{\mathbf{xy}}(x,y)}{f_{\mathbf{y}}(y)}=\frac{f_{\mathbf{xy}}(x,y)}{\int_0^{+\infty}f_{\mathbf{xy}}(x,y)dx}=\frac{\dfrac{e^{-\frac{x}{y}}e^{-y}}{y}}{e^{-y}\int_0^{+\infty}\dfrac{e^{-\frac{x}{y}}}{y}dx}=\frac{e^{-\frac{x}{y}}}{y}: x>0, y>0$$

$$\Rightarrow (\mathbf{x}|y)\sim \exp(\frac{1}{y})$$

$$P\{\mathbf{x}>1|y\}=\int_1^{+\infty}f_{\mathbf{x}}(x|y)dx=\int_1^{+\infty}\frac{1}{y}e^{-\frac{x}{y}}dx=-e^{-\frac{x}{y}}\Big|_1^{+\infty}=e^{-\frac{1}{y}}$$ بیشتر برای y بزرگتر :

$$E(\mathbf{x}|y)=y$$

$$var(\mathbf{x}|y)=y^2$$

مثال ۳: $\mathbf{x}$ و $\mathbf{y}$ متغیرهای تصادفی مشترکاً نرمال هستند. $f_{y|x}$ را بیابید.

$$f_y(y|x)=\frac{f_{xy}(x,y)}{f_x(x)}=\frac{\dfrac{1}{2\pi\sigma_x\sigma_y\sqrt{1-r^2}}\exp\left\{-\dfrac{1}{2(1-r^2)}\left[\dfrac{(x-\eta_x)^2}{\sigma_x^2}-2r\dfrac{(x-\eta_x)(y-\eta_y)}{\sigma_x\sigma_y}+\dfrac{(y-\eta_y)^2}{\sigma_y^2}\right]\right\}}{\dfrac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\dfrac{(x-\eta_x)^2}{2\sigma_x^2}\right\}}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-r^2}}\exp\left\{-\frac{\left[y-\eta_y-r\dfrac{\sigma_y}{\sigma_x}(x-\eta_x)\right]^2}{2\sigma_y^2(1-r^2)}\right\}$$

یعنی $(\mathbf{y}|x)\sim N(\eta_y+r\dfrac{\sigma_y}{\sigma_x}(x-\eta_x),\sigma_y\sqrt{1-r^2})$ و داریم:

$$\eta_{y|x}=\eta_y+r\frac{\sigma_y}{\sigma_x}(x-\eta_x)=\eta_y+\frac{\mu_{xy}}{\sigma_x^2}(x-\eta_x)$$

$$\sigma_{y|x}^2=\sigma_y^2(1-r^2)=\sigma_y^2-\frac{\mu_{xy}^2}{\sigma_x^2}$$

(اهمیت زیادی در بحث تخمین دارد)

مثال دیگری از کاربرد توزیع شرطی:

روش دیگری برای تعیین توزیع $\mathbf{z} = g(\mathbf{x}, \mathbf{y})$ با استفاده از چگالی شرطی به صورت زیر است:

$$f_{\mathbf{z}}(z) = \int_{-\infty}^{+\infty} f_{\mathbf{z}}(z \mid x) f_{\mathbf{x}}(x) dx$$

برای $\mathbf{x} = x$ داده شده، $\mathbf{z} = g(x, \mathbf{y})$ تابعی از $\mathbf{y}$ است و لذا به راحتی، اگر ریشهٔ $z = g(x, y)$ نقطهٔ y_0 باشد، داریم:

$$f_{\mathbf{z}}(z \mid x) = \frac{f_{\mathbf{y}}(y_0 \mid x)}{\left| \frac{\partial g(x, y_0)}{\partial y} \right|}$$

(یا می‌توان $\mathbf{y} = y$ را مشروط گرفت.)

مثال: اگر $\mathbf{z}=\mathbf{xy}$ باشد، برای $\mathbf{x}=x$ داده شده، $\mathbf{z}$ ضریبی از $\mathbf{y}$ است و داریم:

$$f_{\mathbf{z}}(z|x)=\frac{1}{|x|}f_{\mathbf{y}}(\frac{z}{x}|x)$$

$$f_{\mathbf{z}}(z)=\int_{-\infty}^{+\infty}\frac{1}{|x|}f_{\mathbf{y}}(\frac{z}{x}|x)f_{\mathbf{x}}(x)dx=\int_{-\infty}^{+\infty}\frac{1}{|x|}f_{\mathbf{xy}}(x,\frac{z}{x})dx$$

که قبلاً هم این را به دست آورده بودیم.

مثلاً اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال با متوسطهای صفر و واریانسهای یک و ضریب همبستگی r باشند، داریم:

$$f_{\mathbf{xy}}(x,y)=\frac{1}{2\pi\sqrt{1-r^2}}\exp\left[-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)\right]$$

$$f_{\mathbf{z}}(z)=\frac{1}{2\pi\sqrt{1-r^2}}\int_{-\infty}^{+\infty}\frac{1}{|x|}e^{\left[-\frac{1}{2(1-r^2)}\left(x^2-2rx(\frac{z}{x})+(\frac{z}{x})^2\right)\right]}dx=\frac{e^{\frac{rz}{1-r^2}}}{\pi\sqrt{1-r^2}}\int_{0}^{+\infty}\frac{1}{x}e^{-\frac{x^2+\frac{z^2}{x^2}}{2(1-r^2)}}dx$$

$$=\frac{e^{\frac{rz}{1-r^2}}}{\pi\sqrt{1-r^2}}\int_{0}^{+\infty}\frac{1}{2u}e^{-\frac{u+\frac{z^2}{u}}{2(1-r^2)}}du=\frac{e^{\frac{rz}{1-r^2}}}{\pi\sqrt{1-r^2}}K_0(\frac{|z|}{1-r^2})$$

(تابع بسل اصلاح شده نوع دوم مرتبه صفر)

امید ریاضی شرطی:

ابتدا یک متغیر تصادفی را در نظر می‌گیریم.

می‌دانیم که:

$$E(\mathbf{x}) = \int_{-\infty}^{+\infty} x f_{\mathbf{x}}(x) dx$$

$$E(g(\mathbf{x})) = \int_{-\infty}^{+\infty} g(x) f_{\mathbf{x}}(x) dx$$

در حالت شرطی، تابع چگالی شرطی جایگزین می‌شود، یعنی:

$$E(\mathbf{x} | M) = \int_{-\infty}^{+\infty} x f_{\mathbf{x}}(x | M) dx$$

$$E(g(\mathbf{x}) | M) = \int_{-\infty}^{+\infty} g(x) f_{\mathbf{x}}(x | M) dx$$

از قضیهٔ احتمال کل می‌دانیم که اگر A_iها $(i = 1,2,\ldots,m)$ افرازی از Ω باشند، داریم:

$$f_{\mathbf{x}}(x) = \sum_{i=1}^{m} f_{\mathbf{x}}(x | A_i) P(A_i)$$

لذا خواهیم داشت:

$$E(\mathbf{x}) = \sum_{i=1}^{m} E(\mathbf{x} | A_i) P(A_i)$$

ممکن است واقعهٔ M، خود، واقعه‌ای در ارتباط با متغیر تصادفی $\mathbf{x}$ باشد.

مثال: اگر $\mathbf{x} \sim N(0,\sigma)$ باشد، مقدار $\mathrm{var}(\mathbf{x}\,|\,\mathbf{x}>0)$ را حساب کنید.

$$f_{\mathbf{x}}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

$$f_{\mathbf{x}}(x\,|\,\mathbf{x}>0) = \begin{cases} \dfrac{f_{\mathbf{x}}(x)}{1-F(0)} = 2f_{\mathbf{x}}(x) & x>0 \\ 0 & x<0 \end{cases}$$

$$E(\mathbf{x}\,|\,\mathbf{x}>0) = \int_{-\infty}^{+\infty} x\, f_{\mathbf{x}}(x\,|\,\mathbf{x}>0)\,dx = \int_0^{+\infty} \frac{2x}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}}\sigma$$

$$E(\mathbf{x}^2\,|\,\mathbf{x}>0) = \int_0^{+\infty} \frac{2x^2}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma^2$$

$$\mathrm{var}(\mathbf{x}\,|\,\mathbf{x}>0) = \sigma^2\left(1-\frac{2}{\pi}\right)$$

(مانند مسألهٔ یکسوساز تمام موج در تمرین سری چهارم)

ممکن است واقعهٔ مشروط کننده در ارتباط با متغیر تصادفی دیگر باشد. می‌دانیم که:

$$E(\mathbf{y}\,|\,M)=\int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y\,|\,M)\,dy$$

اگر $M=\{\mathbf{x}=x\}$ باشد، $f_{\mathbf{y}}(y\,|\,\mathbf{x}=x)$ را باید در انتگرال فوق قرار دهیم:

$$E(\mathbf{y}\,|\,\mathbf{x}=x)=E(\mathbf{y}\,|\,x)=\int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y\,|\,x)\,dy$$: (که فقط تابع x است)

به همین ترتیب خواهیم داشت:

$$E(g(\mathbf{y})\,|\,x))=\int_{-\infty}^{+\infty} g(y)\, f_{\mathbf{y}}(y\,|\,x)\,dy$$: (که فقط تابع x است)

در مثالی که داشتیم، $E(\mathbf{y}|x)$ و $E\left(\left[\mathbf{y}^2 - E(\mathbf{y}|x)\right]|x\right)$ را دیدیم. $E(\mathbf{y}|x)$ در آن مثال تابعی خطی از x بود:

$$\eta_{\mathbf{y}|x} = \eta_{\mathbf{y}} + r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(x - \eta_{\mathbf{x}}) = \eta_{\mathbf{y}} + \frac{\mu_{\mathbf{xy}}}{\sigma_{\mathbf{x}}^2}(x - \eta_{\mathbf{x}})$$

$$\sigma_{\mathbf{y}|x}^2 = \sigma_{\mathbf{y}}^2(1 - r^2) = \sigma_{\mathbf{y}}^2 - \frac{\mu_{\mathbf{xy}}^2}{\sigma_{\mathbf{x}}^2}$$

$E(\mathbf{y})$ یک عدد است. به همین ترتیب $E(\mathbf{y}|\mathbf{x} = x)$ دیگر یک متغیر تصادفی نیست، بلکه برای هر x یک عدد است. یعنی تابعی از x است:

$$\Phi(x) = E(\mathbf{y}|x)$$

حال می‌توانیم $\Phi(\mathbf{x})$ را در نظر بگیریم که خود یک متغیر تصادفی است:

$$\Phi(\mathbf{x}) = E(\mathbf{y}|\mathbf{x})$$

خواص متوسط مشروط یک متغیر تصادفی (به شرط متغیر تصادفی دیگر):

۱)

$$E(E(\mathbf{y}\,|\,\mathbf{x})) = E(\mathbf{y})$$

اثبات:

$$E(\underbrace{E(\mathbf{y}\,|\,\mathbf{x})}_{\Phi(\mathbf{x})}) = \int_{-\infty}^{+\infty} \underbrace{E(\mathbf{y}\,|\,x)}_{\Phi(x)} f_{\mathbf{x}}(x)\,dx = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y\,|\,x)\,dy \right) f_{\mathbf{x}}(x)\,dx$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} y\, f_{\mathbf{xy}}(x,y)\,dx\,dy = E(\mathbf{y})$$

۲) اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، آنگاه: $E(\mathbf{y}\,|\,\mathbf{x}) = E(\mathbf{y})$.

اثبات: اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، آنگاه داریم:

$$\forall x : E(\mathbf{y}\,|\,x) = \int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y\,|\,x)\,dy = \int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y)\,dy = \eta_{\mathbf{y}}$$

$$\Rightarrow \forall x : \Phi(x) = c \Rightarrow \boldsymbol{\Phi}(\mathbf{x}) = c$$

$$\Rightarrow E(\mathbf{y}\,|\,\mathbf{x}) = \eta_{\mathbf{y}}$$

مثال ۱: در مثال ۲، صفحهٔ ۲۹ دیدیم که:

$$\mathrm{E}(\mathbf{x} \mid \mathrm{y}) = \mathrm{y}$$

پس:

$$\mathrm{E}(\mathbf{x} \mid \mathbf{y}) = \mathbf{y}$$
$$\Rightarrow \mathrm{E}\left[\mathrm{E}(\mathbf{x} \mid \mathbf{y})\right] = \mathrm{E}(\mathbf{y})$$

مثال ۲: در مثال ۳، صفحهٔ ۳۰ دیدیم که:

$$\mathrm{E}(\mathbf{y} \mid \mathrm{x}) = \eta_{\mathbf{y}} + \mathrm{r}\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathrm{x} - \eta_{\mathbf{x}})$$

پس:

$$\mathrm{E}(\mathbf{y} \mid \mathbf{x}) = \eta_{\mathbf{y}} + \mathrm{r}\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x} - \eta_{\mathbf{x}})$$
$$\Rightarrow \mathrm{E}\left[\mathrm{E}(\mathbf{y} \mid \mathbf{x})\right] = \eta_{\mathbf{y}}$$

در همین مثال اگر $\mathrm{r} = 0$ باشد (یعنی $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند)، خواهیم داشت:

$$\forall \mathrm{x} : \mathrm{E}(\mathbf{y} \mid \mathrm{x}) = \eta_{\mathbf{y}}$$

مثال ۳: آزمایشهای ساده و مستقل برنولی با احتمال موفقیت p به طور متوالی انجام می‌شوند. اگر $\mathbf{N}$ تعداد شکستها تا حصول اولین موفقیت باشد، $E(\mathbf{N})$ و $var(\mathbf{N})$ را پیدا کنید.

حل: فرض کنید:

$$\mathbf{z} = \begin{cases} 1 & \text{اگر آزمایش اول موفق باشد} \\ 0 & \text{اگر آزمایش اول موفق نباشد} \end{cases}$$

$$E(\mathbf{N}) = E\left[E(\mathbf{N}\,|\,\mathbf{z})\right]$$
$$E(\mathbf{N}\,|\,\mathbf{z}=1) = 0$$
$$E(\mathbf{N}\,|\,\mathbf{z}=0) = E(1+\mathbf{N})$$

$$E(\mathbf{N}) = E(\mathbf{N}\,|\,\mathbf{z}=1)P(\mathbf{z}=1) + E(\mathbf{N}\,|\,\mathbf{z}=0)P(\mathbf{z}=0) = 0 + q\,E(1+\mathbf{N})$$
$$\Rightarrow E(\mathbf{N}) = q + q\,E(\mathbf{N}) \Rightarrow E(\mathbf{N}) = \frac{q}{p}$$

همچنین:

$$E(\mathbf{N}^2)=E\left[E(\mathbf{N}^2\,|\,\mathbf{z})\right]$$

$$E(\mathbf{N}^2\,|\,\mathbf{z}=1)=0$$

$$E(\mathbf{N}^2\,|\,\mathbf{z}=0)=E\left[(1+\mathbf{N})^2\right]$$

$$E(\mathbf{N}^2)=E(\mathbf{N}^2\,|\,\mathbf{z}=1)P(\mathbf{z}=1)+E(\mathbf{N}^2\,|\,\mathbf{z}=0)P(\mathbf{z}=0)=0+q\,E\left[(1+\mathbf{N})^2\right]$$

$$\Rightarrow E(\mathbf{N}^2)=\frac{pq+2q^2}{p^2}$$

$$\Rightarrow \mathrm{var}(\mathbf{N})=E(\mathbf{N}^2)-\left(E(\mathbf{N})\right)^2=\frac{q}{p^2}$$

(در توزیع دو جمله‌ای منفی داشتیم: $E(\mathbf{x})=\frac{rq}{p}$ و $\mathrm{var}(\mathbf{x})=\frac{rq}{p^2}$)

امید ریاضی مشروط تابعی از دو متغیر تصادفی:

$$E(g(\mathbf{x},\mathbf{y})|M)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{xy}}(x,y|M)\,dx\,dy$$

از جمله می‌توان واقعهٔ $M=\{\mathbf{x}=x\}$ را در نظر گرفت.

ویژگی ۱:

$$E(g(\mathbf{x},\mathbf{y})|\mathbf{x}=x)=E(g(x,\mathbf{y})|\mathbf{x}=x)\overset{\boxed{E(g(\mathbf{y})|x)\text{ بنا به تعریف }}}{=}\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{y}}(y|\mathbf{x}=x)\,dy$$

واضح است (اثبات در کتاب فرآیند Papoulis، فصل ۲، صفحهٔ ۸۰).

حالت خاص:

$$g(\mathbf{x},\mathbf{y})=g_1(\mathbf{x})g_2(\mathbf{y})$$

در این حالت داریم:

$$E(g_1(\mathbf{x})g_2(\mathbf{y})|\mathbf{x}=x)=E(g_1(x)g_2(\mathbf{y})|\mathbf{x}=x)=g_1(x)E(g_2(\mathbf{y})|\mathbf{x}=x)$$

چون تساوی فوق برای هر x برقرار است، پس:

$$E(g_1(\mathbf{x})g_2(\mathbf{y})|\mathbf{x})=g_1(\mathbf{x})E(g_2(\mathbf{y})|\mathbf{x})$$

ویژگی ۲:

$$E\left[E(g(\mathbf{x},\mathbf{y})\,|\,\mathbf{x})\right]=E(g(\mathbf{x},\mathbf{y}))$$

اثبات: چون $E(g(\mathbf{x},\mathbf{y})\,|\,\mathbf{x}=x)$ تابعی از x است، اگر آن را $\theta(x)$ بنامیم، متغیر تصادفی $\theta(\mathbf{x})$ قابل تعریف است.

$$E\left[\underbrace{E(g(\mathbf{x},\mathbf{y})\,|\,\mathbf{x})}_{\theta(\mathbf{x})}\right]=\int_{-\infty}^{+\infty}\underbrace{E(g(x,\mathbf{y})\,|\,\mathbf{x}=x)}_{\theta(x)}\,f_{\mathbf{x}}(x)\,dx$$

حال بنا بر ویژگی ۱ خواهیم داشت:

$$\begin{aligned}E[E(g(\mathbf{x},\mathbf{y})|\mathbf{x})]&=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{y}}(y|x)dy\right)f_{\mathbf{x}}(x)dx\\&=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{xy}}(x,y)dxdy=E(g(\mathbf{x},\mathbf{y}))\end{aligned}$$

حالت خاص:

$$g(\mathbf{x},\mathbf{y})=g_1(\mathbf{x})g_2(\mathbf{y})$$

در این حالت داریم:

$$E(g_1(\mathbf{x})g_2(\mathbf{y}))=E\left[E(g_1(\mathbf{x})g_2(\mathbf{y})\,|\,\mathbf{x})\right]\overset{\boxed{\text{بنا بر حالت خاص در ویژگی ۱}}}{=}E\left[g_1(\mathbf{x})E(g_2(\mathbf{y})\,|\,\mathbf{x})\right]$$

مثالی از کاربرد متوسط مشروط: تخمین یک متغیر تصادفی:

بدون مشاهده:

متغیر تصادفی $\mathbf{y}$ را در نظر بگیرید. اگر بخواهیم $\mathbf{y}$ را با عددی پیش‌بینی کنیم (تخمین بزنیم)، داریم:

$$\hat{y} = c$$

و خطای تخمین برابر خواهد بود با:

$$\mathbf{y} - \hat{y} = \mathbf{y} - c$$

می‌خواهیم به نحوی خطا می‌نیمم شود. مثلاً یک معیار برای این منظور، معیار mae (Mean Absolute Error) است:

$$\text{mae} = E(|\mathbf{y} - c|)$$

می‌توان نشان داد برای اینکه mae می‌نیمم شود، باید داشته باشیم:

$$c = \text{median}(\mathbf{y})$$

معیار متداول‌تر، معیار mse (Mean Square Error) است:

$$\mathrm{mse} = E\left[(\mathbf{y}-c)^2\right]$$

حال باید c را آنچنان بیابیم که mse می‌نیمم شود. در یکی از مسائل نشان دادید که:

$$E\left[(\mathbf{y}-c)^2\right] = (\eta_{\mathbf{y}}-c)^2 + \sigma_{\mathbf{y}}^2$$

پس mse وقتی می‌نیمم می‌شود که:

$$c = \eta_{\mathbf{y}}$$

تخمینی را که mse را می‌نیمم می‌کند، تخمین ls (Least Square) یا LMS (Least Mean Square) می‌نامند.
در اینجا دیدیم که تخمین حداقل مربعات $\mathbf{y}$ (بدون هیچ مشاهده)، همان میانگین آن است:

$$\hat{y}_{ls} = \eta_{\mathbf{y}} \quad \text{(بدون مشاهده)}$$

در این صورت حداقل خطای تخمین mmse (Minimum Mean Square Error) برابر خواهد بود با:

$$\mathrm{mmse} = \sigma_{\mathbf{y}}^2 \quad \text{(بدون مشاهده)}$$

با مشاهده:

حال می‌خواهیم بر مبنای اطلاعاتمان از مقدار متغیر تصادفی دیگری مثل $\mathbf{x}$، $\mathbf{y}$ را تخمین بزنیم (با توجه به اینکه $\mathbf{x}$ و $\mathbf{y}$ به نحوی ارتباط دارند و مستقل نباشند).

مثلاً مقدار $\mathbf{x} = x$ رؤیت شده و می‌خواهیم به وسیله تابع $\Phi(x)$ مقدار متغیر تصادفی $\mathbf{y}$ را تخمین بزنیم (به طوری که mse می‌نیمم شود):

$\hat{y}_{ls} = \Phi(x)$: منحنی رگرسیون

$\hat{\mathbf{y}}_{ls} = \Phi(\mathbf{x})$

خطای تخمین برابر است با: $\mathbf{y} - \hat{\mathbf{y}}$؛

برای به دست آوردن $\hat{\mathbf{y}}_{ls}$، باید تابع Φ را چنان تعیین کنیم که:

$$\text{mse} = E\left[(\mathbf{y} - \hat{\mathbf{y}})^2\right]$$

می‌نیمم شود، یعنی:

$$\text{minimize mse} \leftrightarrow \hat{\mathbf{y}} = \hat{\mathbf{y}}_{ls}$$

قضیه:

$$\hat{\mathbf{y}}_{ls} = E(\mathbf{y} \mid \mathbf{x})$$

اثبات:

$$mse = E\left[(\mathbf{y}-\hat{\mathbf{y}})^2\right] = E\left[(\mathbf{y}-\Phi(\mathbf{x}))^2\right] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y-\hat{y})^2 f_{\mathbf{xy}}(x,y)\,dx\,dy$$

$$= E\left[E\left((\mathbf{y}-\hat{\mathbf{y}})^2 \mid \mathbf{x}\right)\right] = \int_{-\infty}^{+\infty} f_{\mathbf{x}}(x)\left(\int_{-\infty}^{+\infty}(y-\hat{y})\,f_{\mathbf{y}}(y \mid x)\,dy\right)dx$$

برای اینکه mse می‌نیمم شود، عبارت داخل انتگرال را برای هر x می‌نیمم می‌کنیم:

$$A = \int_{-\infty}^{+\infty}(y-\hat{y})^2 f_{\mathbf{y}}(y \mid x)dy = \int_{-\infty}^{+\infty}\left(y-\Phi(x)\right)^2 f_{\mathbf{y}}(y \mid x)dy = E(\mathbf{y}^2 \mid x) - 2\hat{y}\,E(\mathbf{y} \mid x) + \hat{y}^2$$

$$\frac{\partial A}{\partial \hat{y}} = 0 \Rightarrow \hat{y}_{ls} = E(\mathbf{y} \mid x) : \forall x$$

پس:

$$\hat{\mathbf{y}}_{ls} = E(\mathbf{y} \mid \mathbf{x})$$

اصل تعامد:

خطا در تخمین ls بر هر تابعی از داده عمود است، یعنی:

$$E\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{ls}})g(\mathbf{x})\right]=E\left[\left(\mathbf{y}-E(\mathbf{y}\,|\,\mathbf{x})\right)g(\mathbf{x})\right]=0$$

این قضیه استفادۀ زیادی دارد.

اثبات: طبق حالت خاص در ویژگی دوم امید مشروط تابعی از دو متغیر تصادفی داریم:

$$E\left(g(\mathbf{x})h(\mathbf{y})\right)=E\left[g(\mathbf{x})E\left(h(\mathbf{y})\,|\,\mathbf{x}\right)\right]$$

به ازای $h(\mathbf{y})=\mathbf{y}$ داریم:

$$E\left(\mathbf{y}g(\mathbf{x})\right)=E[g(\mathbf{x})E(\mathbf{y}|\,\mathbf{x})]\Rightarrow E\left[g(\mathbf{x})\left(\mathbf{y}-E(\mathbf{y}\,|\,\mathbf{x})\right)\right]=0$$

تعریف: تخمین را بدون بایاس (Unbiased) (ناریب) گویند، هرگاه:

$$E(\hat{\mathbf{y}})=E(\mathbf{y})$$

یعنی $E(\mathrm{error})$ صفر باشد.

قضیه: تخمین ls نااریب است.

زیرا:

$$E(\hat{\mathbf{y}}_{ls}) = E\left[E(\mathbf{y} \mid \mathbf{x})\right] = E(\mathbf{y})$$

قضیه: مقدار حداقل mse که به وسیلهٔ تخمین ls حاصل می‌شود، عبارت است از:

$$\text{mmse} = E(\mathbf{y}^2) - E(\mathbf{y}\,\hat{\mathbf{y}}_{ls}) = E(\mathbf{y}^2) - E(\hat{\mathbf{y}}_{ls}^2) = \sigma_{\mathbf{y}}^2 - \sigma_{\hat{\mathbf{y}}_{ls}}^2$$

اثبات:

$$\text{mmse} = E\left[(\mathbf{y} - \hat{\mathbf{y}}_{ls})^2\right] = E\left[(\mathbf{y} - \hat{\mathbf{y}}_{ls})\mathbf{y}\right] - E\left[(\mathbf{y} - \hat{\mathbf{y}}_{ls})\hat{\mathbf{y}}_{ls}\right]$$

ولی بنا بر اصل تعامد داریم:

$$E\left[(\mathbf{y} - \hat{\mathbf{y}}_{ls})\hat{\mathbf{y}}_{ls}\right] = 0$$

$$\Rightarrow \text{mmse} = E(\mathbf{y}^2 - \mathbf{y}\,\hat{\mathbf{y}}_{ls}) = E(\mathbf{y}^2) - E(\mathbf{y}\,\hat{\mathbf{y}}_{ls})$$

ولی با توجه به اصل تعامد $E(\mathbf{y}\,\hat{\mathbf{y}}_{ls}) = E(\hat{\mathbf{y}}_{ls}^2)$، پس:

$$\text{mmse} = E(\mathbf{y}^2) - E(\hat{\mathbf{y}}_{ls}^2)$$

می دانیم که: $E(\mathbf{y}) = E(\hat{\mathbf{y}}_{ls})$، پس:

$$\text{mmse} = E(\mathbf{y}^2) - E(\hat{\mathbf{y}}_{ls}^2) - \left(E(\mathbf{y})\right)^2 + \left(E(\hat{\mathbf{y}}_{ls})\right)^2 = \sigma_{\mathbf{y}}^2 - \sigma_{\hat{\mathbf{y}}_{ls}}^2$$

نتیجهٔ ۱: mmse نسبت به حالتی که بدون مشاهدهٔ $\mathbf{x}$، $\mathbf{y}$ را تخمین زده بودیم، کمتر شده است:

$$\sigma_{\mathbf{y}}^2 - \sigma_{\hat{\mathbf{y}}_{\text{ls}}}^2 < \sigma_{\mathbf{y}}^2$$

نتیجهٔ ۲ (قضیهٔ رائو- بِلَکوِل): $\sigma_{\hat{\mathbf{y}}_{\text{ls}}}^2 \le \sigma_{\mathbf{y}}^2$، زیرا $\text{mmse} \ge 0$.

مثال: اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً گوسی باشند، تخمین بهینهٔ $\mathbf{y}$ بر اساس مشاهدهٔ $\mathbf{x}$ عبارت است از:

$$\hat{\mathbf{y}}_{\text{ls}} = \mathrm{E}(\mathbf{y}\,|\,\mathbf{x}) = \eta_{\mathbf{y}} + \mathrm{r}\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x} - \eta_{\mathbf{x}}) = \eta_{\mathbf{y}} + \frac{\mu_{\mathbf{xy}}}{\sigma_{\mathbf{x}}^2}(\mathbf{x} - \eta_{\mathbf{x}})$$

مشاهده می‌شود که در اینجا تابع تخمین خطی است. ولی در حالت کلی تابع $\Phi(\mathrm{x}) = \mathrm{E}(\mathbf{y}\,|\,\mathrm{x})$ تابعی خطی نیست و ممکن است تابع پیچیده‌ای باشد (اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال باشند، $\mathrm{E}(\mathbf{y}\,|\,\mathrm{x})$ خطی است، ولی عکس این مطلب صحیح نیست. مثال در کتاب فرآیند Papoulis، فصل ۷، صفحهٔ ۱۹).

تخمین خطی حداقل مربعات (lls):

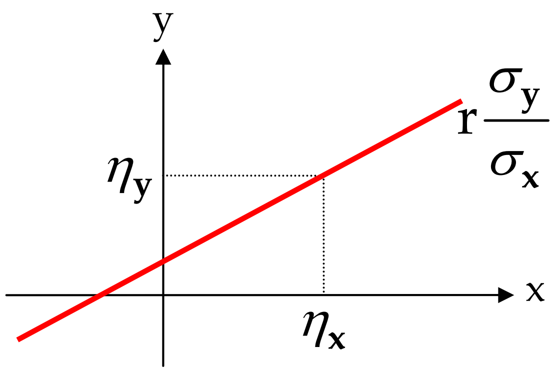
اگر الزام داریم که تابع تخمین $\Phi(x)$ خطی باشد، یعنی: $\Phi(x)=a+bx$ (رگرسیون خطی)، آنگاه باید a و b را چنان تعیین کنیم که mse می‌نیمم شود (بهترین تابع خطی را می یابیم، طبیعتاً ممکن است تابعی غیر خطی وجود داشته باشد که mse را کمتر از این کند. تابع $E(\mathbf{y}|x)$ کمترین mse را در میان تمام توابع می‌داد). پس داریم:

$$mse=E\left[\left(\mathbf{y}-(a+b\mathbf{x})\right)^2\right]=E(\mathbf{y}^2)+a^2+b^2E(\mathbf{x}^2)-2aE(\mathbf{y})-2bE(\mathbf{xy})+2abE(\mathbf{x})$$

$$\left.\begin{array}{l}\dfrac{\partial mse}{\partial a}=0\Rightarrow 2a-2E(\mathbf{y})+2bE(\mathbf{x})=0\\ \dfrac{\partial mse}{\partial b}=0\Rightarrow 2bE(\mathbf{x}^2)-2E(\mathbf{xy})+2aE(\mathbf{x})=0\end{array}\right\}\Rightarrow\left\{\begin{array}{l}a=\eta_y+r\dfrac{\sigma_y}{\sigma_x}\eta_x\\ b=r\dfrac{\sigma_y}{\sigma_x}\end{array}\right.$$

$$\Rightarrow \hat{\mathbf{y}}_{lls}=\eta_y+r\frac{\sigma_y}{\sigma_x}(\mathbf{x}-\eta_x)$$

یعنی خطی که از نقطهٔ (η_x,η_y) می‌گذرد و شیب آن $r\dfrac{\sigma_y}{\sigma_x}$ است.



این همان چیزی است که برای تخمین ls (بدون الزام به خطی بودن) در مورد فرآیند نرمال یافتیم. یعنی اگر فرآیند نرمال باشد ($\mathbf{x}$ و $\mathbf{y}$ مشترکاً نرمال باشند)، تخمین lls بهینه است (همان تخمین ls است).

قضیه: در تخمین lls خطا بر داده عمود است، یعنی:

$$\mathrm{E}\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{lls}})\mathbf{x}\right]=0$$

اثبات:

$$\mathrm{E}(\hat{\mathbf{y}}_{\mathrm{lls}}\mathbf{x})=\mathrm{E}\left[\eta_{\mathrm{y}}\mathbf{x}+\mathrm{r}\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x}-\eta_{\mathrm{x}})\mathbf{x}\right]=\eta_{\mathrm{y}}\eta_{\mathrm{x}}+\mathrm{r}\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathbf{x}}}\underbrace{\left(\mathrm{E}(\mathbf{x}^2)-\eta_{\mathbf{x}}^2\right)}_{\sigma_{\mathbf{x}}^2}=\sigma_{\mathbf{x}}^2\eta_{\mathrm{y}}\eta_{\mathrm{x}}+\underbrace{\mathrm{r}\,\sigma_{\mathbf{x}}\,\sigma_{\mathbf{y}}}_{\mu_{\mathrm{xy}}}$$

$$=\eta_{\mathrm{y}}\eta_{\mathrm{x}}+\mathrm{E}(\mathbf{xy})-\mathrm{E}(\mathbf{x})\mathrm{E}(\mathbf{y})=\mathrm{E}(\mathbf{xy})$$

$$\Rightarrow \mathrm{E}\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{lls}})\mathbf{x}\right]=0$$

توجه کنید که در تخمین ls خطا بر هر تابعی از داده‌ها عمود بود، ولی در اینجا فقط بر خود $\mathbf{x}$ عمود است.

قضیه: تخمین lls ناااریب است.

زیرا:

$$E(\hat{\mathbf{y}}_{\text{lls}}) = \eta_y + r\frac{\sigma_y}{\sigma_x}E(\mathbf{x}-\eta_{\mathbf{x}}) = E(\mathbf{y})$$

قضیه: در تخمین lls داریم:

$$\text{mse} = \sigma_{\mathbf{y}}^2(1-r^2)$$

اثبات:

$$E\left[(\mathbf{y}-\hat{\mathbf{y}}_{\text{lls}})^2\right] = E\left[\left(\mathbf{y}-\eta_{\mathbf{y}}-r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x}-\eta_{\mathbf{x}})\right)^2\right]$$

$$= E\left[(\mathbf{y}-\eta_{\mathbf{y}})^2\right] + r^2\frac{\sigma_{\mathbf{y}}^2}{\sigma_{\mathbf{x}}^2}E\left[(\mathbf{x}-\eta_{\mathbf{x}})^2\right] - 2r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}E\left[(\mathbf{x}-\eta_{\mathbf{x}})(\mathbf{y}-\eta_{\mathbf{y}})\right]$$

$$= \sigma_{\mathbf{y}}^2 + r^2\sigma_{\mathbf{y}}^2 - 2r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}\cdot r\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} = \sigma_{\mathbf{y}}^2 - r^2\sigma_{\mathbf{y}}^2 = \sigma_{\mathbf{y}}^2(1-r^2)$$

(در اینجا هم ملاحظه می‌کنید که mse نسبت به خطای تخمین بدون مشاهده ($\sigma_{\mathbf{y}}^2$) کمتر است.)

فصل ۷: دنبالهٔ متغیرهای تصادفی

Chapter 7

۱. مفاهیم کلی (pmf، CDF و pdf مشترک، استقلال، امید ریاضی، کواریانس، تابع مولد گشتاور، توابعی از n متغیر تصادفی، متغیرهای تصادفی مشترکاً نرمال، توزیع‌های شرطی)

۲. کاربردها (مجموع‌های تصادفی، نمونه‌برداری، آماره‌های رتبه، مجموع متغیرهای تصادفی مشترکاً نرمال)

۳. قضیهٔ حد مرکزی

۴. همگرایی دنبالهٔ متغیرهای تصادفی

۵. قانون اعداد بزرگ

۶. توابع توزیع متداول در آمار

در فصل ۵ دیدیم که چگونه مفاهیمی را که داشتیم به دو متغیر تصادفی توسعه دهیم. به طریق مشابه، وقتی چندتا متغیر تصادفی داشته باشیم، می‌توانیم مفاهیم گذشته (تابع CDF مشترک، pdf مشترک، مشروط و ...) را مطرح کنیم.

در ابتدا بردار $\underline{\mathbf{x}}$ را به صورت زیر تعریف می‌کنیم:

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$$

pmf مشترک:

$$P_{\underline{\mathbf{x}}}(\underline{x}) = P_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = P\{\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \ldots, \mathbf{x}_n = x_n\}$$

که داریم:

$$\underline{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

CDF مشترک:

$$F_{\underline{\mathbf{x}}}(\underline{x}) = F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = P\{\mathbf{x}_1 \le x_1, \mathbf{x}_2 \le x_2, \ldots, \mathbf{x}_n \le x_n\}$$

مقدار این تابع همواره بین صفر و یک است. به ازای تمام آرگومان‌ها صعودی است. در نقطهٔ $(-\infty, -\infty, \cdots, -\infty)$ برابر صفر بوده و در نقطهٔ $(+\infty, +\infty, \cdots, +\infty)$ برابر یک می‌شود. اگر به جای برخی از آرگومان‌ها بی‌نهایت بگذاریم، CDF مشترک سایر متغیرهای تصادفی حاصل می‌شود. مثلاً داریم:

$$F_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(+\infty, x_2, x_3) = F_{\mathbf{x}_2\mathbf{x}_3}(x_2, x_3)$$

pdf مشترک:

$$f_{\underline{\mathbf{x}}}(\underline{x}) = f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\underline{\mathbf{x}}}(\underline{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n)$$

به این ترتیب روابط زیر برقرار خواهند بود:

$$f_{\underline{\mathbf{x}}}(\underline{x}) dx_1 dx_2 \cdots dx_n = P\{x_1 \le \mathbf{x}_1 \le x_1 + dx_1, x_2 \le \mathbf{x}_2 \le x_2 + dx_2, \ldots, x_n \le \mathbf{x}_n \le x_n + dx_n\}$$

$$F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(u_1, u_2, \ldots, u_n) du_n \cdots du_2 du_1$$

مقدار تابع pdf همواره مثبت است. انتگرال nگانهٔ آن از $-\infty$ تا $+\infty$ یک می‌شود. همچنین داریم:

$$\forall D \subset \mathbb{R}^n : P\{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \in D\} = \underset{D}{\int \cdots \int} f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1, x_2, \ldots, x_n) dx_1 \cdots dx_n = \int_D f_{\underline{\mathbf{x}}}(\underline{x}) d\underline{x}$$

اگر از $-\infty$ تا $+\infty$ روی برخی از آرگومان‌ها انتگرال بگیریم، pdf مشترک سایر متغیرهای تصادفی حاصل می‌شود. مثلاً داریم:

$$f_{\mathbf{x}_2\mathbf{x}_3}(x_2, x_3) = \int_{-\infty}^{+\infty} f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1, x_2, x_3) dx_1$$

استقلال:

متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مستقل گویند، هرگاه واقعه‌های $\{\mathbf{x}_1 \le x_1\}$، $\{\mathbf{x}_2 \le x_2\}$، ... و $\{\mathbf{x}_n \le x_n\}$ برای هر x_1، هر x_2، ... و هر x_n مستقل باشند.

معادلاً $\mathbf{x}_i$ها را مستقل گویند، هرگاه برای هر x_1، هر x_2، ... و هر x_n داشته باشیم:

$$F_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=F_{\mathbf{x}_1}(x_1)F_{\mathbf{x}_2}(x_2)\cdots F_{\mathbf{x}_n}(x_n)=\prod_{i=1}^{n}F_{\mathbf{x}_i}(x_i)$$

یا معادلاً:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=f_{\mathbf{x}_1}(x_1)f_{\mathbf{x}_2}(x_2)\cdots f_{\mathbf{x}_n}(x_n)=\prod_{i=1}^{n}f_{\mathbf{x}_i}(x_i)$$

(در این صورت در مورد هر زیرمجموعه از اندیس‌ها نیز برقرار خواهد بود.)

مشابه حالت دو بعدی می‌توان ثابت کرد که اگر $\mathbf{x}_i$ها مستقل باشند، $g_i(\mathbf{x}_i)$ها هم مستقل خواهند بود.

همچنین داریم:

$$\underbrace{\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_m}_{}^{\boxed{\text{گروه } A}},\underbrace{\mathbf{x}_{m+1},\ldots,\mathbf{x}_n}^{\boxed{\text{گروه } B}}$$

گروه A از گروه B مستقل است، اگر:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_m}(x_1,x_2,\ldots,x_m)f_{\mathbf{x}_{m+1}\cdots\mathbf{x}_n}(x_{m+1},\ldots,x_n)$$

امید ریاضی:

اگر $\mathbf{z}=g(\underline{\mathbf{x}})$ باشد، داریم:

$$E(\mathbf{z})=\int_{-\infty}^{+\infty} z f_{\mathbf{z}}(z)dz$$

ولی بدون محاسبهٔ $f_{\mathbf{z}}$ نیز می‌توان $E(\mathbf{z})$ را حساب کرد:

$$E(\mathbf{z})=E[g(\underline{\mathbf{x}})]=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} g(\underline{x})f_{\underline{\mathbf{x}}}(\underline{x})d\underline{x}$$ (این انتگرال n بُعدی است)

و مشابه حالت دوبُعدی، از اینجا می‌توان خطی بودن امید ریاضی را ثابت کرد:

$$E\left[\sum_{i=1}^{n} a_i\, g_i(\underline{\mathbf{x}})\right]=\sum_{i=1}^{n} a_i\, E[g_i(\underline{\mathbf{x}})]$$ (یعنی جای امید و مجموع را می‌توان عوض کرد)

از جمله داریم:

$$E\left[\sum_{i=1}^{n} a_i \mathbf{x}_i\right]=\sum_{i=1}^{n} a_i E(\mathbf{x}_i)$$

برای درک نحوهٔ توزیع هر $\mathbf{x}_i$، η_i و σ_i به ما کمک می‌کنند که ایده‌ای کلی از نحوهٔ توزیع آن به دست آوریم. داریم:

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \quad \rightarrow \quad \eta_{\underline{\mathbf{x}}} = E(\underline{\mathbf{x}}) = \begin{bmatrix} E(\mathbf{x}_1) \\ E(\mathbf{x}_2) \\ \vdots \\ E(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}$$ (بردار میانگین)

و نیز دیده بودیم که کواریانس دو متغیر تصادفی، نحوهٔ بستگی آنها را به هم نشان می‌داد. حال با کواریانس دو به دوی این n تا متغیر تصادفی رو به رو هستیم:

$$\sigma_{ij} \triangleq \mu_{\mathbf{x}_i \mathbf{x}_j} = \operatorname{cov}(\mathbf{x}_i, \mathbf{x}_j) = E\left[(\mathbf{x}_i - \eta_i)(\mathbf{x}_j - \eta_j)\right] : i, j : 1, 2, \ldots, n$$

$$\sigma_{ii} = \operatorname{cov}(\mathbf{x}_i, \mathbf{x}_i) = \operatorname{var}(\mathbf{x}_i) = \sigma_i^2$$

ماتریس کواریانس $\underline{\mathbf{x}}$ (Covariance Matrix) طبق تعریف برابر است با:

$$C_{\underline{\mathbf{x}}} = E\left[(\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}}) \cdot (\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}})^T\right] = E\left(\begin{bmatrix} \mathbf{x}_1 - \eta_1 \\ \mathbf{x}_2 - \eta_2 \\ \vdots \\ \mathbf{x}_n - \eta_n \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x}_1 - \eta_1 & \mathbf{x}_2 - \eta_2 & \cdots & \mathbf{x}_n - \eta_n \end{bmatrix} \right) = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}$$

(گاهی آن را با $C_{\underline{\mathbf{x}}\underline{\mathbf{x}}}$ نیز نشان داده و آن را Auto Covariance هم می‌نامند.)

خواص ماتریس کواریانس:

۱. متقارن است، یعنی: $C = C^T$، زیرا: $\sigma_{ij} = \sigma_{ji}$.

۲. معیّن نامنفی (Nonnegative Definite) است، یعنی: برای هر بردار دلخواه $\underline{a}$ داریم:

$$\underline{a}^T C \underline{a} \geq 0$$

(اثبات این قضیه در تمرین‌ها خواسته شده است.)

از جمله خواص ماتریس‌های معین نامنفی این است که دترمینان آنها نامنفی است، پس داریم:

$$\det(C) \geq 0$$

ماتریس همبستگی $\underline{\mathbf{x}}$ (Correlation Matrix) طبق تعریف برابر است با:

$$R_{\underline{\mathbf{x}}} = E(\underline{\mathbf{x}} \cdot \underline{\mathbf{x}}^T)$$

از تعریف ماتریس‌های کواریانس و همبستگی نتیجه می‌شود:

$$C_{\underline{\mathbf{x}}} = R_{\underline{\mathbf{x}}} - \eta_{\underline{\mathbf{x}}} \cdot \eta_{\underline{\mathbf{x}}}^{T}$$

(شبیه به رابطهٔ $\sigma^2 = E(\mathbf{x}^2) - \left(E(\mathbf{x})\right)^2$)

در حالت کلی‌تر، می‌توانیم ماتریس‌های زیر را تعریف کنیم:

$$C_{\underline{\mathbf{x}}\underline{\mathbf{y}}} = \operatorname{cov}(\underline{\mathbf{x}}, \underline{\mathbf{y}}) = E\left[(\underline{\mathbf{x}} - \eta_{\underline{\mathbf{x}}})(\underline{\mathbf{y}} - \eta_{\underline{\mathbf{y}}})^{T} \right] \quad : \text{Cross Covariance}$$

$$R_{\underline{\mathbf{x}}\underline{\mathbf{y}}} = E(\underline{\mathbf{x}} \cdot \underline{\mathbf{y}}^{T}) \quad : \text{Cross Correlation}$$

اگر $\mathbf{z} = \sum_{i=1}^{n} a_i \mathbf{x}_i$ باشد، خواهیم داشت:

$$E(\mathbf{z}) = \sum_{i=1}^{n} a_i E(\mathbf{x}_i) = \sum_{i=1}^{n} a_i\, \eta_i$$

$$\operatorname{var}(\mathbf{z}) = E\left[(\mathbf{z} - \eta_{\mathbf{z}})^2 \right] = E\left[\left(\sum_{i=1}^{n} a_i (\mathbf{x}_i - \eta_i) \right)^2 \right] = E\left(\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j (\mathbf{x}_i - \eta_i)(\mathbf{x}_j - \eta_j) \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \sigma_{ij}$$

یعنی اگر $\mathbf{z}=\underline{a}^{T}\underline{\mathbf{x}}$ باشد، داریم:

$$\eta_{\mathbf{z}}=\underline{a}^{T}\eta_{\underline{\mathbf{x}}}$$
$$\sigma_{\mathbf{z}}^{2}=\underline{a}^{T}C_{\underline{\mathbf{x}}}\,\underline{a}$$

زیرا:

$$\sigma_{\mathbf{z}}^{2}=E\left[(\mathbf{z}-\eta_{\mathbf{z}})^{2}\right]=E\left[\underline{a}^{T}(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^{T}\underline{a}\right]=\underline{a}^{T}E\left[(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^{T}\right]\underline{a}=\underline{a}^{T}C_{\underline{\mathbf{x}}}\,\underline{a}$$

$\mathbf{x}_i$ها را <u>ناهمبسته</u> گوییم، اگر برای هر $i\neq j$ داشته باشیم: $\sigma_{ij}=0$؛ یعنی $C_{\underline{\mathbf{x}}}$ قطری باشد.

در این صورت خواهیم داشت:

$$\sigma_{\mathbf{z}}^{2}=\sum_{i=1}^{n}a_i^2\sigma_i^2$$ (اصل سوپر پوزیشن قدرت)

$\mathbf{x}_i$ها را <u>متعامد</u> گوییم، اگر برای هر $i\neq j$ داشته باشیم: $E(\mathbf{x}_i\mathbf{x}_j)=0$؛ یعنی $R_{\underline{\mathbf{x}}}$ قطری باشد.

مشابه حالت دوبُعدی به سادگی می‌توان نشان داد که:

اگر $\mathbf{x}_i$ها مستقل باشند، ناهمبسته هم خواهند بود. همچنین $g_i(\mathbf{x}_i)$ها نیز ناهمبسته خواهند بود.

تابع مولد گشتاور و تابع مشخصه:

$$\Phi_{\underline{\mathbf{x}}}(s_1, s_2, \ldots, s_n) = E(e^{s_1\mathbf{x}_1 + s_2\mathbf{x}_2 + \cdots + s_n\mathbf{x}_n}) = E(e^{\sum_{i=1}^{n} s_i\mathbf{x}_i})$$

$$\Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n) = E(e^{j\omega_1\mathbf{x}_1 + j\omega_2\mathbf{x}_2 + \cdots + j\omega_n\mathbf{x}_n}) = E(e^{j\sum_{i=1}^{n} \omega_i\mathbf{x}_i})$$

به این ترتیب خواهیم داشت:

$$\frac{1}{j^{k_1+k_2+\cdots+k_n}} \frac{\partial^{k_1}\partial^{k_2}\cdots\partial^{k_n}}{\partial\omega_1\partial\omega_2\cdots\partial\omega_n} \Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n)\Big|_{\omega_1=\omega_2=\cdots=\omega_n=0} = E(\mathbf{x}_1^{k_1}\mathbf{x}_2^{k_2}\cdots\mathbf{x}_n^{k_n})$$

اگر $\mathbf{z} = \sum_{i=1}^{n} \mathbf{x}_i$ باشد، داریم:

$$\Phi_{\mathbf{z}}(j\omega) = E(e^{j\omega\mathbf{z}}) = E(e^{j\sum_{i=1}^{n}\omega\mathbf{x}_i}) = \Phi_{\underline{\mathbf{x}}}(j\omega, j\omega, \ldots, j\omega)$$

همچنین $\mathbf{x}_i$ها مستقلند، اگر و فقط اگر داشته باشیم:

$$\Phi_{\underline{\mathbf{x}}}(j\omega_1, j\omega_2, \ldots, j\omega_n) = \Phi_{\mathbf{x}_1}(j\omega_1)\Phi_{\mathbf{x}_2}(j\omega_2)\cdots\Phi_{\mathbf{x}_n}(j\omega_n) = \prod_{i=1}^{n}\Phi_{\mathbf{x}_i}(j\omega_i)$$

توابعی از n متغیر تصادفی:

اگر یک تابع از n متغیر تصادفی داشته باشیم، از چهار روشی که قبلاً در حالت دو متغیر تصادفی گفته بودیم، در اینجا نیز می‌توان برای به دست آوردن توزیع آن استفاده کرد.

مثال: متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$ و $\mathbf{x}_3$ مستقل بوده و همگی $N(0,1)$ هستند. اگر $\mathbf{v}=\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2$ باشد، داریم:

$$F_{\mathbf{v}}(v)=P\{\mathbf{v}\le v\}=P\{\mathbf{x}_1^2+\mathbf{x}_2^2+\mathbf{x}_3^2\le v\}$$

برای $v<0$:

$$F_{\mathbf{v}}(v)=0$$

برای $v>0$:

$$F_{\mathbf{v}}(v)=\underset{\text{ناحیهٔ } x_1^2+x_2^2+x_3^2\le v}{\iiint} f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1,x_2,x_3)\,dx_1dx_2dx_3$$

چون $\mathbf{x}_1$، $\mathbf{x}_2$ و $\mathbf{x}_3$ مستقل بوده و همگی $N(0,1)$ هستند، پس:

$$f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3}(x_1,x_2,x_3)=\frac{1}{(2\pi)^{\frac{3}{2}}}e^{-\frac{x_1^2+x_2^2+x_3^2}{2}}$$

(چنین توزیع‌هایی را که فقط تابع r (فاصله از مبدأ) هستند، متقارن کروی گویند. استقلال و تقارن کروی، نرمال بودن را نتیجه می‌دهند.)

پس:

$$F_{\mathbf{v}}(v)=\frac{1}{(2\pi)^{\frac{3}{2}}}\underset{\substack{\text{داخل کره‌ای} \\ \text{به شعاع } \sqrt{v}}}{\iiint} e^{-\frac{x_1^2+x_2^2+x_3^2}{2}}\,dx_1dx_2dx_3$$

با تبدیل به مختصات کروی خواهیم داشت:

$$\begin{cases} x_1 = r\sin\varphi\cos\theta \\ x_2 = r\sin\varphi\sin\theta \\ x_3 = r\cos\varphi \end{cases}$$

پس در نهایت داریم:

$$F_{\mathbf{v}}(v)=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi}\int_0^{\pi}\int_0^{\sqrt{v}} e^{-\frac{r^2}{2}} r^2 \sin\varphi \, dr\, d\varphi\, d\theta=\frac{1}{(2\pi)^{\frac{3}{2}}}\int_0^{2\pi} d\theta\int_0^{\pi}\sin\varphi\, d\varphi\int_0^{\sqrt{v}} r^2 e^{-\frac{r^2}{2}} dr$$

$$=\frac{(2\pi)(2)}{(2\pi)^{\frac{3}{2}}}\int_0^{\sqrt{v}} r^2 e^{-\frac{r^2}{2}} dr$$

$$f_{\mathbf{v}}(v)=\frac{d}{dv}F_{\mathbf{v}}(v)=\frac{2}{\sqrt{2\pi}}\cdot\frac{1}{2\sqrt{v}} v e^{-\frac{v}{2}}=\frac{1}{\sqrt{2\pi}} v^{\frac{1}{2}} e^{-\frac{v}{2}} : v>0$$

که توزیع χ^2 با سه درجه آزادی است.

اگر n تابع از n متغیر تصادفی داشته باشیم، یعنی:

$$\begin{cases} \mathbf{y}_1 = g_1(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \\ \mathbf{y}_2 = g_2(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \\ \qquad \vdots \\ \mathbf{y}_n = g_n(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n) \end{cases} \quad \rightarrow \quad \underline{\mathbf{y}} = \underline{g}(\underline{\mathbf{x}})$$

ابتدا دستگاه $\underline{y} = \underline{g}(\underline{x})$ را حل می‌کنیم. اگر ریشهٔ آن برابر با $\underline{h}(\underline{y}) = \begin{bmatrix} h_1(\underline{y}) \\ h_2(\underline{y}) \\ \vdots \\ h_n(\underline{y}) \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \underline{x}$ باشد، خواهیم داشت:

$$f_{\underline{\mathbf{y}}}(\underline{y}) = \frac{f_{\underline{\mathbf{x}}}(x_1, x_2, \ldots, x_n)}{|J(x_1, x_2, \ldots, x_n)|} = \frac{f_{\underline{\mathbf{x}}}\left(\underline{h}(\underline{y})\right)}{\left|J\left(\underline{h}(\underline{y})\right)\right|}$$

که:

$$J(x_1, x_2, \ldots, x_n) = \det\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_n} \end{bmatrix}$$

یا معادلاً:

$$f_{\underline{y}}(\underline{y}) = f_{\underline{x}}\left(\underline{h}(\underline{y})\right) | \det \begin{bmatrix} \frac{\partial h_1}{\partial y_1} & \cdots & \frac{\partial h_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_n}{\partial y_1} & \cdots & \frac{\partial h_n}{\partial y_n} \end{bmatrix} |$$

(اگر چند ریشه وجود داشته باشد، عبارت فوق را در تمام ریشه‌ها حساب کرده و جمع می‌کنیم.)

مثال: اگر $\underline{\mathbf{y}} = A\underline{\mathbf{x}}$ باشد، یعنی: $\mathbf{y}_i = \sum_{j=1}^{n} a_{ij}\,\mathbf{x}_j : i = 1,2,\ldots,n$؛ $f_{\underline{\mathbf{y}}}$، $\eta_{\underline{\mathbf{y}}}$ و $C_{\underline{\mathbf{y}}}$ را پیدا کنید.

$$\underline{y}=A\underline{x} \quad \Rightarrow \quad \underline{x}=A^{-1}\underline{y}$$

$$J=\det(A)$$

$$f_{\underline{\mathbf{y}}}(\underline{y})=\frac{f_{\mathbf{x}}(A^{-1}\underline{y})}{\left|\det(A)\right|}$$

$$\eta_{\underline{\mathbf{y}}}=E(\underline{\mathbf{y}})=E(A\underline{\mathbf{x}})=A\,\eta_{\underline{\mathbf{x}}}$$

$$C_{\underline{\mathbf{y}}}=E\left[(\underline{\mathbf{y}}-\eta_{\underline{\mathbf{y}}})\cdot(\underline{\mathbf{y}}-\eta_{\underline{\mathbf{y}}})^T\right]=E\left[(A\underline{\mathbf{x}}-A\,\eta_{\underline{\mathbf{x}}})\cdot(A\underline{\mathbf{x}}-A\,\eta_{\underline{\mathbf{x}}})^T\right]=E\left[A(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\cdot(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T A^T\right]=AC_{\underline{\mathbf{x}}}A^T$$

برای مثال داریم:

$$\begin{cases} \mathbf{y}_1 = \mathbf{x}_1 \\ \mathbf{y}_2 = \mathbf{x}_1 + \mathbf{x}_2 \\ \quad\vdots \\ \mathbf{y}_n = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_n \end{cases} \Rightarrow \quad A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 1 & 1 & 0 & \cdots & 0 \\ 1 & 1 & 1 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 0 \\ 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \Rightarrow \quad \det(A) = 1$$

$$\underline{y} = A\underline{x} \quad \Rightarrow \quad \begin{cases} x_1 = y_1 \\ x_2 = y_2 - y_1 \\ x_3 = y_3 - y_2 \\ \quad\vdots \\ x_n = y_n - y_{n-1} \end{cases}$$

$$f_{\underline{\mathbf{y}}}(\underline{y}) = f_{\underline{\mathbf{x}}}(y_1, y_2 - y_1, y_3 - y_2, \ldots, y_n - y_{n-1})$$

حال اگر $\mathbf{x}_i$ها مستقل باشند، $f_{\underline{\mathbf{x}}}(\underline{x}) = f_{\mathbf{x}_1}(x_1) f_{\mathbf{x}_2}(x_2) \cdots f_{\mathbf{x}_n}(x_n)$ و لذا:

$$f_{\underline{\mathbf{y}}}(\underline{y}) = f_{\mathbf{x}_1}(y_1) f_{\mathbf{x}_2}(y_2 - y_1) f_{\mathbf{x}_3}(y_3 - y_2) \cdots f_{\mathbf{x}_n}(y_n - y_{n-1})$$

متغیرهای تصادفی مشترکاً نرمال:

تعریف: متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مشترکاً نرمال گویند، هرگاه $\mathbf{z}=\sum_{i=1}^{n} a_i \mathbf{x}_i$ برای هر a_i نرمال باشد.

توابع $f_{\underline{\mathbf{x}}}$ و $\Phi_{\underline{\mathbf{x}}}$ برای متغیرهای تصادفی مشترکاً نرمال به صورت تابعی نمایی از یک فرم درجهٔ دوم هستند:

$$f_{\underline{\mathbf{x}}}(\underline{x})=\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(C_{\underline{\mathbf{x}}})}}\exp\left[-\frac{1}{2}(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})^T C_{\underline{\mathbf{x}}}^{-1}(\underline{\mathbf{x}}-\eta_{\underline{\mathbf{x}}})\right]$$

یعنی $f_{\underline{\mathbf{x}}}$ به صورت $\alpha\, e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma_{ij}(x_i-\eta_i)(x_j-\eta_j)}$ است که γ_{ij}ها همان درایه‌های $C_{\underline{\mathbf{x}}}^{-1}$ هستند.

$$\Phi_{\underline{\mathbf{x}}}(j\underline{\omega})=e^{j\underline{\omega}^T\eta_{\underline{\mathbf{x}}}}\, e^{-\frac{1}{2}\underline{\omega}^T C_{\underline{\mathbf{x}}}\,\underline{\omega}}$$

یعنی $\Phi_{\underline{\mathbf{x}}}$ به صورت $e^{j\sum_{i=1}^{n}\omega_i\eta_i}\, e^{-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}\,\omega_i\,\omega_j}$ می‌باشد.

(این فرمول‌ها در حالت دوبُعدی به همان فرمول‌هایی که در فصل ۵ داشتیم تبدیل می‌شوند.)

ملاحظه می‌شود که در حالت مشترکاً نرمال، بردار میانگین و ماتریس کواریانس کلیهٔ خواص آماری $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ را مشخص می‌کنند.

قضیه: اگر هر یک از متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ نرمال بوده و مستقل نیز باشند، آنگاه مشترکاً نرمال خواهند بود.

(زیرا $\prod_{i=1}^{n} f_{\mathbf{x}_i}$ همان شکل فوق را خواهد داشت.)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مشترکاً نرمال بوده و ناهمبسته نیز باشند، آنگاه مستقل خواهند بود.

اثبات:

$$\sigma_{ij} = 0 : i \neq j \quad \Rightarrow \quad C_{\underline{\mathbf{x}}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix}$$

$$\Rightarrow f_{\underline{\mathbf{x}}}(\underline{x}) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\prod_{i=1}^{n} \sigma_i^2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n} \frac{(x_i - \eta_i)^2}{\sigma_i^2}\right\} = \prod_{i=1}^{n} f_{\mathbf{x}_i}(x_i)$$

(یا می‌توان نشان داد که: $\Phi_{\underline{\mathbf{x}}}(j\underline{\omega}) = \prod_{i=1}^{n} \Phi_{\mathbf{x}_i}(j\omega_i)$)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مشترکاً نرمال بوده و $\underline{\mathbf{y}} = A\underline{\mathbf{x}}$ باشد، یعنی: $\mathbf{y}_i = \sum_{j=1}^{n} a_{ij}\,\mathbf{x}_j : i = 1,2,\ldots,m$، $\mathbf{y}_i$ها هم مشترکاً نرمال خواهند بود.

اثبات:

$$\mathbf{z} = \sum_{i=1}^{m} c_i \mathbf{y}_i = \sum_{i=1}^{m} c_i \sum_{j=1}^{n} a_{ij}\,\mathbf{x}_j = \sum_{j=1}^{n} \underbrace{\left(\sum_{i=1}^{m} c_i\, a_{ij} \right)}_{b_j} \mathbf{x}_j = \sum_{j=1}^{n} b_j\,\mathbf{x}_j \rightarrow$$
بنا به فرض دارای توزیع نرمال است

یا می‌توانیم به صورت برداری نیز نشان دهیم:

$$\mathbf{z} = C^T \underline{\mathbf{y}} = C^T (A\underline{\mathbf{x}}) = (C^T A)\underline{\mathbf{x}} = b^T \underline{\mathbf{x}} \rightarrow$$
بنا به فرض دارای توزیع نرمال است

کاربرد:

نویز گوسی ← مدار خطی ← نویز گوسی

توزیع‌های شرطی:

مشابه حالت دوبُعدی در اینجا نیز داریم:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_k}(x_1,x_2,\ldots,x_k\,|\,\mathbf{x}_{k+1}=x_{k+1},\ldots,\mathbf{x}_n=x_n)=\frac{f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)}{f_{\mathbf{x}_{k+1}\cdots\mathbf{x}_n}(x_{k+1},\ldots,x_n)}\quad(*)$$

مثلاً داریم:

$$f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2\,|\,x_3,x_4)=\frac{f_{\mathbf{x}_1\mathbf{x}_2\mathbf{x}_3\mathbf{x}_4}(x_1,x_2,x_3,x_4)}{f_{\mathbf{x}_3\mathbf{x}_4}(x_3,x_4)}$$

با استفاده از رابطهٔ (*) می‌توان نشان داد که:

قاعدهٔ زنجیری:

$$f_{\mathbf{x}_1\mathbf{x}_2\cdots\mathbf{x}_n}(x_1,x_2,\ldots,x_n)=f_{\mathbf{x}_1}(x_1)f_{\mathbf{x}_2}(x_2\,|\,x_1)f_{\mathbf{x}_3}(x_3\,|\,x_2,x_1)\cdots f_{\mathbf{x}_n}(x_n\,|\,x_1,x_2,\ldots,x_{n-1})$$

حذف متغیر سمت چپ در چگالی شرطی:

می‌دانیم که اگر $f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2)$ را داشته باشیم و بخواهیم $f_{\mathbf{x}_1}$ را به دست آوریم، کافی است روی $f_{\mathbf{x}_1\mathbf{x}_2}$ از $-\infty$ تا $+\infty$ نسبت به x_2 انتگرال بگیریم. در چگالی شرطی نیز به همین ترتیب عمل می‌کنیم. یعنی مثلاً اگر بخواهیم از $f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)$ مقدار $f_{\mathbf{x}_1}(x_1|x_3)$ را به دست آوریم (توجه کنید که عین خواص pdf معمولی برقرار است)، خواهیم داشت:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)\,dx_2$$

(برای اثبات از تعریف چگالی شرطی و رابطهٔ چگالی مشترک حاشیه‌ای استفاده کنید.)

حذف متغیر سمت راست در چگالی شرطی:

اگر بخواهیم از $f_{\mathbf{x}_1}(x_1|x_2,x_3)$ به $f_{\mathbf{x}_1}(x_1|x_3)$ برسیم، داریم:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)dx_2$$

زیرا:

$$f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)=f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3)$$

لذا:

$$f_{\mathbf{x}_1}(x_1|x_3)=\int_{-\infty}^{+\infty} f_{\mathbf{x}_1}(x_1|x_2,x_3)f_{\mathbf{x}_2}(x_2|x_3)dx_2$$ (معادلهٔ چاپمن- کولموگروف)

اگر $\mathbf{x}_i$ها مستقل باشند، چگالی مشروط بعضی از آنها بر بعضی دیگر با چگالی غیرمشروط یکسان است. مثلاً:

$$f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2|x_3,x_4)=f_{\mathbf{x}_1\mathbf{x}_2}(x_1,x_2)$$

متوسط مشروط:

$$E(\mathbf{y}|\mathbf{x}_1 = x_1, \mathbf{x}_2 = x_2, \dots, \mathbf{x}_n = x_n) = E(\mathbf{y}|x_1, x_2, \dots, x_n) = \int_{-\infty}^{+\infty} y\, f_{\mathbf{y}}(y|x_1, x_2, \dots, x_n)\, dy$$

(که تابعی از x_1، x_2، ... و x_n است، نه y.)

$$E(g(\mathbf{y})|x_1, x_2, \dots, x_n) = \int_{-\infty}^{+\infty} g(y)\, f_{\mathbf{y}}(y|x_1, x_2, \dots, x_n)\, dy$$

(که تابعی از x_1، x_2، ... و x_n است، نه y.)

حال اگر فرض کنیم:

$$E(\mathbf{y}|\underline{x}) = \Phi(\underline{x})$$

متناظر با این تابع می‌توانیم متغیر تصادفی $\Phi(\underline{\mathbf{x}})$ را تعریف کنیم:

$$E(\mathbf{y}|\underline{\mathbf{x}}) = \Phi(\underline{\mathbf{x}})$$

مشابه قبل به سادگی می‌توان نشان داد که:

$$\mathrm{E}\left[\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})\right]=\mathrm{E}(\mathbf{y})$$

و اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، خواهیم داشت:

$$\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})=\mathrm{E}(\mathbf{y})$$

$\mathrm{E}(\mathbf{y}\,|\,\underline{\mathbf{x}})$ در واقع تخمین بهینه (به مفهوم mse) برای $\mathbf{y}$ بر مبنای مشاهدهٔ $\underline{\mathbf{x}}$ است.

برخی کاربردهای توزیع شرطی و دنبالهٔ متغیرهای تصادفی:

۱. مجموع‌های تصادفی (Random Sums):

فرض کنید $\mathbf{y}=\sum_{i=1}^{N}\mathbf{x}_i$ باشد که در آن، $\mathbf{x}_i$ها متغیرهای تصادفی مستقل بوده و نیز داریم:

$$\forall i: E(\mathbf{x}_i)=\eta_{\mathbf{x}}\ ,\ \mathrm{var}(\mathbf{x}_i)=\sigma_{\mathbf{x}}^2$$

از طرف دیگر، خود $\mathbf{N}$ نیز یک متغیر تصادفی با میانگین $\eta_{\mathbf{N}}$ و واریانس $\sigma_{\mathbf{N}}^2$ باشد و $\mathbf{x}_i$ها از $\mathbf{N}$ نیز مستقل باشند.

مثلاً $\mathbf{y}$ می‌تواند سود روزانهٔ یک مغازه، $\mathbf{x}_i$ سود حاصل از هر مشتری و $\mathbf{N}$ تعداد مشتری‌هایی که در روز مراجعه می‌کنند، باشد.
یا در مسألهٔ Multipath Fading یا انرژی کل الکترون‌های منتشر شده از یک مادهٔ رادیواکتیو ($\mathbf{N}$ دارای توزیع پواسون).

حال می‌خواهیم میانگین و واریانس $\mathbf{y}$ را به دست آوریم:

$$E(\mathbf{y}) = E\left[E(\mathbf{y}|\mathbf{N})\right]$$

$$E(\mathbf{y}|\mathbf{N}=n) = E\left(\sum_{i=1}^{n}(\mathbf{x}_i|n)\right) = \sum_{i=1}^{n}E(\mathbf{x}_i|n) = \sum_{i=1}^{n}E(\mathbf{x}_i)$$

ولی داشتیم:

$$\forall i : E(\mathbf{x}_i) = \eta_{\mathbf{x}}$$

پس داریم:

$$E(\mathbf{y}|n) = n\eta_{\mathbf{x}}$$

$$\Rightarrow E(\mathbf{y}|\mathbf{N}) = \mathbf{N}\eta_{\mathbf{x}}$$

$$\Rightarrow E(\mathbf{y}) = E\left[E(\mathbf{y}|\mathbf{N})\right] = E(\mathbf{N}\eta_{\mathbf{x}}) = \eta_{\mathbf{x}}E(\mathbf{N}) = \eta_{\mathbf{x}}\eta_{\mathbf{N}}$$

$$E(\mathbf{y}^2|n)=E\left[\left(\sum_{i=1}^{n}\mathbf{x}_i\right)^2\right]=E\left(\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{x}_i\mathbf{x}_j\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}E(\mathbf{x}_i\mathbf{x}_j)$$

به دلیل استقلال $\mathbf{x}_i$ها، برای $\mathbf{i}\neq\mathbf{j}$ داریم:

$$E(\mathbf{x}_i\mathbf{x}_j)=E(\mathbf{x}_i)E(\mathbf{x}_j)$$

در نتیجه:

$$E(\mathbf{y}^2|n)=\sum_{i=1}^{n}E(\mathbf{x}_i^2)+\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}E(\mathbf{x}_i)E(\mathbf{x}_j)$$

ولی داشتیم:

$$\forall \mathbf{i}:\mathrm{var}(\mathbf{x}_i)=\sigma_{\mathbf{x}}^2$$

پس خواهیم داشت:

$$E(\mathbf{y}^2|n)=n(\sigma_{\mathbf{x}}^2+\eta_{\mathbf{x}}^2)+(n^2-n)\eta_{\mathbf{x}}^2=n\sigma_{\mathbf{x}}^2+n^2\eta_{\mathbf{x}}^2$$
$$\Rightarrow E(\mathbf{y}^2|\mathbf{N})=\mathbf{N}\sigma_{\mathbf{x}}^2+\mathbf{N}^2\eta_{\mathbf{x}}^2$$
$$E(\mathbf{y}^2)=E\left[E(\mathbf{y}^2|\mathbf{N})\right]=\sigma_{\mathbf{x}}^2E(\mathbf{N})+\eta_{\mathbf{x}}^2E(\mathbf{N}^2)=\sigma_{\mathbf{x}}^2\eta_{\mathbf{N}}+\eta_{\mathbf{x}}^2(\sigma_{\mathbf{N}}^2+\eta_{\mathbf{N}}^2)$$
$$\Rightarrow\sigma_{\mathbf{y}}^2=E(\mathbf{y}^2)-\left(E(\mathbf{y})\right)^2=\sigma_{\mathbf{x}}^2\eta_{\mathbf{N}}+\eta_{\mathbf{x}}^2\sigma_{\mathbf{N}}^2+\eta_{\mathbf{x}}^2\eta_{\mathbf{N}}^2-\eta_{\mathbf{x}}^2\eta_{\mathbf{N}}^2=\sigma_{\mathbf{x}}^2\eta_{\mathbf{N}}+\eta_{\mathbf{x}}^2\sigma_{\mathbf{N}}^2$$

۲. نمونه‌برداری:

متغیر تصادفی $\mathbf{x}$ را طول قد مردان ایرانی در نظر گرفته و فرض کنیم (بنابر استدلالات یا مشاهدهٔ طولانی) که دارای چگالی $f_{\mathbf{x}}$ و توزیع $F_{\mathbf{x}}$ باشد (مثلاً $N(170,10)$). حال مردی را به عنوان نمونه (به طور کاملاً تصادفی) انتخاب کنیم و از شما سؤال کنیم احتمال اینکه قد این مرد کوچکتر از x باشد، چقدر است؟

خواهید گفت:

$$P\{\mathbf{x}_1 \le x\} = F_{\mathbf{x}}(x)$$

به همین ترتیب برای سایر نمونه‌ها نیز داریم:

$$F_{\mathbf{x}_i}(x) = F_{\mathbf{x}}(x)\ ,\ f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

از سوی دیگر، با فرض استقلال آزمایشها، $\mathbf{x}_i$ها مستقلند (تعداد نمونه‌ها باید نسبت به تعداد افراد جامعه کم باشد یا نمونه‌برداری با جایگزینی باشد).

یا مثلاً اگر متغیر تصادفی $\mathbf{x}$، قطر پیچ‌های تولیدی یک کارخانه باشد و فرض کنیم دارای چگالی $f_{\mathbf{x}}$ باشد، این مدلی است که طبق فرض برای کلیهٔ $\mathbf{x}$ها صادق است. پس اگر $\mathbf{x}_i$ قطر پیچ نمونهٔ i ام باشد، داریم:

$$f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

و $\mathbf{x}_i$ها (با فرض استقلال آزمایشها) مستقلند.

به طور کلی، متغیرهای تصادفی که مستقل و دارای توزیع یکسان باشند، i.i.d. (Independent Identically Distributed) نامیده می‌شوند.

اگر $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$ و $\mathbf{x}_i$ها i.i.d. باشند، داریم:

$$f_{\underline{\mathbf{x}}}(\underline{x}) \underset{\boxed{\text{استقلال}}}{\underset{\downarrow}{=}} f_{\mathbf{x}_1}(x_1) \cdots f_{\mathbf{x}_n}(x_n) \underset{\boxed{\text{یکسان بودن توزیع}}}{\underset{\downarrow}{=}} f_{\mathbf{x}}(x_1) \cdots f_{\mathbf{x}}(x_n)$$

با نمونه‌برداری مکرر از یک متغیر تصادفی در یک جامعه (Population) (مثلاً جامعهٔ پیچ‌ها در مثال بالا)، n متغیر تصادفی i.i.d. به دست می‌آیند. n را <u>اندازهٔ نمونه (Sample Size)</u> می‌گویند.

چون $\mathbf{x}_i$ها مستقلند، پس:

$$f_{\mathbf{x}_i}(x) = f_{\mathbf{x}}(x)$$

میانگین جامعه : $\Rightarrow E(\mathbf{x}_i) = E(\mathbf{x}) = \eta$

واریانس جامعه : $\Rightarrow \mathrm{var}(\mathbf{x}_i) = \mathrm{var}(\mathbf{x}) = \sigma^2$

بحث نمونه‌برداری نقش اساسی در آمار دارد.

میانگین نمونه (Sample Mean):

طبق تعریف، برابر است با:

$$\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$$

داریم:

$$E(\bar{\mathbf{x}}) = \frac{1}{n}\sum_{i=1}^{n}E(\mathbf{x}_i) = \frac{1}{n}\cdot n\eta = \eta$$

$$\sigma_{\bar{\mathbf{x}}}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_{\mathbf{x}_i}^2 = \frac{1}{n^2}\cdot n\sigma^2 = \frac{\sigma^2}{n}$$

پس هر چقدر n زیادتر شود، مقدار $\bar{\mathbf{x}}$ به η واقعی نزدیکتر خواهد بود. $\bar{\mathbf{x}}$ میانگین نمونه است، در حالی که η میانگین جامعه می‌باشد (پارامتر مدل مفروض).

با توجه به قضیهٔ چبیشف داریم:

$$P\{|\overline{\mathbf{x}}-\eta|<\varepsilon\}\geq 1-\frac{\sigma_{\overline{\mathbf{x}}}^{2}}{\varepsilon^{2}}$$

یعنی:

$$P\{|\overline{\mathbf{x}}-\eta|<\varepsilon\}\geq 1-\frac{\sigma^{2}}{n\varepsilon^{2}}$$

پس این احتمال با افزایش n به یک نزدیک می‌شود.

به همین ترتیب می‌توان در مورد واریانس نمونه بحث کرد که بعداً آن را خواهیم دید.

حال فرض کنید که یک آزمایش تصادفی را n بار انجام دهیم ($\mathbf{x}_i$ها طبق تعریفی که داشتیم تعریف شده باشند) و مقادیر x_i در این آزمایشها ظاهر شوند $(i=1,2,\ldots,n)$. این اعداد را به ترتیب صعودی مرتب می‌کنیم:

$$x_{r_1}\leq x_{r_2}\leq\cdots\leq x_{r_n}$$

اکنون نام مرتب شدهٔ آنها را y_i می‌نامیم، یعنی: $y_i=x_{r_i}$. پس خواهیم داشت:

$$y_1\leq y_2\leq\cdots\leq y_n$$

مثلاً اگر اعداد x_i به ترتیب زیر به دست آمده باشند، داریم:

$$\begin{cases} x_1 = 12 \\ x_2 = 7 \\ x_3 = 13 \\ x_4 = 8 \\ x_5 = 9 \end{cases} \Rightarrow \begin{cases} y_1 = x_{r_1} = x_2 = 7 \\ y_2 = x_{r_2} = x_4 = 8 \\ y_3 = x_{r_3} = x_5 = 9 \\ y_4 = x_{r_4} = x_1 = 12 \\ y_5 = x_{r_5} = x_3 = 13 \end{cases}$$

متغیرهای تصادفی $\mathbf{y}_i$ را آماره‌های رتبهٔ (آمارگان رتبهٔ) $\mathbf{x}_i$ها می‌نامند (آماره‌های رتبه کاربرد بسیاری در آشکارسازی و تخمین پارامتری دارند و در رادار و جنگ الکترونیک مورد استفاده قرار می‌گیرند).

آماره (Statistic):

طبق تعریف، تابعی از یک یا چند متغیر تصادفی را که به پارامتر نامعلومی بستگی نداشته باشند، آماره گویند. در بحث نمونه‌برداری، تابعی از صرفاً بردار متغیرهای تصادفی نمونهٔ $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$ را آماره گویند. مثلاً آمارهٔ رتبه یک نوع آماره است (زیرا تابعی از $\underline{\mathbf{x}}$ می‌باشد).

به عنوان حالت خاص برای $\mathbf{y}_1$ و $\mathbf{y}_n$ داریم:

$$\mathbf{y}_1 = \mathbf{x}_{\min} = \min(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$
$$\mathbf{y}_n = \mathbf{x}_{\max} = \max(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

پس:

$$\begin{aligned} F_{\mathbf{y}_1}(y) &= P\{\mathbf{y}_1 \le y\} = P\{\mathbf{x}_{\min} \le y\} = 1 - P\{\mathbf{x}_{\min} > y\} = 1 - P\{\mathbf{x}_1 > y, \mathbf{x}_2 > y, \ldots, \mathbf{x}_n > y\} \\ &= 1 - P\{\mathbf{x}_1 > y\} P\{\mathbf{x}_2 > y\} \cdots P\{\mathbf{x}_n > y\} \\ &= 1 - \left[1 - F_{\mathbf{x}}(y)\right]^n \end{aligned}$$
$$\Rightarrow f_{\mathbf{y}_1}(y) = n\left[1 - F_{\mathbf{x}}(y)\right]^{n-1} f_{\mathbf{x}}(y)$$

$$\begin{aligned} F_{\mathbf{y}_n}(y) &= P\{\mathbf{y}_n \le y\} = P\{\mathbf{x}_{\max} \le y\} = P\{\mathbf{x}_1 \le y, \mathbf{x}_2 \le y, \ldots, \mathbf{x}_n \le y\} \\ &= P\{\mathbf{x}_1 \le y\} P\{\mathbf{x}_2 \le y\} \cdots P\{\mathbf{x}_n \le y\} \\ &= \left[F_{\mathbf{x}}(y)\right]^n \end{aligned}$$
$$\Rightarrow f_{\mathbf{y}_n}(y) = n\left[F_{\mathbf{x}}(y)\right]^{n-1} f_{\mathbf{x}}(y)$$

(به کمک pdf نیز می‌توان روابط بالا را به دست آورد.)

در حالت کلی داریم:

$$f_{\mathbf{y}_k}(y)dy = P\{y < \mathbf{y}_k \le y+dy\}$$

در صورتی این واقعه را خواهیم داشت که از n تا x_i، $k-1$تا کوچکتر از y، $n-k$تا بزرگتر از $y+dy$ و یکی بین y و $y+dy$ باشد.

یادآوری توزیع چندجمله‌ای: احتمال اینکه در n بار آزمایش، واقعهٔ A_1، k_1بار، واقعهٔ A_2، k_2بار و واقعهٔ A_3، k_3بار اتفاق افتد ($k_1+k_2+k_3=n$ و $P(A_i)=p_i$) برابر است با:

$$p = \frac{n!}{k_1!k_2!k_3!}p_1^{k_1}p_2^{k_2}p_3^{k_3}$$

در نتیجه داریم:

$$f_{\mathbf{y}_k}(y)dy = \frac{n!}{(k-1)!(n-k)!1!}\left[F_{\mathbf{x}}(y)\right]^{k-1}\left[1-F_{\mathbf{x}}(y+dy)\right]^{n-k} f_{\mathbf{x}}(y)dy$$

$$dy \to 0 \quad \Rightarrow \quad f_{\mathbf{y}_k}(y) = \frac{n!}{(k-1)!(n-k)!}\left[F_{\mathbf{x}}(y)\right]^{k-1}\left[1-F_{\mathbf{x}}(y)\right]^{n-k} f_{\mathbf{x}}(y)$$

برای $k=1$، توزیع $\mathbf{x}_{\min}$ و برای $k=n$، توزیع $\mathbf{x}_{\max}$ به دست می‌آید:

$$f_{\mathbf{x}_{\min}}(y)=n\left[1-F_{\mathbf{x}}(y)\right]^{n} f_{\mathbf{x}}(y) \quad , \quad 1-F_{\mathbf{x}_{\min}}(y)=\left[1-F_{\mathbf{x}}(y)\right]^{n}$$

$$f_{\mathbf{x}_{\max}}(y)=n\left[F_{\mathbf{x}}(y)\right]^{n-1} f_{\mathbf{x}}(y) \quad , \quad F_{\mathbf{x}_{\max}}(y)=\left[F_{\mathbf{x}}(y)\right]^{n}$$

مجموع متغیرهای تصادفی مستقل:

می‌دانیم که اگر $\mathbf{z}=\mathbf{x}_1+\mathbf{x}_2$ بوده و $\mathbf{x}_1$ و $\mathbf{x}_2$ مستقل باشند، آنگاه داریم:

$$\Phi_{\mathbf{z}}=\Phi_{\mathbf{x}_1}\Phi_{\mathbf{x}_2} \quad , \quad f_{\mathbf{z}}=f_{\mathbf{x}_1}*f_{\mathbf{x}_2}$$

همچنین دیدیم که اگر $\mathbf{z}=\sum_{i=1}^{n}\mathbf{x}_i$ باشد، آنگاه:

$$\Phi_{\mathbf{z}}(j\omega)=\Phi_{\underline{\mathbf{x}}}(j\omega,\ldots,j\omega)$$

و اگر $\mathbf{x}_i$ها مستقل باشند، خواهیم داشت:

$$\Phi_{\mathbf{z}}(j\omega)=\Phi_{\mathbf{x}_1}(j\omega)\Phi_{\mathbf{x}_2}(j\omega)\cdots\Phi_{\mathbf{x}_n}(j\omega)$$
$$f_{\mathbf{z}}=f_{\mathbf{x}_1}*f_{\mathbf{x}_2}*\cdots*f_{\mathbf{x}_n}$$

(با توجه به اینکه عمل کانولوشن دارای خاصیت جابجایی و شرکت‌پذیری است.)

مثال: اگر $\mathbf{x}_i$ها مستقل بوده و هر یک دارای توزیع پواسن با پارامتر λ_i باشند، داریم:

$$\Phi_{\mathbf{z}}(j\omega)=\prod_{i=1}^{n}\Phi_{\mathbf{x}_i}(j\omega)=\prod_{i=1}^{n}e^{\lambda_i(e^{j\omega}-1)}=e^{\lambda(e^{j\omega}-1)}$$

که: $\lambda=\sum_{i=1}^{n}\lambda_i$ (یعنی مجموعشان هم دارای توزیع پواسن است، اما با پارامتر $\sum_{i=1}^{n}\lambda_i$).

اگر $\mathbf{x}_i$ها .i.i.d باشند، یعنی: $\forall x: f_{\mathbf{x}_1}(x) = f_{\mathbf{x}_2}(x) = \cdots = f_{\mathbf{x}_n}(x) = f_{\mathbf{x}}(x)$، در این صورت خواهیم داشت:

$$f_{\mathbf{z}} = f_{\mathbf{x}} * f_{\mathbf{x}} * \cdots * f_{\mathbf{x}}$$

$$\Phi_{\mathbf{z}}(\omega) = \Phi_{\mathbf{x}}^{n}(\omega)$$

$$E(\mathbf{z}) = n\,E(\mathbf{x})$$

$$\sigma_{\mathbf{z}}^2 = n\sigma_{\mathbf{x}}^2$$

مثال: اگر $\mathbf{x}_i$ها مستقل بوده و همگی دارای توزیع نمایی باشند، یعنی: $f_{\mathbf{x}_i}(x) = \lambda e^{-\lambda x}u(x)$ (یعنی .i.i.d باشند) و $\mathbf{z} = \sum_{i=1}^{n} \mathbf{x}_i$ باشد، آنگاه $f_{\mathbf{z}}$ چه خواهد بود؟ مثلاً اگر سیستمی از n المان به صورت Standby استفاده کند و خرابی این المان‌ها از هم مستقل باشد و برای همگی داشته باشیم: $f_{\mathbf{x}}(x) = \lambda e^{-\lambda x}u(x)$، تابع چگالی عمر سیستم چه خواهد بود؟

$$\Phi_{\mathbf{x}}(j\omega) = \frac{\lambda}{\lambda - j\omega} \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega) = \left(\frac{\lambda}{\lambda - j\omega}\right)^n$$

با استفاده از جداول تبدیل فوریه داریم:

$$\frac{x^{n-1}}{(n-1)!}e^{-\lambda x}u(x) \xleftrightarrow{\ \mathfrak{I}\ } \frac{1}{(\lambda + j\omega)^n}$$

لذا:

$$f_{\mathbf{z}}(z)=\lambda^{n}\frac{z^{n-1}}{(n-1)!}e^{-\lambda z}u(z)$$

که همان توزیع ارلانگ است (توزیع زمان لازم برای وقوع n واقعهٔ کاملاً تصادفی یا به عبارتی مجموع n نمایی).

برای توزیع گاما $E(\mathbf{z})=\frac{r}{\lambda}$ است، در نتیجه داریم:

$$E(\mathbf{z})=\frac{n}{\lambda}$$

پس همان طور که انتظار داشتیم: $E(\mathbf{z})=n\,E(\mathbf{x})$؛ یعنی MTTF، n برابر شده است.

اگر این المان‌ها از نظر نقش در خرابی یا صحت سیستم به صورت موازی بودند، خواهیم داشت:

$$\mathbf{z} = \max(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$F_{\mathbf{z}}(z) = F_{\mathbf{x}}^{n}(z)$$

$$f_{\mathbf{z}}(z) = n\, F_{\mathbf{x}}^{n-1}(z) f_{\mathbf{x}}(z)$$

برای $f_{\mathbf{x}}(x) = \lambda e^{-\lambda x} u(x)$ می‌توانید نشان دهید که:

$$E(\mathbf{z}) = \frac{1}{\lambda}(1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n})$$

مثلاً برای $n = 4$، مقدار MTTF، 2.1 برابر می‌شود.

در صورتی که این المان‌ها سری باشند، خواهیم داشت:

$$\mathbf{z} = \min(\mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$F_{\mathbf{z}}(z) = 1 - (1 - F_{\mathbf{x}}(z))^n = 1 - (e^{-\lambda z})^n$$

$$f_{\mathbf{z}}(z) = n\lambda e^{-n\lambda z} u(z)$$

(توجه کنید که مقدار می‌نیمم تعدادی از متغیرهای تصادفی نمایی و i.i.d.، خود نیز نمایی است.)

پس:

$$E(\mathbf{z}) = \frac{1}{n\lambda}$$

یعنی MTTF، $\frac{1}{n}$ برابر شده است.

مثال: اگر متغیر تصادفی $\mathbf{x}$ به صورت زیر تعریف شود:

$$\mathbf{x}(\omega)=\begin{cases}1 & \omega\in A\\ 0 & \text{otherwise}\end{cases} \quad , \quad P(A)=p$$

مثلاً A واقعهٔ آمدن شیر در پرتاب سکه باشد، $\mathbf{x}$ دارای توزیع برنولی (دو جمله‌ای با $n=1$) است. پس:

$$E(\mathbf{x})=p \quad , \quad \sigma_{\mathbf{x}}^2=pq$$

حال با n بار تکرار این آزمایش (نمونه‌برداری از $\mathbf{x}$) و با فرض مستقل بودن آزمایش‌ها، $\mathbf{x}_i$های i.i.d. حاصل می‌شوند.

پس اگر تعریف کنیم:

$$\mathbf{z}=\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_n$$

$\mathbf{z}$ تعداد شیرهای به دست آمده در n بار پرتاب سکه خواهد بود. یعنی:

$$\mathbf{z}\sim \text{Binomial}(n,p)$$

و خواهیم داشت:

$$E(\mathbf{z})=n\,E(\mathbf{x})=np$$

$$\sigma_{\mathbf{z}}^2=n\sigma_{\mathbf{x}}^2=npq$$

(می‌بینیم که این روش خیلی ساده‌تر از روش مستقیم و حتی خیلی ساده‌تر از روش تابع مشخصه است.)

مثال: متغیر تصادفی $\mathbf{x}$ را برابر تعداد شکست‌ها تا حصول اولین موفقیت تعریف می‌کنیم. می‌دانیم که $\mathbf{x}$ دارای توزیع هندسی است (توزیع دو جمله‌ای منفی با $r=1$). در تمرین سری ۵ نشان دادید (از راه تابع مشخصه) که:

$$E(\mathbf{x})=\frac{q}{p} \quad , \quad \sigma_{\mathbf{x}}^2=\frac{q}{p^2} \quad , \quad \Phi_{\mathbf{x}}(j\omega)=\frac{p}{1-q e^{j\omega}}$$

حال اگر تعریف کنیم:

$$\mathbf{z}=\mathbf{x}_1+\mathbf{x}_2+\cdots+\mathbf{x}_r$$

که همهٔ $\mathbf{x}_i$ها دارای توزیع هندسی بوده و مستقلند، آنگاه $\mathbf{z}$ تعداد شکست‌ها تا حصول r امین موفقیت خواهد بود. یعنی:

$$\mathbf{z}\sim \text{Negative Binomial}(r,p)$$

پس:

$$E(\mathbf{z})=\frac{rq}{p} \quad , \quad \sigma_{\mathbf{z}}^2=\frac{rq}{p^2}$$

در حالی که محاسبهٔ این دو از راه مستقیم یا تابع مشخصه خیلی مشکلتر است.

همچنین در مورد $\Phi_{\mathbf{z}}(j\omega)$ داریم:

$$\Phi_{\mathbf{z}}(j\omega)=\left(\frac{p}{1-q e^{j\omega}}\right)^r$$

در حالی که به دست آوردن مستقیم تابع مشخصه مشکلتر است.

قضیهٔ حد مرکزی (Central Limit Theorem):

طبق قضیهٔ حد مرکزی، اگر $\mathbf{x}_i$ها مستقل بوده و $\mathbf{y}=\sum_{i=1}^{n}\mathbf{x}_i$ باشد، وقتی $n\to+\infty$، توزیع $\mathbf{y}$ به توزیع نرمال میل می‌کند (با شرایط سهلی)، ولو $\mathbf{x}_i$ها نرمال نباشند. این در واقع بیانگر خاصیتی از کانولوشن است که کانولوشن تعداد زیادی توابع مثبت به گوسی میل می‌کند. با توجه به CLT معلوم می‌شود که چرا بسیاری از پدیده‌ها در جهان خارج توزیع تقریباً نرمال دارند. اصولاً هرگاه پدیده‌ای تحت تأثیر عوامل متعدد تصادفی باشد (قد یک فرد، خطا در اندازه‌گیری، ولتاژ نویز حرارتی و ...) تقریباً نرمال خواهد بود.

قضیهٔ حد مرکزی لیاپانوف:

اگر $\mathbf{y}=\sum_{i=1}^{n}\mathbf{x}_i$ بوده و $\mathbf{x}_i$ها مستقل باشند و $E(|\mathbf{x}_i-\eta_i|^3)<+\infty$ (یا $E(|\mathbf{x}_i|^3)<+\infty$) و $\sum_{i=1}^{n}\sigma_i^2\xrightarrow[n\to+\infty]{}+\infty$، آنگاه برای $\mathbf{z}=\dfrac{\mathbf{y}-\eta_y}{\sigma_y}$ داریم:

$$F_{\mathbf{z}}(z)\xrightarrow[n\to+\infty]{}G(z)$$

در قضیهٔ لیاپانوف لزوماً توزیع $\mathbf{x}_i$ها یکسان نیست (لزوماً i.i.d. نیستند). صورت دیگری از قضیهٔ حد مرکزی داریم که صرفاً برای $\mathbf{x}_i$های i.i.d. است.

قضیهٔ حد مرکزی لیندبرگ- لوی:

اگر $\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i$ بوده و $\mathbf{x}_i$ها .i.i.d با واریانس محدود $\sigma^2 < +\infty$ باشند، آنگاه برای $\mathbf{z} = \frac{\mathbf{y} - \eta_y}{\sigma_y}$ داریم:

$$F_{\mathbf{z}}(z) \xrightarrow[n \to +\infty]{} G(z)$$

قضیهٔ لیندبرگ- لوی حالت خاصی از قضیهٔ لیاپانوف نمی‌باشد. چون در اینجا دیگر شرط $E\left[(\mathbf{x}_i - \eta_i)^3\right] < +\infty$ لازم نیست و شرط ضعیفتر $\sigma^2 < +\infty$ جایگزین آن شده است (شرط دیگر قضیهٔ لیاپانوف، یعنی $\sum_{i=1}^{n} \sigma_i^2 \xrightarrow[n \to +\infty]{} +\infty$، مسلماً ارضاء می‌شود).

همچنین قضیهٔ لیندبرگ- فلر را داریم که برای حالت کلی ($\mathbf{x}_i$های نه لزوماً .i.i.d) شرط لازم و کافی گوسی بودن توزیع حدی را ارائه می‌کند که تعبیر با تسامحش این است که $\forall j: \lim\limits_{n\to+\infty} \dfrac{\sigma_j^2}{\sum\limits_{i=1}^{n}\sigma_i^2} = 0$. یعنی هیچ جمله‌ای غالب نباشد و در مورد همهٔ جمله‌ها، واریانس آن در مقابل واریانس کل ناچیز باشد. توجه کنید که برقراری شرایط قضیهٔ لیاپانوف، برقراری شرایط قضیهٔ لیندبرگ- فلر را نتیجه می‌دهد.

وقتی n بی‌نهایت نبوده ولی بزرگ باشد هم می‌توانیم نرمال را به عنوان تقریب به کار ببریم. اگر توزیع $\mathbf{x}$ هموار باشد، حتی برای $n=5$، بسیار به توزیع نرمال نزدیک خواهیم بود. در اغلب کاربردها $n=30$ کاملاً کافی است (در کتاب، $n=3$ را برای توزیع یکنواخت به کار برده و با نرمال مقایسه کرده است).

مثال: جعبه ای شامل ۱۰۰ مقاومت $1\text{k}\Omega \pm 5\%$ در اختیار داریم. اگر توزیع مقدار مقاومت‌ها را یکنواخت فرض کنیم، احتمال اینکه مجموع مقدار این مقاومت‌ها در محدودهٔ $100\text{k}\Omega \pm 0.5\%$ باشد، چیست؟

$$\mathbf{y} = \sum_{i=1}^{100} \mathbf{x}_i$$

$$\sigma_{\mathbf{x}}^2 = \frac{(b-a)^2}{12} = \frac{(100)^2}{12} < +\infty$$

پس شرط قضیهٔ لیندبرگ- لوی برقرار است (اگر توزیع را نمی‌دانستیم و فقط همین واریانس را می‌دانستیم هم کافی بود).

$$E(\mathbf{x}) = 1000 \quad \Rightarrow \quad E(\mathbf{y}) = 100000$$

$$\sigma_{\mathbf{y}}^2 = n\sigma_{\mathbf{x}}^2 = 100\frac{(100)^2}{12} = \frac{10^6}{12}$$

$$P\{99500 < \mathbf{y} < 100500\} = G(\frac{100500-100000}{\sqrt{\frac{10^6}{12}}}) - G(\frac{99500-100000}{\sqrt{\frac{10^6}{12}}}) = 2G(1.732) - 1 = 0.917$$

در بیان قضیه‌های فوق از تابع توزیع انباشته استفاده شد تا برای حالت گسسته هم صادق باشد. چون در حالت گسسته، تابع چگالی یک سری δ است و نرمال که پیوسته است، نمی‌تواند آن را تقریب بزند (اگر چه نرمال، پوش مقدار وزنهٔ این δها را تقریب می‌زند).

به عنوان مثال قضیهٔ دموار- لاپلاس حالت خاصی از CLT است.

مثال: $\mathbf{x}_i$ها متغیرهای تصادفی i.i.d. و دارای توزیع برنولی هستند، یعنی:

$$P\{\mathbf{x}_i = 1\} = p \quad , \quad P\{\mathbf{x}_i = 0\} = q = 1 - p$$

حال اگر تعریف کنیم:

$$\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i$$

$\mathbf{y}$ دارای توزیع دوجمله‌ای است، یعنی:

$$P\{\mathbf{y} = k\} = \binom{n}{k} p^k q^{n-k}$$

چون $\mathbf{x}_i$ها i.i.d. هستند و $\sigma_x^2 = pq < +\infty$، پس طبق قضیهٔ لیندبرگ- لوی تابع توزیع $\mathbf{y}$ توسط نرمال قابل تقریب است (برای $n \to \infty$ به نرمال میل می‌کند). یعنی قضیهٔ دموار- لاپلاس را می‌توان حالت خاصی از CLT دانست.

اثبات قضیهٔ حد مرکزی در حالت .i.i.d بودن $\mathbf{x}_i$ها و محدود بودن تمامی گشتاورها:

$$\mathbf{y} = \sum_{i=1}^{n} \mathbf{x}_i \quad \Rightarrow \quad \eta_{\mathbf{y}} = n\eta \quad , \quad \sigma_{\mathbf{y}}^2 = n\sigma^2$$

$$\mathbf{z} = \frac{\mathbf{y} - \eta_{\mathbf{y}}}{\sigma_{\mathbf{y}}} = \frac{\sum_{i=1}^{n} (\mathbf{x}_i - \eta)}{\sqrt{n}\,\sigma} = \frac{1}{\sqrt{n}} \overbrace{\sum_{i=1}^{n} (\underbrace{\frac{\mathbf{x}_i - \eta}{\sigma}}_{\mathbf{u}_i})}^{\mathbf{v}}$$

$$\Phi_{\mathbf{z}}(j\omega) = E(e^{j\omega \mathbf{z}}) = E(e^{\frac{j\omega}{\sqrt{n}}\mathbf{v}}) = \Phi_{\mathbf{v}}(\frac{j\omega}{\sqrt{n}})$$

اگر فرض کنیم $\mathbf{u}_i \triangleq \frac{\mathbf{x}_i - \eta}{\sigma}$، چون $\mathbf{v} = \sum_{i=1}^{n} \mathbf{u}_i$ و $\mathbf{u}_i$ها .i.i.d هستند، پس:

$$\Phi_{\mathbf{v}}(j\omega) = \Phi_{\mathbf{u}}^{n}(j\omega) \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega) = \Phi_{\mathbf{u}}^{n}(\frac{j\omega}{\sqrt{n}})$$

اما اگر تابع مشخصه را بسط تیلور دهیم، خواهیم داشت:

$$\Phi(j\omega)=1+j\omega\Phi'(0)+\frac{(j\omega)^2}{2}\Phi''(0)+\frac{(j\omega)^3}{3!}\Phi'''(0)+\cdots$$

که: $\Phi^{(n)}(0)=m_n$.

پس در مورد $\mathbf{u}$ داریم:

$$\Phi'(0)=\eta_{\mathbf{u}}=0 \quad , \quad \Phi''(0)=\sigma_{\mathbf{u}}^2=1 \quad , \quad \Phi'''(0)=m_3(\mathbf{u})=E\left[\left(\frac{\mathbf{x}_i-\eta}{\sigma}\right)^3\right]$$

در نتیجه:

$$\Phi_{\mathbf{u}}(j\omega)=1+\frac{(j\omega)^2}{2}+\frac{(j\omega)^3}{3!}\Phi'''(0)+\cdots \quad \Rightarrow \quad \Phi_{\mathbf{z}}(j\omega)=\left[1+\frac{(j\omega)^2}{2n}+\frac{(j\omega)^3}{3!n^{\frac{3}{2}}}\Phi'''(0)+\cdots\right]^n$$

می‌دانیم که اگر $\lim\limits_{n\to+\infty} a_n = b$ ، آنگاه:

$$\lim_{n\to+\infty}\left(1+\frac{a_n}{n}\right)^n = e^b$$

(یا با توجه به $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \mp \cdots$)

پس:

$$\Phi_{\mathbf{z}}(j\omega) = \left[1 + \frac{(j\omega)^2}{2n} + \frac{(j\omega)^3}{3!n^{\frac{3}{2}}}\Phi'''(0) + \cdots\right]^n$$

$$\to a_n = 1 + \frac{(j\omega)^2}{2} + \frac{(j\omega)^3}{3!n^{\frac{1}{2}}}\Phi'''(0) + \cdots \quad \Rightarrow \quad \lim_{n\to+\infty}\Phi_{\mathbf{z}}(j\omega) = e^{-\frac{\omega^2}{2}}$$

که تابع مشخصهٔ نرمال استاندارد است (قضیه‌ای داریم که اگر $\Phi_n \to \Phi^*$، آنگاه $F_n \to F^*$).

همگرایی دنبالهٔ متغیرهای تصادفی:

می‌دانیم که: $x_1, x_2, \ldots, x_n, \ldots$ یک دنباله از اعداد است و همگرایی این دنباله به عدد a، یعنی برای هر $\varepsilon > 0$، n_0ای وجود داشته باشد که:

$$|x_n - a| < \varepsilon : n > n_0$$

در این صورت می‌نویسیم:

$$\lim_{n \to +\infty} x_n = a$$

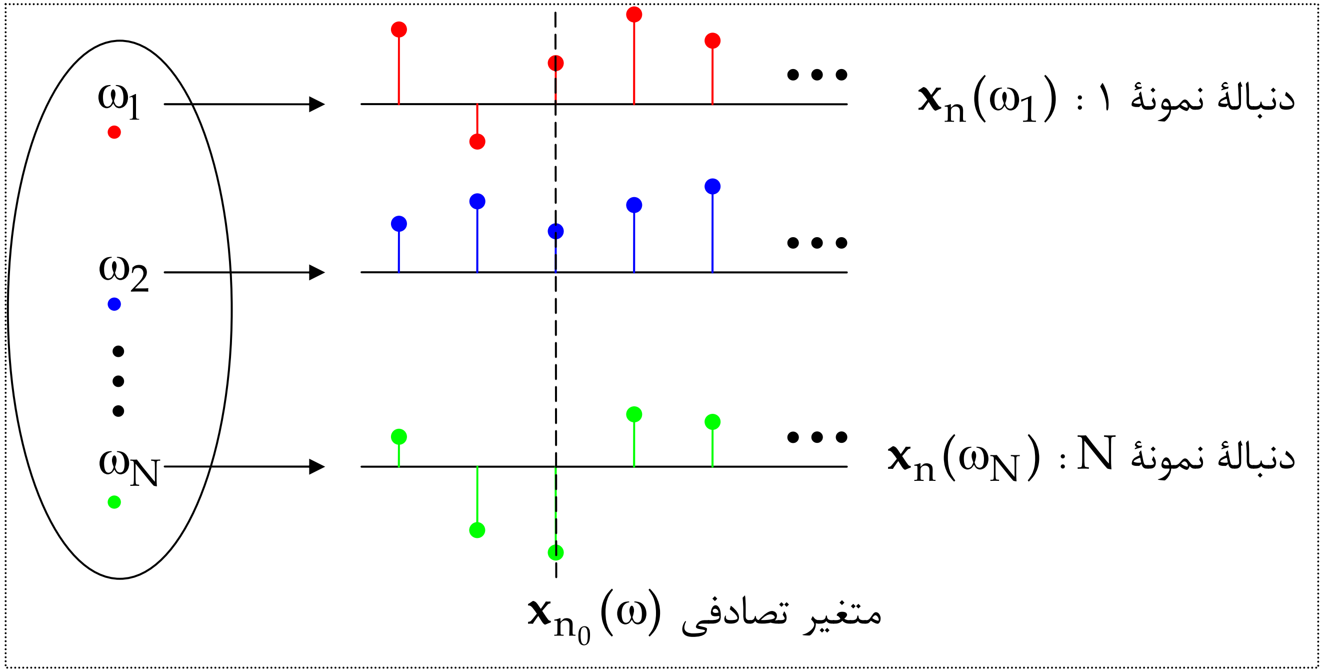
حال در مورد متغیرهای تصادفی مفهوم همگرایی به چه صورت خواهد بود؟

تعریف فرآیند تصادفی (Random Process or Stochastic Process):

دنبالهٔ نامحدود $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \ldots$ از متغیرهای تصادفی را فرآیند تصادفی گویند.

توجه دارید که هر $\mathbf{x}_i$ تابعی از نقاط فضای نمونه است، یعنی برای هر نقطهٔ ω_i یک دنبالهٔ عددی به صورت زیر داریم:

$$\mathbf{x}_1(\omega_i), \mathbf{x}_2(\omega_i), \ldots, \mathbf{x}_n(\omega_i), \ldots$$



توجه داشته باشید که:

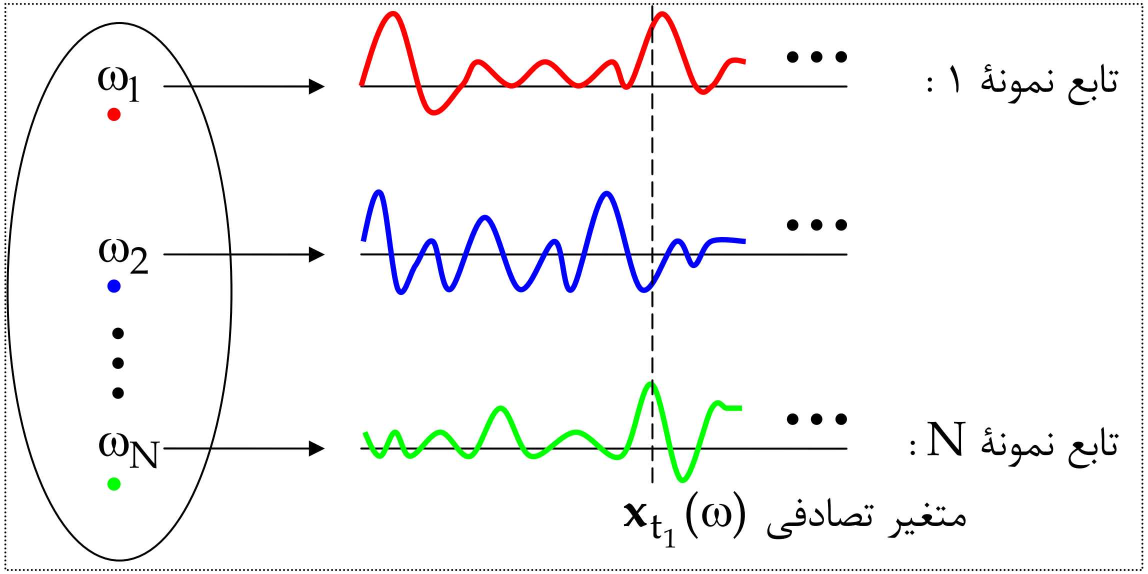
فرآیند تصادفی : $\mathbf{x}_n(\omega)$

دنبالهٔ عددی : $\mathbf{x}_n(\omega_i)$

متغیر تصادفی : $\mathbf{x}_{n_0}(\omega)$

یک عدد : $\mathbf{x}_{n_0}(\omega_i)$

آنچه گفتیم فرآیند تصادفی گسسته بود (سیگنال گسسته در زمان). به طور مشابه می‌توانیم فرآیند تصادفی پیوستهٔ $\mathbf{x}_t(\omega)$ را تعریف کنیم.



مثال: شیر یا خط می‌اندزیم. اگر شیر آمد، سیگنال سینوسی و اگر خط آمد سیگنال مربعی می‌فرستیم، یعنی به احتمال $\frac{1}{2}$ سینوسی و به احتمال $\frac{1}{2}$ مربعی.

مثال: برای نویز روی مقاومت، مقدار ولتاژ نویز در یک لحظه، متغیر تصادفی است. ولی شکل موج روی آن در یک مدت زمانی، فرآیند تصادفی است.

در مورد همگرایی دنبالۀ متغیرهای تصادفی ممکن است $\mathbf{x}_n(\omega)$ برای بعضی ωها همگرا باشد و برای بعضی دیگر نباشد. یک تعریف برای همگرایی دنبالۀ متغیر تصادفی می‌تواند این باشد که کلیۀ دنباله‌های نمونه، یعنی $\mathbf{x}_n(\omega)$ برای هر ω همگرا باشد (everywhere). ولی در بسیاری موارد ممکن است این طور نباشد و همگرایی ضعیف‌تری داشته باشیم.

۱. همگرایی تقریباً همه‌جا (Almost Everywhere):

به این نوع همگرایی، همگرایی .a.e، همگرایی به احتمال یک و نیز همگرایی Almost Sure گفته می‌شود.

$\mathbf{x}_n(\omega)$ ممکن است برای بعضی ωها حد داشته و برای بعضی دیگر حد نداشته باشد. اگر مجموعۀ ωهایی‌که برای آنها $\mathbf{x}_n(\omega)$ حد ندارد احتمالش صفر باشد، یعنی:

$$P\{\omega : \mathbf{x}_n(\omega) \xrightarrow[n\to+\infty]{} \mathbf{x}\} = 1$$

گوییم $\mathbf{x}_n$ دارای همگرایی .a.e است. در حالت کلی $\mathbf{x}$ می‌تواند تابع ω باشد (یعنی عدد ثابتی نباشد، بلکه متغیر تصادفی باشد).

در این صورت می‌نویسیم:

$$\mathbf{x}_n \xrightarrow[n\to+\infty]{a.e.} \mathbf{x}$$

۲. همگرایی در احتمال:

یعنی برای هر $\varepsilon > 0$ داشته باشیم:

$$P\{|\mathbf{x}_n - \mathbf{x}| \geq \varepsilon\} \xrightarrow[n\to+\infty]{} 0$$

اگر $\mathbf{x} = c$، یعنی عددی ثابت باشد، این به آن معناست که توزیع $\mathbf{x}_n$ با افزایش n حول c متمرکزتر می‌شود. در این صورت می‌نویسیم:

$$\mathbf{x}_n \xrightarrow[n\to+\infty]{P} \mathbf{x}$$

یا:

$$\underset{n\to+\infty}{P\lim} \mathbf{x}_n = \mathbf{x}$$

روشن است که همگرایی به احتمال یک، همگرایی در احتمال را نتیجه می‌دهد.

۳. همگرایی به مفهوم m.s. (Mean Square) (همگرایی در گشتاور دوم):

برای هر n، $\mathbf{x}_n-\mathbf{x}$ یک متغیر تصادفی است و گشتاور دوم آن $E\left[(\mathbf{x}_n-\mathbf{x})^2\right]$ می‌باشد. پس $E\left[(\mathbf{x}_n-\mathbf{x})^2\right]$ یک دنبالهٔ عددی است.

می‌گوییم $\mathbf{x}_n$ به $\mathbf{x}$ به مفهوم m.s. میل می‌کند و می‌نویسیم: $\mathbf{x}_n \xrightarrow[n\to+\infty]{\text{m.s.}} \mathbf{x}$، هرگاه:

$$E\left[(\mathbf{x}_n-\mathbf{x})^2\right] \xrightarrow[n\to+\infty]{} 0$$

همگرایی به مفهوم m.s.، همگرایی در احتمال را نتیجه می‌دهد.

زیرا با توجه به نامساوی Bienayme (به ازای $a=2$) داریم:

$$P\{|\mathbf{x}_n-\mathbf{x}|>\varepsilon\} \le \frac{E\left[(\mathbf{x}_n-\mathbf{x})^2\right]}{\varepsilon^2}$$

پس وقتی $E\left[(\mathbf{x}_n-\mathbf{x})^2\right]$ به سمت صفر میل می‌کند، این احتمال هم به صفر میل می‌کند.

پس همگرایی m.s. و همگرایی به احتمال یک (a.e.) قویتر از همگرایی در احتمالند.

۴. همگرایی در تابع توزیع:

$\mathbf{x}_n$ را همگرا به $\mathbf{x}$ در تابع توزیع گویند، هرگاه برای هر x (هر xای که $F_{\mathbf{x}}(x)$ پیوسته باشد) داشته باشیم:

$$F_{\mathbf{x}_n}(x) \xrightarrow[n\to+\infty]{} F_{\mathbf{x}}(x)$$

این ضعیف‌ترین نوع همگرایی است. CLT نمونه‌ای از همگرایی در تابع توزیع بود.

قانون اعداد بزرگ:

در فصل ۳ دیدیم که اگر در یک آزمایش تصادفی $P(A)=p$ باشد و در n بار تکرار آزمایش، k بار واقعهٔ A اتفاق افتد، احتمال اینکه $\frac{k}{n}$ بین $p-\varepsilon$ و $p+\varepsilon$ باشد، وقتی $n\to+\infty$ به سمت یک می‌رود، یعنی:

برای هر $\varepsilon>0$ داده شده : $P\{|\frac{k}{n}-p|\leq\varepsilon\}\to 1$

حال در حالت کلی‌تر داریم:

قانون ضعیف اعداد بزرگ (Weak Law of Large Numbers):

اگر $\mathbf{x}_i$ها نمونه‌های تصادفی از جامعه‌ای با میانگین η و .i.i.d باشند و $\overline{\mathbf{x}}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$، میانگین نمونه باشد، داریم:

$$\underset{n\to+\infty}{\mathrm{P}\lim}\,\overline{\mathbf{x}}_n = \eta \qquad \text{یا} \qquad \overline{\mathbf{x}}_n \xrightarrow[n\to+\infty]{P} \eta$$

یعنی:

$$P\{|\overline{\mathbf{x}}_n - \eta| > \varepsilon\} \xrightarrow[n\to+\infty]{} 0$$

اثبات: می‌دانیم که:

$$E(\overline{\mathbf{x}}) = \eta \quad , \quad \sigma_{\overline{\mathbf{x}}}^2 = \frac{\sigma^2}{n}$$

لذا با توجه به قضیهٔ چبیشف خواهیم داشت:

$$P\{|\overline{\mathbf{x}} - \eta| > \varepsilon\} < \frac{\sigma^2}{n\varepsilon^2} \xrightarrow[n\to+\infty]{} 0$$

یا می‌توانیم بگوییم که:

$$E\left[(\overline{\mathbf{x}}_n-\eta)^2\right]=\sigma_{\overline{\mathbf{x}}_n}^2=\frac{\sigma^2}{n}\xrightarrow[n\to+\infty]{}0$$

پس:

$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{m.s.}}\eta$$

در نتیجه:

$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{P}}\eta$$

قانون قوی اعداد بزرگ (Strong Law of Large Numbers):

اگر $\mathbf{x}_i$ها متغیرهای تصادفی .i.i.d باشند و $\overline{\mathbf{x}}_n=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i$ باشد، $\overline{\mathbf{x}}_n$ با احتمال یک به η میل می‌کند.

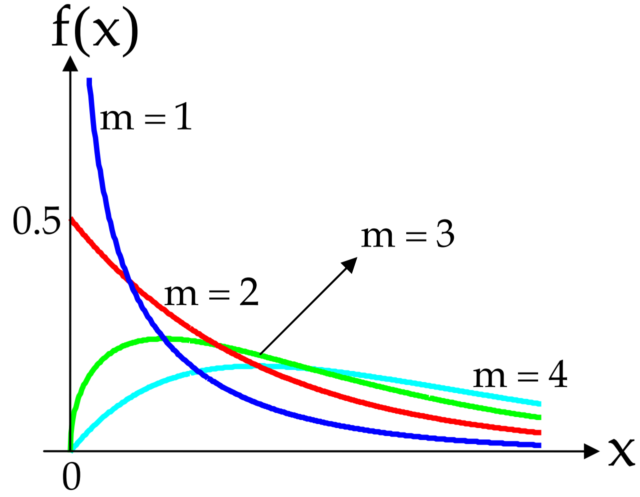
$$\overline{\mathbf{x}}_n\xrightarrow[n\to+\infty]{\text{a.e.}}\eta$$

توابع توزیع متداول در آمار:

توزیع χ^2: این توزیع کاربرد زیادی در آمار دارد. بنا بر تعریف، توزیع χ^2 با m درجه آزادی برابر است با:

$$\mathbf{x} \sim \chi^2(m) = \text{Gamma}(r = \frac{m}{2}, \lambda = \frac{1}{2})$$

$$f_{\mathbf{x}}(x) = \frac{1}{2^{\frac{m}{2}} \Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}} u(x)$$



به ازای $m = 2$، این توزیع همان توزیع نمایی می‌شود.

داریم:

$$\Gamma(\frac{m}{2}) = \begin{cases} (k-1)! & m = 2k \\ (k-\frac{1}{2})(k-\frac{3}{2})\cdots\frac{1}{2}\sqrt{\pi} & m = 2k+1 \end{cases} \quad (\text{چون } \Gamma(\frac{1}{2}) = \sqrt{\pi})$$

قبلاً گشتاور مرتبهٔ nام توزیع گاما را به دست آورده بودیم:

$$m_n = \frac{\Gamma(n+r)}{\Gamma(r)\lambda^n}$$

پس برای توزیع χ^2 داریم:

$$E(\mathbf{x}^n) = \frac{\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\,(\frac{1}{2})^n}$$

پس خواهیم داشت:

$$E(\mathbf{x}) = m \quad , \quad E(\mathbf{x}^2) = m(m+2) \quad \Rightarrow \quad var(\mathbf{x}) = 2m$$

رابطهٔ $m_n = \frac{\Gamma(n+\frac{m}{2})}{\Gamma(\frac{m}{2})\,(\frac{1}{2})^n}$ حتی برای n منفی صادق است (مشروط بر اینکه $0 < n+\frac{m}{2}$ باشد). مثلاً داریم:

$$E(\frac{1}{\mathbf{x}}) = \frac{\Gamma(\frac{m}{2}-1)}{2\Gamma(\frac{m}{2})} = \frac{1}{2(\frac{m}{2}-1)} = \frac{1}{m-2} : m > 2$$

تابع مشخصهٔ توزیع گاما را نیز قبلاً به دست آورده بودیم (در تمرین سری ۵):

$$\Phi(j\omega)=\left(\frac{\lambda}{\lambda-j\omega}\right)^{r}$$

پس برای توزیع χ^2 داریم:

$$\Phi_{\mathbf{x}}(j\omega)=\left(\frac{\frac{1}{2}}{\frac{1}{2}-j\omega}\right)^{\frac{m}{2}}=\frac{1}{(1-2j\omega)^{\frac{m}{2}}}$$

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل باشند و $\mathbf{x}_i \sim \chi^2(k_i)$ بوده و $\mathbf{z}=\sum_{i=1}^{n}\mathbf{x}_i$ باشد، در این صورت داریم:

$$\mathbf{z}\sim\chi^2\left(\sum_{i=1}^{n}k_i\right)$$

در تمرین سری هشتم این قضیه را برای توزیع گاما دیدیم. پس برای توزیع χ^2 نیز که حالت خاصی از توزیع گاما است نیز صادق می‌باشد $(\lambda=\frac{1}{2})$.

قضیهٔ خاصیت اساسی توزیع χ^2: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل و نرمال استاندارد باشند و $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2$ باشد، در این صورت داریم:

$$\mathbf{Q}\sim\chi^2(n)$$

اثبات:

$$\begin{cases}\mathbf{y}_i=\mathbf{x}_i^2\\ \mathbf{x}_i\sim N(0,1)\end{cases}\quad\Rightarrow\quad f_{\mathbf{y}_i}(y)=\frac{1}{\sqrt{y}}e^{-\frac{y}{2}}:y>0\quad\Rightarrow\quad \mathbf{y}_i\sim\chi^2(1)$$

پس طبق قضیهٔ قبل چون: $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2=\sum_{i=1}^{n}\mathbf{y}_i$، داریم:

$$\mathbf{Q}\sim\chi^2(n)$$

(به طور مستقیم هم می‌توان نشان داد، ولی قدری مشکل است. اثبات در کتاب فرآیند Papoulis، صفحهٔ ۲۵۱.)

قضیه: اگر متغیرهای تصادفی $\mathbf{x}_1$، $\mathbf{x}_2$، ... و $\mathbf{x}_n$ مستقل و نرمال استاندارد باشند، فرم درجهٔ دوم $\mathbf{Q}=\underline{\mathbf{x}}^T A\underline{\mathbf{x}}=\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}\mathbf{x}_i\mathbf{x}_j$ (که در آن A یک ماتریس متقارن $n\times n$ است) دارای توزیع $\chi^2(r)$ خواهد بود، اگر و تنها اگر r مقدار ویژهٔ ماتریس A برابر یک و سایر $n-r$ مقدار ویژهٔ آن برابر صفر باشند.

اثبات: از جبر خطی می‌دانیم که ماتریس A را می‌توان به صورت زیر نوشت:

$$A=U\Lambda U^{-1}=U\Lambda U^T$$

که در آن:

$$U=\left[\underline{u}_1\quad \underline{u}_2\quad \cdots\quad \underline{u}_n\right]\quad,\quad \Lambda=\begin{bmatrix}\lambda_1 & 0 & \cdots & 0\\ 0 & \lambda_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \lambda_n\end{bmatrix}$$

زیرا از تعریف بردار ویژه و مقدار ویژه داریم:

$$A\underline{u}_i=\lambda_i\underline{u}_i$$

$$\Rightarrow \begin{bmatrix} A\underline{u}_1 & A\underline{u}_2 & \cdots & A\underline{u}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\underline{u}_1 & \lambda_2\underline{u}_2 & \cdots & \lambda_n\underline{u}_n \end{bmatrix}$$

$$\Rightarrow A\begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_n \end{bmatrix} = \begin{bmatrix} \underline{u}_1 & \underline{u}_2 & \cdots & \underline{u}_n \end{bmatrix}\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix}$$

$$\Rightarrow AU = U\Lambda \quad \Rightarrow \quad A = U\Lambda U^{-1}$$

البته اگر A متقارن باشد، U ماتریس Unitary خواهد بود، یعنی: $U^{-1} = U^T$. زیرا:

$$\underline{u}_i \cdot \underline{u}_j^T = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

به عبارت دیگر:

$$\Lambda = U^T A U$$

اگر تعریف کنیم:

$$\underline{\mathbf{z}} = U^T \underline{\mathbf{x}} \quad \Rightarrow \quad \underline{\mathbf{x}} = U\underline{\mathbf{z}}$$

$$E(\underline{\mathbf{z}}) = U^T E(\underline{\mathbf{x}}) = 0$$

$$C_{\underline{\mathbf{z}}} = U^T C_{\underline{\mathbf{x}}} U = U^T I U = U^T U = I$$

یعنی $\mathbf{z}_i$ها نرمال استاندارد و مستقل هستند.

$$\mathbf{Q} = \underline{\mathbf{x}}^T A \underline{\mathbf{x}} = \underline{\mathbf{z}}^T U^T A U \underline{\mathbf{z}} = \underline{\mathbf{z}}^T \Lambda \underline{\mathbf{z}} = \sum_{i=1}^{n} \lambda_i \mathbf{z}_i^2$$

پس $\mathbf{Q}$ توزیع $\chi^2(r)$ خواهد داشت، اگر و تنها اگر λ_iها صرفاً صفر یا یک باشند (r تعداد مقادیر ویژهٔ غیرصفر $(\lambda = 1)$ است).

قضیه: اگر برای سه فرم درجه دوم $\mathbf{Q}_1$، $\mathbf{Q}_2$ و $\mathbf{Q}_3$ داشته باشیم: $\mathbf{Q}_3 = \mathbf{Q}_2 - \mathbf{Q}_1$ که: $\mathbf{Q}_1 \sim \chi^2(r)$، $\mathbf{Q}_2 \sim \chi^2(n)$ و $\mathbf{Q}_3 \geq 0$ باشند، آنگاه $\mathbf{Q}_1$ و $\mathbf{Q}_3$ مستقلند و $\mathbf{Q}_3 \sim \chi^2(n-r)$ است.

از کاربردهای مهم توزیع χ^2 در بررسی واریانس نمونه از توزیع نرمال است.

واریانس نمونه:

قبلاً دیدیم که میانگین نمونه برابر است با:

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \rightarrow \quad E(\overline{\mathbf{x}}) = \eta \quad , \quad \sigma_{\overline{\mathbf{x}}}^2 = \frac{\sigma^2}{n}$$

طبق تعریف، واریانس نمونه برابر است با:

$$\mathscr{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2$$

در تمرین (سری هشتم) نشان دادید که:

$$E\left[(\mathbf{x}_i - \overline{\mathbf{x}})^2\right] = \frac{n-1}{n}\sigma^2$$

پس:

$$E(\mathscr{s}^2) = \frac{1}{n}\cdot n(\frac{n-1}{n}\sigma^2) = \frac{n-1}{n}\sigma^2$$

یعنی $\mathscr{s}^2$ تخمینی نااریب از σ^2 نیست. البته اگر η را می‌دانستیم، تخمینی نااریب از σ^2 می‌شد، زیرا:

$$\mathscr{s}^2 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \eta)^2 \quad \rightarrow \quad E(\mathscr{s}^2) = \frac{1}{n}\cdot n\sigma^2 = \sigma^2$$

ولی وقتی η و σ^2 هیچکدام معلوم نیستند، معمولاً از تخمین زیر استفاده می‌شود که نااریب است (اگر چه بعداً خواهیم دید که واریانس این تخمین قدری بیشتر است):

$$\mathbf{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2$$: واریانس نمونه

(بعضاً این تعریف را با $\mathbf{s}_{n-1}^2$ و تعریف اولی را با $\mathbf{s}_n^2$ نشان می‌دهند. برای n بزرگ، این دو تعریف یکسان می‌شوند. برای n کوچک معمولاً از تعریف دومی استفاده می‌شود.)

با این تعریف خواهیم داشت:

$$E(\mathbf{s}^2) = \frac{1}{n-1}\cdot n \cdot \frac{n-1}{n}\sigma^2 = \sigma^2$$

قضیه: میانگین نمونه و واریانس نمونهٔ یک متغیر تصادفی نرمال، مستقل از هم هستند و داریم:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2}=\frac{n\mathscr{s}^2}{\sigma^2}\sim\chi^2(n-1)$$

اثبات:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i-\overline{\mathbf{x}}}{\sigma}\right)^2=\underbrace{\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i-\eta}{\sigma}\right)^2}_{\mathbf{Q}_2}-\underbrace{\left(\frac{\overline{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}}\right)^2}_{\mathbf{Q}_1}\quad\rightarrow\quad\mathbf{Q}_2\sim\chi^2(n)\ ,\ \mathbf{Q}_1\sim\chi^2(1)$$

پس با توجه به قضیهٔ قبلی، $\mathbf{Q}_3=\mathbf{Q}_2-\mathbf{Q}_1$ دارای توزیع $\chi^2(n-1)$ است، یعنی:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2}\sim\chi^2(n-1)$$

و نیز $\mathbf{Q}_1$ و $\mathbf{Q}_3$ مستقلند. در نتیجه میانگین نمونه ($\overline{\mathbf{x}}$) و واریانس نمونه ($\mathscr{s}^2$ یا $\mathbf{s}^2$) مستقلند.

حال که فهمیدیم $\frac{(n-1)\mathbf{s}^2}{\sigma^2} = \frac{n\mathscr{s}^2}{\sigma^2} \sim \chi^2(n-1)$ است و می‌دانیم که واریانس متغیر تصادفی $\chi^2(m)$، $2m$ است، داریم:

$$\operatorname{var}(\mathbf{s}^2) = \left(\frac{\sigma^2}{n-1}\right)^2 \cdot 2(n-1) = \frac{2\sigma^4}{n-1}$$

$$\operatorname{var}(\mathscr{s}^2) = \left(\frac{\sigma^2}{n}\right)^2 \cdot 2(n-1) = \frac{2(n-1)}{n^2}\sigma^4$$

(یعنی $\mathbf{s}^2$ اگر چه ناریب است، ولی واریانس بیشتری دارد. همچنین ملاحظه می‌کنید که وقتی $n \to +\infty$، واریانس به سمت صفر می‌رود.)

توزیع χ^2 غیر مرکزی (Noncentral χ^2 Distribution):

اگر $\mathbf{x}_i$ها به جای اینکه $N(0,1)$ باشند، $N(\eta_i,1)$ بوده و مستقل نیز باشند، $\mathbf{Q}=\sum_{i=1}^{n}\mathbf{x}_i^2$ دیگر توزیع χ^2 ندارد، بلکه توزیع آن χ^2 غیرمرکزی است:

$\mathbf{Q}\sim\chi^2(n,e) \quad \rightarrow \quad$ n : درجهٔ آزادی / e : خروج از مرکز (Eccentricity)

$$e=\sum_{i=1}^{n}\eta_i^2 \quad , \quad E(\mathbf{Q})=n+e$$

همچنین اگر $\mathbf{x}_i\sim N(\eta_i,1)$، $\mathbf{Q}=\underline{\mathbf{x}}^T A\underline{\mathbf{x}}$ و ماتریس A دارای r مقدار ویژهٔ یک و $n-r$ مقدار ویژهٔ صفر باشد، آنگاه $\mathbf{Q}\sim\chi^2(r,e)$ خواهد بود که:

$$e=\eta_{\underline{\mathbf{x}}}^T A\,\eta_{\underline{\mathbf{x}}}$$

اثبات در کتاب، صفحات ۲۲۷ و ۲۲۸.

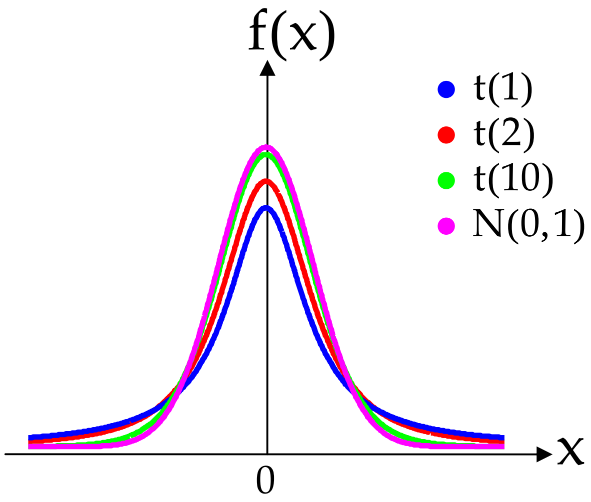
توزیع t (Student t) :

Gosset در سال ۱۹۰۸ تحت نام مستعار Student این توزیع را معرفی کرد:

$\mathbf{x} \sim t(m) \quad \rightarrow \quad m$: درجهٔ آزادی

$$f_{\mathbf{x}}(x) = \gamma \frac{1}{(1+\frac{x^2}{m})^{\frac{m+1}{2}}} \quad \rightarrow \quad \gamma = \frac{\Gamma(\frac{m+1}{2})}{\sqrt{\pi m}\,\Gamma(\frac{m}{2})}$$

این توزیع برای $m = 1$ همان توزیع کوشی است و برای $n \rightarrow +\infty$ می‌توان نشان داد که توزیع نرمال می‌شود. یعنی هر چه m کوچکتر باشد، دُم pdf درازتر است.



برای توزیع $t(n)$، m_k برای $k \geq n$ وجود ندارد، چون $E(|\mathbf{x}|^k)$ نامحدود می‌شود. مثلاً دیده بودیم که در توزیع کوشی، $E(\mathbf{x})$ نداریم. پس داریم:

$$E(\mathbf{x}) = 0 : m > 1$$

$$E(\mathbf{x}^2) = \mathrm{var}(\mathbf{x}) = \frac{m}{m-2} : m > 2$$

قضیهٔ خاصیت اساسی توزیع t: اگر $\mathbf{z} \sim N(0,1)$ و $\mathbf{w} \sim \chi^2(m)$ بوده و $\mathbf{z}$ و $\mathbf{w}$ مستقل باشند، در این صورت $\mathbf{x} = \dfrac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ دارای توزیع $t(m)$ خواهد بود، یعنی:

$$\mathbf{x} \sim t(m)$$

برای اثبات، مثلاً به روش متغیر کمکی می‌توانیم از روی $f_{\mathbf{zw}}$، $f_{\mathbf{x}}$ را به دست آوریم ($\mathbf{x} = \dfrac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ و $\mathbf{y} = \mathbf{w}$).

اثبات در کتاب، صفحهٔ ۲۳۳.

ضمناً اگر $\mathbf{z} \sim N(e,1)$ باشد، $\mathbf{x} \sim t(m,e)$ خواهد بود، یعنی توزیع t غیرمرکزی (Noncentral t Distribution) با m درجه آزادی و خروج از مرکز e خواهد داشت.

با توجه به قضیهٔ فوق برای توزیع t داریم:

$$E(\mathbf{x}^2) = m \cdot E(\mathbf{z}^2)E(\frac{1}{\mathbf{w}}) = m \cdot 1 \cdot \frac{1}{m-2} = \frac{m}{m-2}$$

همچنین داریم:

$$t(m) \xrightarrow[m \to +\infty]{} N(0,1)$$

زیرا اگر $\mathbf{x} = \frac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{m}}}$ باشد که $\mathbf{z} \sim N(0,1)$ و $\mathbf{w} \sim \chi^2(m)$ هستند، در مورد متغیر تصادفی $\mathbf{w}$ که دارای توزیع $\chi^2(m)$ است، خواهیم داشت:

$$\begin{cases} E(\mathbf{w}) = m \\ var(\mathbf{w}) = 2m \end{cases} \Rightarrow \begin{cases} E(\frac{\mathbf{w}}{m}) = 1 \\ var(\frac{\mathbf{w}}{m}) = \frac{2}{m} \end{cases}$$

پس: $\frac{\mathbf{w}}{m} \xrightarrow[m \to +\infty]{m.s.} 1$، یعنی: $\mathbf{x} \to \mathbf{z}$.

برای mهای کمتر $(m > 20)$ داریم:

$$\mathbf{x} \simeq \mathbf{z}\sqrt{\frac{m}{m-2}}$$

زیرا: $E(\mathbf{x}^2) = \frac{m}{m-2}$، ولی: $E(\mathbf{z}^2) = 1$.

کاربرد توزیع t: اگر $\overline{\mathbf{x}}$ و $\mathbf{s}^2$ میانگین و واریانس نمونه از متغیرهای تصادفی نرمال $N(\eta,\sigma)$ باشند و داشته باشیم:

$$\mathbf{T} = \frac{\overline{\mathbf{x}} - \eta}{\frac{\mathbf{s}}{\sqrt{n}}}$$

در این صورت داریم:

$$\mathbf{T} \sim t(n-1)$$

(در حالی که می‌دانیم $\frac{\overline{\mathbf{x}} - \eta}{\frac{\sigma}{\sqrt{n}}}$ دارای توزیع نرمال استاندارد است.)

اثبات: با توجه به قضیه‌ای که داشتیم، $\mathbf{z}=\dfrac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}}$ و $\mathbf{w}=(n-1)\dfrac{\mathbf{s}^2}{\sigma^2}$ مستقلند و $\mathbf{z}\sim N(0,1)$ و $\mathbf{w}\sim\chi^2(n-1)$ هستند.

پس طبق قضیهٔ بالا خواهیم داشت:

$$\mathbf{T}=\frac{\bar{\mathbf{x}}-\eta}{\frac{\mathbf{s}}{\sqrt{n}}}=\frac{\frac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)\frac{\mathbf{s}^2}{\sigma^2}}{n-1}}}=\frac{\mathbf{z}}{\sqrt{\frac{\mathbf{w}}{n-1}}}\sim t(n-1)$$

توجه کنید که مقدار $\mathbf{T}$ و توزیع آن ربطی به σ^2 پروسهٔ نرمال مورد نمونه‌گیری ندارد.

نکتهٔ دیگر اینکه برای $n\to+\infty$ که تخمین $\mathbf{s}$ از σ بهتر و بهتر می‌شود، توزیع متغیر تصادفی فوق به نرمال استاندارد میل می‌کند، یعنی:

$$t(n)\xrightarrow[n\to+\infty]{}N(0,1)$$

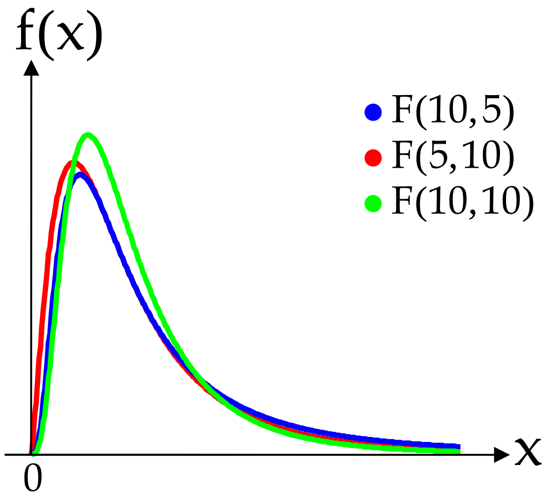
توزیع F (توزیع فیشر):

این توزیع در مسائل مختلف تست فرضیه در آمار ظاهر می‌شود.

گوییم **x** دارای توزیع F با درجه‌های آزادی k و m است: $\mathbf{x} \sim F(k,m)$، هرگاه:

$$f_{\mathbf{x}}(x) = \gamma \frac{x^{\frac{k}{2}-1}}{(1+\frac{kx}{m})^{\frac{k+m}{2}}} : x > 0 \quad \rightarrow \quad \gamma = \frac{\Gamma(\frac{k+m}{2})}{\Gamma(\frac{k}{2})\Gamma(\frac{m}{2})}\left(\frac{k}{m}\right)^{\frac{k}{2}}$$

(ترتیب k و m مهم است.)



قضیهٔ خاصیت اساسی توزیع F: اگر $\mathbf{z} \sim \chi^2(k)$ و $\mathbf{w} \sim \chi^2(r)$ بوده و مستقل باشند و داشته باشیم:

$$\mathbf{x} = \frac{\frac{\mathbf{z}}{k}}{\frac{\mathbf{w}}{r}}$$

آنگاه:

$$\mathbf{x} \sim F(k,r)$$

مشابه توزیع t، مثلاً با روش متغیر تصادفی کمکی می‌توان $f_{\mathbf{x}}$ را به دست آورد و نشان داد که فرم فوق‌الذکر را دارد.

اثبات در کتاب، صفحات ۲۲۴ و ۲۲۵.

ضمناً اگر $\mathbf{z} \sim \chi^2(k,e)$ و $\mathbf{w} \sim \chi^2(r)$ باشند، در این صورت: $\mathbf{x} \sim F(k,r,e)$، یعنی دارای توزیع F غیرمرکزی با درجه‌های آزادی k و r و خروج از مرکز e خواهد بود.

به توجه به قضیهٔ فوق برای توزیع F داریم:

$$E(\mathbf{x}) = \frac{r}{k} E(\mathbf{z}) E(\frac{1}{\mathbf{w}}) = \frac{r}{k} \cdot k \cdot \frac{1}{r-2} = \frac{r}{r-2} : r > 2$$

نتیجهٔ ۱:

$$\mathbf{x} \sim F(k,r) \quad \Rightarrow \quad \frac{1}{\mathbf{x}} \sim F(r,k)$$

نتیجهٔ ۲:

$$\mathbf{x} \sim t(m) \quad \Rightarrow \quad \mathbf{x}^2 \sim F(1,m)$$

زیرا:

$$\begin{cases} \mathbf{x} \sim \dfrac{\mathbf{z}}{\sqrt{\dfrac{\mathbf{w}}{m}}} \\ \\ \mathbf{z} \sim N(0,1) \\ \mathbf{w} \sim \chi^2(m) \end{cases} \Rightarrow \begin{cases} \mathbf{x}^2 \sim \dfrac{\mathbf{z}^2}{\dfrac{\mathbf{w}}{m}} \\ \\ \mathbf{z}^2 \sim \chi^2(1) \\ \mathbf{w} \sim \chi^2(m) \end{cases} \Rightarrow \quad \mathbf{x}^2 \sim F(1,m)$$

نتیجهٔ ۳:

$$\mathbf{x} \sim F(k,r) \quad \Rightarrow \quad \underset{r\to+\infty}{k\mathbf{x}} \sim \chi^2(k)$$

زیرا:

$$\begin{cases} k\mathbf{x} = \dfrac{\mathbf{z}}{\frac{\mathbf{w}}{r}} \\ \mathbf{z} \sim \chi^2(k) \\ \mathbf{w} \sim \chi^2(r) \Rightarrow \dfrac{\mathbf{w}}{r} \xrightarrow[r\to+\infty]{} 1 \end{cases} \quad \Rightarrow \quad k\mathbf{x} \xrightarrow[r\to+\infty]{} \mathbf{z}$$

بخش دوم: آمار

دیدیم که تئوری احتمال یک تئوری کاملاً ریاضی بود که بر اساس یک اصول اولیه و استنتاجات ریاضی از یک مدل مفروض (احتمال یک سری وقایع)، احتمال وقایع دیگری را به دست می‌آوردیم. مدلها فرض شده‌اند و کاری به اینکه با واقعیت تطابق داشته باشند یا نه، نداریم. آمار در مورد کاربرد تئوری احتمال در مسائل واقعی بحث می‌کند و کمک می‌کند که به وسیلهٔ نتایج مشاهدات واقعی (نمونه‌هایی از جامعه) بتوانیم استنتاجاتی (که به احتمال نزدیک یک درست هستند) انجام دهیم. البته این آمار استنتاجی است که بحث ما در مورد آن خواهد بود. شاخهٔ دیگر آمار، آمار توصیفی است که در مورد نحوهٔ دسته‌بندی و ارائهٔ اطلاعات آماری بحث می‌کند (رجوع کنید به کتاب Kreyszig، صفحات ۸۳۸ تا ۸۴۹).

مبحث عمدهٔ آمار استنتاجی، تخمین و آزمون فرضیه است که هر دو کاربرد زیادی در مهندسی برق دارند. مانند سیستم هدایت و کنترل موشک، آشکارسازی هدف در رادار، تعقیب اهداف راداری و ...

آمار
- توصیفی
- استنتاجی
  - تخمین
  - آزمون فرضیه

فصل ۸:

تولید اعداد تصادفی و شبیه‌سازی کامپیوتری

۱. تعریف دنبالهٔ اعداد تصادفی

۲. تولید دنبالهٔ اعداد تصادفی با توزیع یکنواخت

۳. تولید دنبالهٔ اعداد تصادفی با توزیع دلخواه

۴. شبیه‌سازی مونت‌کارلو

تعریف دنبالهٔ اعداد تصادفی:

به جای اینکه سیستم فیزیکی را که دارای ورودی‌های تصادفی یا رفتار تصادفی است عیناً مشاهده کنیم و اثرات پارامترهای مختلف را ببینیم یا با تحمل هزینهٔ فراوان بسازیم (و بعد بفهمیم آن طور که مطلوب است کار نمی‌کند!)، می‌توانیم از شبیه‌سازی کامپیوتری استفاده کنیم (مانند مراحل گیرندهٔ مخابراتی، آشکارساز رادار و ...). امروزه شبیه‌سازی در سراسر دنیا و در رشته‌های مختلف مهندسی کاربرد فراوان دارد. مثلاً طرح اتومبیل ابتدا در کامپیوتر، تعیین محل قراردادن دکلهای مخابرات سیار با شبیه‌سازی امواج و ... .

خصوصاً در مواردی که محاسبهٔ مستقیم مشکل است یا پارامترهای متعددی در مسأله دخیل هستند یا به عنوان مؤیدی برای محاسبات.

به منظور شبیه‌سازی احتیاج به مولد اعداد تصادفی داریم. همچنین برای نمونه‌برداری تصادفی از یک جامعه از اعداد تصادفی استفاده می‌شود. بعضاً برای حل مسائل عددی مشکل از اعداد تصادفی (شبیه‌سازی مونت‌کارلو) استفاده می‌شود. تولید اعداد تصادفی کاربرد بسیار مهمی در ایمن ساختن مخابرات (رمزنگاری) دارد.

دنبالهٔ اعداد تصادفی:

دنباله‌ای از نمونه‌های متغیرهای تصادفی i.i.d. را دنبالهٔ اعداد تصادفی گویند: $x_1, x_2, x_3, \ldots, x_n, \ldots$ (هر دنبالهٔ نمونهٔ فرآیند $\mathbf{x}_n(\omega)$ مشروط بر i.i.d. بودن $\mathbf{x}_i$ها).

طبیعتاً تولید اعداد تصادفی فقط توسط یک پدیدهٔ فیزیکی تصادفی امکان‌پذیر است. آنچه به وسیلهٔ کامپیوتر می‌توان تولید کرد، دنبالهٔ اعداد شبه تصادفی (Pseudo Random) است، یعنی توسط الگوریتمی به طور Deterministic تولید می‌شوند، ولی خواص دنبالهٔ اعداد تصادفی را دارا هستند و لذا تصادفی به نظر می‌رسند. آزمونهایی وجود دارد که تصادفی بودن دنباله را بررسی می‌کنند و در فصل ۸ کتاب آمده‌اند.

برای اینکه دنبالهٔ اعداد شبه تصادفی (از این به بعد تسامحاً می‌گوییم تصادفی) با توزیع f_x داده شده را تولید کنیم، ابتدا تولید دنبالهٔ اعداد تصادفی با توزیع یکنواخت را بررسی می‌کنیم.

تولید دنبالهٔ اعداد تصادفی با توزیع یکنواخت:

تولید دنبالهٔ اعداد تصادفی با توزیع یکنواخت روشهای گوناگونی دارد و در این زمینه بسیار کار شده است. یک روش که در عین ساده بودن، دنبالهٔ حاصل از آن از نظر تصادفی بودن خیلی مطلوب است، الگوریتم Lehmer می‌باشد.

یادآوری: گوییم a همنهشت (Conquence) b است در هنگ n، هرگاه $a-b=kn$ باشد. در این صورت می‌نویسیم:

$$a = b \mathit{mod}\ n$$

مثلاً داریم:

$$17 = 1 \mathit{mod}\ 4$$
$$19 = 3 \mathit{mod}\ 16$$

الگوریتم **Lehmer**:

$$z_n = az_{n-1} \, mod \; m$$

یعنی:

$$z_n = a^n z_0 \, mod \; m$$

فرض می‌کنیم m یک عدد اول بسیار بزرگ بوده و z_0 عددی دلخواه و کوچکتر از m باشد. عدد $a < m$ طوری انتخاب می‌شود که: $\forall n < m-1 : a^n \neq 1 \, mod \; m$ (در این صورت: $a^{m-1} = 1 \, mod \; m$ خواهد بود). به این ترتیب پریود تکرار دنباله ماکزیمم می‌شود $(m-1)$ (و کلیهٔ اعداد 1 تا $m-1$ یک و فقط یک بار ظاهر می‌شوند).

مثلاً می‌توان $m = 2^{31} - 1$ و $a = 2^7 - 1$ را در نظر گرفت.

اگر بگیریم: $u_i = \frac{z_i}{m}$، با توجه به بزرگی m، u_i دنباله‌ای پیوسته با توزیع یکنواخت $u(0,1)$ به نظر می‌رسد. یعنی می‌توان فرض کزد که: $\mathbf{u} \sim u(0,1)$.

نرم‌افزار MATLAB دنبالهٔ تصادفی یکنواخت برای شما تولید می‌کند.

تولید دنبالهٔ اعداد تصادفی با توزیع دلخواه:

۱. روش تبدیل معکوس

۲. روش رد کردن

۳. روش‌های خاص

۱. روش تبدیل معکوس (Inverse Transform Method):

در فصل ۴ و تمرینهای آن دیدیم که اگر متغیر تصادفی $\mathbf{x}$ دارای توزیع انباشتهٔ F باشد، در این صورت $\mathbf{u}=\mathrm{F}(\mathbf{x})$ دارای توزیع یکنواخت $\mathrm{u}(0,1)$ خواهد بود و به عکس اگر $\mathbf{u}\sim\mathrm{u}(0,1)$ باشد، در این صورت $\mathbf{x}=\mathrm{F}^{-1}(\mathbf{u})$ دارای تابع توزیع انباشتهٔ F خواهد بود.

و نیز اگر $\mathbf{x}$ دارای توزیع F باشد و بخواهیم از روی آن متغیر تصادفی $\mathbf{y}$ با تابع توزیع داده شدهٔ G را بسازیم، کافی است بگیریم:

$$\mathbf{y}=\underbrace{\mathrm{G}^{-1}(\overbrace{\mathrm{F}(\mathbf{x})}^{\text{یکنواخت}})}_{\text{دارای توزیع انباشتهٔ } \mathrm{G}}$$

مثال ۱: تولید متغیر تصادفی نمایی (از روی یکنواخت):

$$F(x) = 1 - e^{-\lambda x}$$

$$\mathbf{x} = F^{-1}(\mathbf{u})$$

$$\mathbf{u} = F(\mathbf{x}) = 1 - e^{-\lambda \mathbf{x}} \Rightarrow -\lambda \mathbf{x} = \ln(1 - \mathbf{u}) \Rightarrow \mathbf{x} = -\frac{1}{\lambda}\ln(1 - \mathbf{u})$$

$$\mathbf{u} \sim u(0,1) \Rightarrow 1 - \mathbf{u} \sim u(0,1)$$

پس:

$$x_i = -\frac{1}{\lambda}\ln u_i$$

مثال ۲: تولید متغیر تصادفی رایلی:

$$f(y)=\frac{y}{\sigma^2}e^{-\frac{y^2}{2\sigma^2}} \quad , \quad F(y)=1-e^{-\frac{y^2}{2\sigma^2}}$$

$$\mathbf{u}=F(\mathbf{y})=1-e^{-\frac{\mathbf{y}^2}{2\sigma^2}} \Rightarrow \mathbf{y}^2=-2\sigma^2\ln(1-\mathbf{u})$$

پس:

$$y_i=\sqrt{-2\sigma^2\ln u_i}$$

ضمناً اگر x_i دارای توزیع نمایی با پارامتر λ باشد، $y_i=\sqrt{x_i}$ دارای توزیع رایلی با پارامتر $\sigma^2=\frac{1}{2\lambda}$ خواهد بود.

۲. روش رد کردن (Rejection Method):

اکنون می‌خواهیم دنبالهٔ تصادفی y_i با تابع چگالی g (توزیع انباشتهٔ G) را تولید کنیم.

فرض: دنبالهٔ x_i با تابع چگالی f (توزیع انباشتهٔ F) قابل تولید است و :

$$\forall x : \frac{g(x)}{f(x)} \le \alpha \quad , \quad f(x)=0 \Rightarrow g(x)=0$$

(چون **x** در منطقهٔ $f(x)=0$ ظاهر نمی‌شود، **y** نیز در آن منطقه ظاهر نخواهد شد.)

روش: ابتدا داریم: $i=j=1$؛

۱. x_i را (با توزیع f) تولید می‌کنیم:

$$f_{\mathbf{x}}(x)=f(x)$$

۲. u_i را با توزیع یکنواخت ($u(0,1)$) تولید می‌کنیم (**u** مستقل از **x** است):

$$f_{\mathbf{u}}(u)=\begin{cases}1 & 0<u<1\\ 0 & \text{otherwise}\end{cases}$$

۳.اگر $u_i \le \dfrac{g(x_i)}{\alpha f(x_i)}$ باشد، $y_j = x_i$ قرار می‌دهیم ($i=i+1$ و $j=j+1$، بازگشت به قدم اول). در غیر این صورت این x_i را رد می‌کنیم ($i=i+1$، بازگشت به قدم اول).

اثبات: $F_{\mathbf{y}}(y) \overset{?}{=} G(y) \Leftrightarrow f_{\mathbf{y}}(y) \overset{?}{=} g(y)$؛

داریم:

$$F_{\mathbf{y}}(y) = P\{\mathbf{y} \le y\} = P\{\mathbf{x} \le y \,|\, \mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\} = \frac{P\{\mathbf{x} \le y, \mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}}{P\{\mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\}}$$

$$\to f_{\mathbf{xu}}(x,u) = f_{\mathbf{x}}(x) f_{\mathbf{u}}(u) = \begin{cases} f(x) & 0 < u < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\Rightarrow P\{\mathbf{x} \le y, \mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\} = \int_{-\infty}^{y} \int_{0}^{\frac{g(x)}{\alpha f(x)}} \underbrace{f_{\mathbf{xu}}(x,u)}_{\boxed{0<u<1 \text{ برای } f(x)}} du\, dx$$

با توجه به اینکه $\frac{g(x)}{\alpha f(x)} \le 1$ است:

$$= \int_{-\infty}^{y} f(x) \frac{g(x)}{\alpha f(x)} dx = \frac{G(y)}{\alpha}$$

لذا:

$$P\{\mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\} = P\{\mathbf{x} \le +\infty, \mathbf{u} \le \frac{g(\mathbf{x})}{\alpha f(\mathbf{x})}\} = \frac{G(+\infty)}{\alpha} = \frac{1}{\alpha} \Rightarrow F_{\mathbf{y}}(y) = \frac{\frac{G(y)}{\alpha}}{\frac{1}{\alpha}} = G(y)$$

تعداد تکرار حلقه برای یک بار تولید اعداد تصادفی y_j توزیع هندسی با پارامتر $p=\frac{1}{\alpha}$ است، لذا:

$$\frac{q}{p}=\frac{1-\frac{1}{\alpha}}{\frac{1}{\alpha}}=\alpha-1=$$ (برای یک بار پیروزی) متوسط تعداد شکستها

$$\Rightarrow \alpha=$$ (برای یک بار پیروزی) متوسط تعداد کل

مثال: تولید دنبالهٔ تصادفی نرمال (از روی نمایی) به روش رد کردن و مخلوط کردن:

$$f(x)=e^{-x}:x>0 \quad \rightarrow \quad h(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$$

ولی:

$$f(x)=0 \not\Rightarrow h(x)=0$$

پس ابتدا دنباله با توزیع g (نرمال یکطرفه) را تولید می‌کنیم:

$$g(x)=\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}:x>0$$

$$\frac{g(x)}{f(x)}=\frac{\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}}{e^{-x}}=\frac{2}{\sqrt{2\pi}}e^{-\frac{x^2-2x}{2}}=\sqrt{\frac{2}{\pi}}\,e^{\frac{1}{2}}\,e^{-\frac{x^2-2x+1}{2}}=\sqrt{\frac{2e}{\pi}}\,e^{-\frac{(x-1)^2}{2}}$$

$$\Rightarrow \frac{g(x)}{f(x)}\le\sqrt{\frac{2e}{\pi}}=\alpha=1.32\simeq\frac{4}{3}$$

یعنی تقریباً در هر ۴ بار، ۳ بار موفقیت.

پس ابتدا x_i نمایی تولید می‌کنیم (که چگونگی آن را قبلاً دیدیم):

$$\mathbf{x}\sim \exp(1)$$

سپس u_i یکنواخت تولید می‌کنیم ($\mathbf{u}$ مستقل از $\mathbf{x}$ است):

$$\mathbf{u}\sim u(0,1)$$

اگر $u_i\le e^{-\frac{(x_i-1)^2}{2}}$ باشد، $y_j=x_i$ قرار می‌دهیم. در غیر این صورت x_i را رد می‌کنیم.

حال برای اینکه از روی **y** که نرمال یکطرفه است، متغیر تصادفی **z** نرمال (استاندارد) بسازیم، توجه می‌کنیم که:

$$h(z)=\frac{1}{2}g(z)+\frac{1}{2}g(-z)$$

پس v_i یکنواخت تولید می‌کنیم (**v** مستقل از **x** و **u** است):

$$\mathbf{v}\sim u(0,1)$$

اگر $0<v_i<\frac{1}{2}$ بود، $z_i=y_i$ قرار داده و اگر $\frac{1}{2}<v_i<1$ بود، $z_i=-y_i$ قرار می‌دهیم.

(با توجه به اینکه: $f(x)=\sum_k f(x|A_k)P(A_k)$)

نرم‌افزار MATLAB می‌تواند دنبالهٔ تصادفی نرمال نیز تولید کند.

۳. روش‌های خاص:

متغیرهای تصادفی مختلف را با توجه به خواص ویژهٔ هر یک به روشهایی خاص می‌توان تولید کرد.

مثال ۱: تولید دنبالهٔ نرمال: با توجه به قضیهٔ حد مرکزی، مجموع تعداد زیادی متغیر تصادفی یکنواخت، نرمال می‌شود. پس:

$$z_1 = u_1 + \cdots + u_n \quad , \quad z_2 = u_{n+1} + \cdots + u_{2n} \quad , \quad \cdots$$

یعنی:

$$z_i = \sum_{j=1}^{n} u_{(n-1)i+j}$$

مقدار ۱۰ تا ۱۵ برای n کافی است.

مثال ۲: روش دیگری برای تولید دنبالهٔ نرمال؛ تولید دو دنبالهٔ نرمال استاندارد و مستقل از هم:

یادآوری: اگر متغیرهای تصادفی $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند، داریم:

$$\begin{cases}\mathbf{x} \sim \mathrm{N}(0,1) \\ \mathbf{y} \sim \mathrm{N}(0,1)\end{cases} \Leftrightarrow \begin{cases}\mathbf{R} = \sqrt{\mathbf{x}^2 + \mathbf{y}^2} \sim \mathrm{Rayleigh}(1) \\ \Phi = \tan^{-1} \dfrac{\mathbf{y}}{\mathbf{x}} \sim \mathrm{u}(0,2\pi)\end{cases}$$

اگر u_i و v_i دنباله‌های تصادفی یکنواخت و مستقل باشند، آنگاه $r_i = \sqrt{-2\ln u_i}$ دارای توزیع رایلی و $2\pi v_i$ دارای توزیع یکنواخت $(0,2\pi)$ است. پس دنباله‌های زیر:

$$x_i = \sqrt{-2\ln u_i}\cos(2\pi v_i)$$
$$y_i = \sqrt{-2\ln u_i}\sin(2\pi v_i)$$

نرمال و مستقل هستند.

این هم توسط نرم‌افزار MATLAB مستقیماً قابل تولید است.

مثال ۳: تولید دنبالهٔ تصادفی ارلانگ:

می‌دانیم که مجموع n تا متغیر تصادفی نمایی مستقل با پارامتر λ دارای توزیع ارلانگ با پارامتر n و λ است. پس x_i نمایی با پارامتر λ تولید می‌کنیم و سپس قرار می‌دهیم:

$$y_i = \sum_{j=1}^{n} x_{(n-1)i+j}$$

$$\mathbf{x} \sim \exp(\lambda) \Rightarrow \mathbf{y} \sim \mathrm{Erlang}(n,\lambda)$$

شبیه‌سازی مونت‌کارلو (Monte Carlo Simulation):

به دست آوردن یک کمیت غیرتصادفی توسط روش آماری (با تولید دنبالهٔ تصادفی) را شبیه‌سازی مونت‌کارلو گویند.

مثال ۱: فرض کنید می‌خواهیم انتگرال زیر را حساب کنیم:

$$I = \int_0^1 g(u)du$$

(اگر انتگرال در محدودهٔ دیگری را بخواهیم با تغییر متغیر می‌توانیم به این محدوده ببریم.)

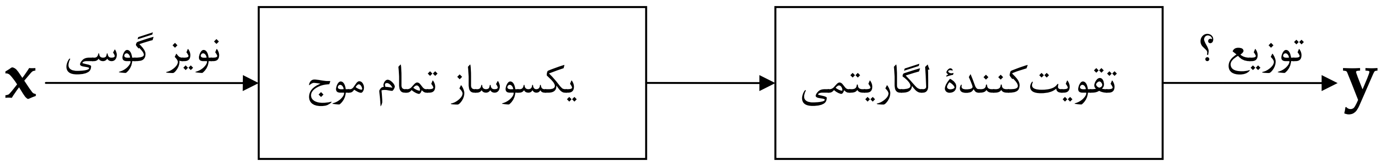
اگر متغیر تصادفی $\mathbf{u}$ را در نظر بگیریم که: $\mathbf{u} \sim u(0,1)$ و فرض کنیم: $\mathbf{x} = g(\mathbf{u})$ باشد، آنگاه:

$$\eta_{\mathbf{x}} = E(g(\mathbf{u})) = \int_{-\infty}^{+\infty} g(u)f(u)du = \int_0^1 g(u)du = I$$

در شبیه‌سازی مونت‌کارلو، دنبالهٔ تصادفی یکنواخت u_i را تولید کرده و سپس $x_i = g(u_i)$ را تولید می‌کنیم. آنگاه با توجه به اینکه: $\eta_{\mathbf{x}} \simeq \frac{1}{n}\sum_{i=1}^{n} x_i$، داریم:

$$I \simeq \frac{1}{n}\sum_{i=1}^{n} g(u_i)$$

مثال ۲: می‌خواهیم $F(y) = P\{\mathbf{y} \le y\}$ (کمیتی غیرتصادفی) را تخمین بزنیم. مثلاً:



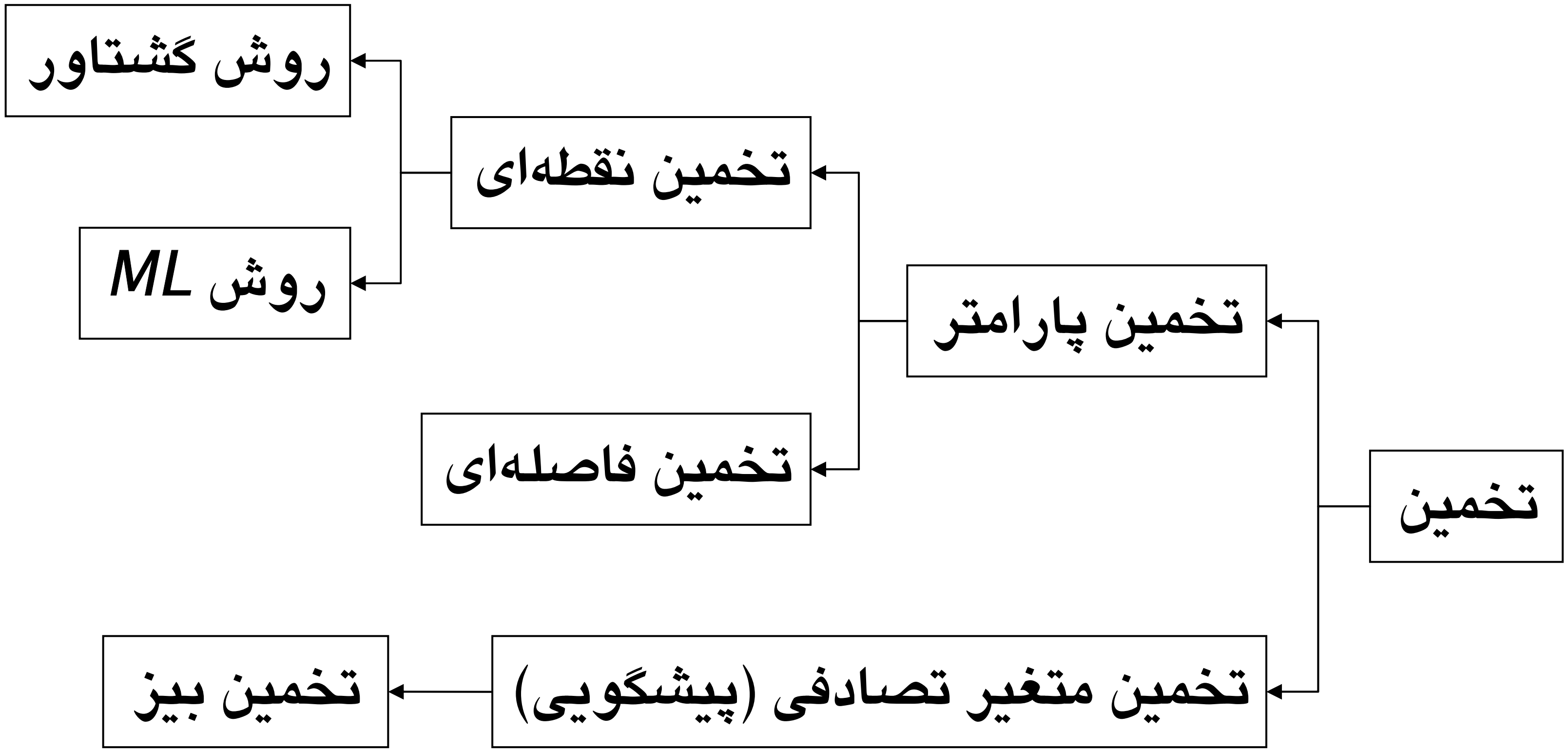
اصولاً روش تحلیلی می‌تواند دشوار یا بعضاً بدون جواب بسته باشد.

دنبالهٔ y_i را تولید می‌کنیم (با تولید x_i و طی مراحل سیستم). اگر تعداد y_iهایی که کوچکتر از y هستند را n_y بنامیم و تعداد کل y_iهای تولید شده برابر n باشد، داریم:

$$F(y) \simeq \frac{n_y}{n}$$

فصل ۹: تخمین (برآورد)

متناظر با Chapter 9، به غیر از Section 9.6



یکی از مباحث مهم آمار که کاربرد زیادی هم در مهندسی برق دارد، تخمین (Estimation) است. مثلاً از روی سیگنالهایی که به آنتن‌های آرایه رسیده است می‌خواهیم زاویهٔ ورود منبع یا منابع ارسال سیگنال را تخمین بزنیم (Direction Finding).

تخمین پارامتر:

فرض کنید توزیع متغیر تصادفی $\mathbf{x}$ دارای نوع مشخصی است، ولی بستگی به پارامتر نامعلوم θ (اسکالر یا بردار) دارد، یعنی: $f(x;\theta)$. مثلاً نرمال با میانگین صفر و واریانس نامعلوم (θ اسکالر) یا با میانگین و واریانس نامعلوم (θ برداری). هدف ما در اینجا این است که با رؤیت نمونه‌های این متغیر تصادفی θ را تخمین بزنیم (تخمین پارامتر).

تخمین نقطه‌ای (Point Estimation):

نمونهٔ $\mathbf{x}$ با اندازهٔ n : $\hat{\theta} = g(\underline{\mathbf{x}}) \quad , \quad \underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$

مشاهدات : $\hat{\theta} = g(\underline{x}) \quad , \quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$

$g(\underline{\mathbf{x}})$ تابعی از متغیرهای تصادفی است (و لذا خود یک متغیر تصادفی می‌باشد).

(اصولاً تابعی از یک یا چند متغیر تصادفی (در اینجا بردار متغیرهای تصادفی .i.i.d نمونه‌ها) را که به پارامتر نامعلومی بستگی نداشته باشد، آماره (Statistic) گویند.)

تخمین ما بدون بایاس (ناریب) خواهد بود، اگر:

$$E(\hat{\theta}) = \theta$$

($E(\hat{\theta}) - \theta$ را بایاس تخمین می‌نامند.)

روش گشتاور (Method of Moments):

می‌دانیم گشتاور kام متغیر تصادفی $\mathbf{x}$ برابر است با:

$$m_k = E(\mathbf{x}^k)$$

در روش گشتاور می‌گیریم:

$$\hat{m}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

و از اینجا تخمینی برای θ (یا θها) به دست می‌آوریم.

مثال ۱: فرض کنید متغیر تصادفی $\mathbf{x}$ دارای توزیع نمایی با پارامتر λ باشد:

$$f_{\mathbf{x}}(x;\lambda) = \lambda e^{-\lambda x} : x > 0$$

می‌دانیم که:

$$m_1 = E(\mathbf{x}) = \frac{1}{\lambda}$$

یا:

$$\lambda = \frac{1}{m_1}$$

پس با رؤیت مقادیر $\mathbf{x}_i$ها داریم:

$$\hat{\lambda} = \frac{1}{\hat{m}_1} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$$

مثال ۲: فرض کنید می‌دانیم که سیگنال دریافتی ما نمونه‌هایی از یک سطح dc ثابت آغشته به نویز گوسی $N(0,\sigma)$ است:

$$\mathbf{x}_i = \theta + \mathbf{n}_i$$

برای اینکه مقدار θ را تخمین بزنیم، داریم:

$$E(\mathbf{x}) = \theta + E(\mathbf{n}) = \theta$$

$$\Rightarrow \hat{\theta} = \hat{m}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

مثال ۳: متغیر تصادفی **x** دارای توزیع گوسی با میانگین و واریانس نامعلوم است. می‌خواهیم با رؤیت مقادیر نمونه‌های آن، میانگین و واریانس را تخمین بزنیم. می‌دانیم که:

$$\eta = m_1 \quad , \quad \sigma^2 = m_2 - m_1^2$$

پس:

$$\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\eta})^2$$

(یعنی همان s^2 که داشتیم.)

در این روش توجه کنید که اگر $\hat{\theta}$ تخمینی از θ باشد، $q(\hat{\theta})$ نیز تخمینی از $q(\theta)$ خواهد بود.

روش ماکزیمم بخت (Method of Maximum Likelihood):

این روش که به آن روش ماکزیمم درست‌نمایی نیز گفته می‌شود، بیش از روش گشتاور متداول است. اگر چه محاسبهٔ آن قدری مشکلتر است. روی خواص ریاضی آن نیز خیلی زیاد کار شده است.

منحنی $f(x;\theta)$ بر حسب θ را تابع بخت (Likelihood Function) (برای θ) می‌گویند.

اگر θ معلوم بود و از ما می‌پرسیدند که کدام x بیشترین بخت برای وقوع را دارد، یعنی $f(x;\theta)dx$ برای کدام x ماکزیمم است، محل پیک منحنی $f(x;\theta)$ جواب (پیشگویی) ما بود (یعنی: $\hat{x}_{ML} = x_{mod}$؛ ولی در معیار mse، میانگین و در معیار mae، میانه می‌شد). حال به عکس θ نامعلوم است و مقدار x خاصی رؤیت شده است. محتمل‌ترین θ با توجه به این x رؤیت شده، محل ماکزیمم تابع بخت است:

$$\theta = \hat{\theta}_{ML} \Leftrightarrow \forall\theta : f(x;\hat{\theta}_{ML}) \geq f(x;\theta)$$

برای این منظور باید داشته باشیم:

$$\frac{\partial f(x;\theta)}{\partial\theta} = 0 \quad \rightarrow \quad \hat{\theta}_{ML}$$

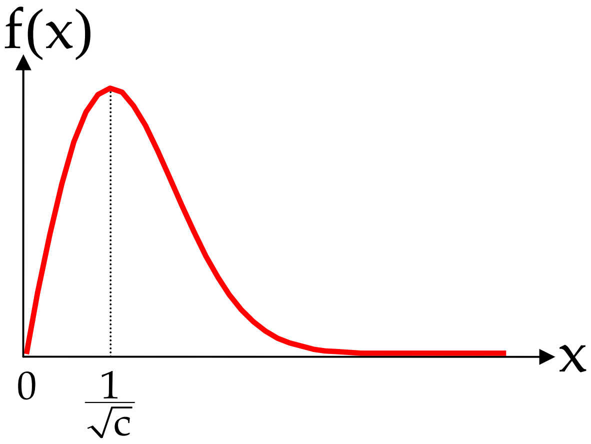
یعنی $\hat{\theta}_{ML}$ محل ماکزیمم تابع $f(x;\theta)$ (برای x داده شده) بر حسب θ می‌باشد (محل ماکزیمم تابع بخت).

به سادگی می‌توان نشان داد که:

$$[\hat{g(\theta)}]_{ML} = g(\hat{\theta}_{ML})$$

(یعنی ماکزیمم بودن $f(x;g(\hat{\theta}))$ بر حسب θ معادل است با ماکزیمم شدن $f(x;g(\theta))$ به ازای $g = g(\hat{\theta})$ برای g یکنوا.)

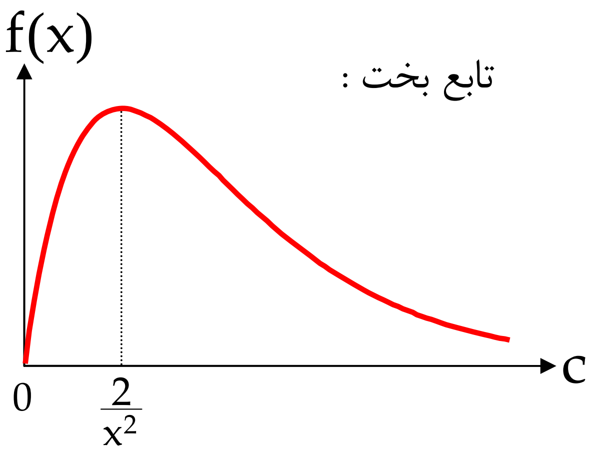
مثال: طول عمر نوعی لامپ دارای تابع چگالی ویبول با $b = 2$ است:



$$f_{\mathbf{x}}(x;c) = cxe^{-c\frac{x^2}{2}}$$

اگر $\mathbf{x} = x$ رؤیت شده باشد، $\hat{c}_{ML}$ چقدر است؟

تابع بخت :



$$\frac{\partial f_{\mathbf{x}}(x;c)}{\partial c} = (x - c\frac{x^3}{2})e^{-c\frac{x^2}{2}} = 0 \quad \Rightarrow \quad \hat{c}_{ML} = \frac{2}{x^2}$$

در حالی که اگر c معلوم بود و از ما می‌خواستند تا $\hat{x}_{ML}$ را به دست آوریم (Prediction)، داشتیم:

$$\frac{\partial f_{\mathbf{x}}(x;c)}{\partial x} = 0 \quad \Rightarrow \quad \hat{x}_{ML} = \frac{1}{\sqrt{c}}$$

توجه: معمولاً فقط یک مشاهده نداریم، بلکه مشاهدات .i.i.d داریم:

$$\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix} \quad \rightarrow \quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$: مقادیر مشاهده شده

پس باید محل ماکزیمم شدن $f_{\underline{\mathbf{x}}}(\underline{x};\theta)$ بر حسب θ را یافت. با توجه به .i.i.d بودن $\mathbf{x}_i$ها داریم:

$$f_{\underline{\mathbf{x}}}(\underline{x};\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

$$\frac{\partial}{\partial\theta}\prod_{i=1}^{n} f(x_i;\theta) = 0 \quad \rightarrow \quad \hat{\theta}_{ML}$$

یا با توجه به یکنوا بودن $\ln$ می‌توان محل ماکزیمم لگاریتم تابع بخت را یافت:

$$\ln f_{\underline{\mathbf{x}}}(\underline{x};\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

$$\frac{\partial}{\partial\theta}\ln f_{\underline{\mathbf{x}}}(\underline{x};\theta) = \sum_{i=1}^{n} \frac{\dfrac{\partial f(x_i;\theta)}{\partial\theta}}{f(x_i;\theta)} = 0 \quad \rightarrow \quad \hat{\theta}_{ML}$$

در مثال قبل داریم:

$$f_{\underline{\mathbf{x}}}(\underline{x};\theta)=c^n\left(\prod_{i=1}^{n}x_i\right)e^{-\frac{c}{2}\sum_{i=1}^{n}x_i^2}$$

$$\Rightarrow \ln f_{\underline{\mathbf{x}}}(\underline{x};\theta)=n\ln c+\sum_{i=1}^{n}\ln x_i-\frac{c}{2}\sum_{i=1}^{n}x_i^2$$

پس خواهیم داشت:

$$\frac{\partial \ln f_{\underline{\mathbf{x}}}(\underline{x};\theta)}{\partial c}=\frac{n}{c}-\frac{1}{2}\sum_{i=1}^{n}x_i^2=0 \quad \Rightarrow \quad \hat{c}_{ML}=\frac{2n}{\sum_{i=1}^{n}x_i^2}$$

مثال: محاسبهٔ تخمین ML برای میانگین و واریانس توزیع نرمال براساس n بار مشاهدهٔ مقدار متغیر تصادفی:

$$f_{\underline{\mathbf{x}}}(\underline{x};\eta,\sigma^2)=\prod_{i=1}^{n} f(x_i;\eta,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\eta)^2}{2\sigma^2}}$$

$$\Rightarrow \ln f_{\underline{\mathbf{x}}}(\underline{x};\eta,\sigma^2)=\sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{(x_i-\eta)^2}{2\sigma^2}\right)$$

$$\rightarrow\begin{cases}\dfrac{\partial \ln f_{\underline{\mathbf{x}}}}{\partial \eta}=0\\[2ex] \dfrac{\partial \ln f_{\underline{\mathbf{x}}}}{\partial \sigma^2}=0\end{cases}\quad\Rightarrow\quad\begin{cases}\displaystyle\sum_{i=1}^{n}-\frac{(x_i-\hat{\eta})^2}{\hat{\sigma}^2}=0\\[2ex] \displaystyle\sum_{i=1}^{n}\left(-\frac{1}{2}\frac{1}{\hat{\sigma}^2}+\frac{(x_i-\hat{\eta})^2}{2(\hat{\sigma}^2)^2}\right)=0\end{cases}\quad\Rightarrow\quad\begin{cases}\displaystyle\hat{\eta}_{ML}=\frac{1}{n}\sum_{i=1}^{n}x_i\\[2ex] \displaystyle\hat{\sigma}^2_{ML}=\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\eta}_{ML})^2=s^2\end{cases}$$

که البته در مورد σ^2، تخمین بایاس‌دار است. البته برای $n\rightarrow+\infty$ بایاس از بین می‌رود (در روش گشتاور نیز به همین نتایج رسیده بودیم).

$$E(s^2)=\frac{n-1}{n}\sigma^2 \quad , \quad var(s^2)=\frac{2(n-1)}{n^2}\sigma^4$$

اصولاً در تخمین ML برای $n\rightarrow+\infty$، بایاس تخمین و واریانس تخمین به سمت صفر میل می‌کنند.

تخمین فاصله‌ای (Interval Estimation):

در اینجا به جای اینکه برای $\underline{x}$ رؤیت شده، یک نقطهٔ $g(\underline{x})$ را به عنوان تخمین θ بدهیم، یک فاصله را ارائه می‌کنیم که θ به احتمال زیاد در آن فاصله قرار دارد. مثلاً برای η به جای اینکه یک نقطهٔ $\overline{x}$ را به عنوان تخمین در نظر بگیریم، فاصله‌ای را معرفی می‌کنیم که η به احتمال زیاد داخل آن است. یعنی دو مقدار $\theta_1 = g_1(\underline{x})$ و $\theta_2 = g_2(\underline{x})$ را پیدا می‌کنیم که θ بین آنها باشد. بسته به اینکه مشاهدات چه باشند، θ_1 و θ_2 متفاوت خواهند بود. پس دو آمارهٔ θ_1 و θ_2 به صورت زیر قابل بیان هستند:

$$\begin{cases}\theta_1 = g_1(\underline{\mathbf{x}}) \\ \theta_2 = g_2(\underline{\mathbf{x}})\end{cases}$$

و لذا فاصله نیز فاصله‌ای تصادفی است (بسته به اینکه $\underline{\mathbf{x}}$ چه باشد) که عدد θ به احتمال زیادی داخل این فاصلهٔ تصادفی است.

اگر $P\{\theta_1 < \theta < \theta_2\} = 1-\alpha$ باشد، فاصلهٔ (θ_1, θ_2) را فاصلهٔ اطمینانِ $1-\alpha$ (Confidence Interval) و α را سطح اطمینان (Confidence Level) گویند. α معمولاً برابر ۵٪ یا ۱٪ یا ۰/۱٪ اختیار می‌شود.

ضمناً اگر $\tau = q(\theta)$ تابعی یکنوا باشد، واقعهٔ $\{\theta_1 < \theta < \theta_2\}$ با $\{\tau_1 < \tau < \tau_2\}$ که $\tau_1 = q(\theta_1)$ و $\tau_2 = q(\theta_2)$ معادل خواهد بود و لذا:

$$P\{\tau_1 < \tau < \tau_2\} = P\{\theta_1 < \theta < \theta_2\} = 1 - \alpha$$

پس اگر فاصلهٔ اطمینان برای θ، (θ_1, θ_2) باشد، برای $q(\theta)$، $(q(\theta_1), q(\theta_2))$ خواهد بود.

هدف در تخمین فاصله‌ای آن است که توابع θ_1 و θ_2 را چنان بیابیم که: $P\{\theta_1 < \theta < \theta_2\} = 1 - \alpha$ شده و مقدار $\theta_2 - \theta_1$ حداقل باشد. در حالت کلی حل این مسئله ساده نیست و ما در موارد خاص مهمی آن را حل می‌کنیم.

فاصلهٔ اطمینان برای میانگین در صورت معلوم بودن واریانس:

اگر متغیر تصادفی $\mathbf{x}$ دارای میانگین نامعلوم η و واریانس σ^2 باشد و نمونه‌های .i.i.d $\mathbf{x}_i$ $(i=1,2,\ldots,n)$ از این متغیر تصادفی برداشته باشیم، تخمین نقطه‌ای زیر را دیدیم:

$$\overline{\mathbf{x}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \text{ : میانگین نمونه} \qquad \rightarrow E(\overline{\mathbf{x}})=\eta \quad , \quad \sigma_{\overline{\mathbf{x}}}^2=\frac{\sigma^2}{n}$$

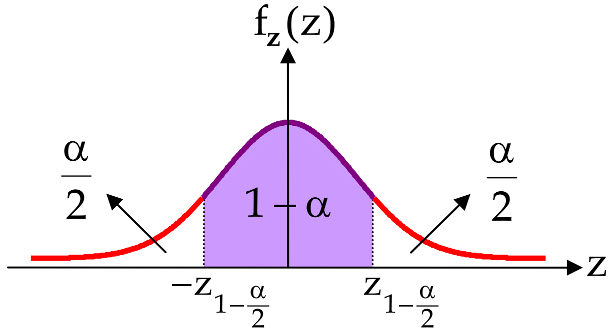
$$\overline{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \text{ : میانگین نمونهٔ مشاهده شده}$$

حال می‌خواهیم یک فاصلهٔ $(\overline{\mathbf{x}}-a,\overline{\mathbf{x}}+a)$ ارائه دهیم که به احتمال $1-\alpha$، η داخل این فاصله باشد.

حتی اگر $\mathbf{x}$ نرمال نباشد، با توجه به قضیهٔ حد مرکزی برای n بزرگ، $\bar{\mathbf{x}}$ تقریباً نرمال است، یعنی: $\bar{\mathbf{x}} \sim N(\eta, \frac{\sigma^2}{n})$. پس اگر

تعریف کنیم: $\mathbf{z} = \frac{\bar{\mathbf{x}} - \eta}{\frac{\sigma}{\sqrt{n}}}$، $\mathbf{z}$ نرمال استاندارد خواهد بود (تقریباً) و لذا:

$$P\{-z < \mathbf{z} < z\} = 1 - \alpha \quad \Rightarrow \quad z = z_{1-\frac{\alpha}{2}}$$



مثلاً برای $\alpha = 0.05$، با توجه به جدول داریم:

$$G^{-1}(0.975) = D^{-1}(0.95) = 1.960$$

یعنی:

$$z_{0.975} = 1.960$$

پس در نهایت خواهیم داشت:

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

تخمین پارامتر : $\boxed{P\{\bar{\mathbf{x}} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \eta < \bar{\mathbf{x}} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\} = 1-\alpha}$

همچنین داریم:

پیش‌بینی : $P\{\eta - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \bar{\mathbf{x}} < \eta + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\} = 1-\alpha$

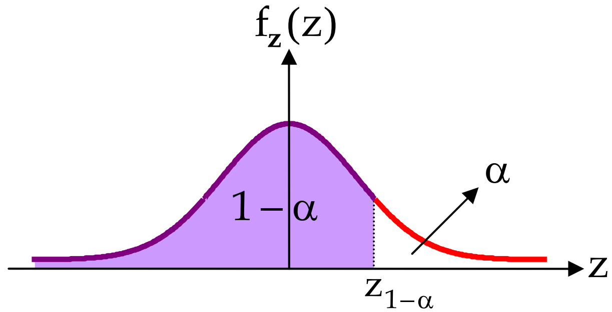
پس فاصلهٔ اطمینانِ $1-\alpha$ برای η در حالت واریانس معلوم عبارت است از:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \eta < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

مثال: طول یک محصول دارای توزیع نرمال با انحراف معیار $\sigma = 4\text{mm}$ است. فاصلهٔ اطمینان ۹۵٪ را برای میانگین به دست آورید اگر در یک نمونهٔ ۳۰ تایی، میانگین نمونه 101mm باشد.

$$101 - \frac{4}{\sqrt{30}}1.96 < \eta < 101 + \frac{4}{\sqrt{30}}1.96 \quad \Rightarrow \quad 99.57 < \eta < 102.43$$

گاهی برای ما مهم این است که η از مقدار خاصی بیشتر است یا نه یا به عکس η از مقدار خاصی کمتر است یا نه. در این صورت به فاصلهٔ یکطرفه احتیاج داریم:



$$P\{\mathbf{z} < z\} = 1-\alpha \quad \Rightarrow \quad z = z_{1-\alpha}$$

$$\Rightarrow P\{\frac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}} < z_{1-\alpha}\} = 1-\alpha$$

پس فاصلهٔ اطمینان $1-\alpha$ سمت راست عبارت است از:

$$P\{\eta > \bar{\mathbf{x}} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\} = 1-\alpha$$

به طور مشابه، فاصلهٔ اطمینان $1-\alpha$ سمت چپ عبارت است از:

$$P\{\eta < \bar{\mathbf{x}} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\} = 1-\alpha$$

(اگر $\mathbf{x}$ نرمال نباشد و n هم بزرگ نباشد، دیگر $\bar{\mathbf{x}}$ تقریباً نرمال نخواهد بود و برای به دست آوردن فاصلهٔ اطمینان می‌توان از قضایای چبیشف یا چرنیوف استفاده کرد.)

مثال: در مثال قبل، فاصلهٔ اطمینان $1-\alpha$ سمت راست را بیابید.

$$\eta > 101 - \frac{40}{\sqrt{30}} 1.645 \quad \Rightarrow \quad \eta > 99.80$$

محاسبهٔ تخمین فاصله‌ای (فاصلهٔ اطمینان) برای میانگین بدون معلوم بودن واریانس:

یکی از راههایی که به نظر می‌رسد این است که خود واریانس را هم تخمین زده و در رابطهٔ فاصلهٔ اطمینان فوق قرار دهیم:

$$\bar{x}-\frac{s}{\sqrt{n}}z_{1-\frac{\alpha}{2}}<\eta<\bar{x}+\frac{s}{\sqrt{n}}z_{1-\frac{\alpha}{2}} \quad \rightarrow \quad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

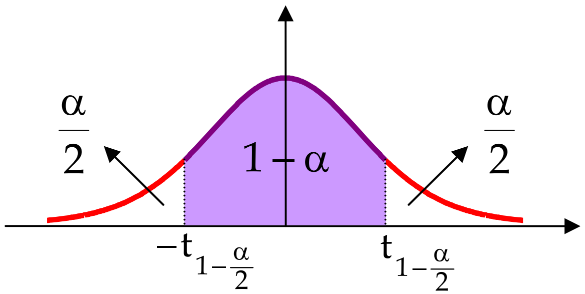
برای n خیلی بزرگ $(n>30)$ این فاصله، فاصلهٔ اطمینان خوبی است و احتمال تقریباً همان $1-\alpha$ است و برای n های کوچکتر احتمال اینکه η داخل این محدوده باشد، کمتر از $1-\alpha$ است. به عبارت دیگر فاصلهٔ اطمینان $1-\alpha$ واقعی قدری بیش از این است.

توجه می‌کنیم که مستقل از مقدار σ داریم:

$$\mathbf{T}=\frac{\bar{\mathbf{x}}-\eta}{\frac{\mathbf{s}}{\sqrt{n}}}\sim t(n-1)$$

پس:

$$P\{-t<\mathbf{T}<t\}=1-\alpha \quad \Rightarrow \quad t=t_{1-\frac{\alpha}{2}}(n-1)$$



لذا:

$$P\{\bar{\mathbf{x}}-\frac{\mathbf{s}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)<\eta<\bar{\mathbf{x}}+\frac{\mathbf{s}}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)\}=1-\alpha$$

پس فاصلهٔ اطمینان $1-\alpha$ برای میانگین جامعه در حالت واریانس نامعلوم عبارت است از:

$$\boxed{\bar{x}-\frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)<\eta<\bar{x}+\frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)}$$

مثال: اگر در مثال قبل، انحراف معیار جامعه نامعلوم بوده و انحراف معیار نمونه برابر $3.91\,mm$ به دست آمده باشد، فاصلهٔ اطمینان ٪۹۵ برای میانگین جامعه را به دست آورید.

با توجه به جدول داریم:

$t_{0.975}(29)=2.05$ (می‌بینیم که 2.05 بزرگتر از 1.96 است)

$$101-\frac{3.91}{\sqrt{30}}2.05<\eta<101+\frac{3.91}{\sqrt{30}}2.05 \quad\Rightarrow\quad 99.54<\eta<102.46$$

(با وجود اینکه s نسبت به σ کوچکتر شده است، فاصلهٔ اطمینان بزرگتر شد.)

فاصلهٔ اطمینان برای احتمال یک واقعه:

واقعهٔ A را در نظر بگیرید که: $P(A)=p$. آزمایش تصادفی را n بار تکرار می‌کنیم و k بار واقعهٔ A اتفاق می‌افتد. تخمین نقطه‌ای در اینجا $\hat{p}=\frac{k}{n}$ است که با توجه به قانون اعداد بزرگ برای $n\to+\infty$ به p میل می‌کند. حال می‌خواهیم فاصلهٔ اطمینان را برای p به دست آوریم.

اگر $\mathbf{x}$ را متغیر تصادفی یک- صفر متناظر با واقعهٔ A تعریف کنیم، داریم:

$$\eta_{\mathbf{x}}=p \quad , \quad \sigma_{\mathbf{x}}^2=pq$$

$$\eta_{\overline{\mathbf{x}}}=p \quad , \quad \sigma_{\overline{\mathbf{x}}}^2=\frac{pq}{n} \quad , \quad \overline{x}=\frac{k}{n}$$

پس طبق آنچه در مورد تخمین میانگین داشتیم (با فرض n نسبتاً بزرگ):

$$P\{\overline{\mathbf{x}}-\sqrt{\frac{pq}{n}}\,z_{1-\frac{\alpha}{2}}<p<\overline{\mathbf{x}}+\sqrt{\frac{pq}{n}}\,z_{1-\frac{\alpha}{2}}\}=1-\alpha$$

ولی واریانس، یعنی $\frac{p(1-p)}{n}$، خود به مقدار میانگین (مورد تخمین) مرتبط است. برای حل این مشکل چند روش داریم:

فاصلهٔ اطمینان محافظه‌کارانه:

pq حداکثر برابر است با $\frac{1}{4}$. پس خواهیم داشت:

$$\frac{k}{n}-\frac{1}{2\sqrt{n}}z_{1-\frac{\alpha}{2}}<p<\frac{k}{n}+\frac{1}{2\sqrt{n}}z_{1-\frac{\alpha}{2}}$$

(فاصلهٔ اطمینان $1-\alpha$ واقعی کمتر از این است.)

فاصلهٔ اطمینان تقریبی:

در این روش به جای p از تخمین آن استفاده می‌کنیم. چون در روش قبل pq ممکن است خیلی کوچکتر از $\frac{1}{4}$ باشد (وقتی p نزدیک $\frac{1}{2}$ نباشد)، تقریب بهتر این است که بگوییم برای n بزرگ، $\hat{p}=\overline{x}=\frac{k}{n}$ به p نزدیک است. پس در فرمول به جای p از $\hat{p}$ استفاده می‌کنیم:

$$\frac{k}{n}-\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}z_{1-\frac{\alpha}{2}}<p<\frac{k}{n}+\sqrt{\frac{\frac{k}{n}(1-\frac{k}{n})}{n}}z_{1-\frac{\alpha}{2}}$$

(فاصلهٔ اطمینان واقعی قدری بیش از این است.)

فاصلهٔ اطمینان دقیق:

$$P\{\bar{\mathbf{x}}-\sqrt{\frac{pq}{n}}z_{1-\frac{\alpha}{2}}<p<\bar{\mathbf{x}}+\sqrt{\frac{pq}{n}}z_{1-\frac{\alpha}{2}}\}=1-\alpha$$

یعنی:

$$P\{(p-\bar{\mathbf{x}})^2<\frac{z_{1-\frac{\alpha}{2}}^2}{n}p(1-p)\}=1-\alpha$$

معادلهٔ درجه دوم زیر را حل می‌کنیم:

$$(p-\frac{k}{n})^2<\frac{z_{1-\frac{\alpha}{2}}^2}{n}p(1-p) \quad \rightarrow \quad p_1,p_2$$

فاصلهٔ اطمینان عبارت است از:

$$p_1<p<p_2$$

مثال: با بررسی ۵۰۰ خانوار در تهران ملاحظه شده است که ۱۵۰ خانوار (۳۰٪) زیر خط فقر زندگی می‌کنند. فاصلهٔ اطمینان ۹۵٪ را برای درصد خانواده‌هایی که زیر خط فقر زندگی می‌کنند، به دست آورید.

$$k = 150 \quad , \quad n = 500 \quad , \quad \alpha = 0.05$$

فاصلهٔ اطمینان محافظه‌کارانه:

$$\frac{150}{500} \pm \frac{1.96}{2\sqrt{500}} \quad \rightarrow \quad 0.3 \pm 0.044 \quad \rightarrow \quad \%30 \pm \%4.4$$

یعنی بین ۲۵/۶ درصد و ۳۴/۴ درصد مردم زیر خط فقر زندگی می‌کنند.

فاصلهٔ اطمینان تقریبی:

$$\frac{150}{500} \pm 1.96\sqrt{\frac{0.3 \times 0.7}{500}} \quad \rightarrow \quad 0.3 \pm 0.040 \quad \rightarrow \quad \%30 \pm \%4.0$$

یعنی بین ۲۶ درصد و ۳۴ درصد مردم زیر خط فقر زندگی می‌کنند.

فاصلهٔ اطمینان دقیق:

$$(p - 0.3)^2 = \frac{1.96^2}{500} p(1-p) \quad \Rightarrow \quad p_1 = 0.261, p_2 = 0.342 \quad \Rightarrow \quad 0.261 < p < 0.342$$

فاصلهٔ اطمینان برای تابع توزیع یک متغیر تصادفی:

$$F(x)=P\{\mathbf{x}\le x\}$$

می‌دانیم تخمین نقطه‌ای توزیع انباشته به صورت: $\hat{F}(x)=\frac{n_x}{n}$ است. در اینجا نیز طبق آنچه در تخمین احتمال داشتیم، عمل می‌کنیم. با استفاده از فاصلهٔ اطمینان تقریبی داریم:

$$\frac{n_x}{n}-\sqrt{\frac{\frac{n_x}{n}(1-\frac{n_x}{n})}{n}}\,z_{1-\frac{\alpha}{2}}<F(x)<\frac{n_x}{n}+\sqrt{\frac{\frac{n_x}{n}(1-\frac{n_x}{n})}{n}}\,z_{1-\frac{\alpha}{2}}$$

$F(x)$ واقعی به احتمال $1-\alpha$ در فاصلهٔ (تصادفی) فوق قرار دارد.

البته این فاصلهٔ اطمینان به x بستگی دارد.

فاصلهٔ اطمینان دیگری برای تابع توزیع:

$$P\{\max_{x}\left|\hat{F}(x)-F(x)\right|\le c\}=1-\alpha$$

تخمین کولموگروف برای n بزرگ (مثلاً لااقل ۱۰۰):

$$c\simeq\sqrt{-\frac{1}{2n}\ln\frac{\alpha}{2}}$$

فاصلهٔ اطمینان برای تفاوت میانگین دو جامعه:

دو متغیر تصادفی $\mathbf{x}$ و $\mathbf{y}$ با میانگین‌های نامعلوم $\eta_{\mathbf{x}}$ و $\eta_{\mathbf{y}}$ و واریانس‌های نامعلوم $\sigma_{\mathbf{x}}^2$ و $\sigma_{\mathbf{y}}^2$ داریم. می‌خواهیم تفاوت میانگین‌ها، یعنی $\eta_{\mathbf{x}} - \eta_{\mathbf{y}}$ را تخمین بزنیم (مثلاً تفاوت درآمد قشر تحصیل کرده و قشر تحصیل نکرده یا تفاوت میزان جنایت در افراد باسواد و بی‌سواد؛ یا مثلاً تغییری در سیستمی داده‌ایم و می‌خواهیم اثر آن را روی η ببینیم). اگر تعریف کنیم:

$$\mathbf{w} = \mathbf{x} - \mathbf{y}$$

می‌خواهیم $\eta_{\mathbf{w}}$ را تخمین بزنیم.

اگر $\mathbf{x}$ و $\mathbf{y}$ نرمال باشند یا تعداد نمونه‌ها زیاد باشد، $\bar{\mathbf{w}}$ نرمال بوده و خواهیم داشت:

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\bar{\mathbf{w}} - \eta_{\mathbf{w}}}{\sigma_{\bar{\mathbf{w}}}} < z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$

$$\bar{w} - \sigma_{\bar{\mathbf{w}}}\, z_{1-\frac{\alpha}{2}} < \eta_{\mathbf{w}} < \bar{w} + \sigma_{\bar{\mathbf{w}}}\, z_{1-\frac{\alpha}{2}}$$

که:

$$\bar{w} = \bar{x} - \bar{y}$$

الف) اگر نمونه‌برداری جفتی باشد (یعنی x_i و y_i در نمونه‌گیری iام به دست آمده باشند)، داریم:

$$\sigma_{\bar{\mathbf{w}}} = \frac{\sigma_{\mathbf{w}}}{\sqrt{n}} \quad , \quad \sigma_{\mathbf{w}}^2 = \sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 - 2\mu_{\mathbf{xy}}$$

یعنی فاصلهٔ اطمینان $1-\alpha$ برای $\eta_{\mathbf{x}} - \eta_{\mathbf{y}}$ عبارت است از:

$$\bar{x} - \bar{y} \pm z_{1-\frac{\alpha}{2}} \frac{\sigma_{\mathbf{w}}}{\sqrt{n}} = \bar{x} - \bar{y} \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_{\mathbf{x}}^2 + \sigma_{\mathbf{y}}^2 - 2\mu_{\mathbf{xy}}}{n}}$$

اگر $\sigma_{\mathbf{w}}$ نامعلوم باشد، خواهیم داشت:

$$\mathbf{T} = \frac{\bar{\mathbf{w}} - \eta_{\mathbf{w}}}{\frac{\mathbf{s}_{\mathbf{w}}}{\sqrt{n}}} \sim t(n-1)$$

فاصلهٔ اطمینان عبارت است از:

$$\bar{x} - \bar{y} \pm t_{1-\frac{\alpha}{2}} \frac{s_{\mathbf{w}}}{\sqrt{n}}$$

ب) با فرض اینکه نمونه‌های $\mathbf{x}$ (n تا) و نمونه‌های $\mathbf{y}$ (m تا) در نمونه‌گیری‌های مختلف به دست آیند، نمونه‌های $\mathbf{x}$ و نمونه‌های $\mathbf{y}$ مستقلند و داریم:

$$\sigma_{\overline{\mathbf{w}}}^2 = \sigma_{\overline{\mathbf{x}}}^2 + \sigma_{\overline{\mathbf{y}}}^2 = \frac{\sigma_{\mathbf{x}}^2}{\mathrm{n}} + \frac{\sigma_{\mathbf{y}}^2}{\mathrm{m}}$$

پس فاصلهٔ اطمینان $1-\alpha$ برای $\eta_{\mathbf{x}} - \eta_{\mathbf{y}}$ عبارت است از:

$$\boxed{\overline{\mathrm{x}} - \overline{\mathrm{y}} \pm \mathrm{z}_{1-\frac{\alpha}{2}} \sqrt{\frac{\sigma_{\mathbf{x}}^2}{\mathrm{n}} + \frac{\sigma_{\mathbf{y}}^2}{\mathrm{m}}}}$$

اگر $\mathbf{x}$ و $\mathbf{y}$ مستقل بوده و واریانس‌ها نامعلوم باشند، می‌توانیم (برای n و m بزرگ)، $\mathrm{s}_{\mathbf{x}}$ و $\mathrm{s}_{\mathbf{y}}$ را جایگزین $\sigma_{\mathbf{x}}$ و $\sigma_{\mathbf{y}}$ کنیم که البته فاصلهٔ اطمینان واقعی قدری بیشتر است.

اگر واریانس‌ها نامعلوم ولی برابر باشند، فاصلهٔ اطمینان دقیق را می‌توان به دست آورد.

در تمرین نشان دادید (مسألهٔ 7.30) که (تخمین ls از σ^2 با فرض $\mathbf{x} \sim N(\eta_x, \sigma)$ و $\mathbf{y} \sim N(\eta_y, \sigma)$)؛

تخمین تلفیقی از σ^2:

$$s_p^2 = \hat{\sigma}^2 = \frac{(n-1)\mathbf{s}_\mathbf{x}^2 + (m-1)\mathbf{s}_\mathbf{y}^2}{n+m-2} = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})^2 + \sum_{i=1}^{n}(\mathbf{y}_i - \bar{\mathbf{y}})^2\right)$$

$$\hat{\sigma}_{\bar{\mathbf{w}}}^2 = \frac{(n-1)\mathbf{s}_\mathbf{x}^2 + (m-1)\mathbf{s}_\mathbf{y}^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right) = s_p^2\left(\frac{1}{n} + \frac{1}{m}\right)$$

$$\mathbf{T} = \frac{\bar{\mathbf{w}} - \eta_\mathbf{w}}{\hat{\sigma}_{\bar{\mathbf{w}}}} \sim t(n+m-2)$$

(متغیر تصادفی فوق مستقل از پارامترهای نامعلوم است.)

$$P\{-t_{1-\frac{\alpha}{2}} < \mathbf{T} < t_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

$$P\{\bar{\mathbf{w}} - \hat{\sigma}_{\bar{\mathbf{w}}}\, t_{1-\frac{\alpha}{2}} < \eta_\mathbf{w} < \bar{\mathbf{w}} + \hat{\sigma}_{\bar{\mathbf{w}}}\, t_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

فاصلهٔ اطمینان $1-\alpha$ برای $\eta_x - \eta_y$ عبارت است از:

$$\boxed{\bar{x} - \bar{y} \pm t_{1-\frac{\alpha}{2}}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}}$$

فاصلهٔ اطمینان برای واریانس جامعهٔ نرمال:

اگر η معلوم باشد، تخمین نقطه‌ای واریانس به صورت: $\mathbf{v}=\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\eta)^2$ است (که بدون بایاس نیز خواهد بود).

حال در اینجا می‌خواهیم فاصلهٔ اطمینان را به دست آوریم:

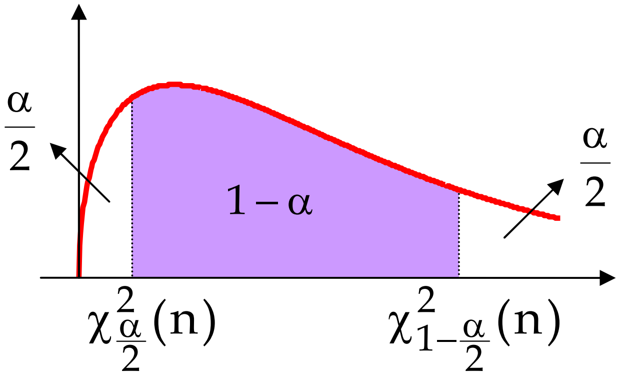
$$\frac{n\mathbf{v}}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i-\eta}{\sigma}\right)^2\sim\chi^2(n)$$

$$P\{\chi^2_{\frac{\alpha}{2}}(n)<\frac{n\mathbf{v}}{\sigma^2}<\chi^2_{1-\frac{\alpha}{2}}(n)\}=1-\alpha$$

پس فاصلهٔ اطمینان $1-\alpha$ برای σ^2 عبارت است از:

$$\boxed{P\{\frac{n\mathbf{v}}{\chi^2_{1-\frac{\alpha}{2}}(n)}<\sigma^2<\frac{n\mathbf{v}}{\chi^2_{\frac{\alpha}{2}}(n)}\}=1-\alpha}$$

(البته این فاصله مینیمم نیست، چون متقارن نیست. ولی برای راحتی استفاده می‌شود.)



مثال: برای اندازه‌گیری قدرت نویز ۱۰ نمونه از دامنهٔ نویز برداشته‌ایم و اعداد زیر حاصل شده‌اند. فاصلهٔ اطمینان ۹۵٪ را برای قدرت نویز به دست آورید ($\eta = 0$ است؛ تخمین فاصله‌ای را برای σ^2 می‌خواهیم).

0.020 -0.099 -0.178 0.059 0.071 -0.111 -0.026 0.053 -0.096 0.067

$$v = \frac{1}{10}\sum_{i=1}^{10} x_i^2 = 0.0066$$

با توجه به جدول داریم:

$$\begin{cases} \chi^2_{0.975}(10) = 20.48 \\ \chi^2_{0.025}(10) = 3.25 \end{cases} \Rightarrow \frac{10 \times 0.0066}{20.48} < \sigma^2 < \frac{10 \times 0.0066}{3.25} \Rightarrow 0.0032 < \sigma^2 < 0.0204$$

اگر η معلوم نباشد، تخمین نقطه‌ای σ^2 برابر خواهد بود با:

$$\mathbf{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2$$

می‌دانیم که:

$$\frac{(n-1)\mathbf{s}^2}{\sigma^2} \sim \chi^2(n-1)$$

پس داریم:

$$P\{\chi^2_{\frac{\alpha}{2}}(n-1) < \frac{(n-1)\mathbf{s}^2}{\sigma^2} < \chi^2_{1-\frac{\alpha}{2}}(n-1)\} = 1-\alpha$$

پس فاصلهٔ اطمینان $1-\alpha$ برای σ^2 عبارت است از:

$$\boxed{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}(n-1)}}$$

فاصلهٔ اطمینان برای ضریب همبستگی:

$$r = \frac{\mu_{\mathbf{xy}}}{\sigma_{\mathbf{x}}\sigma_{\mathbf{y}}}$$

تخمین نقطه‌ای:

$$\hat{r} = \frac{s_{\mathbf{xy}}}{s_{\mathbf{x}}s_{\mathbf{y}}} = \frac{\frac{1}{n-1}\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\frac{1}{n-1}\left[\sum(x_i-\bar{x})^2\right]\cdot\frac{1}{n-1}\left[\sum(y_i-\bar{y})^2\right]}} = \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\left[\sum(x_i-\bar{x})^2\right]\cdot\left[\sum(y_i-\bar{y})^2\right]}}$$

با فرض نمونه‌برداری حقیقی (یعنی در اینجا x_i و y_i حاصل یک آزمایش هستند)، باید آماره‌های r_1 و r_2 را چنان بیابیم که:

$$P\{r_1 < r < r_2\} = 1-\alpha$$

می‌توان (برای n بزرگ) نشان داد که:

$$\begin{cases} r_1 = \tanh\left[\tanh^{-1}(\hat{r}) - \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right] \\ r_2 = \tanh\left[\tanh^{-1}(\hat{r}) + \dfrac{z_{1-\frac{\alpha}{2}}}{\sqrt{n-3}}\right] \end{cases}$$

اشاره به اثبات: صفحهٔ ۲۹۶ کتاب.

تخمین بیز (Bayesian Estimation):

در تخمین پارامتر می‌خواستیم بر مبنای مشاهداتمان از یک متغیر تصادفی، پارامتری نامعلوم (عددی مثل θ) را تخمین بزنیم. در آنجا f_θ مفهوم نداشت، چون θ اصولاً تصادفی نبود. در تخمین بیزی می‌خواهیم بر مبنای مشاهداتمان از یک متغیر تصادفی، متغیر تصادفی دیگری را تخمین بزنیم. برای این منظور در تخمین بیز، تابع ریسک را مینیمم می‌کنیم. یعنی $\hat{\theta} = g(\underline{\mathbf{x}})$ آنچنان پیدا می‌شود که:

$$R = E\left[L(\theta - \hat{\theta})\right]$$

مینیمم گردد (R را تابع ریسک و L را تابع زیان (Loss Function) می‌نامند).

متداول‌ترین نوع تابع زیان، نوع مربعی است:

$$L(e) = e^2 \quad \rightarrow \quad L(\theta - \hat{\theta}) = (\theta - \hat{\theta})^2$$

در این صورت باید $mse = E\left[(\theta-\hat{\theta})^2\right]$ را مینیمم کنیم که همان تخمین ls خواهد بود که قبلاً به عنوان کاربردی از متوسط مشروط دیدیم و ملاحظه کردیم که اگر مشاهداتمان $\underline{\mathbf{x}}$ باشد (در اینجا بردار $\underline{\mathbf{x}}$ که $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$)، داریم:

$$\hat{\theta}_{ls} = E(\theta \mid \underline{\mathbf{x}})$$

$$\hat{\theta}_{ls} = E(\theta \mid \underline{x}) = \int_{-\infty}^{+\infty} \theta \, f_{\theta}(\theta \mid \underline{x}) \, d\theta$$

$f_{\theta}(\theta)$ را تابع چگالی پیشین و $f_{\theta}(\theta \mid \underline{x})$ را تابع چگالی پسین گویند.

ولی معمولاً داده‌های مسئله شامل $f_{\theta}(\theta \mid \underline{x})$ نیست، بلکه $f_{\theta}(\theta)$ و $f_{\underline{\mathbf{x}}}(\underline{x} \mid \theta)$ را داریم ($f_{\underline{\mathbf{x}}}(\underline{x} \mid \theta)$ را تابع بخت (Likelihood Function) می‌نامند).

با توجه به قضیهٔ بیز داریم:

$$f_\theta(\theta|\underline{x}) = \frac{f_{\underline{\mathbf{x}}}(\underline{x}|\theta)f_\theta(\theta)}{\underbrace{f_{\underline{\mathbf{x}}}(\underline{x})}_{\boxed{\text{مستقل از } \theta}}} = \gamma f_{\underline{\mathbf{x}}}(\underline{x}|\theta)f_\theta(\theta)$$

(پس از نظر نوع تابعیت از θ، حاصلضرب چگالی پیشین و تابع بخت، چگالی پسین می‌شود.)

$$\hat{\theta}_{ls} = \int_{-\infty}^{+\infty}\theta f_\theta(\theta|\underline{x})d\theta = \frac{1}{f_{\underline{\mathbf{x}}}(\underline{x})}\int_{-\infty}^{+\infty}\theta f_{\underline{\mathbf{x}}}(\underline{x}|\theta)f_\theta(\theta)d\theta$$

تابع زیان دیگر:

$$L(e) = |e| \quad \to \quad L(\theta-\hat{\theta}) = |\theta-\hat{\theta}|^2$$

یعنی باید $mae = E(|\theta-\hat{\theta}|^2)$ مینیمم شود که منجر می‌شود به:

$$\hat{\theta}_{abs} = \text{median}\left(f_\theta(\theta|\underline{x})\right)$$

یعنی:

$$\int_{-\infty}^{\hat{\theta}_{abs}} f_\theta(\theta|\underline{x})d\theta = \int_{\hat{\theta}_{abs}}^{+\infty} f_\theta(\theta|\underline{x})d\theta$$

تابع زیان دیگر:

$$L(e)=\begin{cases}0 & |e|<\varepsilon\\ 1 & \text{otherwise}\end{cases} \quad\rightarrow\quad L(\theta-\hat{\theta})=\begin{cases}0 & |\theta-\hat{\theta}|<\varepsilon\\ 1 & \text{otherwise}\end{cases}$$

که منجر می‌شود به تخمین MAP (Maximum Apestriori Probability). این θای است که $f_\theta(\theta|\underline{x})$ را ماکزیمم می‌کند، یعنی:

$$\theta=\hat{\theta}_{MAP}\Leftrightarrow \forall\theta: f_\theta(\hat{\theta}_{MAP}|\underline{x})\geq f_\theta(\theta|\underline{x})$$

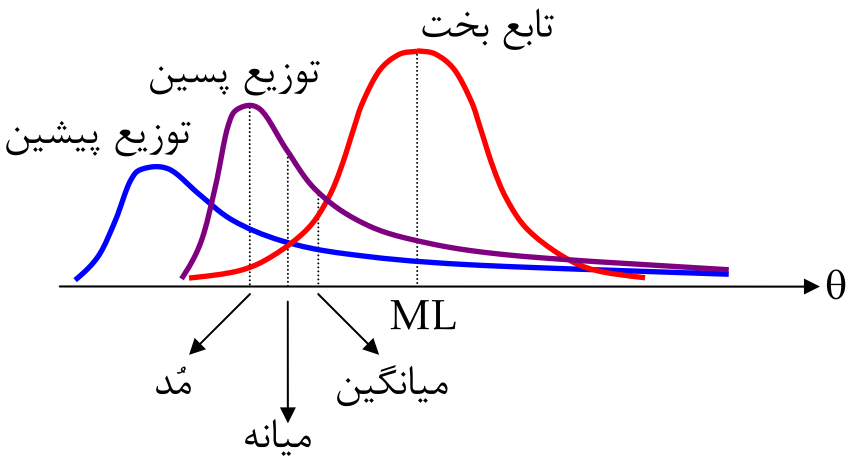
یا:

$$\theta=\hat{\theta}_{MAP}\Leftrightarrow \forall\theta: f_{\underline{x}}(\underline{x}|\hat{\theta}_{MAP})f_\theta(\hat{\theta}_{MAP})\geq f_{\underline{x}}(\underline{x}|\theta)f_\theta(\theta)$$

(پس $\hat{\theta}_{ML}$، θای است که $f(x|\theta)$ را ماکزیمم می‌کند، چون در آنجا $f_\theta(\theta)$ نداشتیم. ولی $\hat{\theta}_{MAP}$، θای است که $f(\theta|x)=\gamma f(x|\theta)f(\theta)$ را ماکزیمم کند.)

پس این سه نوع تخمین (ls، abs و MAP)، یکی میانگین، دیگری میانه و سومی مُد (نِمای) تابع چگالی پسین $f_\theta(\theta|\underline{x})$ را به عنوان تخمین انتخاب می‌کنند.

با افزایش n منحنی likelihood تیزتر شده و تابع چگالی پسین به تابع likelihood نزدیکتر می‌شود. یعنی با افزایش مشاهدات، اطلاعات پیشین ارزش خود را از دست می‌دهند.



مثال ۱: در فصل ۶ این مثال را داشتیم. اگر $\mathbf{x} \sim \text{Binomial}(n,\mathbf{p})$ بوده و خود $\mathbf{p}$ یک متغیر تصادفی با توزیع $u(0,1)$ باشد، می‌خواهیم بر مبنای مشاهدهٔ $\mathbf{x}=k$ مقدار $\mathbf{p}$ را تخمین ls بزنیم.

$$\hat{p}_{ls}=?$$

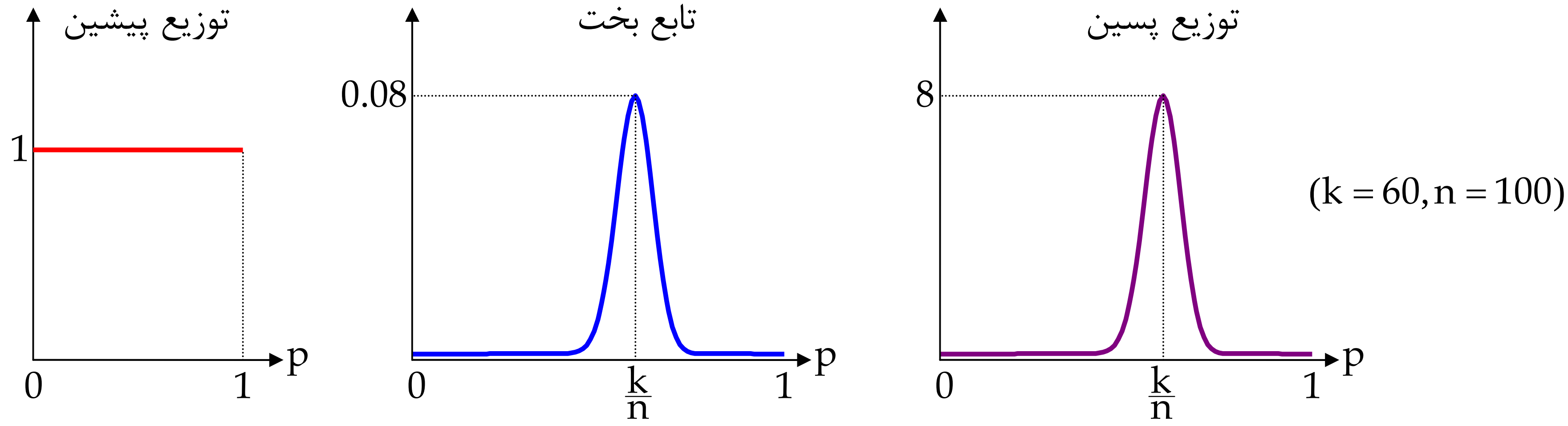
در آنجا دیدیم که:

$$\overbrace{f_{\mathbf{p}}(p|\mathbf{x}=k)}^{\text{پسین}}=\frac{\overbrace{P\{\mathbf{x}=k|\mathbf{p}=p\}}^{\text{تابع بخت}}\overbrace{f_{\mathbf{p}}(p)}^{\text{پیشین}}}{P\{\mathbf{x}=k\}}=\frac{P\{\mathbf{x}=k|\mathbf{p}=p\}f_{\mathbf{p}}(p)}{\int_0^1 P\{\mathbf{x}=k|\mathbf{p}=p\}f_{\mathbf{p}}(p)dp}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}f_{\mathbf{p}}(p)}{\int_0^1\binom{n}{k}p^k(1-p)^{n-k}f_{\mathbf{p}}(p)dp}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\frac{1}{n+1}}=\frac{(n+1)!}{k!(n-k)!}p^k(1-p)^{n-k}:0\le p\le 1$$: یکنواخت $f_{\mathbf{p}}(p)$ برای

که توزیع بتای $\text{Beta}(k+1,n-k+1)$ بوده و هر چقدر n بزرگتر باشد، در حوالی $\frac{k}{n}$ تیزتر است.



$$\hat{p}_{ls} = E(\mathbf{p} \mid \mathbf{x}=k) = \int_0^1 p\, f_{\mathbf{p}}(p \mid \mathbf{x}=k)\,dp = \frac{\int_0^1 p^{k+1}(1-p)^{n-k} f_{\mathbf{p}}(p)\,dp}{\int_0^1 p^{k}(1-p)^{n-k} f_{\mathbf{p}}(p)\,dp}$$

برای $f_{\mathbf{p}}(p)$ یکنواخت : $$= \frac{(n+1)!}{k!(n-k)!} \cdot \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k+1}{n+2}$$

در حالی که بدون مشاهده (بر مبنای اطلاعات پیشین $f_{\mathbf{p}}$)، $\hat{p} = \frac{1}{2}$ بود، اکنون برابر $\frac{k+1}{n+2}$ است که برای n بزرگ تقریباً برابر $\frac{k}{n}$ می‌شود.

اما تخمین MAP (و نیز تخمین ML) برابر $\frac{k}{n}$ هستند.

مثال ۲: اگر $\underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$ مشاهدات i.i.d. از یک متغیر تصادفی با توزیع $N(\eta,\sigma)$ باشند و فرض کنیم که η خود یک متغیر تصادفی نرمال با میانگین صفر و واریانس σ_η^2 باشد، یعنی: $\eta \sim N(0,\sigma_\eta^2)$، $\hat{\eta}_{ls}$ را به دست آورید.

$$f_\eta(\eta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{\eta^2}{2\sigma_\eta^2}}$$

$$f_{\underline{\mathbf{x}}}(\underline{x}\,|\,\eta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\eta)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i-\eta)^2}{2\sigma^2}}$$

$$f_\eta(\eta\,|\,\underline{x}) = \frac{f_{\underline{\mathbf{x}}}(\underline{x}\,|\,\eta)\, f_\eta(\eta)}{f_{\underline{\mathbf{x}}}(\underline{x})} = \frac{f_{\underline{\mathbf{x}}}(\underline{x}\,|\,\eta)\, f_\eta(\eta)}{\int_{-\infty}^{+\infty} f_{\underline{\mathbf{x}}}(\underline{x}\,|\,\eta)\, f_\eta(\eta)\, d\eta}$$

که پس از قدری محاسبات نتیجه می‌شود که:

$$f_\eta(\eta \mid \underline{x}) = N(\underbrace{\frac{\sum x_i}{n + \frac{\sigma^2}{\sigma_\eta^2}}}_{\hat{\eta}}, \underbrace{\sqrt{\frac{\sigma_\eta^2 \cdot \frac{\sigma^2}{n}}{\sigma_\eta^2 + \frac{\sigma^2}{n}}}}_{\sigma_0})$$

چون در توزیع نرمال، میانگین و مُد و میانه بر هم منطبق هستند، داریم:

$$\hat{\eta}_{\text{ls}} = \hat{\eta}_{\text{abs}} = \hat{\eta}_{\text{MAP}} = \frac{n}{n + \frac{\sigma^2}{\sigma_\eta^2}} \overline{x}$$

ملاحظه می‌شود که برای $n \to +\infty$، اثر اطلاعات پیشین کمتر و کمتر می‌شود و $\hat{\eta} = \overline{x}$ می‌گردد.

اگر تخمین فاصله‌ای را هم بخواهیم، با توجه به نرمال بودن $\eta \mid \underline{x}$ (یا تقریباً نرمال بودن آن) داریم:

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\eta - \hat{\eta}}{\sigma_0} < z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$

$$\Rightarrow P\{\hat{\eta} - \sigma_0\, z_{1-\frac{\alpha}{2}} < \eta < \hat{\eta} + \sigma_0\, z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$