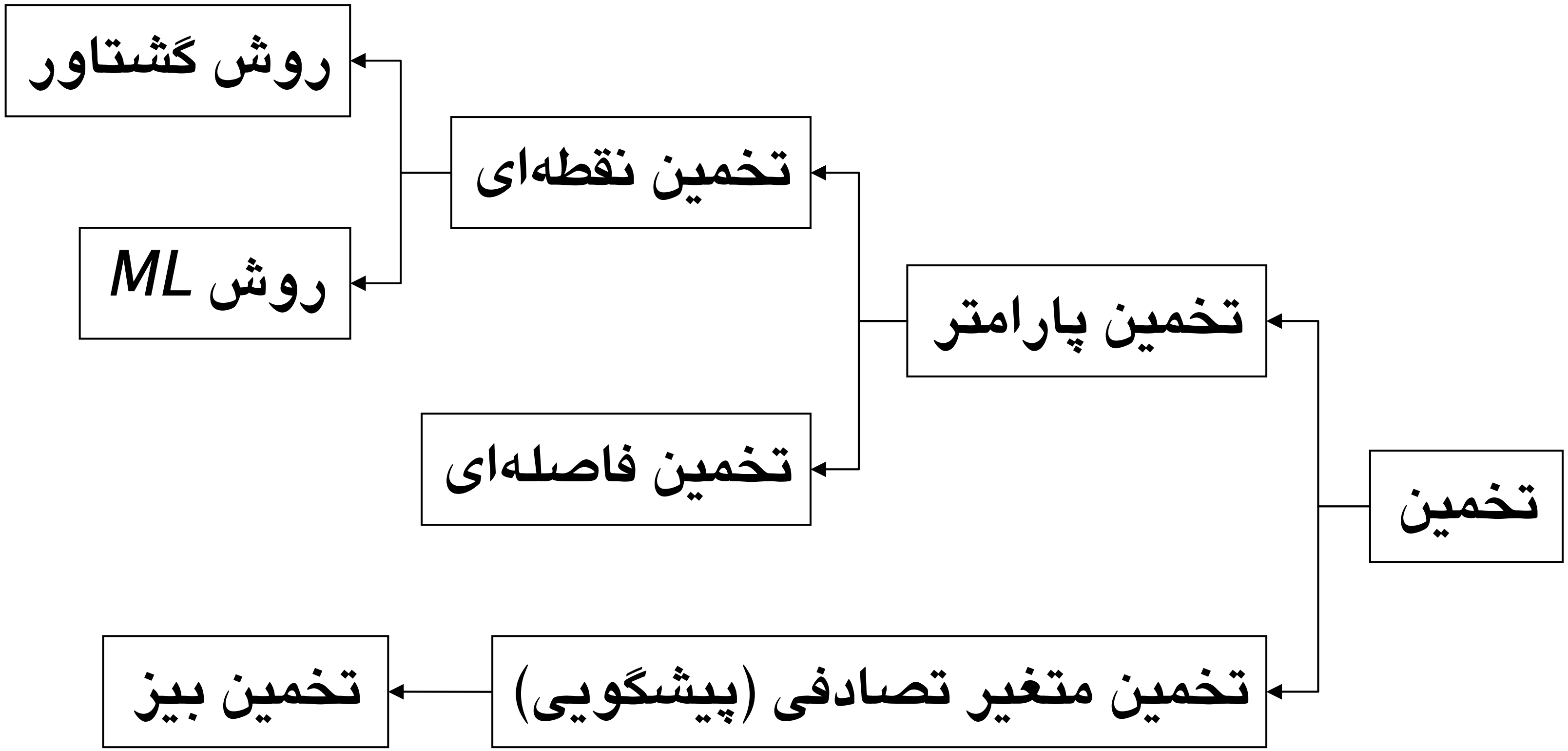
فصل ۹: تخمین (برآورد)

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 **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):

$$\hat{\theta} = g(\underline{\mathbf{x}}) \quad , \quad \underline{\mathbf{x}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_n \end{bmatrix}$$: نمونهٔ $\mathbf{x}$ با اندازهٔ n :

$$\hat{\theta} = g(\underline{x}) \quad , \quad \underline{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$: مشاهدات :

$g(\underline{\mathbf{x}})$ 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 محاسبۀ 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:

$$\frac{\partial f(x;\theta)}{\partial\theta} = 0 \quad \rightarrow \quad \hat{\theta}_{ML}$$

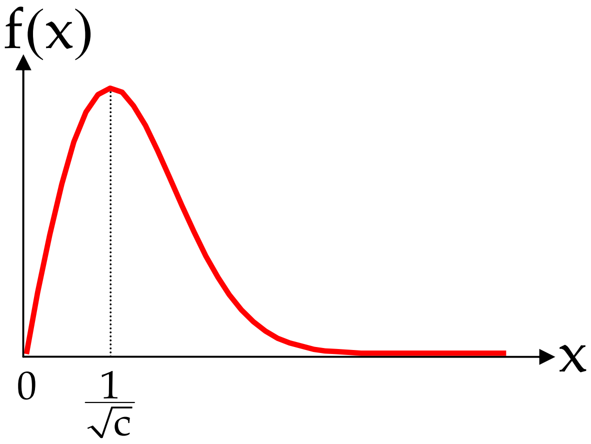
یعنی $\hat{\theta}_{ML}$ محل ماکزیمم تابع $f(x;\theta)$ (برای x داده شده) بر حسب θ می‌باشد (محل ماکزیمم تابع بخت).

به سادگی می‌توان نشان داد که:

$$[g(\hat{\theta})]_{ML} = g(\hat{\theta}_{ML})$$

(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.)

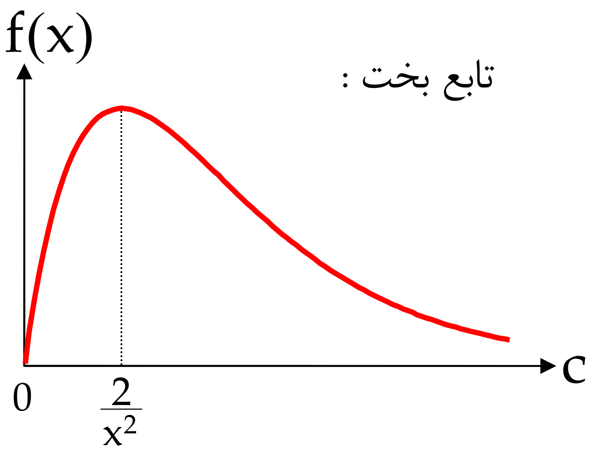
مثال: طول عمر نوعی لامپ دارای تابع چگالی ویبول با $b=2$ است:



$$f_{\mathbf{x}}(x;c)=cxe^{-c\frac{x^2}{2}}$$

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:

$$\ln f_{\underline{\mathbf{x}}}(\underline{x};\theta) = \sum_{i=1}^{n} \ln f(x_i;\theta)$$

$$\frac{\partial}{\partial\theta}\ln f_{\underline{\mathbf{x}}}(\underline{x};\theta) = \sum_{i=1}^{n} \frac{\frac{\partial f(x_i;\theta)}{\partial\theta}}{f(x_i;\theta)} = 0 \quad \rightarrow \quad \hat{\theta}_{ML}$$

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:

$$f_{\underline{\mathbf{x}}}(\underline{x};\eta,\sigma^2)=\prod_{i=1}^{n} f(x_i;\eta,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x_i-\eta)^2}{2\sigma^2}}$$

$$\Rightarrow \ln f_{\underline{\mathbf{x}}}(\underline{x};\eta,\sigma^2)=\sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\pi\sigma^2)-\frac{(x_i-\eta)^2}{2\sigma^2}\right)$$

$$\rightarrow\begin{cases}\dfrac{\partial \ln f_{\underline{\mathbf{x}}}}{\partial\eta}=0\\[2ex] \dfrac{\partial \ln f_{\underline{\mathbf{x}}}}{\partial\sigma^2}=0\end{cases}\quad\Rightarrow\quad\begin{cases}\displaystyle\sum_{i=1}^{n}-\frac{(x_i-\hat{\eta})^2}{\hat{\sigma}^2}=0\\[2ex] \displaystyle\sum_{i=1}^{n}\left(-\frac{1}{2}\frac{1}{\hat{\sigma}^2}+\frac{(x_i-\hat{\eta})^2}{2(\hat{\sigma}^2)^2}\right)=0\end{cases}\quad\Rightarrow\quad\begin{cases}\hat{\eta}_{ML}=\displaystyle\frac{1}{n}\sum_{i=1}^{n}x_i\\[2ex] \hat{\sigma}^2_{ML}=\displaystyle\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\eta}_{ML})^2=s^2\end{cases}$$

که البته در مورد σ^2، تخمین بایاس‌دار است. البته برای $n\to+\infty$ بایاس از بین می‌رود (در روش گشتاور نیز به همین نتایج رسیده بودیم).

$$E(s^2)=\frac{n-1}{n}\sigma^2\quad,\quad \mathrm{var}(s^2)=\frac{2(n-1)}{n^2}\sigma^4$$

اصولاً در تخمین ML برای $n\to+\infty$، 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:

$$\bar{\mathbf{x}}=\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \text{ : میانگین نمونه} \qquad \to E(\bar{\mathbf{x}})=\eta \quad , \quad \sigma_{\bar{\mathbf{x}}}^2=\frac{\sigma^2}{n}$$

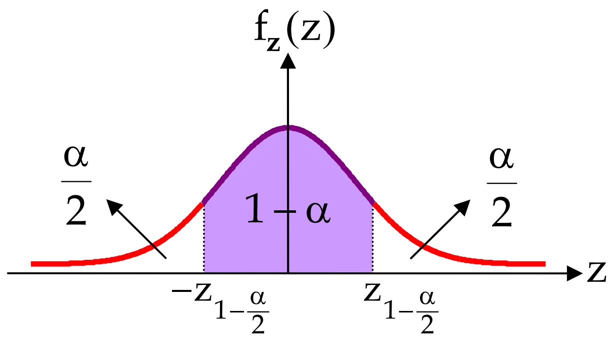
$$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i \text{ : میانگین نمونهٔ مشاهده شده}$$

حال می‌خواهیم یک فاصلهٔ $(\bar{\mathbf{x}}-a,\bar{\mathbf{x}}+a)$ 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: $\mathbf{z} = \dfrac{\bar{\mathbf{x}} - \eta}{\frac{\sigma}{\sqrt{n}}}$، $\mathbf{z}$ 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:

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\bar{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

$$\boxed{\text{تخمین پارامتر :}\quad P\{\bar{\mathbf{x}} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \eta < \bar{\mathbf{x}} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}\} = 1-\alpha}$$

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:

$$\bar{x} - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} < \eta < \bar{x} + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}$$

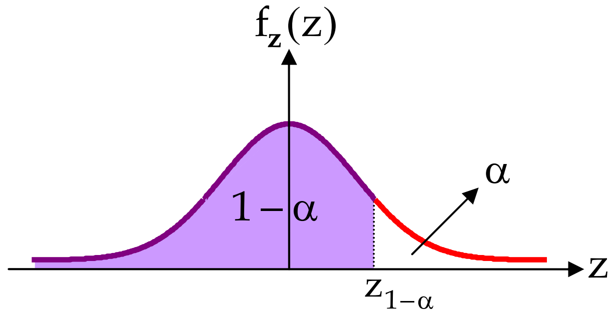
مثال: طول یک محصول دارای توزیع نرمال با انحراف معیار $\sigma = 4\text{mm}$ است. فاصلهٔ اطمینان ۹۵٪ را برای میانگین به دست آورید اگر در یک نمونهٔ ۳۰ تایی، میانگین نمونه 101mm باشد.

$$101 - \frac{4}{\sqrt{30}}1.96 < \eta < 101 + \frac{4}{\sqrt{30}}1.96 \quad \Rightarrow \quad 99.57 < \eta < 102.43$$

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:

$$P\{\mathbf{z} < z\} = 1-\alpha \quad \Rightarrow \quad z = z_{1-\alpha}$$

$$\Rightarrow P\{\frac{\overline{\mathbf{x}}-\eta}{\frac{\sigma}{\sqrt{n}}} < z_{1-\alpha}\} = 1-\alpha$$



پس فاصلهٔ اطمینان $1-\alpha$ سمت راست عبارت است از:

$$P\{\eta > \overline{\mathbf{x}} - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\} = 1-\alpha$$

به طور مشابه، فاصلهٔ اطمینان $1-\alpha$ سمت چپ عبارت است از:

$$P\{\eta < \overline{\mathbf{x}} + \frac{\sigma}{\sqrt{n}} z_{1-\alpha}\} = 1-\alpha$$

(اگر $\mathbf{x}$ نرمال نباشد و n هم بزرگ نباشد، دیگر $\overline{\mathbf{x}}$ 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:

$$\bar{x}-\frac{s}{\sqrt{n}}z_{1-\frac{\alpha}{2}}<\eta<\bar{x}+\frac{s}{\sqrt{n}}z_{1-\frac{\alpha}{2}} \quad \rightarrow \quad s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

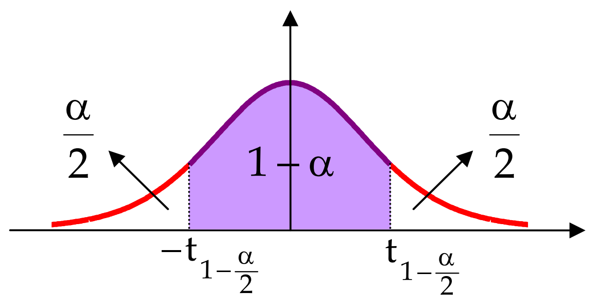
برای n خیلی بزرگ $(n>30)$ این فاصله، فاصلهٔ اطمینان خوبی است و احتمال تقریباً همان $1-\alpha$ است و برای n های کوچکتر احتمال اینکه η داخل این محدوده باشد، کمتر از $1-\alpha$ است. به عبارت دیگر فاصلهٔ اطمینان $1-\alpha$ واقعی قدری بیش از این است.

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:

$$\bar{x}-\frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)<\eta<\bar{x}+\frac{s}{\sqrt{n}}t_{1-\frac{\alpha}{2}}(n-1)$$

مثال: اگر در مثال قبل، انحراف معیار جامعه نامعلوم بوده و انحراف معیار نمونه برابر 3.91mm 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: $\hat{F}(x) = \dfrac{n_x}{n}$ 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:

$$\boxed{\bar{x} - \bar{y} \pm t_{1-\frac{\alpha}{2}} \frac{s_{\mathbf{w}}}{\sqrt{n}}}$$

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:

$$\overline{\mathrm{x}}-\overline{\mathrm{y}}\pm \mathrm{z}_{1-\frac{\alpha}{2}}\sqrt{\frac{\sigma_{\mathbf{x}}^2}{\mathrm{n}}+\frac{\sigma_{\mathbf{y}}^2}{\mathrm{m}}}$$

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:

$$s_p^2 = \hat{\sigma}^2 = \frac{(n-1)\mathbf{s}_{\mathbf{x}}^2 + (m-1)\mathbf{s}_{\mathbf{y}}^2}{n+m-2} = \frac{1}{n+m-2}\left(\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})^2 + \sum_{i=1}^{n}(\mathbf{y}_i - \overline{\mathbf{y}})^2\right)$$

$$\hat{\sigma}_{\overline{\mathbf{w}}}^2 = \frac{(n-1)\mathbf{s}_{\mathbf{x}}^2 + (m-1)\mathbf{s}_{\mathbf{y}}^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right) = s_p^2\left(\frac{1}{n} + \frac{1}{m}\right)$$

$$\mathbf{T} = \frac{\overline{\mathbf{w}} - \eta_{\mathbf{w}}}{\hat{\sigma}_{\overline{\mathbf{w}}}} \sim t(n+m-2)$$

(متغیر تصادفی فوق مستقل از پارامترهای نامعلوم است.)

$$P\{-t_{1-\frac{\alpha}{2}} < \mathbf{T} < t_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

$$P\{\overline{\mathbf{w}} - \hat{\sigma}_{\overline{\mathbf{w}}}\, t_{1-\frac{\alpha}{2}} < \eta_{\mathbf{w}} < \overline{\mathbf{w}} + \hat{\sigma}_{\overline{\mathbf{w}}}\, t_{1-\frac{\alpha}{2}}\} = 1-\alpha$$

فاصلهٔ اطمینان $1-\alpha$ برای $\eta_{\mathbf{x}} - \eta_{\mathbf{y}}$ عبارت است از:

$$\overline{x} - \overline{y} \pm t_{1-\frac{\alpha}{2}}\, s_p \sqrt{\frac{1}{n} + \frac{1}{m}}$$

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:

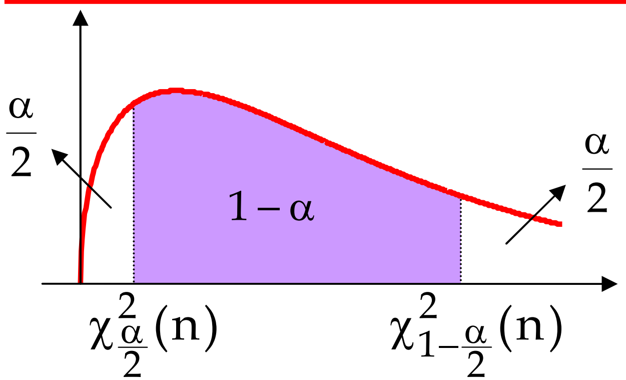
$$\frac{n\mathbf{v}}{\sigma^2}=\sum_{i=1}^{n}\left(\frac{\mathbf{x}_i-\eta}{\sigma}\right)^2\sim\chi^2(n)$$

$$P\{\chi^2_{\frac{\alpha}{2}}(n)<\frac{n\mathbf{v}}{\sigma^2}<\chi^2_{1-\frac{\alpha}{2}}(n)\}=1-\alpha$$

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:

$$\hat{\theta}_{ls} = E(\theta \,|\, \underline{\mathbf{x}})$$

$$\hat{\theta}_{ls} = E(\theta \,|\, \underline{x}) = \int_{-\infty}^{+\infty} \theta \, f_\theta(\theta \,|\, \underline{x}) \, d\theta$$

$f_\theta(\theta)$ را تابع چگالی پیشین و $f_\theta(\theta \,|\, \underline{x})$ را تابع چگالی پسین گویند.

ولی معمولاً داده‌های مسئله شامل $f_\theta(\theta \,|\, \underline{x})$ نیست، بلکه $f_\theta(\theta)$ و $f_{\underline{\mathbf{x}}}(\underline{x} \,|\, \theta)$ را داریم ($f_{\underline{\mathbf{x}}}(\underline{x} \,|\, \theta)$ را تابع بخت (Likelihood Function) می‌نامند).

با توجه به قضیهٔ بیز داریم:

$$f_\theta(\theta|\underline{x}) = \frac{f_{\underline{x}}(\underline{x}|\theta)\, f_\theta(\theta)}{\underbrace{f_{\underline{x}}(\underline{x})}_{\boxed{\text{مستقل از } \theta}}} = \gamma\, f_{\underline{x}}(\underline{x}|\theta)\, f_\theta(\theta)$$

(پس از نظر نوع تابعیت از θ، حاصلضرب چگالی پیشین و تابع بخت، چگالی پسین می‌شود.)

$$\hat{\theta}_{ls} = \int_{-\infty}^{+\infty} \theta\, f_\theta(\theta|\underline{x})\, d\theta = \frac{1}{f_{\underline{x}}(\underline{x})} \int_{-\infty}^{+\infty} \theta\, f_{\underline{x}}(\underline{x}|\theta)\, f_\theta(\theta)\, d\theta$$

تابع زیان دیگر:

$$L(e) = |e| \quad \rightarrow \quad L(\theta - \hat{\theta}) = |\theta - \hat{\theta}|^2$$

یعنی باید $mae = E(|\theta - \hat{\theta}|^2)$ مینیمم شود که منجر می‌شود به:

$$\hat{\theta}_{abs} = \text{median}\big(f_\theta(\theta|\underline{x})\big)$$

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:

$$\theta=\hat{\theta}_{MAP} \Leftrightarrow \forall\theta: f_\theta(\hat{\theta}_{MAP}|\underline{x}) \geq f_\theta(\theta|\underline{x})$$

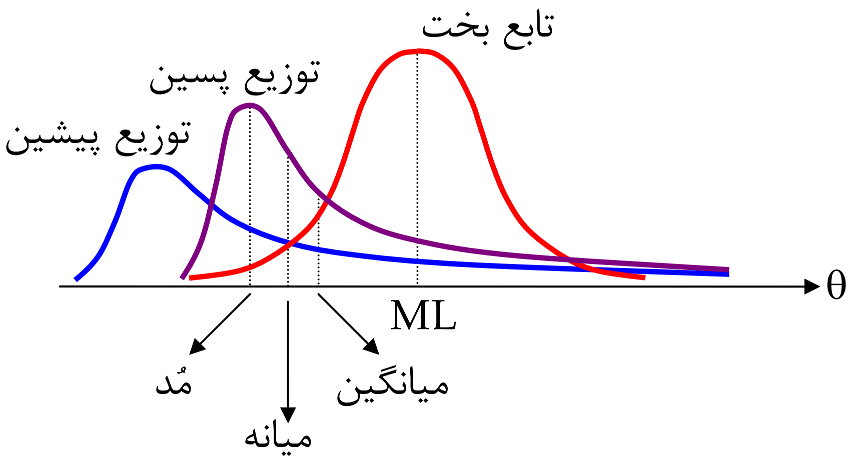
یا:

$$\theta=\hat{\theta}_{MAP} \Leftrightarrow \forall\theta: f_{\underline{x}}(\underline{x}|\hat{\theta}_{MAP})f_\theta(\hat{\theta}_{MAP}) \geq f_{\underline{x}}(\underline{x}|\theta)f_\theta(\theta)$$

(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.

$$\hat{p}_{ls}=?$$

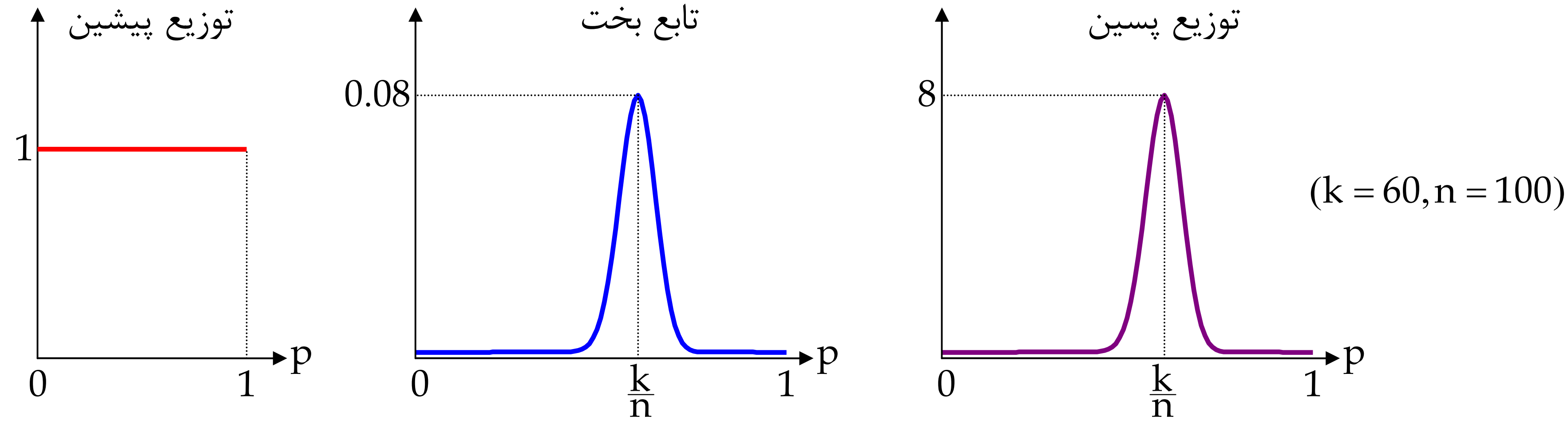
در آنجا دیدیم که:

$$\overbrace{f_{\mathbf{p}}(p|\mathbf{x}=k)}^{\text{پسین}}=\frac{\overbrace{P\{\mathbf{x}=k|\mathbf{p}=p\}}^{\text{تابع بخت}}\overbrace{f_{\mathbf{p}}(p)}^{\text{پیشین}}}{P\{\mathbf{x}=k\}}=\frac{P\{\mathbf{x}=k|\mathbf{p}=p\}f_{\mathbf{p}}(p)}{\int_0^1 P\{\mathbf{x}=k|\mathbf{p}=p\}f_{\mathbf{p}}(p)dp}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}f_{\mathbf{p}}(p)}{\int_0^1\binom{n}{k}p^k(1-p)^{n-k}f_{\mathbf{p}}(p)dp}$$

$$=\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\frac{1}{n+1}}=\frac{(n+1)!}{k!(n-k)!}p^k(1-p)^{n-k}:0\le p\le 1$$: 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.



$$\hat{p}_{ls} = E(\mathbf{p} \,|\, \mathbf{x} = k) = \int_0^1 p\, f_{\mathbf{p}}(p \,|\, \mathbf{x} = k)\, dp = \frac{\int_0^1 p^{k+1}(1-p)^{n-k} f_{\mathbf{p}}(p)\, dp}{\int_0^1 p^{k}(1-p)^{n-k} f_{\mathbf{p}}(p)\, dp}$$

$$= \frac{(n+1)!}{k!(n-k)!} \cdot \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k+1}{n+2}$$ برای $f_{\mathbf{p}}(p)$ یکنواخت :

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 $N(\eta,\sigma)$ از یک متغیر تصادفی با توزیع که η باشند و فرض کنیم که خود یک متغیر η 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.

$$f_\eta(\eta) = \frac{1}{\sqrt{2\pi\sigma_\eta^2}} e^{-\frac{\eta^2}{2\sigma_\eta^2}}$$

$$f_{\underline{\mathbf{x}}}(\underline{x} \mid \eta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\eta)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\sum_{i=1}^{n}(x_i-\eta)^2}{2\sigma^2}}$$

$$f_\eta(\eta \mid \underline{x}) = \frac{f_{\underline{\mathbf{x}}}(\underline{x} \mid \eta) f_\eta(\eta)}{f_{\underline{\mathbf{x}}}(\underline{x})} = \frac{f_{\underline{\mathbf{x}}}(\underline{x} \mid \eta) f_\eta(\eta)}{\int_{-\infty}^{+\infty} f_{\underline{\mathbf{x}}}(\underline{x} \mid \eta) f_\eta(\eta) d\eta}$$

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:

$$\hat{\eta}_{ls} = \hat{\eta}_{abs} = \hat{\eta}_{MAP} = \frac{n}{n + \frac{\sigma^2}{\sigma_\eta^2}} \overline{x}$$

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:

$$P\{-z_{1-\frac{\alpha}{2}} < \frac{\eta - \hat{\eta}}{\sigma_0} < z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$

$$\Rightarrow P\{\hat{\eta} - \sigma_0 z_{1-\frac{\alpha}{2}} < \eta < \hat{\eta} + \sigma_0 z_{1-\frac{\alpha}{2}}\} = 1 - \alpha$$