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 $\mathbf{x}$ یک متغیر تصادفی گسسته باشد که مقادیر x_i $(i = 0,1,2,\dots)$ را اختیار می‌کند، تابع احتمال شرطی $\mathbf{x}$ 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 انباشتۀ شرطی (CDF شرطی):

طبق تعریف، تابع توزیع انباشتۀ شرطی $\mathbf{x}$ به شرط واقعۀ M برابر است با:

$$F_{\mathbf{x}}(x|M) = P\{\mathbf{x} \le x|M\}$$

این تابع نیز تمام خواص CDF غیرشرطی را دارد، از جمله داریم:

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

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

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

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 $(i = 0,1,2,\dots,m)$ 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:

$$P(A) = \sum_i P(A \mid \mathbf{x} = x_i) P\{\mathbf{x} = x_i\} = \sum_i P(A \mid x_i) P_{\mathbf{x}}(x_i)$$

همچنین داریم:

$$P\{\mathbf{x} = x_k \mid A\} = \frac{P(A \mid \mathbf{x} = x_k) P\{\mathbf{x} = x_k\}}{P(A)}$$

پس:

۲. قضیهٔ بِیز:

$$P_{\mathbf{x}}(x_k \mid A) = \frac{P(A \mid x_k) P_{\mathbf{x}}(x_k)}{P(A)} = \frac{P(A \mid x_k) P_{\mathbf{x}}(x_k)}{\sum_i P(A \mid x_i) P_{\mathbf{x}}(x_i)}$$

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:

$$P(A)=\int_{-\infty}^{+\infty}P(A|x)f_{\mathbf{x}}(x)dx$$

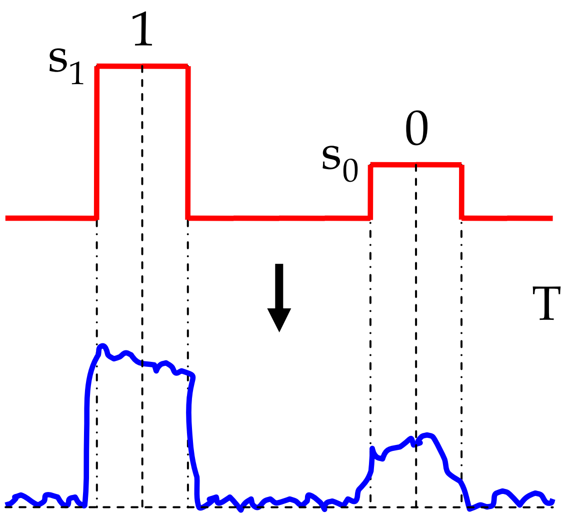
۲.قضیهٔ بیز:

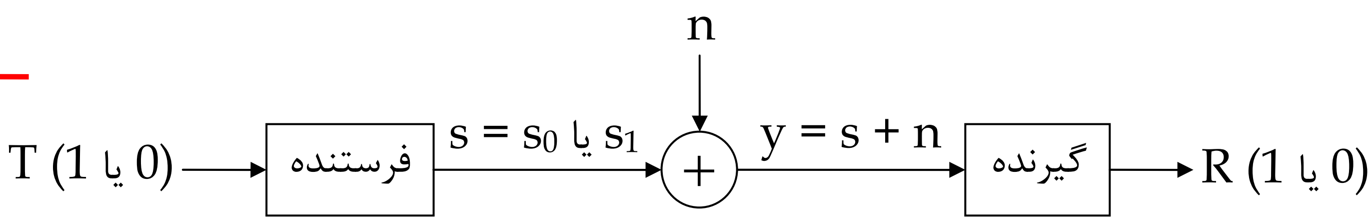
$$f_{\mathbf{x}}(x|A)=\frac{P(A|x)f_{\mathbf{x}}(x)}{P(A)}=\frac{P(A|x)f_{\mathbf{x}}(x)}{\int_{-\infty}^{+\infty}P(A|x)f_{\mathbf{x}}(x)dx}$$

(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.

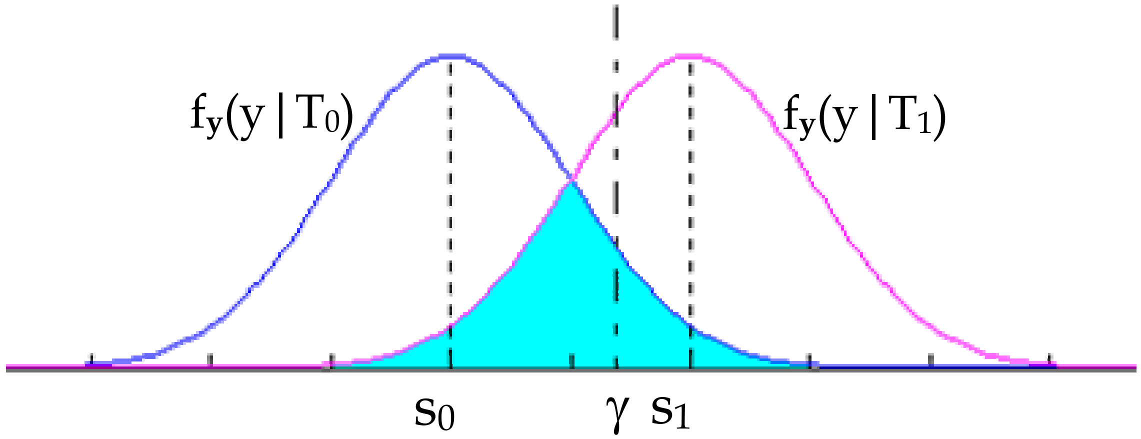




فرض کنید نویز دارای توزیع گوسی $N(0,\sigma)$ است. می‌توان نشان داد برای اینکه گیرنده از روی این داده‌های نویزی بتواند به طور بهینه (کمترین احتمال خطا) تصمیم بگیرد که 0 یا 1 ارسال شده بوده، باید ولتاژ دریافتی را با سطح آستانه‌ای مقایسه کند. اگر از این سطح آستانه بزرگتر بود، 1 و اگر کوچکتر بود، 0 فرض می‌شود. 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:

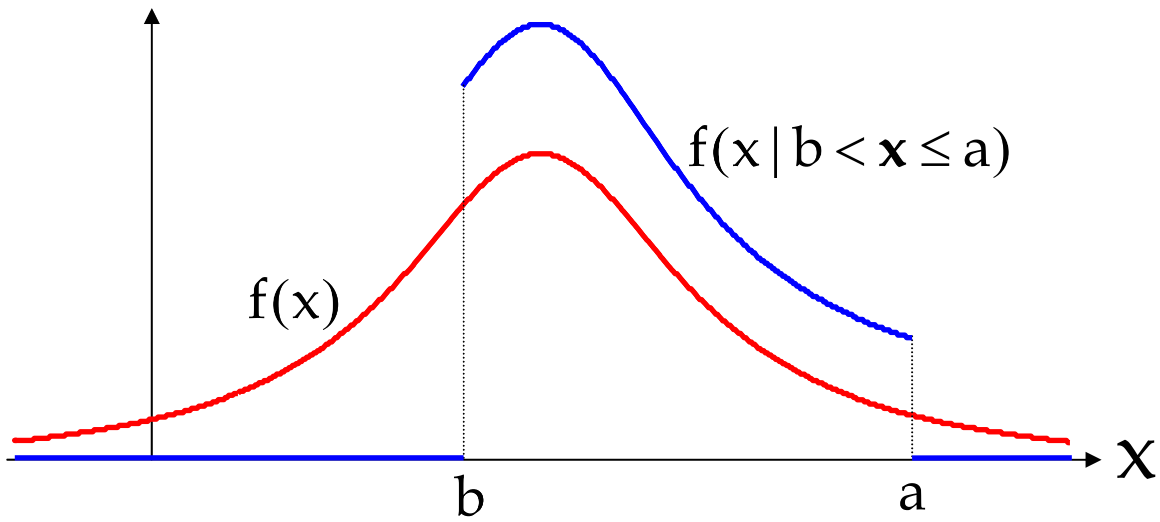
$$M = \{b < \mathbf{x} \le a\}$$

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:

$$f_{\mathbf{x}}(x \mid b < \mathbf{x} \le a) = \begin{cases} 0 & x \ge a \\ 0 & x < b \\ \dfrac{f_{\mathbf{x}}(x)}{F_{\mathbf{x}}(a) - F_{\mathbf{x}}(b)} = \dfrac{f_{\mathbf{x}}(x)}{\int_b^a f_{\mathbf{x}}(x)dx} & b \le x < a \end{cases}$$



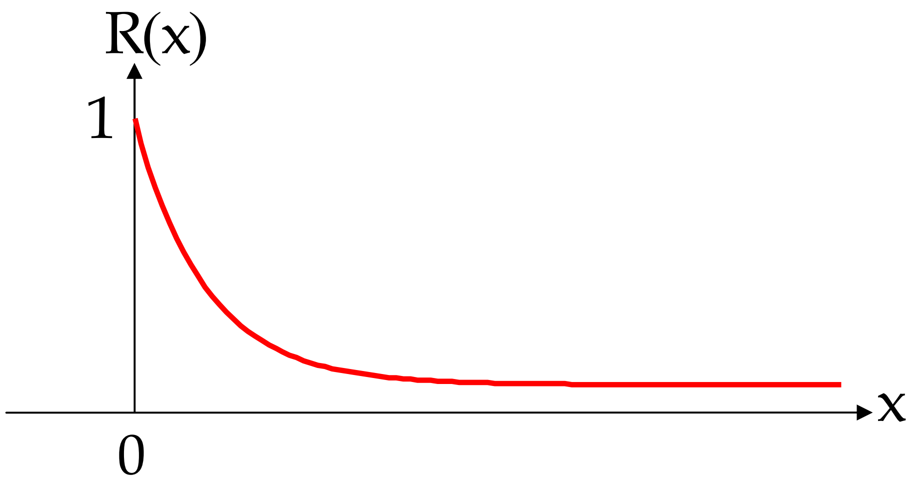
مثالی از کاربرد توزیع مشروط: قابلیت اعتماد

فرض کنید متغیر تصادفی **x** زمان خراب شدن سیستمی باشد که از $t=0$ شروع به کار کرده است (متغیر تصادفی **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 **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:

$$F(x|\mathbf{x}\ge t)=\frac{P\{\mathbf{x}\le x,\mathbf{x}\ge t\}}{P\{\mathbf{x}\ge t\}}=\begin{cases}0 & x<t\\ \dfrac{F(x)-F(t)}{1-F(t)} & x\ge t\end{cases}$$

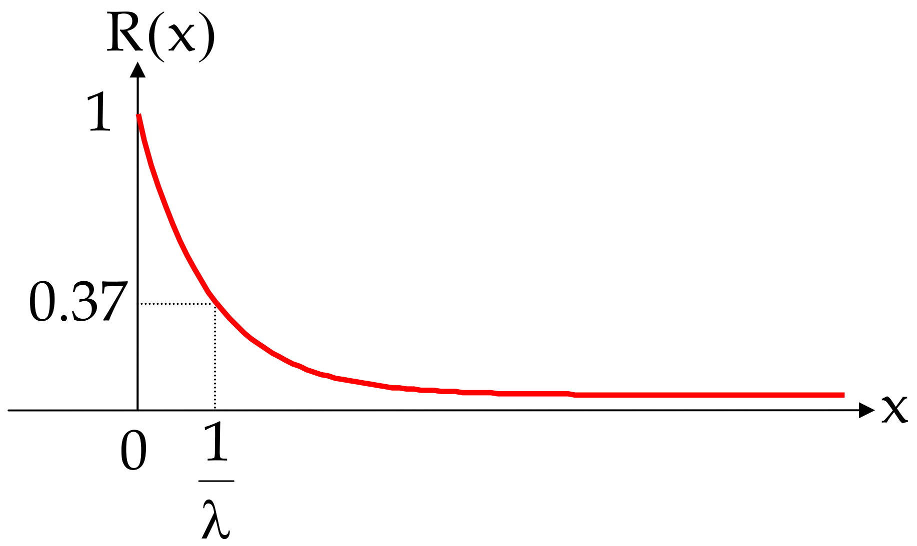
با مشتق‌گیری، $f(x|\mathbf{x}\ge t)$ به دست می‌آید. تابع فوق در $x=t$ پیوسته است، پس f شامل δ نخواهد شد:

$$f(x|\mathbf{x}\ge t)=\begin{cases}0 & x<t\\ \dfrac{f(x)}{1-F(t)} & x\ge t\end{cases}$$

$f(x|\mathbf{x}\ge t)dx$ 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:

$$-\int_0^x \beta(t)dt = \ln R(x)$$

$$\Rightarrow R(x) = e^{-\int_0^x \beta(t)dt} \ : x \ge 0$$

$$\Rightarrow F(x) = 1 - e^{-\int_0^x \beta(t)dt} \ : x \ge 0$$

$$\Rightarrow f(x) = \beta(x) e^{-\int_0^x \beta(t)dt} \ : x \ge 0$$

خواص $\beta(t)$:

$$1)\ \ \beta(t)=\frac{f(t)}{1-F(t)}\Rightarrow\beta(t)\geq 0$$

$$2)\ \ F(+\infty)=1\Rightarrow\int_0^{+\infty}\beta(t)dt\to+\infty$$

مثلاً اگر $\beta(t)=\lambda$، یعنی مقدار ثابت باشد، داریم:

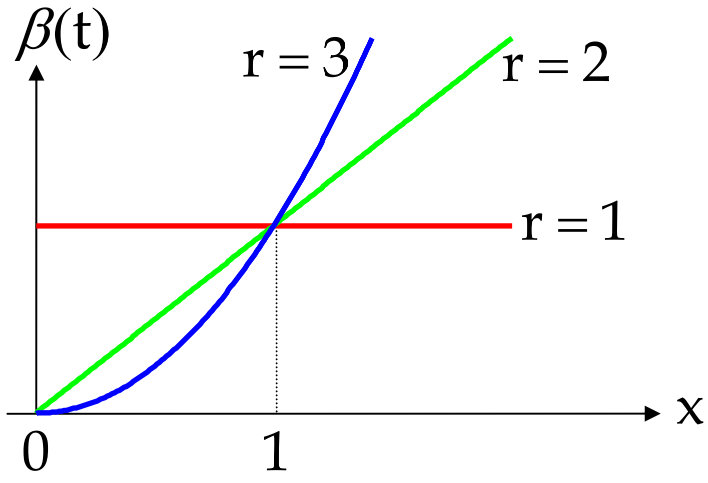
$$f(x)=\lambda e^{-\lambda x}u(x)$$

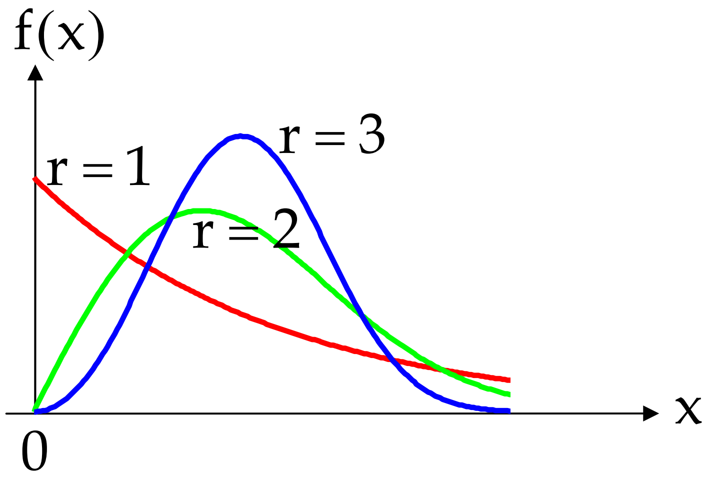
اگر $\beta(t)=\lambda t^{r-1}$ باشد، در بسیاری موارد تابع مناسبی است و داریم:

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

این توزیع را <u>توزیع ویبول (Weibull)</u> می‌گویند که برای $r=1$ همان توزیع نمایی و برای $r=2$ همان توزیع رایلی می‌شود.

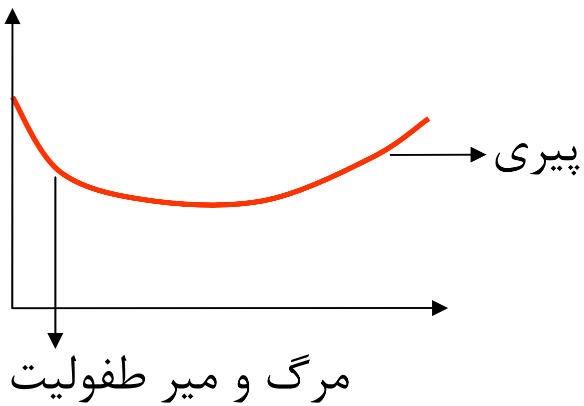
برای $r=1$، $\beta(t)$ ثابت است و برای rهای بزرگتر، فرسودگی را مدل می‌کند:





$$MTTF = E(\mathbf{x}) = (\frac{r}{\lambda})^{\frac{1}{r}}\,\Gamma(\frac{r+1}{r})$$

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.

$$P_{\mathbf{x}}(k \mid \mathbf{p}=p) = P\{\mathbf{x}=k \mid \mathbf{p}=p\} = \binom{n}{k} p^k (1-p)^{n-k} : k=0,1,\ldots,n$$

$$f_{\mathbf{p}}(p) = \begin{cases} 1 & 0 \le p < 1 \\ 0 & \text{otherwise} \end{cases}$$

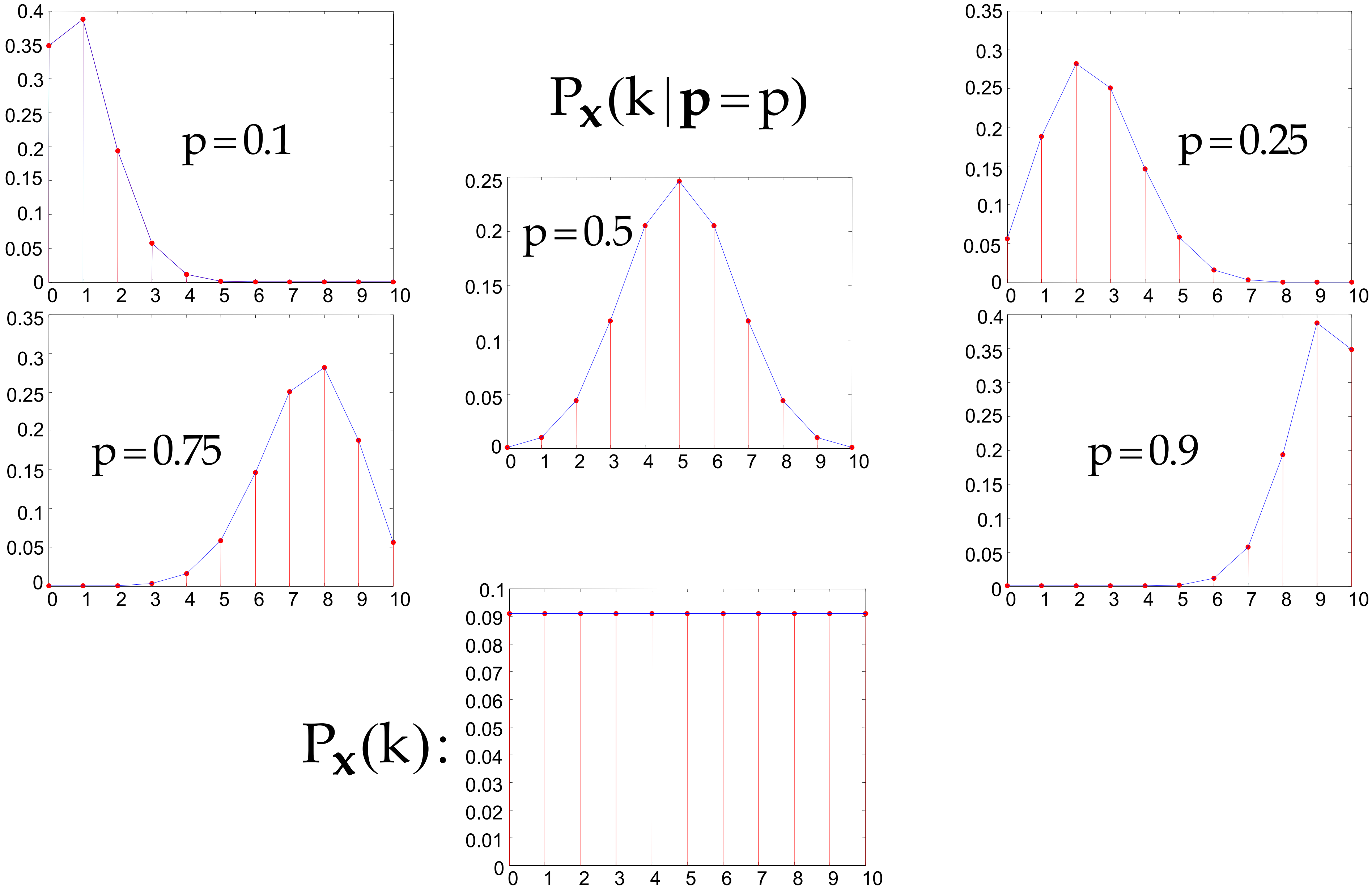
توابع $P_{\mathbf{x}}(k) = P\{\mathbf{x}=k\}$ و $f_{\mathbf{p}}(p \mid \mathbf{x}=k)$ را بیابید.

$$P_{\mathbf{x}}(k) = \int_{-\infty}^{+\infty} P_{\mathbf{x}}(k \mid \mathbf{p}=p) f_{\mathbf{p}}(p)\,dp : k=0,1,\ldots,n$$

$$= \int_0^1 \binom{n}{k} p^k (1-p)^{n-k}\,dp \qquad \rightarrow \beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$= \binom{n}{k} \beta(k+1, n-k+1) = \frac{n!}{k!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!}$$

$$= \frac{1}{n+1} : k=0,1,\ldots,n$$



$P_{\mathbf{x}}(k|\mathbf{p}=p)$
$p=0.1$
$p=0.5$
$p=0.25$
$p=0.75$
$p=0.9$
$P_{\mathbf{x}}(k):$

$$f_{\mathbf{p}}(p|\mathbf{x}=k)=\frac{P\{\mathbf{x}=k|\mathbf{p}=p\}f_{\mathbf{p}}(p)}{P\{\mathbf{x}=k\}}=\frac{P_{\mathbf{x}}(k|p)f_{\mathbf{p}}(p)}{P_{\mathbf{x}}(k)}=\frac{\binom{n}{k}p^k(1-p)^{n-k}}{\frac{1}{n+1}}:0\le p<1$$

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

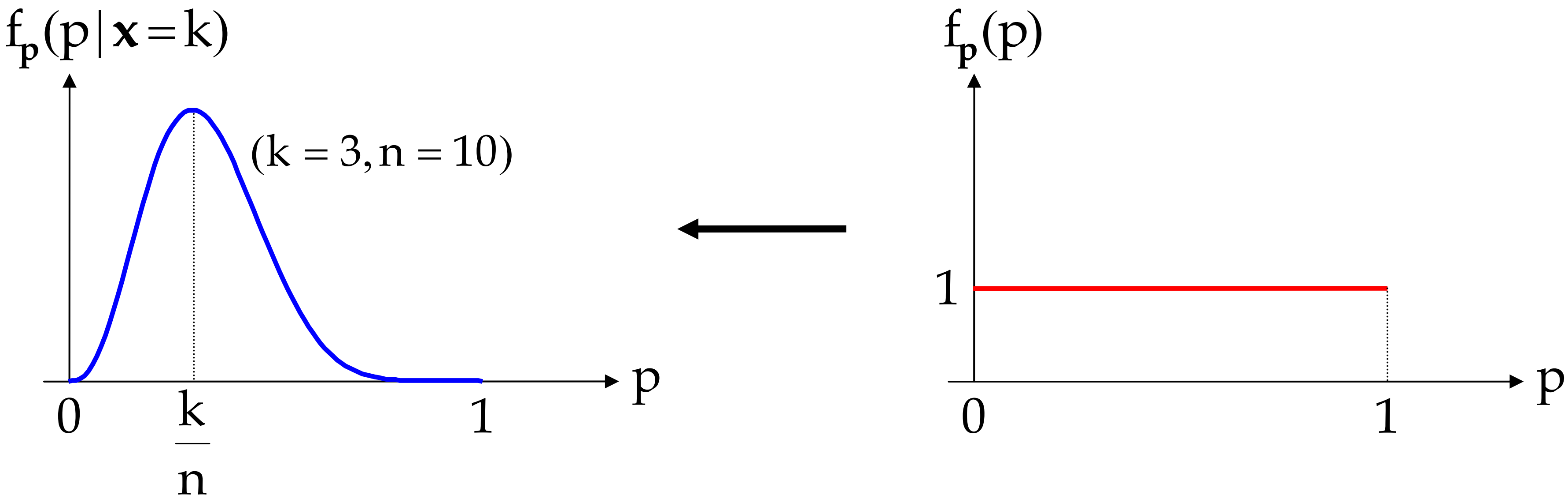
$$\Rightarrow(\mathbf{p}|\mathbf{x}=k)\sim \text{Beta}(k+1,n-k+1)$$

یادآوری: **توزیع بتا:**

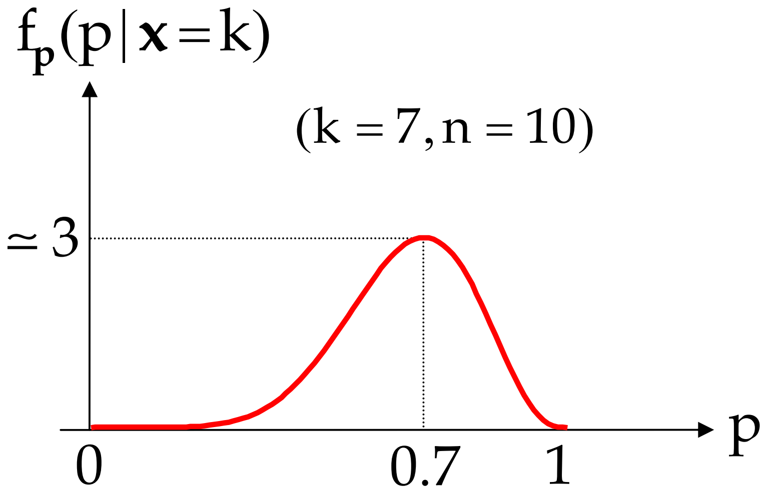
$$f(x)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}:0<x<1$$

یعنی $\mathbf{p}$ که توزیع یکنواخت داشت، با مشاهدهٔ مقدار $\mathbf{x}$ (مثلاً تعداد شیرهای آمده در nبار پرتاب سکه) توزیع بتا پیدا می‌کند.

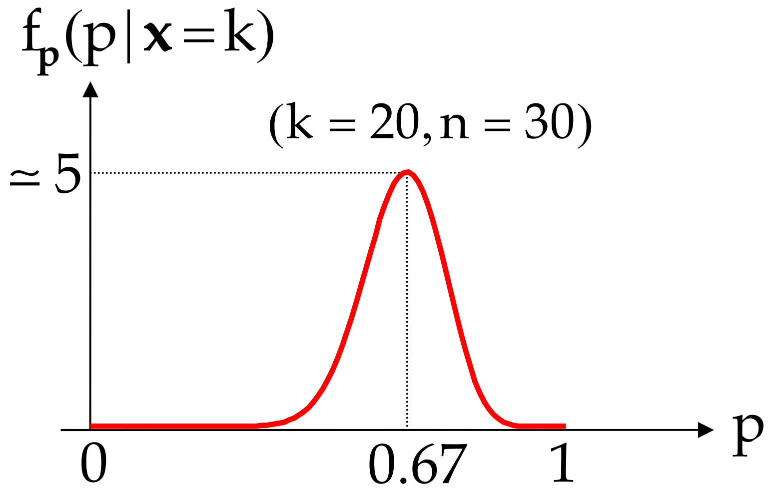
ماکزیمم توزیع بتا در $\frac{a-1}{a+b-2}$ یعنی $\frac{k}{n}$ اتفاق می‌افتد. هر چه n بزرگتر باشد، نمودار تیزتر می‌شود.



کتاب در فصل ۶ (Section 6.1) در مورد تخمین بیز نیز بحث کرده است که ما بعداً مفصل دربارهٔ آن صحبت خواهیم کرد.



$f_{\mathbf{p}}(p|\mathbf{x}=k)$
$(k=7, n=10)$
$\simeq 3$
0
0.7
1
p



$f_{\mathbf{p}}(p|\mathbf{x}=k)$
$(k=20, n=30)$
$\simeq 5$
0
0.67
1
p

مثال ۲: مقادیر $P\{\mathbf{x}>1\,|\,\mathbf{y}=y\}$، $E(\mathbf{x}\,|\,\mathbf{y}=y)$ و $\mathrm{var}(\mathbf{x}\,|\,\mathbf{y}=y)$ 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 :

$$E(\mathbf{x}\,|\,y)=y$$

$$\mathrm{var}(\mathbf{x}\,|\,y)=y^2$$

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:

$$f_{\mathbf{z}}(z)=\int_{-\infty}^{+\infty} f_{\mathbf{z}}(z\,|\,x)f_{\mathbf{x}}(x)dx$$

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:

$$f_{\mathbf{z}}(z\,|\,x)=\frac{f_{\mathbf{y}}(y_0\,|\,x)}{\left|\dfrac{\partial g(x,y_0)}{\partial y}\right|}$$

(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 $var(\mathbf{x} \,|\, \mathbf{x} > 0)$ را حساب کنید.

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

$$f_{\mathbf{x}}(x \,|\, \mathbf{x} > 0) = \begin{cases} \dfrac{f_{\mathbf{x}}(x)}{1 - F(0)} = 2f_{\mathbf{x}}(x) & x > 0 \\ 0 & x < 0 \end{cases}$$

$$E(\mathbf{x} \,|\, \mathbf{x} > 0) = \int_{-\infty}^{+\infty} x\, f_{\mathbf{x}}(x \,|\, \mathbf{x} > 0)\, dx = \int_{0}^{+\infty} \frac{2x}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}}\sigma$$

$$E(\mathbf{x}^2 \,|\, \mathbf{x} > 0) = \int_{0}^{+\infty} \frac{2x^2}{\sqrt{2\pi}\,\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sigma^2$$

$$var(\mathbf{x} \,|\, \mathbf{x} > 0) = \sigma^2 (1 - \frac{2}{\pi})$$

(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:

$$\forall x: E(\mathbf{y}|x)=\int_{-\infty}^{+\infty}y\,f_{\mathbf{y}}(y|x)\,dy=\int_{-\infty}^{+\infty}y\,f_{\mathbf{y}}(y)\,dy=\eta_{\mathbf{y}}$$

$$\Rightarrow \forall x: \Phi(x)=c \Rightarrow \mathbf{\Phi(x)}=c$$

$$\Rightarrow E(\mathbf{y}|\mathbf{x})=\eta_{\mathbf{y}}$$

مثال ۱: در مثال ۲، صفحهٔ ۲۹ دیدیم که:

$$E(\mathbf{x} \mid y) = y$$

پس:

$$E(\mathbf{x} \mid \mathbf{y}) = \mathbf{y}$$
$$\Rightarrow E\left[E(\mathbf{x} \mid \mathbf{y})\right] = E(\mathbf{y})$$

مثال ۲: در مثال ۳، صفحهٔ ۳۰ دیدیم که:

$$E(\mathbf{y} \mid x) = \eta_{\mathbf{y}} + r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(x - \eta_{\mathbf{x}})$$

پس:

$$E(\mathbf{y} \mid \mathbf{x}) = \eta_{\mathbf{y}} + r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x} - \eta_{\mathbf{x}})$$
$$\Rightarrow E\left[E(\mathbf{y} \mid \mathbf{x})\right] = \eta_{\mathbf{y}}$$

در همین مثال اگر $r = 0$ باشد (یعنی $\mathbf{x}$ و $\mathbf{y}$ مستقل باشند)، خواهیم داشت:

$$\forall x : E(\mathbf{y} \mid x) = \eta_{\mathbf{y}}$$

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:

$$E(\mathbf{N}^2) = E\left[E(\mathbf{N}^2 \mid \mathbf{z})\right]$$

$$E(\mathbf{N}^2 \mid \mathbf{z}=1) = 0$$

$$E(\mathbf{N}^2 \mid \mathbf{z}=0) = E\left[(1+\mathbf{N})^2\right]$$

$$E(\mathbf{N}^2) = E(\mathbf{N}^2 \mid \mathbf{z}=1)P(\mathbf{z}=1) + E(\mathbf{N}^2 \mid \mathbf{z}=0)P(\mathbf{z}=0) = 0 + q\,E\left[(1+\mathbf{N})^2\right]$$

$$\Rightarrow E(\mathbf{N}^2) = \frac{pq+2q^2}{p^2}$$

$$\Rightarrow \mathrm{var}(\mathbf{N}) = E(\mathbf{N}^2) - \left(E(\mathbf{N})\right)^2 = \frac{q}{p^2}$$

(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:

$$E(g(\mathbf{x},\mathbf{y})|\mathbf{x}=x)=E(g(x,\mathbf{y})|\mathbf{x}=x)\overset{\substack{\boxed{\text{بنا به تعریف } E(g(\mathbf{y})|x)}\\ \uparrow}}{=}\int_{-\infty}^{+\infty} g(x,y)f_{\mathbf{y}}(y|\mathbf{x}=x)\,dy$$

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:

$$E\left[E(g(\mathbf{x},\mathbf{y})|\mathbf{x})\right]=\int_{-\infty}^{+\infty}\left(\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{y}}(y|x)dy\right)f_{\mathbf{x}}(x)dx$$

$$=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x,y)f_{\mathbf{xy}}(x,y)dxdy=E(g(\mathbf{x},\mathbf{y}))$$

حالت خاص:

$$g(\mathbf{x},\mathbf{y})=g_1(\mathbf{x})g_2(\mathbf{y})$$

در این حالت داریم:

بنا بر حالت خاص در ویژگی ۱

$$E(g_1(\mathbf{x})g_2(\mathbf{y}))=E\left[E(g_1(\mathbf{x})g_2(\mathbf{y})\,|\,\mathbf{x})\right]=E\left[g_1(\mathbf{x})E(g_2(\mathbf{y})\,|\,\mathbf{x})\right]$$

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 **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):

منحنی رگرسیون : $\hat{y}_{ls} = \Phi(x)$

$$\hat{\mathbf{y}}_{ls} = \Phi(\mathbf{x})$$

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:

$$\hat{\mathbf{y}}_{ls} = E(\mathbf{y} \mid \mathbf{x})$$

<u>اصل تعامد:</u>

خطا در تخمین ls بر هر تابعی از داده عمود است، یعنی:

$$E\left[(\mathbf{y}-\hat{\mathbf{y}}_{ls})g(\mathbf{x})\right]=E\left[\left(\mathbf{y}-E(\mathbf{y}\,|\,\mathbf{x})\right)g(\mathbf{x})\right]=0$$

این قضیه استفادهٔ 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 <u>بدون بایاس</u> (Unbiased) (ناریب) گویند، هرگاه:

$$E(\hat{\mathbf{y}})=E(\mathbf{y})$$

یعنی E(error) صفر باشد.

قضیه: تخمین ls ناااریب است.

زیرا:
$$E(\hat{\mathbf{y}}_{ls}) = E\left[E(\mathbf{y}\,|\,\mathbf{x})\right] = E(\mathbf{y})$$

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:

$$\sigma_{\mathbf{y}}^2 - \sigma_{\hat{\mathbf{y}}_{\mathrm{ls}}}^2 < \sigma_{\mathbf{y}}^2$$

نتیجهٔ ۲ (قضیهٔ رائو- بِلَکوِل): $\sigma_{\hat{\mathbf{y}}_{\mathrm{ls}}}^2 \le \sigma_{\mathbf{y}}^2$، زیرا $\mathrm{mmse} \ge 0$.

مثال: اگر $\mathbf{x}$ و $\mathbf{y}$ مشترکاً گوسی باشند، تخمین بهینهٔ $\mathbf{y}$ بر اساس مشاهدهٔ $\mathbf{x}$ عبارت است از:

$$\hat{\mathbf{y}}_{\mathrm{ls}} = \mathrm{E}(\mathbf{y}\,|\,\mathbf{x}) = \eta_{\mathbf{y}} + \mathrm{r}\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x} - \eta_{\mathbf{x}}) = \eta_{\mathbf{y}} + \frac{\mu_{\mathbf{xy}}}{\sigma_{\mathbf{x}}^2}(\mathbf{x} - \eta_{\mathbf{x}})$$

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).

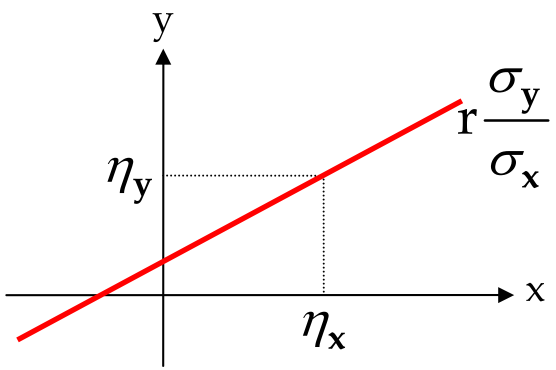
تخمین خطی حداقل مربعات (lls):

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:

$$mse=E\left[\left(\mathbf{y}-(a+b\mathbf{x})\right)^2\right]=E(\mathbf{y}^2)+a^2+b^2E(\mathbf{x}^2)-2aE(\mathbf{y})-2bE(\mathbf{xy})+2abE(\mathbf{x})$$

$$\left.\begin{array}{l}\dfrac{\partial mse}{\partial a}=0\Rightarrow 2a-2E(\mathbf{y})+2bE(\mathbf{x})=0\\ \dfrac{\partial mse}{\partial b}=0\Rightarrow 2bE(\mathbf{x}^2)-2E(\mathbf{xy})+2aE(\mathbf{x})=0\end{array}\right\}\Rightarrow\begin{cases}a=\eta_y+r\dfrac{\sigma_y}{\sigma_x}\eta_x\\ b=r\dfrac{\sigma_y}{\sigma_x}\end{cases}$$

$$\Rightarrow \hat{\mathbf{y}}_{lls}=\eta_y+r\frac{\sigma_y}{\sigma_x}(\mathbf{x}-\eta_x)$$



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:

$$E\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{lls}})\mathbf{x}\right]=0$$

اثبات:

$$E(\hat{\mathbf{y}}_{\mathrm{lls}}\mathbf{x})=E\left[\eta_y\mathbf{x}+r\frac{\sigma_y}{\sigma_x}(\mathbf{x}-\eta_x)\mathbf{x}\right]=\eta_y\eta_x+r\frac{\sigma_y}{\sigma_x}\underbrace{\left(E(\mathbf{x}^2)-\eta_x^2\right)}_{\sigma_x^2}=\sigma_x^2\eta_y\eta_x+\underbrace{r\sigma_x\sigma_y}_{\mu_{xy}}$$

$$=\eta_y\eta_x+E(\mathbf{xy})-E(\mathbf{x})E(\mathbf{y})=E(\mathbf{xy})$$

$$\Rightarrow E\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{lls}})\mathbf{x}\right]=0$$

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 نااریب است.

زیرا:

$$E(\hat{\mathbf{y}}_{\mathrm{lls}}) = \eta_{\mathrm{y}} + r\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathrm{x}}}E(\mathbf{x}-\eta_{\mathbf{x}}) = E(\mathbf{y})$$

قضیه: در تخمین lls داریم:

$$\mathrm{mse} = \sigma_{\mathbf{y}}^2(1-r^2)$$

اثبات:

$$E\left[(\mathbf{y}-\hat{\mathbf{y}}_{\mathrm{lls}})^2\right] = E\left[\left(\mathbf{y}-\eta_{\mathrm{y}} - r\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathbf{x}}}(\mathbf{x}-\eta_{\mathbf{x}})\right)^2\right]$$

$$= E\left[(\mathbf{y}-\eta_{\mathrm{y}})^2\right] + r^2\frac{\sigma_{\mathbf{y}}^2}{\sigma_{\mathbf{x}}^2}E\left[(\mathbf{x}-\eta_{\mathbf{x}})^2\right] - 2r\frac{\sigma_{\mathrm{y}}}{\sigma_{\mathbf{x}}}E\left[(\mathbf{x}-\eta_{\mathbf{x}})(\mathbf{y}-\eta_{\mathrm{y}})\right]$$

$$= \sigma_{\mathbf{y}}^2 + r^2\sigma_{\mathbf{y}}^2 - 2r\frac{\sigma_{\mathbf{y}}}{\sigma_{\mathbf{x}}}\cdot r\sigma_{\mathbf{x}}\sigma_{\mathbf{y}} = \sigma_{\mathbf{y}}^2 - r^2\sigma_{\mathbf{y}}^2 = \sigma_{\mathbf{y}}^2(1-r^2)$$

(در اینجا هم ملاحظه می‌کنید که mse نسبت به خطای تخمین بدون مشاهده ($\sigma_{\mathbf{y}}^2$) کمتر است.)