类似可定义在 $X=x$ 条件下随机变量 Y 的条件概率密度和分布函数分别为

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}\qquad F_{Y|X}(y|x)=\int_{-\infty}^{y}\frac{f(x,v)}{f_X(x)}dv.$$

下面给出条件概率的一种解释, 这里以 $f_{X|Y}=\frac{f(x,y)}{f_Y(y)}$ 为例, 首先分布函数有

$$F_{X|Y}(x|y)=\lim_{\epsilon\to 0^+}P\{X\leqslant x|y\leqslant Y\leqslant y+\epsilon\}=\lim_{\epsilon\to 0^+}\frac{P\{X\leqslant x,y\leqslant Y\leqslant y+\epsilon\}}{P\{y\leqslant Y\leqslant y+\epsilon\}}.$$

根据积分中值定理有

$$\frac{P\{X\leqslant x,y\leqslant Y\leqslant y+\epsilon\}}{P\{y\leqslant Y\leqslant y+\epsilon\}}=\frac{\int_{-\infty}^{x}\int_{y}^{y+\epsilon}f(u,v)dudv}{\int_{y}^{y+\epsilon}f_Y(u)dv}=\frac{\epsilon\int_{-\infty}^{x}f(u,y+\theta_1\epsilon)du}{\epsilon f_Y(y+\theta_2\epsilon)}$$

其中$\theta_1,\theta_2\in(0,1)$. 当$\epsilon\to 0^+$时, 有

$$F_{X|Y}(x|y)=\lim_{\epsilon\to 0^+}P\{X\leqslant x|y\leqslant Y\leqslant y+\epsilon\}=\frac{\int_{-\infty}^{x}f(u,y)du}{f_Y(y)}=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du,$$

由此可得条件概率密度 $f_{X|Y}(x|y)=f(x,y)/f_Y(y)$. 下面给出条件概率的性质:

引理 5.3 (乘法公式) 对于随机变量 X 和 Y, 有

$$\begin{aligned}f(x,y)&=f_X(x)f_{Y|X}(y|x),\quad (f_X(x)>0),\\ f(x,y)&=f_Y(y)f_{X|Y}(x|y),\quad (f_Y(y)>0).\end{aligned}$$

若随机变量 X 和 Y 相互独立, 则有联合概率密度 $f(x,y)=f_X(x)f_Y(y)$, 由此可得

引理 5.4 如果随机变量 X 和 Y 相互独立, 则有

$$f_{Y|X}(y|x)=f_Y(y)\quad\text{和}\quad f_{X|Y}(x|y)=f_X(x).$$

由此可根据条件概率来可判别随机变量 (X,Y) 的独立性. 下面看几个条件概率的例子:

例 5.23 设二维随机变量 (X,Y) 的概率密度

$$f(x,y)=\begin{cases}\frac{e^{-x/y}e^{-y}}{y} & 0<x<+\infty,0<y<+\infty\\ 0 & \text{其它},\end{cases}$$

求 $P(X>1|Y=y)$.

解 首先求解随机变量 Y 的边缘分布为

$$f_Y(y)=\int_0^{+\infty}\frac{e^{-x/y}e^{-y}}{y}dx=e^{-y}[-e^{-\frac{x}{y}}]_0^{+\infty}=e^{-y}\qquad (y>0).$$

进而得到在 $Y=y$ 条件下 X 的条件概率密度为

$$f_{X|Y}(x|y)=e^{-x/y}/y.$$

最后求解得到

$$P(X>1|Y=y)=\int_1^{\infty}\frac{e^{-x/y}}{y}dx=-e^{-x/y}|_1^{\infty}=e^{-\frac{1}{y}}.$$

例 5.24 已知随机变量 $X\sim U(0,1)$, 当观察到 $X=x$ 的条件下, 随机变量 $Y\sim U(x,1)$. 求 Y 的概率密度.

解 根据题意可知 $X\sim U(0,1)$, 在随机变量 $X=x$ 的条件下 $Y\sim U(x,1)$, 即 $f_{Y|X}(y|x)=1/(1-x)$. 根据条件概率乘积公式有

$$f(x,y)=f_X(x)f_{Y|X}(y|x)=\begin{cases}\frac{1}{1-x} & 0<x<y<1,\\ 0 & \text{其它}.\end{cases}$$

根据联合分布求解随机变量 Y 的边缘分布

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_0^y\frac{1}{1-x}dx=-\ln(1-y) & y>0,\\ 0 & \text{其它}.\end{cases}$$

例 5.25 设随机变量 (X,Y) 的概率密度为

$$f(x,y)=\begin{cases}e^{-y} & y>x>0\\ 0 & \text{其它}\end{cases}$$

求条件概率密度 $f_{X|Y}(x|y)$.

定理 5.23 多维正太分布的条件分布是正太分布.

证明 为简单起见仅给出二维正太分布的详细证明. 设随机变量 $(X,Y)\sim\mathcal{N}(\mu,\Sigma)$, 其中

$$\mu=\begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}\quad \text{和}\quad \Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}.$$

下面证明在 $Y=y$ 的条件下随机变量 X 服从正态分布. 首先给出二维正太分布的联合分布

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}]}.$$

以及随机变量 Y 的边缘分布 $Y\sim\mathcal{N}(\mu_2,\sigma_2^2)$. 于是得到条件概率

$$\begin{aligned}
f_{X|Y}(x|y) &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}+\rho^2\frac{(y-\mu_2)^2}{\sigma_2^2}]}\\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{1}{2(1-\rho^2)}[\frac{x-\mu_1}{\sigma_1}+\frac{\rho(y-\mu_2)}{\sigma_2}]^2}\\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}}e^{-\frac{1}{2\sigma_1^2(1-\rho^2)}[x-\mu_1+\sigma_1^2\rho^2(y-\mu_2)/\sigma_2^2]^2}
\end{aligned}$$

由此可知在 $Y=y$ 的条件下 $X\sim\mathcal{N}(\mu_1-\sigma_1^2\rho(y-\mu_2)/\sigma_2^2,\sigma_1^2(1-\rho^2))$.

5.6.1 条件期望

定义 5.17 对二维离散随机变量 (X,Y), 在 $Y=y$ 条件下随机变量 X 的期望为

$$E[X|Y=y]=\sum_{i=1}^{\infty}x_iP(X=x_i|Y=y);$$

对二维连续随机变量 (X,Y), 在 $Y=y$ 条件下随机变量 X 的期望为

$$E[X|Y=y]=\int_{-\infty}^{+\infty}xf(x|y)dx.$$

注意 $E[X|Y=y]$ 是 y 的函数, 对条件期望有如下重要性质:

定理 5.24 对离散随机变量 $X_1,X_2,\cdots,X_n$ 及常数 $c_1,c_2,\cdots,c_n$ 有

$$E\left[\sum_{i=1}^{n}c_iX_i|Y=y\right]=\sum_{i=1}^{n}c_iE[X_i|Y=y].$$

定理 5.25 (全期望公式, law of total expectation) 对随机变量 X 和事件 A 有

$$E[X]=E[X|A]P(A)+E[X|\bar{A}](1-P(A))$$

其中事件 $\bar{A}$ 为事件 A 的补.

全期望公式对应于全概率公式的期望版本, 在很多应用中有重要的性质。

证明 此定理对离散和连续随机变量都成立, 为证明简单起见, 这里给出离散情况下的详细证明. 根据概率的性质有

$$\begin{aligned}
E[X] &= \sum_i x_i P(X=x_i) = \sum_i x_i [P(X=x_i, A) + P(X=x_i, \bar{A})] \\
&= \sum_i x_i P(X=x_i|A)P(A) + \sum_i x_i P(X=x_i|\bar{A})P(\bar{A}) \\
&= P(A)\sum_i x_i P(X=x_i|A) + P(\bar{A})\sum_i x_i P(X=x_i|\bar{A}) \\
&= P(A)E[X|A] + P(\bar{A})E[X|\bar{A}].
\end{aligned}$$

该定理有一个关于随机变量的定理:

定理 5.26 对二维随机变量 (X, Y) 有

$$E[X] = E_Y[E(X|Y)].$$

特别地, 对二维离散随机变量有

$$E[X] = E[E(X|Y)] = \sum_{Y=y_j} P[Y=y_j]E[X|Y=y_j].$$

证明 利用全概率公式有

$$\begin{aligned}
E[X] &= \sum_i x_i P(X=x_i) = \sum_i \sum_j x_i P(X=x_i, Y=y_j) \\
&= \sum_i \sum_j x_i P(X=x_i, Y=y_j) \\
&= \sum_j P(Y=y_j) \sum_i x_i \frac{P(X=x_i, Y=y_j)}{P(Y=y_j)} \\
&= \sum_j P(Y=y_j) \sum_i x_i P(X=x_i|Y=y_j) \\
&= \sum_j P(Y=y_j)E[X|Y=y_j] = E_Y[E[X|Y]].
\end{aligned}$$

待加入连续随机变量的证明.