称随机变量 X 的协方差矩阵为

$$Cov(X)=\Sigma=\begin{pmatrix} Cov(X_1,X_1) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & \cdots & Cov(X_2,X_n) \\ \vdots & & \vdots \\ Cov(X_n,X_1) & \cdots & Cov(X_n,X_n) \end{pmatrix}.$$

定理5.22. 随机变量 X 的协方差矩阵是对称半正定的矩阵.

Proof. 证明根据函数的

$$\begin{aligned} f(t_1,t_2,\dots,t_n) &= (t_1,t_2,\dots,t_n)\begin{pmatrix} Cov(X_1,X_1) & \cdots & Cov(X_1,X_n) \\ Cov(X_2,X_1) & \cdots & Cov(X_2,X_n) \\ \vdots & & \vdots \\ Cov(X_n,X_1) & \cdots & Cov(X_n,X_n) \end{pmatrix}(t_1,t_2,\dots,t_n)^\top \\ &= (t_1(X_1-E[X_1])+t_2(X_2-E[X_2])+\cdots+t_n(X_n-E[X_n]))^2 \geq 0. \end{aligned}$$

□

定理5.23. 设多维正态分布 $X=(X_1,X_2,\dots,X_n)^\top \sim N(\mu,\Sigma)$, 则有

$$\mu=(E[X_1],E[X_2],\dots,E[X_n])^\top \quad \text{和} \quad \Sigma=[Cov(X_i,X_j)]_{n\times n}.$$

对多维随机变量 $X=(X_1,X_2,\cdots,X_n)\ \sim\mathcal{N}(\mu,\Sigma)$, 有

- 每个变量X_i的边缘分布是正态分布;
- $X_1,X_2,\cdots,X_n$ 相互独立 $\Longleftrightarrow$ $X_1,X_2,\cdots,X_n$ 相互不相关;
- $X\sim\mathcal{N}(\mu,\Sigma)\Longleftrightarrow\sum_{i=1}^n a_iX_i$ 是正态分布(对任意非全为0常数$a_1,a_2,\dots,a_n$).

5.6 条件分布与条件期望

前面学过随机事件的条件概率, 即在事件 B 发生的条件下事件 A 发生的条件概率

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

相关概念可推广到随机变量: 给定随机变量 Y 取值条件下求随机变量 X 的概率分布, 即条件分布.

首先考虑离散型随机变量:

定义5.15. 设二维离散型随机变量 (X,Y) 的分布列为 $\{p_{ij}\}$, 若 Y 的边缘分布 $P(Y=y_j)=p_{\cdot j}>0$, 称

$$P(X=x_i|Y=y_j)=\frac{P(X=x_i,Y=y_j)}{P(Y=y_j)}=\frac{p_{ij}}{p_{\cdot j}} \quad (i=1,2,\cdots)$$

在 $Y=y_j$ 条件下随机变量 X 的条件分布列.

类似可定义在 $X=x_i$ 条件下随机变量 Y 的条件分布列. 条件分布是一种概率分布, 具有分布的性质. 例如, 非负性 $P(X=x_i|Y=y_j)\geq 0$, 规范性 $\sum_{i=1}^{\infty}P(X=x_i|Y=y_j)=1$ 等性质.

例5.16. 一个选手击中目标的概率为 p, 射击进行到击中两次目标为止, 用 X 表示首次击中目标所进行的射击次数, 用 Y 表示第二次射中目标所进行的射击次数, 求 X 和 Y 的联合分布和条件分布.

解. 随机变量 $X=m$ 表示首次击中目标射击了 m 次, $Y=n$ 表示第二次次击中目标射击了 n 次, 则 X 和 Y 的联合分布列为:

$$P\{X=m,Y=n\}=f(x,y)=\begin{cases}p^2(1-p)^{n-2} & 1\leq m<n<\infty\\ 0 & \text{其它}\end{cases}$$

由此可得 X 的边缘分布列为

$$P\{X=m\}=\sum_{n=m+1}^{\infty}P\{X=m,Y=n\}=\sum_{n=m+1}^{\infty}p^2(1-p)^{n-2}=p^2\frac{(1-p)^{m-1}}{1-(1-p)}=p(1-p)^{m-1}.$$

同理得到随机变量 Y 的边缘分布列为

$$P\{Y=n\}=\sum_{m=1}^{n-1}P\{X=m,Y=n\}=\sum_{m=1}^{n-1}p^2(1-p)^{n-2}=(n-1)p^2(1-p)^{n-2}(n=2,3,\cdots).$$

因此, 当 $n=2,3,\cdots$ 时, 随机变量 X 在 $Y=n$ 条件下的分布列为:

$$P\{X=m|Y=n\}=\frac{P\{X=m,Y=n\}}{P\{Y=n\}}=\frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}}=\frac{1}{n-1}\quad m=1,2,\cdots,n-1.$$

当 $m=1,2,3,\cdots$ 时, 随机变量 Y 在 $X=m$ 条件下的分布列为:

$$P\{Y=n|X=m\}=\frac{P\{X=m,Y=n\}}{P\{X=m\}}=\frac{p^2(1-p)^{n-2}}{p(1-p)^{m-1}}=p(1-p)^{n-m-1}\quad n=m+1,m+2,\cdots$$

□

对于连续型随机变量 (X,Y), 对任意 $x,y\in(-\infty,+\infty)$, 有 $P(X=x)=0$ 和 $P(Y=y)=0$ 成立, 因此不能利用离散随机变量的条件概率推导连续随机变量的条件分布. 下面给出条件概率的定义:

定义5.16. 设连续随机变量 (X,Y) 的联合概率密度为 $f(x,y)$, 以及 Y 的边缘概率密度为 $f_Y(y)$, 对任意 $f_Y(y)>0$, 称

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}$$

在 $Y=y$ 条件下随机变量 X 的条件概率密度; 称

$$F_{X|Y}(x|y)=P\{X\leq x|Y=y\}=\int_{-\infty}^{x}f_{X|Y}(u|y)du$$

为 $Y=y$ 条件下 X 的条件分布函数.

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

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

根据积分中值定理有

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

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

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

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

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

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5.17. 设二维随机变量 (X,Y) 的概率密度

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

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5.18. 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

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

□

课题练习. 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5.24. 多维正太分布的条件分布是正太分布.

Proof. 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

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

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

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

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5.25. 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

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

定理5.26 (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。

Proof. 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5.27. 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

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

Proof. 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. □