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: 在给定随机变量Y取值条件下, 求X的概率分布, 即条件分布.

首先考虑离散型随机变量:

定义5.12. 设(X,Y)为二维离散型随机变量, 其分布列为$\{p_{ij}\}$. 若Y的边缘分布$P(Y=y_j)=p_{\cdot j}>0$, 称

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

为在$Y=y_j$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, 具有分布的一切性质, 例如:

- 非负性: $P(X=x_i|Y=y_j)\geq 0$;
- 规范性: $\sum_{i=1}^{\infty} P(X=x_i|Y=y_j)=1$.

例5.14. 一射手进行射击, 击中目标的概率为p, 射击进行到击中两次目标为止. X表示首次击中目标所进行的射击次数, 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:

$$P\{X=m, Y=n\} = p^2(1-p)^{n-2} \qquad 1\leq m<n<\infty.$$

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, 有$P(X=x)=0$和$P(Y=y)=0$成立, 因此不能利用条件概率得到条件分布. 下面给出条件概率的定义:

定义5.13. 设连续型X和Y的联合概率密度为$f(x,y)$, 随机变量Y的边缘概率密度为$f_Y(y)$, 对任意y, $f_Y(y)>0$, 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

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

下面解释条件概率的含义, 这里以$f_{X|Y}=\frac{f(x,y)}{f_Y(y)}$为例, 首先分布函数有

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

根据积分中值定理有

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

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

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

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5.2 (乘法公式)**.** 对于随机变量X和Y, 有

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

引理5.3. 如果随机变量X和Y相互独立, 则有

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

由条件概率可判别随机变量(X,Y)是否独立. 下面看几个例子:

例5.15. 设(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5.16. 设$X\sim U(0,1)$, 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$f_{Y|X}(y|x)=\frac{1}{1-x}$. 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X的边缘分布

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

□

习题. 设(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$f_{X|Y}(x|y)$.

定理5.17. 多维正太分布的条件分布是正太分布.

Proof. 为简单起见, 这里仅仅给出二维正太分布的详细证明. 设$X=(X_1,X_2)\sim\mathcal{N}(\mu,\Sigma)$, $\mu=(\mu_1,\mu_2)$, $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$, 在$X_2=x_2$的条件下证明$X_1\sim\mathcal{N}(\mu_1-\sigma_1^2\rho(x_2-\mu_2)/\sigma_2^2,\sigma_1^2(1-\rho^2))$. 首先给出二维正太分布的联合分布

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}]}.$$

以及随机变量X_2的边缘分布为$X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$. 于是得到条件概率

$$
\begin{aligned}
f_{X_1|X_2}(x_1|x_2) &= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \rho^2\frac{(x_2-\mu_2)^2}{\sigma_2^2}]} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{(1-\rho)^2}[\frac{x_1-\mu_1}{\sigma_1} + \frac{\rho(x_2-\mu_2)}{\sigma_2}]^2} \\
&= \frac{1}{\sqrt{2\pi}\sigma_1\sqrt{1-\rho^2}} e^{-\frac{1}{2}\frac{1}{\sigma_1^2(1-\rho)^2}[x_1-\mu_1+\sigma_1^2\rho^2(x_2-\mu_2)/\sigma_2^2]^2}.
\end{aligned}
$$

因此在$X_2 = x_2$条件下, $X_1 \sim \mathcal{N}(\mu_1 - \sigma_1^2\rho(x_2-\mu_2)/\sigma_2^2, \sigma_1^2(1-\rho^2))$. □

5.6.1 条件期望

定义5.14. 若X, Y为离散型随机变量,

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

若X, Y为连续型随机变量,

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

注: $E[X|Y=y]$是y的函数. 对条件期望, 有如下重要性质:

定理5.18. 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

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

定理5.19 (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.

Proof. 利用全概率公式有

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

□

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。该定理有一个关于随机变量的定理:

定理5.20. 对随机变量X,Y, 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

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

□

通过全期望公式, 可以证明Markov不等式:

定理5.21 (Markov不等式)**.** 设随机变量$X\geq 0$, 对任意$\epsilon>0$, 有

$$P(X\geq\epsilon)\leq\frac{E(X)}{\epsilon}.$$

Proof. 利用全期望公式考虑随机事件$X\geq\epsilon$, 有

$$E[X]=E[X|X\geq\epsilon]P(X\geq\epsilon)+E[X|X\leq\epsilon]P(X\leq\epsilon)\geq Pr(X\geq\epsilon)\epsilon$$

从而完成证明. □

利用Markov不等式可以推导Chebyshev不等式:

定理5.22 (Chebyshev不等式)**.** 设随机变量X的均值为μ, 则

$$P(|X-\mu|>\epsilon)\leq\frac{Var(X)}{\epsilon^2}.$$

Proof. 根据Markov不等式有

$$P(|X-\mu|>\epsilon)=P((X-\mu)^2\geq\epsilon^2)\leq\frac{E(X-\mu)^2}{\epsilon^2}=\frac{Var(X)}{\epsilon^2}.$$

□