

第八周 条件分布与条件期望

8.1. 条件分布

离散随机变量的条件分布

二元离散型随机变量 (X, Y) 的联合分布列： $P(X = x_i, Y = y_j) = p_{ij}$,

$P(Y = y_j) = \sum_i p_{ij} = p_{\cdot j}$, 对一切使得 $p_{\cdot j} > 0$ 的 y_j ,

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

称为给定 $Y = y_j$ 条件下 X 的分布列。

在 $Y = y_j$ 条件下 X 的分布函数 $F(x | y_j) = \sum_{x_i \leq x} P(X = x_i | Y = y_j)$ 。

例 8.1.1 设离散型随机变量 (X, Y) 联合分布列为 $P(X = i, Y = j) = \frac{\lambda_1^i}{i!} \frac{\lambda_2^j}{j!} e^{-(\lambda_1 + \lambda_2)}$,

$(i = 0, 1, \dots; j = 0, 1, \dots)$, 求 $X + Y = n$ 条件下 X 的分布。

解：首先计算 $P(X + Y = n) = \sum_{k=0}^n P(X = k, Y = n - k) = \sum_{k=0}^n \frac{\lambda_1^k}{k!} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-(\lambda_1 + \lambda_2)}$

$$= \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} = \frac{e^{-(\lambda_1 + \lambda_2)}}{n!} (\lambda_1 + \lambda_2)^n$$

$$= \frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)} \sim P(\lambda_1 + \lambda_2)$$

$$P(X = k | X + Y = n) = \frac{P(X = k, Y = n - k)}{P(X + Y = n)} = \frac{\frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2}}{\frac{(\lambda_1 + \lambda_2)^n}{n!} e^{-(\lambda_1 + \lambda_2)}}$$

$$= \frac{n!}{k!(n-k)!} \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n} = C_n^k \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1 + \lambda_2)^n}$$

$$= C_n^k \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left(1 - \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^{n-k} \sim b \left(n, \frac{\lambda_1}{\lambda_1 + \lambda_2} \right).$$

实际上，如果我们更充分地利用已有的概率分布和独立性的知识，可以观察到本题中的二元离散型随机变量 X 和 Y 的边缘分布均服从泊松分布， X 服从参数为 λ_1 的泊松分布， Y 服从参数为 λ_2 的泊松分布，且 X, Y 相互独立。上一周我们曾经介绍过相互独立的泊松分布随机变量具有可加性，即可知道 $X+Y$ 服从参数为 $\lambda_1 + \lambda_2$ 的泊松分布。

连续随机变量的条件分布

二元连续型随机变量 (X, Y) 的联合密度函数为 $f(x, y)$ ，对一切使 Y 的边缘密度函数 $f_Y(y) > 0$ 的 y ，给定 $Y = y$ 条件下 X 的条件分布函数为：

$$F(x|y) = P(X \leq x | Y = y)$$

$$= \lim_{h \rightarrow 0} P(X \leq x | y \leq Y \leq y+h) = \lim_{h \rightarrow 0} \frac{P(X \leq x, y \leq Y \leq y+h)}{P(y \leq Y \leq y+h)}$$

$$= \lim_{h \rightarrow 0} \frac{\int_{-\infty}^x \int_y^{y+h} f(u, v) du dv}{\int_y^{y+h} f_Y(v) dv} = \frac{\int_{-\infty}^x f(u, y) du}{f_Y(y)} = \int_{-\infty}^x \frac{f(u, y)}{f_Y(y)} du$$

$$\text{条件密度函数 } f(x|y) = \frac{d \int_{-\infty}^x \frac{f(u, y)}{f_Y(y)} du}{dx} = \frac{f(x, y)}{f_Y(y)}$$

例 8.1.2 设 (X, Y) 服从 $G = \{(x, y); x^2 + y^2 \leq 1\}$ 内的均匀分布，求 $f(x|y)$

$$\text{解: } f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1, \\ 0, & \text{其他} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} \frac{2}{\pi} \sqrt{1-y^2}, & -1 \leq y \leq 1 \\ 0, & \text{其他} \end{cases}$$

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} = \frac{1}{2\sqrt{1-y^2}}, \quad |x| \leq \sqrt{1-y^2}, \quad |y| < 1.$$

连续场合下的全概率公式

$$f(x, y) = f_Y(y) f(x|y) \Rightarrow \int_{-\infty}^{+\infty} f(x, y) dy = f_X(x) = \int_{-\infty}^{+\infty} f_Y(y) f(x|y) dy$$

这一公式，可以理解为将 X 在某一个固定点 x 的密度，分解为 Y 的所有取值可能条件下 x 的密度用 y 出现的密度加权求和。一定程度上是有限或可列多种情况下的分情况讨论到不可数种情况的推广形式，因此称为连续型随机变量场合下的全概率公式。

连续场合下的贝叶斯公式

$$f(x|y) = \frac{f(x, y)}{f_Y(y)} \Rightarrow f(x|y) = \frac{f_X(x) f(y|x)}{\int_{-\infty}^{+\infty} f_X(x) f(y|x) dx}.$$

公式中将 Y 条件下 X 的密度函数表达为 X 条件下 Y 的条件概率，交换了条件与目标的顺序，称为连续场合下的贝叶斯公式。

8.2 条件期望

条件数学期望

$$E(X|Y=y) = \begin{cases} \sum_i x_i P(X=x_i|Y=y), & (X, Y) \text{ 为离散随机变量} \\ \int_{-\infty}^{+\infty} xf(x|y)dx, & (X, Y) \text{ 为连续随机变量} \end{cases}$$

例 8.2.1 假设我们独立地抛掷两枚均匀的六面色子，令 X 表示第一枚色子抛出的点数， Y 表示第二枚色子掷出的点数， Z 表示两枚色子的点数和。求 $X=2$ 条件下 Z 的

期望, 以及 $Z=5$ 条件下 X 的期望。

$$\text{解: } E(Z|X=2) = \sum_{k=3}^8 k \cdot P(Z=k|X=2) = \sum_{k=3}^8 k \cdot \frac{1}{6} = \frac{11}{2}$$

$$\begin{aligned} E(X|Z=5) &= \sum_{k=1}^4 k P(X=k|Z=5) = \sum_{k=1}^4 k \frac{P(X=k, Z=5)}{P(Z=5)} \\ &= \sum_{k=1}^4 k \frac{P(X=k)P(Y=5-k)}{4/36} = \sum_{k=1}^4 k \frac{1/36}{4/36} = \frac{5}{2} \end{aligned}$$

例 8.2.2 随机变量 X 的密度函数为阶梯形函数, $f(x) = \begin{cases} 2/3, & 0 \leq x < 1 \\ 1/3, & 1 \leq x < 2, \\ 0, & \text{其他} \end{cases}$

设事件 $A = \{X \text{ 落入区间 } [1, 2)\}$, 计算条件期望 $E(X|A)$ 。

解: 当 $x < 1$ 或 $x \geq 2$ 时, 事件 A 不发生, 所以 $f(x|A) = 0$;

$$\text{当 } 1 \leq x < 2 \text{ 时, 事件 } A \text{ 发生, } P(A) = P(1 \leq X < 2) = \int_1^2 f(x) dx = \int_1^2 \frac{1}{3} dx = \frac{1}{3}$$

$$\text{此时, } f(x|A) = \frac{f(x)}{P(A)} = \frac{1/3}{1/3} = 1;$$

$$E(X|A) = \int_{-\infty}^{+\infty} x \cdot f(x|A) dx = \int_1^2 x \cdot 1 dx = \frac{3}{2}.$$

$E(X|Y)$ 是随机变量,

例如 Y 是离散型随机变量, 则

$$E(X|Y) \sim \begin{pmatrix} E(X|Y=y_1) & E(X|Y=y_2) & \cdots & E(X|Y=y_n) & \cdots \\ P(Y=y_1) & P(Y=y_2) & \cdots & P(Y=y_n) & \cdots \end{pmatrix}$$

$$Y \sim \begin{pmatrix} Y=y_1 & Y=y_2 & \cdots & Y=y_n & \cdots \\ P(Y=y_1) & P(Y=y_2) & \cdots & P(Y=y_n) & \cdots \end{pmatrix} \xRightarrow{E(X|Y)=g(Y)}$$

$$E(X|Y) \sim \begin{pmatrix} E(X|Y=y_1) & E(X|Y=y_2) & \cdots & E(X|Y=y_n) & \cdots \\ P(Y=y_1) & P(Y=y_2) & \cdots & P(Y=y_n) & \cdots \end{pmatrix}$$

X, Y 为两个随机变量，表达式 $E(X|Y)$ 表示一个随机变量。

以 Y 为离散型随机变量的情形进行说明。这里强调的是 $E(X|Y)$ 不是一个实数值，而是一个取值依赖于 Y 的随机变量。也可以这样理解， $E(X|Y)$ 是随机变量 Y 的一个函数，具体的映射关系就是 $g(y_n) = E(X|Y=y_n)$ 。

例 8.2.3 假设我们独立地抛掷两枚均匀的六面色子，令 X 表示第一枚色子抛出的点数， Y 表示第二枚色子掷出的点数， Z 表示两枚色子的点数和。求 $E(Z|X)$ 。

解：对所有 X 可能的取值 $k=1, 2, \dots, 6$ ，计算 $X=k$ 条件下 Y 的期望

$$E(Z|X=k) = E(X+Y|X=k) = E(k+Y) = k + E(Y) = k + \frac{7}{2}$$

$$E(Z|X) \sim \begin{pmatrix} 1+7/2 & 2+7/2 & 3+7/2 & 4+7/2 & 5+7/2 & 6+7/2 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{pmatrix}$$

也可以直接将 $E(Z|X)$ 表示为 X 的函数的形式：

$$E(Z|X) = \sum_{z=X+1}^{z=X+6} z \cdot P(Z=z|X) = \sum_{k=1}^6 (X+k) \cdot P(Z=X+k|X)$$

$$= \sum_{k=1}^6 (X+k) \cdot \frac{1}{6} = X + \frac{7}{2}$$

例 8.2.4 设随机变量 $X \sim Ge(p)$ ， $0 < p < 1$ ， $P(X=k) = p \cdot (1-p)^{k-1}$ ， $k=1, 2, \dots$ ，

随机变量 $Y = \begin{cases} 1, & X=1 \\ 0, & X>1 \end{cases}$, 求条件期望 $E(X|Y)$

解: 分别计算随机变量 $Y=1$ 和 $Y=0$ 条件下期望 X 的期望

$$E(X|Y=1) = E(X|X=1) = 1$$

$$E(X|Y=0) = E(X|X>1) = \sum_{k=2}^{\infty} k \cdot P(X=k|X>1)$$

$$= \sum_{k=2}^{\infty} k \cdot P(X=k-1) = \sum_{k=2}^{\infty} [(k-1)+1] \cdot P(X=k-1)$$

$$= \sum_{i=1}^{\infty} i \cdot P(X=i) + \sum_{i=1}^{\infty} P(X=i) = E(X) + 1 = 1 + \frac{1}{p}$$

$$\text{所以 } E(X|Y) \sim \begin{pmatrix} 1+\frac{1}{p} & 1 \\ P(Y=0) & P(Y=1) \end{pmatrix} = \begin{pmatrix} 1+\frac{1}{p} & 1 \\ 1-p & p \end{pmatrix}.$$

8.3 全期望公式(上)

全期望公式(重期望公式) $E(X) = E(E(X|Y))$

以离散型为例证明:

$$\begin{aligned} E(X) &= \sum_{i=1}^n x_i P(X=x_i) \\ &= \sum_{i=1}^n x_i \left(\sum_{j=1}^m P(X=x_i, Y=y_j) \right) = \sum_{i=1}^n x_i \left(\sum_{j=1}^m P(Y=y_j) P(X=x_i|Y=y_j) \right) \\ &= \sum_{j=1}^m P(Y=y_j) \left(\sum_{i=1}^n x_i P(X=x_i|Y=y_j) \right) \\ &= \sum_{j=1}^m P(Y=y_j) E(X|Y=y_j) = E(E(X|Y)) \end{aligned}$$

随机变量 $E(X|Y)$ 的期望, 相当于计算 Y 的不同取值下, X 的条件期望, 再用 Y 的各个取值概率对条件期望值加权求和。相当于将 X 的期望分解为 Y 的各个取值条件下 X 在

此局部的期望，然后再将各局部的期望加权汇总为 X 整个的期望。类似于全概率公式将事件 A 的概率，分解为一个样本空间分割下不同事件条件下 A 的局部概率，再汇总的到 A 的总的概率。因此这一公式被称为全期望公式。又因为该公式刻画期望的期望，此公式也往往被形象地称为重期望公式。

例 8.3.1 口袋里编号 $1, 2, \dots, n$ 的 n 个球，任取 1 个，若取到为 1 号则得 1 分停止，若为 i ($i \geq 2$) 号，则得到 i 分，放回继续摸球。求总得分的期望。

解：设 X 为总得分， Y 为第一次抽到的号码

$$\text{则 } P(Y = k) = \frac{1}{n} \quad (k = 1, \dots, n),$$

$$E(X | Y = 1) = 1, \quad E(X | Y = k, k \neq 1) = k + E(X)$$

$$E(X) = E(E(X | Y)) = \sum_{k=1}^n E(X | Y = k) P(Y = k) = \frac{1}{n} + \frac{1}{n} \sum_{k=2}^n [k + E(X)]$$

$$\Rightarrow E(X) = \frac{n(n+1)}{2}.$$

例 8.2.4 设随机变量 $X \sim Ge(p)$, $0 < p < 1$, $P(X = k) = p \cdot (1-p)^{k-1}$, $k = 1, 2, \dots$, 计算 X 的期望和方差。

解： 设定义随机变量 $Y = \begin{cases} 1, & X = 1 \\ 0, & X > 1 \end{cases}$, 则由全期望公式

$$\begin{aligned} E(X) &= E(E(X | Y)) = P(Y = 1) \cdot E(X | Y = 1) + P(Y = 0) \cdot E(X | Y = 0) \\ &= P(X = 1) \cdot E(X | X = 1) + P(X > 1) \cdot E(X | X > 1) \end{aligned}$$

$$E(X | X = 1) = 1, \quad E(X | X > 1) = 1 + E(X)$$

$$E(X) = P(X = 1) \cdot E(X | X = 1) + P(X > 1) \cdot E(X | X > 1) = p \cdot 1 + (1-p) \cdot (1 + E(X)),$$

解得, $E(X) = \frac{1}{p}$ 。

例 8.2.4 设随机变量 $X \sim Ge(p)$, $0 < p < 1$, $P(X=k) = p \cdot (1-p)^{k-1}$, $k=1,2,\dots$, 计算 X 的期望和方差。

解: $E(X^2) = E(E(X^2|Y)) = P(X=1) \cdot E(X^2|X=1) + P(X>1) \cdot E(X^2|X>1)$

$$E(X^2|X=1) = 1, \quad E(X^2|X>1) = E((1+X)^2) = 1 + 2E(X) + E(X^2)$$

$$E(X^2) = \frac{1 + 2(1-p)E(X)}{p} = \frac{2}{p^2} - \frac{1}{p},$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{1-p}{p^2}$$

例 8.2.5. 投掷一枚公平的硬币, 直至首次出现相继的两个正面停止, 求投掷次数的期望。

解: 设 X 为投掷次数, Y 定义为 $\begin{pmatrix} T & HH & HT \\ 1/2 & 1/4 & 1/4 \end{pmatrix}$, 其中 T 表示第一次掷出反面,

HH 表示前两次依次掷出正、正, HT 表示前两次依次掷出正、反。

$$E(X) = E(E(X|Y))$$

$$= P(Y=T)E(X|Y=T) + P(Y=HH)E(X|Y=HH) + P(Y=HT)E(X|Y=HT)$$

$$= \frac{1}{2}(1+E(X)) + \frac{1}{4} \cdot 2 + \frac{1}{4}(2+E(X))$$

$$E(X) = \frac{1}{2}(1+E(X)) + \frac{1}{4} \cdot 2 + \frac{1}{4}(2+E(X)) \Rightarrow E(X) = 6。$$

思考: X 的方差如何计算?

8.4 全期望公式（下）

例 8.4.1 随机变量 X 的密度函数为阶梯形函数, $f(x) = \begin{cases} 2/3, & 0 \leq x < 1 \\ 1/3, & 1 \leq x < 2 \\ 0, & x \geq 2 \end{cases}$,

设定事件 $A = \{X \text{ 落入区间 } [0, 1)\}$, $B = \{X \text{ 落入区间 } [1, 2)\}$, 试利用全期望公式计算期望 $E(X)$ 和方差 $Var(X)$ 。

解: 先计算事件 A 、 B 发生的概率, 得到

$$P(A) = \int_0^1 f(x) dx = \int_0^1 \frac{2}{3} dx = \frac{2}{3}, \quad P(B) = \int_1^2 f(x) dx = \int_1^2 \frac{1}{3} dx = \frac{1}{3}$$

$$f(x|A) = \begin{cases} 1, & 0 \leq x < 1 \\ 0, & \text{其他} \end{cases}, \quad f(x|B) = \begin{cases} 1, & 1 \leq x < 2 \\ 0, & \text{其他} \end{cases}$$

$$E(X|A) = \int_0^1 x \cdot f(x|A) dx = \int_0^1 x \cdot 1 dx = \frac{1}{2},$$

$$E(X|B) = \int_1^2 x \cdot f(x|B) dx = \int_1^2 x \cdot 1 dx = \frac{3}{2}$$

$$E(X^2|A) = \int_0^1 x^2 \cdot f(x|A) dx = \int_0^1 x^2 \cdot 1 dx = \frac{1}{3},$$

$$E(X^2|B) = \int_1^2 x^2 \cdot f(x|B) dx = \int_1^2 x^2 \cdot 1 dx = \frac{7}{3}$$

$$E(X) = P(A)E(X|A) + P(B)E(X|B) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{3}{2} = \frac{7}{6}$$

$$E(X^2) = P(A)E(X^2|A) + P(B)E(X^2|B) = \frac{1}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{7}{3} = \frac{5}{3}$$

$$Var(X) = E(X^2) - E(X)^2 = \frac{5}{3} - \left(\frac{7}{6}\right)^2 = \frac{11}{36}$$

例 8.4.2 $(X, Y) \sim N(0, 0, \sigma_1^2, \sigma_2^2, \rho)$, 证明: $E(XY) = \rho\sigma_1\sigma_2$ 。

证明: $E(XY) = E[E(XY|X)] = E[X \cdot E(Y|X)]$

$$\begin{aligned} f(y|x) &= \frac{f(x, y)}{f_X(x)} = \frac{\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{x^2}{\sigma_1^2} - 2\rho\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2}\right]\right\}}{\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{x^2}{2\sigma_1^2}\right\}} \\ &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2\sigma_2^2(1-\rho^2)}\left(y - \rho\frac{\sigma_2}{\sigma_1}x\right)^2\right\} \end{aligned}$$

$$E(Y|X=x) = \rho\frac{\sigma_2}{\sigma_1}x \Rightarrow E(Y|X) = \rho\frac{\sigma_2}{\sigma_1}X \Rightarrow$$

$$E(XY) = E[E(XY|X)] = E[X \cdot E(Y|X)] = E\left(\rho\frac{\sigma_2}{\sigma_1}X^2\right) = \rho\frac{\sigma_2}{\sigma_1}\sigma_1^2 = \rho\sigma_1\sigma_2。$$

例 8.4.3 (随机多个独立随机变量的和的期望) 假设某医生每天门诊挂号的病人数为 N , 是服从参数为 a 的泊松分布随机变量。又假设每位病人门诊看病的时间也为随机的, 均服从参数为 b 的指数分布随机变量, 且相互独立。这名医生总的门诊看病时间记为 T , 求 $E(T)$ 和 $Var(T)$

解: 设 N 位病人的看病时间分别为 X_1, X_2, \dots, X_N , $X_k \sim \text{Exp}(b)$ ($k=1, 2, \dots, N$)

则 $T = X_1 + X_2 + \dots + X_N$, 利用全期望公式,

$$\begin{aligned} E(T) &= E(E(T|N)) = E\left(E\left(\sum_{k=1}^N X_k \middle| N\right)\right) = \sum_{n=1}^{\infty} P(N=n) \cdot E\left(\sum_{k=1}^N X_k \middle| N=n\right) \\ &= \sum_{n=1}^{\infty} P(N=n) \cdot E\left(\sum_{k=1}^n X_k\right) = \sum_{n=1}^{\infty} P(N=n) \cdot n \cdot E(X_1) \\ &= E(X_1) \sum_{n=1}^{\infty} n \cdot P(N=n) = E(X_1)E(N) \end{aligned}$$

$$\begin{aligned}
E(T^2) &= E(E(T^2|N)) = \sum_{n=1}^{\infty} P(N=n) \cdot E\left(\left(\sum_{k=1}^N X_k\right)^2 \middle| N=n\right) \\
&= \sum_{n=1}^{\infty} P(N=n) \cdot E\left(\left(\sum_{k=1}^n X_k\right)^2\right) = \sum_{n=1}^{\infty} P(N=n) \cdot E\left(\sum_{k=1}^n X_k^2 + \sum_{1 \leq i, j \leq n, i \neq j} X_i X_j\right) \\
&= \sum_{n=1}^{\infty} P(N=n) \cdot (n \cdot E(X_1^2) + n \cdot (n-1) \cdot E(X_1)^2) \\
&= \sum_{n=1}^{\infty} P(N=n) \cdot (n \cdot E(X_1^2) - n \cdot E(X_1)^2 + n^2 \cdot E(X_1)^2) \\
&= (E(X_1^2) - E(X_1)^2) \cdot \sum_{n=1}^{\infty} n \cdot P(N=n) + E(X_1)^2 \cdot \sum_{n=1}^{\infty} n^2 \cdot P(N=n) \\
&= \text{Var}(X_1) \cdot E(N) + E(X_1)^2 \cdot E(N^2)
\end{aligned}$$

因为 N 服从参数为 a 的泊松分布, X_1 服从参数为 b 的指数分布, 所以

$$E(N) = \text{Var}(N) = a, \quad E(N^2) = a^2 + a, \quad E(X_1) = \frac{1}{b}, \quad \text{Var}(X_1) = \frac{1}{b^2}$$

$$E(T) = E(X_1)E(N) = \frac{a}{b},$$

$$E(T^2) = \text{Var}(X_1) \cdot E(N) + E(X_1)^2 \cdot E(N^2) = \frac{a^2 + 2a}{b^2}, \quad \text{Var}(T) = E(T^2) - E(T)^2 = \frac{2a}{b^2}$$
