Part III-B: Probability Theory and Mathematical Statistics

Lecture by 李漫漫

Note by THF

2024 年 10 月 17 日

目录

Lecture 8

10.15

0.0.1 条件分布

Notation. 离散型条件分布：

$$P\left\{X=a_i|Y=b_j\right\}=\frac{P\left\{X=a_i,Y=b_j\right\}}{P\left\{Y=b_j\right\}}.$$

$$P\left\{Y=b_j\right\}=\sum_{i=1}^{+\infty}P\left\{Y=b_j,X=a_i\right\}.$$

性质：

1. $P\left\{X=a_i|Y=b_j\right\}\geq 0$

2. $\sum_{i=1}^{+\infty} P = 1$

Notation. 连续型条件分布:

$$P\left\{X = a|Y = b\right\} = 0.$$

(无穷多个点)

通过微元法:

$$\begin{aligned} P\left\{X \leq x|Y = y\right\} &= \lim_{\varepsilon \to 0^+} P\left\{X \leq x|y - \varepsilon < Y \leq y\right\} \\ &= \lim_{\varepsilon \to 0^+} \frac{P\left\{X \leq x, Y \in (y-\varepsilon, y]\right\}}{P\left\{Y \in (y-\varepsilon, y]\right\}} \\ &= \lim_{\varepsilon \to 0^+} \frac{F\left(x,y\right) - F\left(x, y-\varepsilon\right)}{F_Y\left(y\right) - F_Y\left(y-\varepsilon\right)} \\ &= \frac{\dfrac{\partial F\left(x,y\right)}{\partial y}}{\dfrac{\mathrm{d}F_Y\left(y\right)}{\mathrm{d}y}} \\ &= \frac{\displaystyle\int_{-\infty}^{x} f\left(u,y\right)\mathrm{d}u}{f_Y\left(y\right)} \\ &= \int_{-\infty}^{x} \frac{f\left(u,y\right)}{f_Y\left(y\right)}\mathrm{d}u \end{aligned}$$

.

0.1 二维随机变量函数的分布

$$\begin{cases} aX + bY + c \\ \max\{X, Y\} \\ \min\{X, Y\} \end{cases}.$$

重点公式: 3.4.5 ~ 3.4.8

假设随机变量 Z=aX+bY+c，有分布函数 $(X, Y) \to F(x, y)$

$$
\begin{aligned}
F_Z(z) &= P\{Z \le z\} \\
&= P\{aX + bY + c \le z\} \\
&= \begin{cases} \sum\limits_{ax_i + by_j + c \le z} \sum P\{X = x_i, Y = y_j\}, & (X,Y)\text{ 离散} \\ \iint\limits_{ax + by + c \le z} P\{X = x, Y = y\}\,\mathrm{d}x\mathrm{d}y, & (X,Y)\text{ 连续} \end{cases}
\end{aligned}
$$

.

Notation. 二项分布可加性：有两个相互独立的试验 $X \sim B(m,p), Y \sim B(n,p)$ ，相当于一个试验 $Z \sim B(m+n,p)$

泊松分布可加性：相互独立的两个随机变量 $X \sim P(\lambda_1), Y \sim P(\lambda_2)$ ，相当于一个分布 $Z \sim P(\lambda_1 + \lambda_2)$

Notation. 极值公式：

两个随机变量（连续）X, Y 相互独立，求 $Z_1 = \max(X,Y), Z_2 = \min(X,Y)$ 的分布函数和密度函数

1. $F_{Z_1}(z)$ ：最大的不超过 z 等价于每一个都不超过 z

$$
\begin{aligned}
F_{Z_1}(z) &= P\{X \le z, Y \le z\} \\
&= P\{X \le z\} \cdot P\{Y \le z\} \\
&= F_X(z) \cdot F_Y(z)
\end{aligned}
$$

.

2. $F_{Z_2}(z)$：最小的不超过 z 不等价于每一个都不超过 z ，但最小的超过 z 等价与每一个都超过了 z

$$
\begin{aligned}
F_{Z_2}(z) &= 1 - P\{X > z, Y > z\} \\
&= 1 - [1 - F_X(z)] \cdot [1 - F_Y(z)]
\end{aligned}
$$

.

Notation. 独立同分布：变量相互独立且分布律相同

对极值公式扩展：$(X_1, X_2, \ldots, X_n)$ 独立同分布：

$$F_{Z_1}(z) = \prod_{i=1}^{n} F_{X_i}(z).$$

由于同分布，因此 $F_{X_i}(z) = F_X(z)$

$$F_{Z_1}(z) = (F_X(z))^n.$$

$$f_{Z_1}(z) = (F_{Z_1}(z))' = n(F_X(z))^{n-1}.$$

Notation. 3.4.5:

$$\begin{aligned}
F_Z(z) &= P\{X+Y \le z\} \\
&= \iint_D f(x,y)\,\mathrm{d}x\mathrm{d}y \\
&= \int_{-\infty}^{+\infty} \mathrm{d}x \int_{-\infty}^{z-x} f(x,y)\,\mathrm{d}y \\
&= \int_{-\infty}^{z} \mathrm{d}t \int_{-\infty}^{+\infty} f(x,t-x)\,\mathrm{d}x \\
f_Z(z) &= F_Z'(z) \\
&= \int_{-\infty}^{+\infty} f(x,z-x)\,\mathrm{d}x
\end{aligned}$$

.

同理：

$$f_Z(z) = \int_{-\infty}^{+\infty} f(z-y,y)\,\mathrm{d}y.$$

称以上两个公式为卷积公式：$f_Z = f_X * f_Y$

Notation. 正态分布的可加性：

$X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ ，且 X, Y 相互独立，则：

$$X+Y = Z \sim N\left(\mu_1+\mu_2, \sigma_1^2, \sigma_2^2\right).$$

多元：

$$Z = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right).$$

Lecture 8

0.2 二元正态分布

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

Lecture 9

10.17

Example. $X\sim U[-1,1]$ ，对应的密度函数:

$$f_X\left(x\right)=\begin{cases}0, & x<-1\text{ or }x>1\\ \frac{1}{2}, & x\in[-1,1]\end{cases}.$$

同理 $Y\sim U[-1,1]$，求 $Z=|X-Y|$ 的分布函数

解: Z 的取值: [0,2]

$$\begin{aligned}F_Z\left(z\right)&=P\left\{Z\leq z\right\}\\&=P\left\{|X-Y|\leq z\right\}\\&=\begin{cases}0, & z<0\\ 1, & z\geq 2\\ P\left\{-z\leq X-Y\leq z\right\}, & z\in[0,2)\end{cases}.\end{aligned}$$

其中

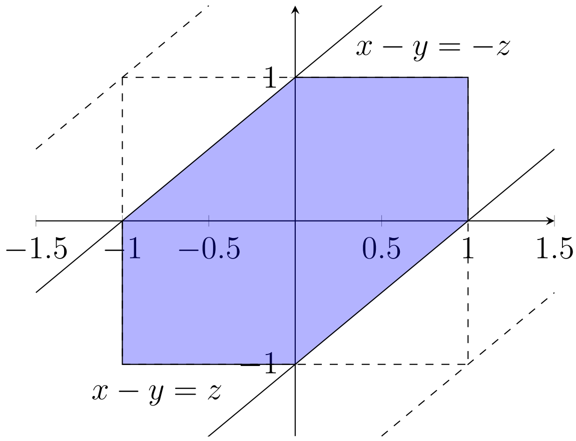
$$P\left\{z-z\leq X-Y\leq z\right\}=\iint_{-z\leq x-y\leq z}f\left(x,y\right)\mathrm{d}x\mathrm{d}y.$$

通过 f_X和f_Y 求联合密度函数: X,Y 独立，即 $f\left(x,y\right)=f_X\left(x\right)f_Y\left(y\right)$

$$f\left(x,y\right)=\begin{cases}0, & x,y<-1\text{ or }x,y>1\\ \frac{1}{4}, & x,y\in[-1,1]\end{cases}.$$

$$P\left\{-z\leq X-Y\leq z\right\}=\iint\limits_{-z\leq x-y\leq z,x,y\in[-1,1]}\frac{1}{4}\mathrm{d}x\mathrm{d}y.$$

画图确认积分区域:



Notation. i.i.d. : 独立同分布

Notation. 伽马分布 $\Gamma(\alpha,\beta)$ 的密度函数:

$$f(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}\mathrm{e}^{-\beta x}, & x\geq 0\\ 0, & x<0\end{cases}.$$

Notation. 重点题目: 3.6.2

1 数字特征

$$\text{数字特征}\begin{cases}\text{数学期望: } E(X)\\ \text{方差: } D(X) \text{ or } \mathrm{Var}(X)\\ \text{协方差: } \mathrm{cov}(X,Y)\\ \text{相关系数: } \rho(X,Y)\\ \text{矩: } E(X)^k \text{ and } E(X-EX)^k\end{cases}.$$

Notation. $D(X)=E(X-EX)^2$

$\mathrm{cov}(X,Y)=E((X-EX)(Y-EY))$

$\rho(X,Y)=\dfrac{\mathrm{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}$

Notation. $|\rho_{X,Y}| \in [0,1]$

ρ 越大越线性相关，$\rho > 0.8$ 时基本可以确定为线性相关

Notation. 矩 (moment) 是最一般的概念

矩分为两大类：k 阶原点矩和 k 阶中心矩

原点矩：$E(X)^k$

中心矩：$E(X-EX)^k$

$k+l$ 阶混合中心矩：$E\left((X-EX)^k (Y-EY)^l\right)$

Example. 数学期望为一阶原点矩

方差为一阶中心矩

协方差为二阶混合中心矩

Notation. 可以写出无穷阶的中心矩等同于通过泰勒原理得出分布函数

本章重点：如何计算任意随机变量有关函数的数学期望

唯一计算公式：4.1.5 和 4.1.6

1.1 数学期望

Definition. 离散型随机变量 X 的分布律：$P\{X = x_i\} = p_i, i = 1, 2, \ldots$

若级数 $\sum\limits_{i=1}^{+\infty} x_i p_i$ **绝对收敛**（$\sum\limits_{i=1}^{+\infty} |x_i| p_i < +\infty$），则 X 的数学期望**存在**（x_i 为取值，p_i 为权重，$p_i \geq 0$）

$$E(X) = EX = \sum_{i=1}^{+\infty} x_i P\{X = x_i\} = \sum_{i=1}^{+\infty} x_i p_i.$$

Rule. 当一个随机变量的密度函数与分布律已知：$X \to f(x), P\{X = x_i\} = p_i$

即可以求关于 X 函数的数学期望（公式 4.1.5）：

$$E(g(X)) = \begin{cases} \sum\limits_{i=1}^{+\infty} g(x_i) P\{X = x_i\}, & X\text{离散} \\ \int_{-\infty}^{+\infty} g(x) f(x)\, \mathrm{d}x, & X\text{连续} \end{cases}.$$

Rule. 扩展至二阶：$(X, Y) \to P\{X = x_i, Y = y_j\}, f(x, y)$

关于 (X,Y) 的函数的数学期望（公式 4.1.6）:

$$E\left(g\left(X,Y\right)\right)=\begin{cases}\displaystyle\sum_{i=1}^{+\infty}\sum_{j=1}^{+\infty}g\left(x_i,y_j\right)P\left\{X=x_i,Y=y_j\right\}, & \text{离散}\\ \displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g\left(x,y\right)f\left(x,y\right)\mathrm{d}x\mathrm{d}y, & \text{连续}\end{cases}.$$

Notation. 柯西分布:

$$f\left(x\right)=\frac{1}{\pi\left(1+x^2\right)},x\in\mathbb{R}.$$

常见分布数学期望:

Notation. 伯努利分布 $X\sim B\left(n,p\right)$： $EX=np$

泊松分布 $X\sim P\left(\lambda\right)\left(\lambda>0\right)$： $EX=\lambda$

柯西分布: EX 不存在（柯西分布不绝对收敛）

Notation. 柯西活了 68 岁，21 岁成名（导师拉格朗日）,27 岁当选法国科学院院士