

## 第十周 独立随机变量和的分布与顺序统计量

### 10.1. 独立随机变量和的分布

#### 泊松分布和二项分布的可加性

若  $X_1 \sim P(\lambda_1)$ ,  $X_2 \sim P(\lambda_2), \dots, X_m \sim P(\lambda_m)$ , 且  $X_1, X_2, \dots, X_m$  相互独立,

则  $X_1 + X_2 + \dots + X_m \sim P(\lambda_1 + \lambda_2 + \dots + \lambda_m)$

$X_1 \sim B(n_1, p)$ ,  $X_2 \sim B(n_2, p), \dots, X_m \sim B(n_m, p)$ , 且  $X_1, X_2, \dots, X_m$  相互独立,

则  $X_1 + X_2 + \dots + X_m \sim B(n_1 + n_2 + \dots + n_m, p)$

回顾一下, 第7周第3讲的最后, 我们曾经介绍过泊松分布和二项分布的可加性。本节我们更一般地考虑独立随机变量和的分布, 并引入卷积公式。下一周课我们要介绍大数定律和中心极限定理等概率论的极限理论, 在极限理论的研究中, 经常需要考虑独立随机变量和的分布问题。本节中, 我们主要讨论两个独立随机变量的和的分布问题。

\*\*\*\*\*

#### 相互独立的离散型随机变量和的分布

离散型随机变量  $X$  和  $Y$  相互独立, 分布律分别为

$$P(X = x_i), x_i \in X(\Omega) \text{ 和 } P(Y = y_j), y_j \in Y(\Omega),$$

$Z = X + Y$  也是一个离散型随机变量, 其值域  $Z(\Omega) = \{x_i + y_j : x_i \in X(\Omega), y_j \in Y(\Omega)\}$ ,

对所有的  $z_k \in Z(\Omega)$  有

$$P(Z = z_k) = P(X + Y = z_k) = \sum_{x_i \in X(\Omega)} P(X = x_i, Y = z_k - x_i) = \sum_{x_i \in X(\Omega)} P(X = x_i)P(Y = z_k - x_i)$$

或互换  $X$  和  $Y$  的位置, 得到  $P(Z = z_k) = \sum_{y_j \in Y(\Omega)} P(X = z_k - y_j)P(Y = y_j)$ .

\*\*\*\*\*

例 10.1.1 两位射手各向自己的靶子独立射击，直到自己有一次命中时，停止射击。假设两位射手每次命中概率分别为  $p_1$  和  $p_2$ 。求两射手均停止射击时，他们脱靶（未命中）总数的分布。

解：记  $X_1$  和  $X_2$  分别为两位射手首次命中自己的靶子时所射击的次数，

则  $X_1 \sim Ge(p_1)$ ,  $X_2 \sim Ge(p_2)$ , 且  $X_1, X_2$  相互独立。

记  $X$  为两射手均停止射击时的脱靶总数，则  $X = (X_1 - 1) + (X_2 - 1)$ , 对  $n = 0, 1, 2, \dots$

$$\begin{aligned} P(X = n) &= P(X_1 + X_2 = n + 2) = \sum_{k=1}^{n+1} P(X_1 = k, X_2 = n + 2 - k) = \sum_{k=1}^{n+1} P(X_1 = k) \cdot P(X_2 = n + 2 - k) \\ &= \sum_{k=1}^{n+1} q_1^{k-1} p_1 \cdot q_2^{n+1-k} p_2 = p_1 p_2 q_2^n \sum_{k=1}^{n+1} \left(\frac{q_1}{q_2}\right)^{k-1} = p_1 p_2 (1 - p_2)^n \sum_{k=0}^n \left(\frac{1 - p_1}{1 - p_2}\right)^k \end{aligned}$$

当  $p_1 \neq p_2$  时,  $P(X = n) = \frac{p_1 p_2}{p_1 - p_2} [(1 - p_2)^{n+1} - (1 - p_1)^{n+1}]$

当  $p_1 = p_2$  时,  $P(X = n) = (n + 1) p_1^2 (1 - p_1)^n$

\*\*\*\*\*

### 卷积公式

相互独立的连续型随机变量  $X$ 、 $Y$  的概率密度函数分别为  $f_X(x)$  和  $f_Y(y)$ ,

则  $Z = X + Y$  的概率密度函数为

$$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z - x) dx = \int_{-\infty}^{+\infty} f_X(z - y) f_Y(y) dy。$$

这一公式的证明需要二重积分，这里略去。本课程中我们就不加证明地使用这一公式。

\*\*\*\*\*

例 10.1.2 设随机变量  $X_1, X_2, \dots, X_n$  独立同分布，且  $X_1 \sim Exp(\lambda)$ ，记  $S_n = \sum_{i=1}^n X_i$ ，求

$S_n$  的分布。

解：先计算  $S_2 = X_1 + X_2$ 。当  $z < 0$  时， $f_{S_2}(z) = 0$ ；当  $z \geq 0$  时，利用卷积公式，

$$f_{S_2}(z) = \int_{-\infty}^{+\infty} f_{X_1}(x) \cdot f_{X_2}(z-x) dx = \int_0^z \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda(z-x)} dx = \lambda^2 \int_0^z e^{-\lambda z} dx = \lambda^2 z e^{-\lambda z}。$$

再计算  $S_3 = S_2 + X_3$ ，当  $z < 0$  时， $f_{S_3}(z) = 0$ ；当  $z \geq 0$  时，对  $S_2$  和  $X_3$  再使用卷积公式，

$$f_{S_3}(z) = \int_{-\infty}^{+\infty} f_{X_1+X_2}(x) f_{X_3}(z-x) dx = \int_0^z \lambda^2 x e^{-\lambda x} \lambda e^{-\lambda(z-x)} dx = \lambda^3 \int_0^z x e^{-\lambda z} dx = \frac{1}{2} \lambda^3 z^2 e^{-\lambda z}$$

最后，可以归纳证明， $S_n = \sum_{i=1}^n X_i$  的概率密度函数为  $f_{S_n}(z) = \begin{cases} \frac{\lambda^n z^{n-1} e^{-\lambda z}}{(n-1)!}, & z \geq 0, \\ 0, & z < 0. \end{cases}$

直接用卷积公式来计算多个独立随机变量的和的分布往往是非常繁琐的，数学家引入了矩母函数和特征函数等工具简化计算，而理解和应用这些工具，还需要多元积分、复变函数等更多的数学知识。

\*\*\*\*\*

## 10.2 独立正态分布和的分布

定理：设  $X \sim N(0,1)$ ， $Y \sim N(0,1)$ ，且  $X$  和  $Y$  相互独立，则  $Z = X + Y \sim N(0,2)$ 。

证明：利用卷积公式，

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-(x^2 - zx + \frac{z^2}{2})} dx = \frac{1}{2\pi} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} e^{-(x-\frac{z}{2})^2} dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{4}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \times \frac{1}{\sqrt{2}}} e^{-\frac{(x-\frac{z}{2})^2}{2 \times (\frac{1}{\sqrt{2}})^2}} dx \end{aligned}$$

$$\frac{1}{\sqrt{2\pi}/\sqrt{2}} e^{-\frac{(x-\frac{z}{2})^2}{2 \times (\frac{1}{\sqrt{2}})^2}} \text{ 恰好为 } N(\frac{z}{2}, \frac{1}{2}) \text{ 的概率密度函数, } \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \times \frac{1}{\sqrt{2}}} e^{-\frac{(x-\frac{z}{2})^2}{2 \times (\frac{1}{\sqrt{2}})^2}} dx = 1$$

所以  $f_Z(z) = \frac{1}{\sqrt{4\pi}} e^{-\frac{z^2}{4}} \Rightarrow Z = X + Y \sim N(0,2)$ 。

\*\*\*\*\*

定理：  $X \sim N(0, \sigma_1^2)$  ,  $Y \sim N(0, \sigma_2^2)$  , 且  $X$  、  $Y$  相互独立, 则  $Z = X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)$  。

$$\begin{aligned}
 \text{证明: } f_Z(z) &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}} \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(z-x)^2}{2\sigma_2^2}} dx \\
 &= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-\left[\frac{x^2}{2\sigma_1^2} + \frac{(z-x)^2}{2\sigma_2^2}\right]} dx = \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{+\infty} e^{-\left[\frac{\left(x - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z\right)}{2\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)}\right] - \frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}} dx \\
 &= \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}} \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi} \cdot \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}} e^{-\left[\frac{\left(x - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z\right)}{2\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)}\right]} dx \\
 &\quad \frac{1}{\sqrt{2\pi} \cdot \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2 + \sigma_2^2}}} e^{-\left[\frac{\left(x - \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z\right)}{2\sigma_1^2\sigma_2^2/(\sigma_1^2 + \sigma_2^2)}\right]} \text{ 恰好为 } N\left(\frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z, \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\right) \text{ 的密度函数} \\
 f_Z(z) &= \frac{1}{\sqrt{2\pi} \cdot \sqrt{\sigma_1^2 + \sigma_2^2}} e^{-\frac{z^2}{2(\sigma_1^2 + \sigma_2^2)}} \Rightarrow Z = X + Y \sim N(0, \sigma_1^2 + \sigma_2^2)
 \end{aligned}$$

\*\*\*\*\*

定理： 设随机变量  $X_1, \dots, X_n$  相互独立, 且  $X_i \sim N(\mu_i, \sigma_i^2)$  ,  $(i=1, \dots, n)$  , 则有

$$\sum_{i=1}^n a_i X_i \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right), \quad (a_1, a_2, \dots, a_n \in R).$$

$$\begin{aligned}
 \text{证明: } a_1(X_1 - \mu_1) &\sim N(0, a_1^2 \sigma_1^2), \quad a_2(X_2 - \mu_2) \sim N(0, a_2^2 \sigma_2^2) \\
 \Rightarrow a_1 X_1 + a_2 X_2 - (a_1 \mu_1 + a_2 \mu_2) &\sim N(0, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) \\
 \Rightarrow a_1 X_1 + a_2 X_2 &\sim N(a_1 \mu_1 + a_2 \mu_2, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2)
 \end{aligned}$$

利用归纳法即可得到定理的结论。

\*\*\*\*\*

## 10.3 最大值、最小值分布

### 最大值与最小值的分布

设  $X_1, X_2, \dots, X_n$  相互独立, 分布函数分别为  $F_{X_k}(x) (k=1, 2, \dots, n)$ ,

计算  $Y = \max(X_1, X_2, \dots, X_n)$  与  $Z = \min(X_1, X_2, \dots, X_n)$  的分布。

$$F_Y(y) = P(\max(X_1, \dots, X_n) \leq y) = P(X_1 \leq y, \dots, X_n \leq y) = \prod_{k=1}^n P(X_k \leq y) = \prod_{k=1}^n F_{X_k}(y)$$

$$F_Z(z) = P(\min(X_1, \dots, X_n) \leq z) = 1 - P(\min(X_1, \dots, X_n) > z) = 1 - P(X_1 > z, \dots, X_n > z)$$

$$= 1 - \prod_{k=1}^n P(X_k > z) = 1 - \prod_{k=1}^n [1 - P(X_k \leq z)] = 1 - \prod_{k=1}^n [1 - F_{X_k}(z)]$$

若  $X_1, X_2, \dots, X_n$  为同分布的连续型随机变量, 分布函数和密度函数分别为  $F_X(x)$  和

$$f_X(x), \text{ 则 } f_Y(y) = F_Y'(y) = [F_X(y)^n]' = n[F_X(y)]^{n-1} f_X(y)。$$

类似的, 也可以给出,  $X_1, X_2, \dots, X_n$  同分布条件下, 最小值函数的密度函数, 留作同学们练习。

\*\*\*\*\*

例 10.3.1  $X_1, X_2, \dots, X_n$  相互独立, 分别服从参数为  $\lambda_i$  的指数分布 (即期望为  $\frac{1}{\lambda_i}$ ),

证明:  $Z = \min(X_1, X_2, \dots, X_n)$  也服从指数分布。

$$\text{证明: } P(Z > x) = P(\min(X_1, X_2, \dots, X_n) > x) = P(X_1 > x, X_2 > x, \dots, X_n > x)$$

$$= \prod_{i=1}^n P(X_i > x) = \prod_{i=1}^n e^{-x\lambda_i} = e^{-x \sum_{i=1}^n \lambda_i}$$

$$F_Z(z) = P(Z \leq z) = 1 - P(Z > z) = 1 - e^{-(\lambda_1 + \lambda_2 + \dots + \lambda_n)z} \sim \text{Exp}(\lambda_1 + \lambda_2 + \dots + \lambda_n)。$$

\*\*\*\*\*

## 10.4 顺序统计量

### 顺(次)序统计量

$X_1, X_2, \dots, X_n$  独立同分布, 分布函数  $F(x)$ , 将这  $n$  个随机变量做升序排列

$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ ,  $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$  称为顺(次)序统计量(ordered statistics)。

$X_{(k)}$  ( $k=1, 2, \dots, n$ ): 第  $k$  个顺序统计量

$X_{(1)} = \min(X_1, X_2, \dots, X_n)$ : 最小顺序统计量,  $X_{(n)} = \max(X_1, X_2, \dots, X_n)$ : 最大顺序统计量

\*\*\*\*\*

例 10.4.1 设  $X_1, X_2, X_3$  相互独立, 且均服从  $B\left(2, \frac{1}{2}\right)$ , 求顺序统计量  $(X_{(1)}, X_{(2)}, X_{(3)})$  的联合分布律及  $X_{(1)}, X_{(2)}, X_{(3)}$  各自的分布列。

解:  $X_1 \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \end{pmatrix}$ ,  $X_1, X_2, X_3$  共有  $3^3 = 27$  种不同的取值组合,

将  $X_1, X_2, X_3$  的 27 种组合, 每一组  $X_1, X_2, X_3$  取值的概率和对应  $(X_{(1)}, X_{(2)}, X_{(3)})$  的取值, 列入下表, 如表所示;

$P$	$x_1$	$x_2$	$x_3$	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$P$	$x_1$	$x_2$	$x_3$	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$	$P$	$x_1$	$x_2$	$x_3$	$x_{(1)}$	$x_{(2)}$	$x_{(3)}$
$\frac{1}{64}$	0	0	0	0	0	0	$\frac{1}{32}$	1	0	0	0	0	1	$\frac{1}{64}$	2	0	0	0	0	2
$\frac{1}{32}$	0	0	1	0	0	1	$\frac{1}{16}$	1	0	1	0	1	1	$\frac{1}{32}$	2	0	1	0	1	2
$\frac{1}{64}$	0	0	2	0	0	2	$\frac{1}{32}$	1	0	2	0	1	2	$\frac{1}{64}$	2	0	2	0	2	2
$\frac{1}{32}$	0	1	0	0	0	1	$\frac{1}{16}$	1	1	0	0	1	1	$\frac{1}{32}$	2	1	0	0	1	2
$\frac{1}{16}$	0	1	1	0	1	1	$\frac{1}{8}$	1	1	1	1	1	1	$\frac{1}{16}$	2	1	1	1	1	2
$\frac{1}{32}$	0	1	2	0	1	2	$\frac{1}{16}$	1	1	2	1	1	2	$\frac{1}{32}$	2	1	2	1	2	2
$\frac{1}{64}$	0	2	0	0	0	2	$\frac{1}{32}$	1	2	0	0	1	2	$\frac{1}{64}$	2	2	0	0	2	2
$\frac{1}{32}$	0	2	1	0	1	2	$\frac{1}{16}$	1	2	1	1	1	2	$\frac{1}{32}$	2	2	1	1	2	2
$\frac{1}{64}$	0	2	2	0	2	2	$\frac{1}{32}$	1	2	2	1	2	2	$\frac{1}{64}$	2	2	2	2	2	2

$$X_{(1)}, X_{(2)}, X_{(3)} \text{ 的分布列: } X_{(1)} \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{37}{64} & \frac{13}{32} & \frac{1}{64} \end{pmatrix}, \quad X_{(2)} \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{5}{32} & \frac{11}{16} & \frac{5}{32} \end{pmatrix}, \quad X_{(3)} \sim \begin{pmatrix} 0 & 1 & 2 \\ \frac{1}{64} & \frac{13}{32} & \frac{37}{64} \end{pmatrix}$$

注：希望同学们认真体会随机变量排序的确切概率含义。

\*\*\*\*\*

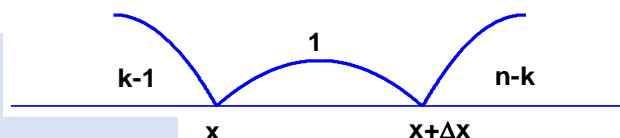
第  $k$  个顺序统计量的分布

$X_1, X_2, \dots, X_n$  独立同分布，分布函数  $F(x)$ ，密度函数  $f(x)$ ，将这  $n$  个随机变量做升序排列  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ ，求  $X_{(k)}$  的分布。

(分析：考虑  $X_{(k)}$  在  $x$  点附近的分布规律， $\Delta x$  非

常小时，只有  $X_{(k)}$  落入区间  $[x, x + \Delta x]$ ，而其余  $n-1$  个随机变

量有  $k-1$  个小于  $x$ ， $n-k$  个大于  $x + \Delta x$ ，可计算出  $X_{(k)}$  落于  $[x, x + \Delta x]$  的概率，进而计算  $X_{(k)}$  的密度函数)



设  $X_{(k)}$  的分布函数为  $F_k(x)$ ，计算  $X_{(k)}$  落于  $[x, \Delta x]$  的概率

$$P(X_{(k)} \in [x, x + \Delta x]) = F_k(x + \Delta x) - F_k(x)$$

$$= \frac{n!}{(k-1)!(n-k)!} \cdot [F(x)]^{k-1} \cdot [F(x + \Delta x) - F(x)] \cdot [1 - F(x + \Delta x)]^{n-k}$$

$$f_k(x) = \lim_{\Delta x \rightarrow 0} \frac{F_k(x + \Delta x) - F_k(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{n!}{(k-1)!(n-k)!} \cdot [F(x)]^{k-1} \cdot \frac{F(x + \Delta x) - F(x)}{\Delta x} \cdot [1 - F(x)]^{n-k}$$

$$= \frac{n!}{(k-1)!(n-k)!} \cdot [F(x)]^{k-1} \cdot f(x) \cdot [1 - F(x)]^{n-k}.$$