第四章 随机变量的数字特征

本章题型

1. 利用概率分布计算数字特征
2. 随机变量相互独立和不相关性的判断
3. 利用重要分布的数字特征

例1 二维连续型随机变量（X,Y）有联合概率密度 $f(x,y)=\begin{cases}\frac{3}{4}, & x^2<y<1\\ 0, & 其他\end{cases}$

求（1）X、Y的条件概率密度；（2）$P(X\le \frac{1}{4}\big|Y=\frac{1}{4})$ （3）X、Y的相关系数。

解（1）$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_{-\sqrt{y}}^{\sqrt{y}}\frac{3}{4}dx=\frac{3}{2}\sqrt{y} & 0<y<1\\ 0 & 其他\end{cases}$$

当$0<y<1$时

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}=\begin{cases}\frac{1}{2\sqrt{y}} & -\sqrt{y}<x<\sqrt{y}\\ 0 & 其他\end{cases}$$

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\int_{x^2}^{1}\frac{3}{4}dy & -1<x<1\\ 0 & 其他\end{cases}$$

当$-1<x<1$时

$$f_{X|Y}(x|y)=\begin{cases}\frac{1}{1-x^2} & x^2<y<1\\ 0 & 其他\end{cases}$$

例1 二维连续型随机变量（X,Y）有联合概率密度 $f(x,y)=\begin{cases}\frac{3}{4}, & x^2<y<1\\ 0, & 其他\end{cases}$

求（1）X、Y的条件概率密度；（2）$P(X\le\frac{1}{4}\Big|Y=\frac{1}{4})$ （3） X、Y的相关系数。

解 （2） 当 $Y=\frac{1}{4}$ 时 $f_{X|Y}(x|y=\frac{1}{4})=\begin{cases}1 & -\frac{1}{2}<x<\frac{1}{2}\\ 0 & 其他\end{cases}$

$$P(X\le\frac{1}{4}\Big|Y=\frac{1}{4})=\int_{-\infty}^{\frac{1}{4}}f_{X|Y}(x|y=\frac{1}{4})dx=\int_{-\frac{1}{2}}^{\frac{1}{4}}dx=\frac{3}{4}$$

（3） $$E(X)=\int_{-\infty}^{+\infty}xf_X(x)dx=\int_{-1}^{1}x\frac{3}{4}(1-x^2)dx=0$$

$$E(Y)=\int_{-\infty}^{+\infty}yf_Y(y)dy=\int_{0}^{1}y\frac{3}{2}\sqrt{y}dy=\frac{3}{5}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\frac{1}{4}\int_{-1}^{1}dx\int_{x^2}^{1}ydy=0$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0$$

$$\therefore\rho_{XY}=0$$

例2 二维连续型随机变量 $(X,Y)\sim U(G), G=\left\{(x,y)\middle|0<y<\sqrt{1-x^2},-1<x<1\right\}$

问X与Y是否相关，是否独立？

解

$$f(x,y)=\begin{cases}\dfrac{2}{\pi} & (x,y)\in G\\ 0 & 其他\end{cases}$$

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=\int_{-1}^{1}dx\int_{0}^{\sqrt{1-x^2}}x\frac{2}{\pi}dy=0$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=\int_{0}^{1}dy\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}xy\frac{2}{\pi}dx=\frac{4}{3\pi}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=\int_{0}^{1}dy\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}xy\frac{2}{\pi}dx=0$$

$$Cov(X,Y)=E(XY)-E(X)E(Y)=0$$

$\therefore$不相关

例2 二维连续型随机变量 $(X,Y)\sim U(G),G=\left\{(x,y)\middle|0<y<\sqrt{1-x^2},-1<x<1\right\}$

问X与Y是否相关，是否独立？

解

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\begin{cases}\int_0^{\sqrt{1-x^2}}\frac{2}{\pi}dy=\frac{2\sqrt{1-x^2}}{\pi} & -1<x<1\\ 0 & \text{其他}\end{cases}$$

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)dx=\begin{cases}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}}\frac{2}{\pi}dx=\frac{4\sqrt{1-y^2}}{\pi} & 0<y<1\\ 0 & \text{其他}\end{cases}$$

$$f(0,\frac{1}{2})=\frac{2}{\pi}\neq f_X(0)f_Y(\frac{1}{2})=\frac{2}{\pi}\cdot\frac{2\sqrt{3}}{\pi}$$

$\therefore$不独立

例3 $X \sim P(4), P(X=\sqrt{D(X)})=?$

解 $D(X)=\lambda=4$

$$P(X=2)=\frac{4^2}{2!}e^{-4}=8e^{-4}$$

例4 $X \sim U(1,2)$ 在 $X=x(1<x<2)$ 条件下，$Y \sim E(x)$ ，求 $E(XY)$

解 $X \sim U(1,2)$ $f_X(x)=\begin{cases} 1 & 1<x<2 \\ 0 & \text{其他} \end{cases}$

$$f_{Y|X}(y|x)=\begin{cases} xe^{-xy} & y>0 \\ 0 & y\le 0 \end{cases}$$

$$f(x,y)=f_X(x)f_{Y|X}(y|x)=\begin{cases} xe^{-x} & 1<x<2, y>0 \\ 0 & \text{其他} \end{cases}$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x,y)dxdy=\int_1^2 dx\int_0^{+\infty} x^2ye^{-xy}dy=1$$

西安電子科技大學
XIDIAN UNIVERSITY

第5章 大数定律及中心极限定理

5.1 Chebyshev（切比雪夫）不等式

（Markov不等式）设 $Y \geq 0$ ，则对 $\forall \varepsilon > 0$，$P(Y \geq \varepsilon) \leq \dfrac{E(Y)}{\varepsilon}$ 。

证明 设 $Y \sim f(y)$，$E(Y) = \int_0^{\infty} yf(y)dy \geq \int_{\varepsilon}^{\infty} yf(y)dy$

$$\geq \int_{\varepsilon}^{\infty} \varepsilon f(y)dy = \varepsilon P(Y \geq \varepsilon)$$

特例： 取 $Y = [X - E(X)]^2, \varepsilon = \varepsilon_1^{\ 2}$

则 $P\left\{|X - E(X)|^2 \geq \varepsilon_1^{\ 2}\right\} \leq \dfrac{D(X)}{\varepsilon_1^{\ 2}}$

$\Rightarrow P\left\{|X - E(X)| \geq \varepsilon_1\right\} \leq \dfrac{D(X)}{\varepsilon_1^{\ 2}}$ Chebyshev（切比雪夫）不等式

或 $P\left\{|X - E(X)| < \varepsilon_1\right\} \geq 1 - \dfrac{D(X)}{\varepsilon_1^{\ 2}}$

条件： $D(X)$、$E(X)$存在。

作用：给出随机变量X分布未知的情况下计算概率下限的估计

例1 调查吸烟率p，问要调查多少人才能保证吸烟频率与p的差不超过0.005的概率不低于95%？

解 假设调查人数为n，吸烟人数n_A，吸烟频率 $f=\dfrac{n_A}{n}$

$$P\{|f-p|\le 0.005\}\ge 0.95$$

$$P\left\{\left|\frac{n_A}{n}-p\right|\le 0.005\right\}\ge 0.95$$

$$P\{|n_A-np|\le 0.005n\}\ge 0.95$$

确定n_A：独立重复做n次实验，服从二项分布。

$$\Leftrightarrow 1-P\{|n_A-np|>0.005n\}\ge 0.95$$

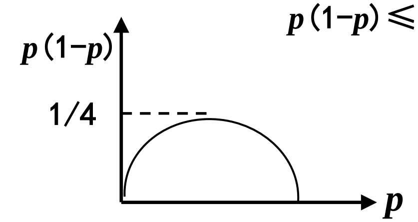
$$P\{|n_A-np|>0.005n\}\le 0.05$$

$$\le \frac{D(n_A)}{0.005^2n^2}$$

$$\frac{np(1-p)}{0.005^2n^2}\le\frac{1}{4\times 0.005^2n^2}\le 0.05$$

确定 $p(1-p)$：

$p(1-p)\leqslant 1/4$



$$\Rightarrow n\approx 40000$$

若 $\dfrac{D(n_A)}{0.005^2n^2}\le 0.05$ ，那么 $P\{|n_A-np|>0.005n\}$ 必然小于0.05。

例2 设X、Y是随机变量且$E(X)=E(Y)$，$D(X)=\frac{1}{4}D(Y), \rho_{XY}=\frac{1}{2}, P(|X-Y|\geq\sqrt{D(Y)})\leq ?$

解

$$E(X-Y)=EX-EY=0$$

$$D(X-Y)=D(X)+D(Y)-2Cov(X,Y)=D(X)+D(Y)-2\rho_{XY}\sqrt{D(X)}\sqrt{D(Y)}$$

$$=\frac{1}{4}D(Y)+D(Y)-2\rho_{XY}\cdot\frac{1}{2}\sqrt{D(Y)}\sqrt{D(Y)}$$

$$=\frac{3}{4}D(Y)$$

$$P(|X-Y|\geq\sqrt{D(Y)})=P(|(X-Y)-E(X-Y)|\geq\sqrt{D(Y)})\leq\frac{D(X-Y)}{D(Y)}=\frac{\frac{3}{4}D(Y)}{D(Y)}=\frac{3}{4}$$

5.2 大数定律

准备知识：

定义 设 $X_1, X_2, \cdots, X_n, \cdots$是随机变量序列，$a$是常数，如果 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|X_n - a\right| \geq \varepsilon\right) = 0$$

则称随机变量序列$\{X_n\}$ 依概率收敛于a，记为 $X_n \xrightarrow{P} a, n \to \infty$

1、*Chebyshev* 大数定律

定理 1 设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，具有相同的数学期望和方差，即 $EX_i = \mu, DX_i = \sigma^2 (i = 1, 2, \cdots)$，$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

则 $\forall \varepsilon > 0$，有 $\lim_{n\to\infty} P\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) = 0$

即 $\bar{X} \xrightarrow{P} \mu, n \to \infty$ 。

证明 由于 $E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu \quad D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} DX_i = \frac{\sigma^2}{n}$

由 *Chebyshev* 不等式，得

$$P\left(\left|\bar{X} - \mu\right| \geq \varepsilon\right) = P\left(\left|\bar{X} - E(\bar{X})\right| \geq \varepsilon\right) \leq \frac{D(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \quad \to 0, n \to \infty$$

因此 $\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$

1、*Chebyshev* 大数定律

Chebyshev 大数定律具有下面更一般的形式：

定理 2 设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的随机变量序列，$E(X_i), D(X_i)(i=1,2,\cdots)$ 若存在常数 $C>0$，使得 $DX_i \le C(i=1,2,\cdots)$，则 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \ge \varepsilon\right) = 0$$

即

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} E(X_i), n \to \infty$$

证明 由于 $E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu \quad D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} DX_i$

由 *Chebyshev* 不等式，得

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \ge \varepsilon\right) \le \frac{D\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]}{\varepsilon^2} \le \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{i=1}^{n} DX_i$$

$$\le \frac{C}{n\varepsilon^2} \to 0, n \to \infty$$

因此

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)\right| \ge \varepsilon\right) = 0$$

2、辛钦大数定律

定理 3 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立且同分布，具有有限的数学期望，即 $EX_i = \mu(i=1,2,\cdots)$，则 $\forall \varepsilon > 0$，有

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \varepsilon\right) = 0$$

即 $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu, n \to \infty$ 。

*Chebyshev*大数定律和*Khintchine*大数定律表明，在随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 满足各自定理的条件下，当n充分大时它前n项的算术平均 $\frac{1}{n}\sum_{i=1}^{n} X_i$，很可能接近于$\mu$；

但是辛钦大数定律要求服从同一分布，不要求方差存在与否；
切比雪夫大数定律不要求同分布，但要求方差存在，方差不同时，要求方差一致有界。

3、伯努利大数定律

定理 4 设n_A表示n重*Bernoulli*试验中事件A发生的次数，p是事件A在一次试验中发生的概率，即 $P(A)=p$，则 $\forall\varepsilon>0$，有 $\lim\limits_{n\to\infty}P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)=0$

即 $$\frac{n_A}{n}\xrightarrow{P}p, n\to\infty$$

证明 设 $X_i=\begin{cases}1, & 若A在第\ i\ 次实验中发生\\ 0, & 若A在第\ i\ 次实验中不发生\end{cases}$ $(i=1,2,\cdots,n)$,

则 $EX_i=p, DX_i=p(1-p), i=1,2,\cdots,n$

且 $n_A=\sum\limits_{i=1}^{n}X_i$, 从而 $\frac{n_A}{n}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i$

由试验的独立性知， $X_1,X_2,\cdots,X_n,\cdots$ 是相互独立的随机变量, 所以：

$$\lim_{n\to\infty}P\left(\left|\frac{n_A}{n}-p\right|\ge\varepsilon\right)=\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-p\right|\ge\varepsilon\right)=0$$