第一章　数字特征

1.1　期望与方差的计算

Remark. 期望与方差

期望的定义

(1) 设随机变量 X 的概率分布为 $P\{X=x_i\}=p_i, i=1,2,\ldots,$ 则 $EX=\sum_i x_i p_i$

推广: 若 $Y=g(X)$ 则 $EY=\sum_i g(x_i)p_i$

(2) 设随机变量 X 的概率密度为 $f(x)$ 则 $EX=\int_{-\infty}^{+\infty} f(x)\mathrm{d}x$

推广: 若 $Y=g(X)$ 则 $EY=\int_{-\infty}^{+\infty} g(x)f(x)\mathrm{d}x$

(3) 设二维随机变量 (X,Y) 的联合概率分布为 $P\{X=x_i,Y=y_j\}=p_{ij}, i,j=1,2,\ldots$ 则

$EZ=\sum_i\sum_j g(x_i,y_j)p_{ij}$

(4) 设二维随机变量 (X,Y) 的联合概率密度为 $f(x,y), Z=g(X,Y)$ 则

$EZ=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$

特别的 $EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x,y)\mathrm{d}x\mathrm{d}y$, $EY=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} yf(x,y)\mathrm{d}x\mathrm{d}y$

期望的性质

(1) $E(aX+bY+c)=aE(X)+bE(Y)+c$

(2) $EXY=EX\cdot EY \iff$ X 与 Y 不相关

特别的若 X 与 Y 相互独立, 由 $EXY=EXEY$

方差的定义

(1) $DX=E(X-EX)^2=EX^2-(EX)^2$

方差的性质

(1) $D(aX+c)=a^2DX$

(2) $D(X \pm Y) = DX + DY \pm 2Cov(X,Y)$

推论 $D(X \pm Y) = D(X) + D(Y) \iff$ X 与 Y 不相关

特别的, 若 X 与 Y 独立, 则有 $D(X \pm Y) = D(X) + D(Y)$

(3) 若 X 与 Y 独立, 则 $DXY = DXDY + (EX)^2DY + (EY)^2DX$

1. 设随机变量 X 的概率密度为 $f(x) = \frac{1}{\pi(1+x^2)}, -\infty < x < \infty$, 则 $E[\min\{|X|, 1\}] =$______.

Solution.

$$
\begin{aligned}
E[\min(|X|,1)] &= \int_{-\infty}^{+\infty} \min(|x|,1) f(x)\mathrm{d}x \\
&= 2\int_{0}^{+\infty} \min(|x|,1) f(x)\mathrm{d}x \\
&= 2(\int_{0}^{1} xf(x)\mathrm{d}x + \int_{1}^{+\infty} f(x)\mathrm{d}x) \\
&= \frac{1}{\pi}\ln(1+x^2)\ |_0^1 + \frac{2}{\pi}\arctan x\ |_1^{+\infty} \\
&= \frac{1}{\pi}\ln 2 + \frac{1}{2}
\end{aligned}
$$

□

2. (2016, 数三) 设随机变量 X 与 Y 相互独立,$X \sim N(1,2)$,$Y \sim N(1,4)$, 则 $D(XY) =$

(A) 6 (B) 8 (C) 14 (D) 15

Solution.

(方法一) 通过计算方法做

$$
\begin{aligned}
DXY &= E(XY)^2 - (EXY)^2 \\
&= EX^2 \cdot EY^2 - (EXEY)^2 \\
&= [DX + (EX)^2][DY + (EY)^2] - (EXEY)^2 \\
&= 3 \times 5 - 1 = 14
\end{aligned}
$$

(方法二) 用结论

$$
\begin{aligned}
DXY &= DXDY + (EX)^2DY + (EY)^2DX \\
&= 8 + 4 + 2 = 14
\end{aligned}
$$

□

3. 设随机变量 X 与 Y 同分布, 则 $E\left(\frac{X^2}{X^2+Y^2}\right)=$______

Solution. 由轮换对称性有

$$E\left(\frac{X^2}{X^2+Y^2}\right)=E\left(\frac{Y^2}{X^2+Y^2}\right)=\frac{1}{2}E\left(\frac{X^2+Y^2}{X^2+Y^2}\right)=\frac{1}{2}$$

□

总结

若 X,Y 同分布, 则 X,Y 具有相同的 F,f,E,D, 上题的推广结论

$$\text{若}X_1,X_2\ldots,X_n\text{同分布, 则}E\left(\frac{X_1^2}{X_1^2+\ldots+X_n^2}\right)=\frac{1}{n}$$

4. 设随机变量 X 与 Y 相互独立,$X\sim P(\lambda_1)$,$Y\sim P(\lambda_2)$, 且 $P\{X+Y>0\}=1-e^{-1}$, 则 $E(X+Y)^2=$______.

Solution. 利用参数可加性可知,$X+Y\sim P(\lambda_1+\lambda_2)$, 由 $P\{X+Y>0\}=1-e^{-1}=1-P\{X=0\}\implies\lambda_1+\lambda_2=1$, 则 $E(X+Y)^2=D(X+Y)+(E(X+Y))^2=1+1=2$ □

5. 设随机变量 X 与 Y 相互独立,$X\sim E(\frac{1}{3})$, $Y\sim E\left(\frac{1}{6}\right)$, 若 $U=\max\{X,Y\}$, $V=\min\{X,Y\}$, 则 $EU=$______,$EV=$______.

Solution. EV 是比较好求的, 由参数可加性有 $V\sim E(\frac{1}{2})$

方法一利用二维概率密度计算:

由 X,Y 独立, 知 $f(x,y)=f_X(x)f_Y(y)$, 则

$$EU=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\max(x,y)f(x,y)\mathrm{d}x\mathrm{d}y=\ldots=7$$

方法二求 U 的概率密度:

由 $U=\max(X,Y)$ 知 $F_U(u)=F_1F_2\implies f_u=f_1F_2+F_1f_2$

$$EU=\int_{-\infty}^{+\infty}uf_u\mathrm{d}u=\ldots=7$$

方法三利用性质

$$E(U+V)=E(X+Y)=EX+EY=3+6=9$$

$EV=2\implies EU=7$ □

总结

若 $U=\max\{X,Y\}$,$V=\min\{X,Y\}$, 则 $E(U+V)=E(X+Y),E(UV)=E(XY)$

独立同分布随机变量的最大值与最小值的分布函数, 由如下结果

令 $Z=\max\left(X_1,X_2,\ldots,X_n\right)$, 则

$$F_Z z=F_{X_1}F_{X_2}\ldots F_{X_n}$$

令 $Z=\min\left(X_1,X_2,\ldots,X_n\right)$, 则

$$F_Z z=1-[(1-F_{(X_2)})][(1-F_{(X_2)})]\ldots[(1-F_{(X_n)})]$$

6. (2017, 数一) 设随机变量 X 的分布函数为 $F(x)=0.5\Phi(x)+0.5\Phi\left(\frac{x-4}{2}\right)$, 其中 $\Phi(x)$ 为标准正态分布函数, 则 $EX=$______

Solution.

(方法一) $f(x)=\frac{1}{2}\phi(x)+\frac{1}{2}\phi(\frac{x-4}{2})$, 则 $EX=\int_{-\infty}^{+\infty}f(x)\mathrm{d}x=2$

(方法二) 考虑 $F(X_1)=0.5\Phi(x),F(X_2)=0.5\Phi(\frac{x-4}{2})$, 则由第二章的结论 $aF_1+bF_2,(a,b>0,a+b=1)$ 的时候也是分布函数, 故 $EX=\frac{1}{2}EX_1+\frac{1}{2}EX_2=0+\frac{4}{2}=2$ □

7. 设随机变量 $X\sim N(0,1)$, 则 $E|X|=$______,$D|X|=$______.

Solution.

$$\begin{aligned}E|X|&=\int_{-\infty}^{+\infty}|x|\phi(x)\mathrm{d}x\\&=2\int_0^{+\infty}x\phi(x)\mathrm{d}x\\&=\frac{-2}{\sqrt{2\pi}}\int_0^{+\infty}e^{-\frac{x^2}{2}}\mathrm{d}(-\frac{x^2}{2})\\&=\sqrt{\frac{2}{\pi}}\end{aligned}$$

$$\begin{aligned}D|X|&=E(|X|)^2-(E|X|)^2\\&=EX^2-(E|X|)^2\\&=DX+(EX)^2-(E|X|)^2\\&=1-\frac{2}{\pi}\end{aligned}$$

□

总结

(1) 若 $X \sim N(0,1)$, 则 $E|X| = \sqrt{\frac{2}{\pi}}, D|X| = 1 - \frac{2}{\pi}$

(2) 若 $X \sim N(0,\sigma^2)$, 则 $E|X| = \sqrt{\frac{2}{\pi}} \cdot \sigma, D|X| = (1 - \frac{2}{\pi}) \cdot \sigma^2$

(3) 若 $X \sim N(\mu,\sigma^2)$, 则 $E|X-\mu| = \sqrt{\frac{2}{\pi}} \cdot \sigma, D|X| = (1 - \frac{2}{\pi}) \cdot \sigma^2$

8. 设随机变量 X 与 Y 相互独立, 均服从 $N(\mu,\sigma^2)$, 求 $E[\max\{X,Y\}]$,$E[\min\{X,Y\}]$.

Solution. 由 X,Y 独立, 有 $X-Y \sim N(0,2\sigma^2)$,$E\,|X-Y| = \frac{2\sigma}{\sqrt{\pi}}$

由下述总结, 可知所求期望为

$$E\,[max\{X,Y\}] = \frac{1}{2}\,[E(X)+E(Y)+E\,|X-Y|] = \mu + \frac{\sigma}{\sqrt{\pi}}$$
$$E\,[min\{X,Y\}] = \frac{1}{2}\,[E(X)+E(Y)-E\,|X-Y|] = \mu - \frac{\sigma}{\sqrt{\pi}}$$

□

总结

关于最大值最小值函数的拆法

$max\{X,Y\} = \frac{X+Y+|X-Y|}{2}$

$min\{X,Y\} = \frac{X+Y-|X-Y|}{2}$

9. 设独立重复的射击每次命中的概率为 p,X 表示第 n 次命中时的射击次数, 求 EX,DX.

Solution. Pascal 分布 (负二项分布), 关键在于分解随机变量, 设 X_i 表示第 $i-1$ 次命中到 i 命中所需要的射击次数, 则有 $X_1, X_2, \ldots$ 之间相互独立, 且 $X_i \sim G(p)$, 对于 $X = X_1 + X_2 \ldots X_n$, 故

$$EX = EX_1 + EX_2 + \ldots + EX_n = \frac{n}{p}$$
$$DX = DX_1 + DX_2 + \ldots + DX_n = \frac{n(1-p)}{p^2}$$

□

10. (2015, 数一、三) 设随机变量 X 的概率密度为 $f(x) = \begin{cases} 2^{-x}\ln 2, & x > 0 \\ 0, & x \le 0 \end{cases}$, 对 X 进行独立的观测, 直到第 2 个大于 3 的观测值出现时停止, 记 Y 为观测次数。

(a) 求 Y 的概率分布;

(b) 求 EY.

Solution. 不妨令 $p = P\{X > 3\} = \int_3^{+\infty} 2^{-x} \ln 2 \mathrm{d}x = \frac{1}{8}$

(1)

$$\begin{aligned} P\{Y = k\} &= C_{k-1}^1 p^2 (1-p)^{k-2} \\ &= (k-1)(\frac{1}{8})^2 (\frac{7}{8})^{k-2}, k = 2, 3, \ldots \end{aligned}$$

(2)

$$\begin{aligned} EY &= \sum_{k=2}^{\infty} k P\{Y = k\} \\ &= p^2 \sum_{k=2}^{\infty} k(k-1)(1-p)^{k-2} \\ &\xlongequal{\text{幂级数求和}} \ldots \\ &= 16 \end{aligned}$$

也可以用 Pascal 分布的结论直接得出 $EX = \frac{2}{\frac{1}{8}} = 16$ □

1.2 协方差的计算

Remark. 协方差

协方差的定义 $Cov(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EX \cdot EY$

协方差的性质

(1) $Cov(X, Y) = Cov(Y, X), Cov(X, X) = DX$

(2) $Cov(aX + bY + c, Z) = aCov(X, Z) + bCov(Y, Z)$

11. 设 $X_1, X_2, \cdots, X_n$ 为来自总体 X 的简单随机样本。若 $DX = 4$, 正整数 $s \leq n$,$t \leq n$, 则

$$\operatorname{Cov}\left(\frac{1}{s}\sum_{i=1}^{s} X_i, \frac{1}{t}\sum_{j=1}^{t} X_j\right) =$$

(A) $4\max\{s, t\}$ (B) $4\min\{s, t\}$ (C) $\dfrac{4}{\max\{s, t\}}$ (D) $\dfrac{4}{\min\{s, t\}}$

Solution.

$$\begin{aligned}
\mathrm{Cov}\left(\frac{1}{s}\sum_{i=1}^{s}X_i,\frac{1}{t}\sum_{j=1}^{t}X_j\right) &= \frac{1}{st}[Cov(X_1,X_1)+Cov(X_1,X_2)+\dots \\
&+Cov(X_2,X_1)+\dots+Cov(X_s,X_t)] \\
&\xlongequal{Cov(X_i,X_i)=DX_i,Cov(X_i,X_j)=0} \frac{\min(s,t)}{st}\cdot DX \\
&= \frac{4}{\max(s,t)}
\end{aligned}$$

来自总体 X 的简单随机样本必然是独立同分布的. □

12. (2005, 数三) 设 $X_1,X_2,\cdots,X_n(n>2)$ 为来自总体 $N(0,\sigma^2)$ 的简单随机样本, 样本均值为 $\bar{X}$。记 $Y_i=X_i-\bar{X}$,$i=1,2,\cdots,n$。
   (1) 求 Y_i 的方差 DY_i,$i=1,2,\cdots,n$;
   (2) 求 Y_1 与 Y_n 的协方差 $Cov(Y_1,Y_n)$;
   (3) 若 $c(Y_1+Y_n)^2$ 为 σ^2 的无偏估计量, 求常数 c.

Solution.

(1) 方法一:

$$\begin{aligned}
DY_i &= D(X_i-\bar{X}) \\
&= DX_i+D\bar{X}-2Cov(X_i,\bar{X}) \\
&\xlongequal{E\bar{X}=\mu,D\bar{X}=\sigma^2/n} \sigma^2+\frac{\sigma^2}{n}-2Cov(X_i,\frac{1}{n}\sum_{i=1}^{n}X_i) \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

方法二:

$$\begin{aligned}
DY_i &= D(\frac{n-1}{n}X_i-\frac{1}{n}\sum_{i=j}^{n}X_j(j\neq i)) \\
&= (\frac{n-1}{n})^2\sigma^2-\frac{n-1}{n^2}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}$$

(2)

$$\begin{aligned}Cov(Y_1,Y_n)&=Cov(X_1,\bar{X},X_n-\bar{X})\\&=Cov(X_1,X_n)-Cov(X_1,\bar{X})-Cov(X_n-\bar{X})+D\bar{X}\\&=\frac{-\sigma^2}{n}\end{aligned}$$

(3) 由无偏性有 $cE(Y_1+Y_n)^2=\sigma^2 \implies c=\frac{\sigma^2}{E(Y_1+Y_n)^2}$

$$\begin{aligned}E(Y_1+Y_n)^2&=D(Y_1+Y_n)+(EY_1EY_n)^2\\&=DY_1+DY_n+2Cov(Y_1,Y_n)+0\\&=\frac{2(n-2)}{n}\sigma^2\end{aligned}$$

故 $c=\frac{n}{2(n-2)}$ □

1.3 相关系数的计算

Remark. 相关系数

相关系数的定义 $\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\sqrt{DY}}=\frac{EXY-EXEY}{\sqrt{DX}\sqrt{DY}}$

相关系数的性质

(1) $|\rho_{XY}|\le 1$

(2) $\rho_{XY}=0 \iff Cov(X,Y)=0 \iff EXY=EXEY \iff D(X+Y)=DX+DY$

(3) $\rho_{XY}=1 \iff P\{Y=aX+b\}=1(a>0);\rho_{XY}=-1 \iff P\{Y=aX+b\}=1(a<0)$

13. (2016, 数一) 设试验有三个两两互不相容的结果 A_1,A_2,A_3, 且三个结果发生的概率均为 $\frac{1}{3}$。将试验独立重复地做两次,X 表示两次试验中 A_1 发生的次数,Y 表示两次试验中 A_2 发生的次数, 则 X 与 Y 的相关系数为

$$(A)\ -\frac{1}{2}\qquad (B)\ -\frac{1}{3}\qquad (C)\ \frac{1}{3}\qquad (D)\ \frac{1}{2}$$

Solution.

(方法一) 由题意有 X,Y 均服从 $B(2,\frac{1}{3})$, 而 $P\{XY=1\}=PX=1,Y=1=C_2^1(\frac{1}{3})^2$, 且 $P\{XY=0\}=\frac{7}{9}$, 故 XY 的概率分布如下所示

XY	0	1
P	$\frac{7}{9}$	$\frac{2}{9}$

故 $EXY=\frac{2}{9}$, 进而可以求出 $\rho_{XY}=\frac{-\frac{2}{9}}{\frac{4}{9}}=-\frac{1}{2}$

(方法二) 设 Z 为"A_3 在两次试验中发生的次数"

由题意有 $Z\sim B(2,\frac{1}{3}), X+Y+Z=2$ 而 $D(X+Y)=DX+DY+2Cov(X,Y)=\frac{8}{9}+2Cov(X,Y)$, 其中 $D(X+Y)=D(2-Z)=DZ=\frac{4}{9}$, 故 $Cov(X,Y)=\frac{-2}{9}$

(方法三)

$$\begin{aligned}Cov(X,X+Y+Z)&=DX+Cov(X,Y)+Cov(X,Z)\\&\xlongequal{\text{轮换对称性}}\frac{4}{9}+2Cov(X,Y)\\&=Cov(X,2)=0\implies Cov(X,Y)=-\frac{2}{9}\end{aligned}$$

□

14. 设随机变量 $X\sim B\left(1,\frac{3}{4}\right)$,$Y\sim B\left(1,\frac{1}{2}\right)$, 且 $\rho_{XY}=\frac{\sqrt{3}}{3}$。

(a) 求 (X,Y) 的联合概率分布;

(b) 求 $P\{Y=1|X=1\}$.

Solution. 这道题比较简单, 直接给答案

X/Y	0	1	P_i
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$
P_j	$\frac{1}{2}$	$\frac{1}{2}$	1

$P\{Y=1|X=1\}=\frac{2}{3}$ □

1.4 相关与独立的判定

Remark. 相关与独立性

(1) 一般来说独立是强于不相关的条件, 即 独立 $\implies$ 不相关

(2) 对于二维正态分布有 独立 $\iff$ 不相关

(3) 对于 0-1 分布有 独立 $\iff$ 不相关

Remark. 判断是否独立的基本方法

(1) $P(AB)=P(A)P(B)$, 对于离散型选点, 对于连续型选区间

(2) 三个充要条件 $\forall(x,y)$或$(i,j)F(x,y)=F_XF_Y, f(x,y)=f_Xf_Y, P(ij)=P_{\cdot i}P_{j\cdot}$

(3) $\rho_{XY}\neq 0 \implies X,Y$ 不独立

15. 设二维随机变量 (X,Y) 服从区域 $D=\{(x,y)|x^2+y^2\leq a^2\}$ 上的均匀分布, 则

(A) X 与 Y 不相关, 也不相互独立 (B) X 与 Y 相互独立

(C) X 与 Y 相关 (D) X 与 Y 均服从 $U(-a,a)$

Solution. 这道题可以记结论, 对于均匀分布若其区域不为 $(a,b)\times(c,d)$ 的矩形, 则必然不独立, 其中 $X\in(a,b), Y\in(c,d)$

正常来做的话, 步骤如下

$$f(x,y)=\begin{cases}\frac{1}{\pi a^2}, & (x,y)\in D\\ 0, & (x,y)\notin D\end{cases}$$

$$EX=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)\mathrm{d}x\mathrm{d}y \xlongequal{\text{对称性}} 0$$

同理根据对称性可知 $EXY=EX=EY=0$, 故 X,Y 一定不相关, 现在求 X,Y 的边缘分布概率密度, 有

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}y=\begin{cases}\frac{2}{\pi a^2}\sqrt{a^2-x^2}, & x\in(-a,a)\\ 0, & x\notin(-a,a)\end{cases}$$

同理可以求出

$$f_Y(y)=\int_{-\infty}^{+\infty}f(x,y)\mathrm{d}x=\begin{cases}\frac{2}{\pi a^2}\sqrt{a^2-y^2}, & y\in(-a,a)\\ 0, & y\notin(-a,a)\end{cases}$$

显然 $f_Yf_X\neq f(x,y)$ 故 X,Y 不独立. □

16. 设随机变量 X 的概率密度为 $f(x)=\frac{1}{2}e^{-|x|}, -\infty<x<+\infty$。

(a) 求 X 的期望与方差;

(b) 求 X 与 $|X|$ 的协方差, 问 X 与 $|X|$ 是否不相关?

(c) 问 X 与 $|X|$ 是否相互独立? 并说明理由.

***Solution*.**

(1)

$$
\begin{aligned}
EX &= \int_{-\infty}^{+\infty} x f(x)\mathrm{d}x = 0 \\
EX^2 &= \int_{-\infty}^{+\infty} x^2 f(x)\mathrm{d}x = \int_0^{+\infty} x^2 e^{-x}\mathrm{d}x = 2 \\
DX &= EX^2 - (EX)^2 = 2
\end{aligned}
$$

(2)

$$E(X|X|) = \int_{-\infty}^{+\infty} |X|X f(x)\mathrm{d}x = 0 = EXE|X| \implies \rho_{X|X|} = 0, Cov(X, |X|) = 0$$

(3) 设 $A = \{0 < X < 1\}, B = \{|X| < 1\}$, 故

$$P(AB) = P\{0 < X < 1, |X| < 1\} = P\{0 < X < 1\} = P(A)$$

而 $P(B) < 1$ 是显然的, 故 $P(AB) \neq P(A)P(B)$, 即 $X|X|$ 不独立 $\square$