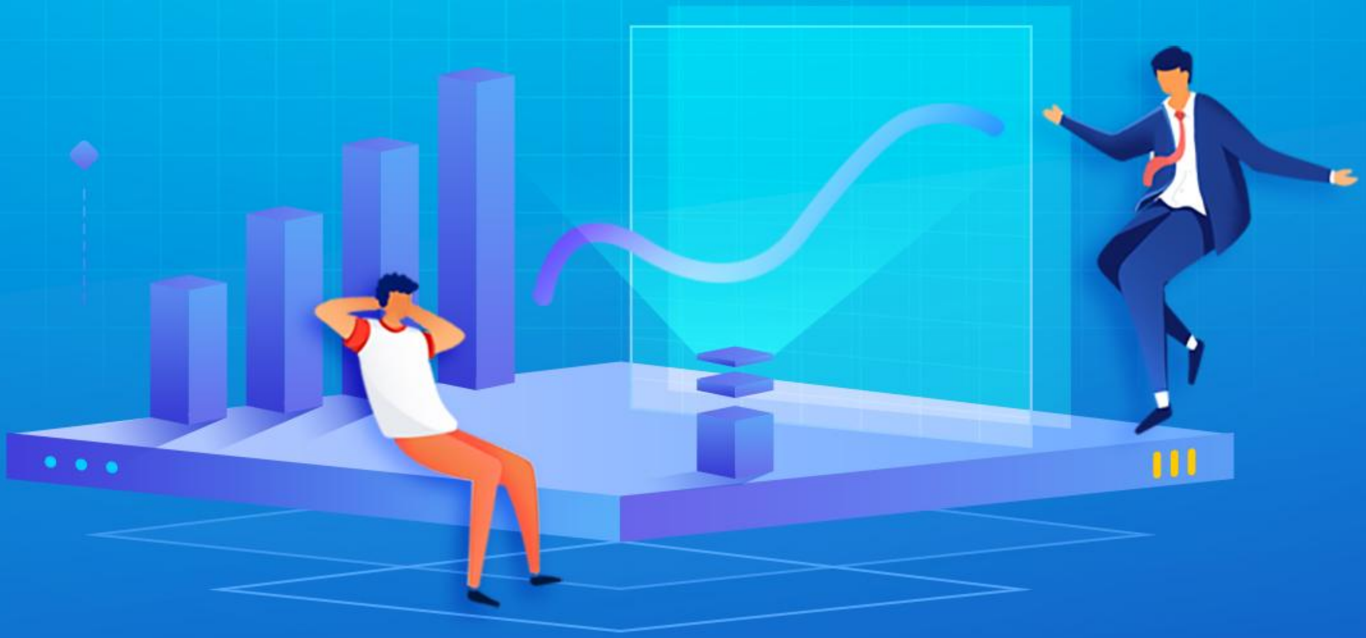
概率论与数理统计



复习

2020 级概率与数理统计试题（A 卷）

一、填空题（14 分，每空 2 分）

1. $\frac{17}{25}$ 2. $\frac{3}{5}$ 3. 一定是 4. 24 5. 9 6. ± 2 7. $t=\frac{(\bar{X}-\mu_0)\sqrt{n(n-1)}}{\sqrt{Q}}$

二、（12 分）

解：设 $A=\{$从乙袋取得白球$\}$，$B_1=\{$从甲袋取一白球放入乙袋中$\}$

$B_2=\{$从甲袋取一红球放入乙袋中$\}$.

1. 由题意知 $P(B_1)=\frac{5}{11}$, $P(B_2)=\frac{6}{11}$, $P(A|B_1)=\frac{11}{20}$, $P(A|B_2)=\frac{10}{20}$.

根据全概率公式可得所求的概率为

$$P(A)=\sum_{i=1}^{2}P(B_i)P(A|B_i)=\frac{5}{11}\times\frac{11}{20}+\frac{6}{11}\times\frac{10}{20}=\frac{23}{44}\approx 0.52.$$

2.根据 Bayes 公式可得所求的概率为

$$P(B_1|A)=\frac{P(B_1)P(A|B_1)}{\sum_{i=1}^{2}P(B_i)P(A|B_i)}=\frac{\frac{5}{11}\times\frac{11}{20}}{\frac{5}{11}\times\frac{11}{20}+\frac{6}{11}\times\frac{10}{20}}=\frac{11}{23}\approx 0.48.$$

三、（12 分）

解：1. （1）已知等待时间 $X\sim E(1/5)$，因此等待时间超过 10 分钟的概率为

$$P\{X>10\}=\int_{10}^{+\infty}\frac{1}{5}e^{-x/5}dx=e^{-2}$$

故顾客去银行一次未等到服务而离开的概率为 e^{-2}。因此 $Y\sim b(5,e^{-2})$，分布律为

$$P\{Y=k\}=C_5^k e^{-2k}(1-e^{-2})^{5-k},\quad k=0,1,\ldots,5$$

（2）$P\{Y\geq 1\}=1-P\{Y=0\}=1-(1-e^{-2})^5$

2. 由于 $Y=|X|$ 是非负取值随机变量，因此

当 $y<0$ 时，$F(y)=0$。

当 $y\geq 0$ 时, $F(y)=P\{|X|\leq y\}=P\{-y\leq X\leq y\}=\Phi(y)-\Phi(-y)=2\Phi(y)-1$.

因此有

$$F(y)=\begin{cases}2\Phi(y)-1 & y\geq 0\\ 0 & y<0\end{cases}$$

因此有

$$f(y)=\begin{cases}\dfrac{dF(y)}{dy}=2\varphi(y)=\sqrt{\dfrac{2}{\pi}}e^{-y^2/2} & y\geq 0\\ 0 & y<0\end{cases}$$

四、（12 分）

解：1. 由

$$1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=\int_0^{\infty}\int_0^{\infty}ke^{-2(x+y)}dxdy=k\int_0^{\infty}e^{-2x}dx\int_0^{\infty}e^{-2y}dy=\frac{k}{4}$$

得

$$k=4$$

2. X 的边缘密度函数为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\begin{cases}2e^{-2x}\int_0^{\infty}2e^{-2y}dy, & x>0\\ 0, & \text{其它}\end{cases}=\begin{cases}2e^{-2x}, & x>0\\ 0, & \text{其它}\end{cases}$$

Y 的边缘密度函数为

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)dx=\begin{cases}2e^{-2y}\int_0^{\infty}2e^{-2x}dx, & y>0\\ 0, & \text{其它}\end{cases}=\begin{cases}2e^{-2y}, & y>0\\ 0, & \text{其它}\end{cases}$$

3. 由于 $f(x,y)=f_X(x)f_Y(y)$，因此 X 和 Y 相互独立.

4. X 和 Y 的分布函数分别为

$$F_X(x)=\begin{cases}1-e^{-2x}, & x>0\\ 0, & \text{其它}\end{cases}\qquad F_Y(y)=\begin{cases}1-e^{-2y}, & y>0\\ 0, & \text{其它}\end{cases}$$

$Z=\min(X,Y)$ 的分布函数为

$$F_Z(z)=P\{Z\le z\}=1-P\{X>z,Y>z\}=1-P\{X>z\}P\{Y>z\}$$

$$=1-(1-F_X(z))(1-F_Y(z))=\begin{cases}1-e^{-4z}, & z>0\\ 0, & \text{其它}\end{cases}$$

$Z=\min(X,Y)$ 的密度函数为

$$f_Z(z)=\begin{cases}4e^{-4z}, & z>0\\ 0, & \text{其它}\end{cases}$$

五、（**8** 分）

解：：设 X_i 为生产第 i 件产品的利润，i=1,2,…,10000. 则

$$X_i=\begin{cases}10, & 第i件产品为正品\\ -5, & 第i件产品为次品\end{cases}$$

由 $P(X_i=10)=0.8,\quad P(X_i=-5)=0.2$，可得

$$EX_i=10\times0.8-5\times0.2=7,\quad EX_i^2=100\times0.8+25\times0.2=85$$

$$DX_i=EX_i^2-E^2X_i=85-49=36.$$

令 $S=\sum_{i=1}^{10000}X_i$，则 $ES=10000\times7=70000, DS=10000\times36=360000$

由中心极限定理，$\dfrac{S-70000}{\sqrt{360000}}$ 近似服从 $N(0,1)$。因此，

$$P(S>69400)=P\left(\frac{S-70000}{600}>\frac{69400-70000}{600}\right)\approx1-\Phi(-1)=\Phi(1)=0.8413.$$

六、（16 分）

解：可知(X,Y)的联合密度函数为

$$f(x,y)=\begin{cases}1, & |x|<y<1\\ 0, & \text{其它}\end{cases}$$

且 $EV=0, DV=\frac{1}{36}$.

于是，有

$$E(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x,y)dxdy=0$$

$$E(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x,y)dxdy=2/3,$$

$$E(X^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}x^2f(x,y)dxdy=1/6,$$

$$E(Y^2)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}y^2f(x,y)dxdy=1/2$$

$$D(X)=E(X^2)-E^2(X)=1/6,$$

$$D(Y)=E(Y^2)-E^2(Y)=1/18,$$

$$E(XY)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xyf(x,y)dxdy=0,$$

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

(2) $EZ=2EX-EY+EV+1=\frac{1}{3}$

由 X 与 Y 不相关，可知 $D(U)=4D(X)+D(Y)=\frac{13}{18}$,

因为 V 与（X，Y）独立，所以 V 与 U 独立

$$DZ=DU+DV=\frac{3}{4}$$

以及 $Cov(Z,U)=D(U)=13/18$,

所以，$\rho_{ZU}=\frac{Cov(Z,U)}{\sqrt{D(Z)D(U)}}=\sqrt{\frac{D(U)}{D(Z)}}=\sqrt{\frac{26}{27}}$ ，

七、(**12** 分)

解：1. 总体 X 的一阶矩为 $\mu_1 = EX = \theta$

解得 $\theta = \mu_1$

以$A_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$代替$\mu_1$，得参数$\theta$的矩估计量为

$$\hat{\theta} = \bar{X}.$$

2.

似然函数为

$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{x_i}{\theta}} = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} x_i}.$$

对数似然函数为

$$\ln L(\theta) = -n\ln\theta - \frac{1}{\theta}\sum_{i=1}^{n} x_i.$$

对θ求导并令导数为零，得对数似然方程为

$$\frac{\partial \ln L(\theta)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n} x_i = 0,$$

解得θ的最大似然估计值为

$$\hat{\theta} = \bar{x}.$$

因为$D(X) = \theta^2$，由最大似然估计的不变性可得$D(X)$的最大似然估计值为

$$\widehat{D(X)} = \bar{x}^2.$$

八、（14 分）

1、第一类错误：原假设成立时拒绝原假设

第二类错误：原假设不成立时，接受原假设

2、提出假设

$$H_0:\sigma^2 \le \sigma_0^2 = 0.016, H_1:\sigma^2 > \sigma_0^2 = 0.016$$

检验统计量：$\chi^2 = \dfrac{(n-1)S^2}{\sigma_0^2}$.

拒绝域：$\dfrac{(n-1)s^2}{\sigma_0^2} \ge \chi_\alpha^2(n-1)$.

查表得 $\chi_\alpha^2(n-1) = \chi_{0.05}^2(24) = 36.415$，计算得 $\dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{24\times 0.025}{0.016} = 37.5$

因为 $37.5 > 36.415$,

所以在显著性水平 $\alpha = 0.05$ 下拒绝原假设，认为这批零件不合格。

2020 级概率与数理统计试题

一、（14 分）

1. 解：$P(\bar{A}\cup B)=P(\bar{A})+P(B)-P(\bar{A}B)=P(\bar{A})+P(B)-[P(B)-P(AB)]$

$$=1-P(A)+P(AB)=1-0.5+0.4=0.9$$

2. 解：设 B_i 表示第一次比赛取到 i 个新球，$i=1,2,3$，A 表示第二次比赛取到 1 个新球. 则

$$P(B_1)=\frac{C_8^1C_2^2}{C_{10}^3}=\frac{1}{15},P(B_2)=\frac{C_8^2C_2^1}{C_{10}^3}=\frac{7}{15},P(B_3)=\frac{C_8^3}{C_{10}^3}=\frac{7}{15}$$

$$P(A\mid B_1)=\frac{7}{10},P(A\mid B_2)=\frac{6}{10},P(A\mid B_3)=\frac{5}{10}$$

（1）由全概率公式得

$$P(A)=\sum_{i=1}^{3}P(B_i)P(A\mid B_i)=\frac{1}{15}\times\frac{7}{10}+\frac{7}{15}\times\frac{6}{10}+\frac{7}{15}\times\frac{5}{10}=\frac{14}{25}.$$

(2) 根据 Bayes 公式可得所求的概率为

$$P(B_1\mid A)=\frac{P(B_1)P(A\mid B_1)}{\sum_{i=1}^{3}P(B_i)P(A\mid B_i)}=\frac{\frac{7}{15}\times\frac{6}{10}}{\frac{14}{25}}=\frac{1}{2}.$$

二、（14 分）

解：1.易知 X 的可能的取值为 0，1，2，3，则

$$P(X=0)=\frac{A_3^1}{A_6^1}=0.5, P(X=1)=\frac{A_3^1A_3^1}{A_6^2}=0.3, P(X=2)=\frac{A_3^2A_3^1}{A_6^3}=0.15, P(X=3)=\frac{A_3^3A_3^1}{A_6^4}=0.05.$$

即 X 的分布律为

X	0	1	2	3
P	0.5	0.3	0.15	0.05

（2）$EX=0\times0.5+1\times0.3+2\times0.15+3\times0.05=0.75$

2. （1）$f_X(x)=\begin{cases}2e^{-2x}, & x>0\\ 0, & x\le 0\end{cases}$

（2）易知当 $x>0$ 时，$y=1-e^{-2x}$ 为单调函数，且 $0<y<1$

解得 $x=-\frac{1}{2}\ln(1-y)$，且 $x'_y=\frac{1}{2(1-y)}$

所以 Y 的密度函数为

$$f_Y(y)=\begin{cases}f_X(-\frac{1}{2}\ln(1-y))\left|\frac{1}{2(1-y)}\right|, & 0<y<1\\ 0, & \text{其它}\end{cases}=\begin{cases}1, & 0<y<1\\ 0, & \text{其它}\end{cases}$$

三、(16 分)

1. 命题 A 成立，

命题 B 不成立，

加 X 与 Y 独立的条件，或者 X 与 Y 的任意非零线性组合仍然是正态分布。

2. 解：(1) 由 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$ 得 $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = \int_0^1 dx\int_x^1 cdy = \frac{1}{2}c = 1$

解得 $c = 2$ 。

(2)

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

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \begin{cases} \int_0^y 2dx, & 0 < y \le 1 \\ 0, & \text{其他} \end{cases} = \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & \text{其他} \end{cases}$$

（3）X 与 Y 不独立，因为 $f(x,y) = f_X(x)f_Y(y)$ 并非几乎处处成立。

（4）$$f_Z(z) = \int_{-\infty}^{+\infty} f(x,z-x)dx = \begin{cases} \int_0^{z/2} 2dx, 0 < z \le 1 \\ \int_{z-1}^{z/2} 2dx, 1 < z \le 2 \\ 0, \quad \text{其它} \end{cases} = \begin{cases} z, & 0 < z \le 1 \\ 2-z, 1 < z \le 2 \\ 0, & \text{其它} \end{cases}$$

四、（14 分）

1. 两个随机变量的相关系数的绝对值|ρ|表达了两个随机变量线性相关的程度，|ρ|越大，线性相关的程度越大。

2. 命题 A 成立。命题 B 不成立，反例：（X,Y）服从圆域上的均匀分布，X 和 Y 不相关，但不独立。

3. 解：由已知：$X \sim N(0,1)$，$DY=4$，且 X 与 Y 相互独立.

因此，$EX=0$， $EY=2$，$D(X)=1$，$D(Y)=4$，$EX^2=1$， $EY^2=8$.

（1）$E(X-2Y)=E(X)-2E(Y)=-4$，$D(X-2Y)=D(X)+4D(Y)=17$.

（2）$E(XY)=EXEY=0$，$D(XY)=E(XY)^2-E^2(XY)=EX^2EY^2=8$

（3）$\rho_{UV}=\dfrac{Cov(U,V)}{\sqrt{D(U)D(V)}}=\dfrac{Cov(X+Y,X-Y)}{\sqrt{D(X+Y)D(X-Y)}}=\dfrac{D(X)-D(Y)}{D(X)+D(Y)}=-\dfrac{3}{5}$.

（4）因为 $\rho_{UV}=-\dfrac{3}{5}\neq 0$，即 U 与 V 相关，所以不独立。

五、（8 分）

解：：设 X_i 为第 i 个零件的质量，i=1,2,…,100. 则

$$EX_i = 0.5 \qquad DX_i = 0.01.$$

令 $S = \sum_{i=1}^{100} X_i$ 表示 100 只零件的总质量

由中心极限定理，$\dfrac{S-100\times0.5}{\sqrt{100\times0.01}} = \dfrac{S-50}{10} \overset{\text{近似}}{\sim} N(0,1)$

因此，$P(S>51) = P(S-50>51-50) \approx 1-\Phi(1) = 1-0.8413 = 0.1587$

六、(8 分)

解：易知 $X_i \sim N(\mu,\sigma^2), i=1,2,\dots,6$，且互相独立

由抽样分布定理知 $\bar{X}=\frac{1}{5}\sum_{i=1}^{5}X_i \sim N(\mu,\frac{\sigma^2}{5})$，$\frac{4S^2}{\sigma^2}\sim\chi^2(4)$。

有正态分布的性质得，$X_6-\bar{X}\sim N(0,\frac{6\sigma^2}{5})$，进而有 $\frac{X_6-\bar{X}}{\sqrt{6\sigma^2/5}}\sim N(0,1)$

易知 $\frac{X_6-\bar{X}}{\sqrt{6\sigma^2/5}}$ 与 $\frac{4S^2}{\sigma^2}$ 互相独立，

由 t 分布的性质得

$$\frac{\dfrac{X_6-\bar{X}}{\sqrt{6\sigma^2/5}}}{\sqrt{\dfrac{4S^2}{\sigma^2}\Big/4}}\sim t(4)$$，即 $\sqrt{\frac{5}{6}}\frac{X_6-\bar{X}}{S}\sim t(4)$

所以 $a=\sqrt{\frac{5}{6}}$

七、（14 分）

解: 1、由于 $\mu_1 = EX = \lambda$

解得 $\lambda = \mu_1$

用 $\bar{X}$ 代替 μ_1,得 λ 的矩估计为 $\hat{\lambda} = \bar{X}$

而样本值为 $\bar{x} = 1.75$

故 λ 的矩估计值为 $\hat{\lambda} = 1.75$

2、先求 θ 的最大似然估计

似然函数为 $L(\theta) = \prod_{i=1}^{n} \theta x_i^{\theta-1} = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta-1}$

对数似然函数为 $\ln L(\theta) = n \ln \theta + (\theta - 1) \ln \left(\prod_{i=1}^{n} x_i \right)$

对 θ 求导并令其为零，得 $\frac{d \ln L(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n} \ln x_i = 0$

解得 θ 的最大似然估计值为 $\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln x_i}$

最大似然估计量为 $\hat{\theta} = \frac{-n}{\sum_{i=1}^{n} \ln X_i}$

$\beta = e^{-1/\theta}$ 的最大似然估计为 $\hat{\beta} = e^{-1/\hat{\theta}} = e^{\sum_{i=1}^{n} \ln X_i / n}$ 。

3. $E\hat{\mu}_1 = E(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3) = \frac{1}{3}EX_1 + \frac{1}{3}EX_2 + \frac{1}{3}EX_3 = \frac{1}{3}EX + \frac{1}{3}EX + \frac{1}{3}EX = \mu$

$E\hat{\mu}_2 = E(\frac{2}{3}X_1 - \frac{5}{9}X_2 + \frac{8}{9}X_3) = \frac{2}{3}EX_1 - \frac{5}{9}EX_2 + \frac{8}{9}EX_3 = \frac{2}{3}EX - \frac{5}{9}EX + \frac{8}{9}EX = \mu$

所以 $\hat{\mu}_1$ 和 $\hat{\mu}_2$ 都是μ的无偏估计。

又因为

$D\hat{\mu}_1 = D(\frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3) = \frac{1}{9}DX_1 + \frac{1}{9}DX_2 + \frac{1}{9}DX_3 = \frac{1}{9}DX + \frac{1}{9}DX + \frac{1}{9}DX = \frac{1}{3}\sigma^2$

$D\hat{\mu}_2 = D(\frac{2}{3}X_1 - \frac{5}{9}X_2 + \frac{8}{9}X_3) = \frac{4}{9}DX_1 + \frac{25}{81}DX_2 + \frac{64}{81}DX_3 = \frac{4}{9}DX + \frac{25}{81}DX + \frac{64}{81}DX = \frac{125}{81}DX$

易知 $D\hat{\mu}_1 < D\hat{\mu}_2$，所以 $\hat{\mu}_1$ 更有效。

八、（12分）

1. 第二类错误；第一类错误

2. 解： $H_0: \mu = \mu_0 = 500 \quad H_1: \mu \neq \mu_0 = 500$

检验统计量 $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

拒绝域 $|z| \geq z_{\alpha/2}$

查表得 $z_{0.025} = 1.96$

本题中 $n = 9, \alpha = 0.05, \bar{x} = 499.5, \sigma = 1.2$

计算得 $|z| = \left| \dfrac{499.5 - 500}{1.2/3} \right| = 1.25 < 1.96$

未落入拒绝域，接受原假设，即认为这天罐装饮料符合要求。

概率论公式

一、随机事件与概率

公式名称	公式表达式
德摩根公式	$\overline{A\cup B}=\bar{A}\cap\bar{B}$，$\overline{A\cap B}=\bar{A}\cup\bar{B}$
古典概型	$P(A)=\frac{m}{n}=\frac{A\text{包含的基本事件数}}{\text{基本事件总数}}$
几何概型	$P(A)=\frac{\mu(A)}{\mu(\Omega)}$，其中μ为几何度量（长度、面积、体积）
求逆公式	$P(\bar{A})=1-P(A)$
加法公式	$P(A\cup B)=P(A)+P(B)-P(AB)$ 当 $P(AB)=0$ 时，$P(A\cup B)=P(A)+P(B)$
减法公式	$P(A-B)=P(A)-P(AB)$，$B\subset A$ 时 $P(A-B)=P(A)-P(B)$
条件概率公式与乘法公式	$P(B\mid A)=\frac{P(AB)}{P(A)}$　　$P(AB)=P(A)P(B\mid A)=P(B)P(A\mid B)$ $P(ABC)=P(A)P(B\mid A)P(C\mid AB)$
全概率公式	$P(A)=\sum_{i=1}^{n}P(B_i)P(A\mid B_i)$
贝叶斯公式（逆概率公式）	$P(B_i\mid A)=\frac{P(B_i)P(A\mid B_i)}{\sum_{i=1}^{n}P(B_i)P(A\mid B_i)}$
两个事件相互独立	$P(AB)=P(A)P(B)$；$P(B\mid A)=P(B)$；$P(B\mid A)=P(B\mid\bar{A})$；

二、随机变量及其分布

1、分布函数

$$F(x)=P(X\le x)=\begin{cases}\sum\limits_{x_k\le x}P(X=x_k)\\ \int_{-\infty}^{x}f(t)dt\end{cases},\quad P(a<X\le b)=F(b)-F(a)$$

2、离散型随机变量及其分布

分布名称	分布律
0-1 分布 X~b(1,p)	$P(X=k)=p^k(1-p)^{1-k},\quad k=0,1$
二项分布(贝努利分布) X~B(n,p)	$P(X=k)=C_n^k p^k(1-p)^{n-k},\quad k=0,1,\cdots,n$
泊松分布 X~p(λ)	$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda},\quad k=0,1,2,\cdots$

3、续型随机变量及其分布

分布名称	密度函数	分布函数
均匀分布 x~U(a,b)	$f(x)=\begin{cases}\frac{1}{b-a}, & a<x<b\\ 0, & 其他\end{cases}$	$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x<b\\ 1, & x\ge b\end{cases}$

分布名称	密度函数	分布函数
指数分布 X~E(λ)	$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$	$F(x)=\begin{cases}1-e^{-\lambda x}, & x>0\\ 0, & x\le 0\end{cases}$
正态分布 x~N(μ,σ^2)	$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<+\infty$	$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\mathrm{d}t$
标准正态分布 x~N(0,1)	$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ $-\infty<x<+\infty$	$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{1}{2}t^2}dt$

分布函数

对连续型随机变量 $F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)dt$

对离散型随机变量 $F(x) = P(X \le x) = \sum_{k \le x} P(X = k)$

分布函数与密度函数的重要关系： $F'(x) = f(x)$

4、随机变量函数 Y=g(X)的分布

离散型： $P(Y = y_i) = \sum_{g(x_j)=y_i} p_j, i = 1, 2, \cdots,$

连续型： ①分布函数法，

②公式法 $f_Y(y) = f_X(h(y)) \cdot |h'(y)| (x = h(y)$单 调)

h(y)是 g(x)的反函数

三、多维随机变量及其分布

1、离散型二维随机变量及其分布

分布律：$P(X=x_i,Y=y_j)=p_{ij},i,j=1,2,\cdots$ 分布函数 $F(X,Y)=\sum_{x_i\le x}\sum_{y_j\le y}p_{ij}$

边缘分布律：$p_{i\cdot}=P(X=x_i)=\sum_j p_{ij}\quad p_{\cdot j}=P(Y=y_j)=\sum_i p_{ij}$

条件分布律：$P(X=x_i|Y=y_j)=\dfrac{p_{ij}}{p_{\cdot j}},i=1,2,\cdots$，$P(Y=y_j|X=x_i)=\dfrac{p_{ij}}{p_{i\cdot}},j=1,2,\cdots$

2、连续型二维随机变量及其分布

①分布函数及性质

分布函数：$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv$

性质：$F(+\infty,+\infty)=1,\dfrac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y),P((x,y)\in G)=\iint_G f(x,y)dxdy$

②边缘分布函数与边缘密度函数

分布函数：$F_X(x)=\int_{-\infty}^{x}\int_{-\infty}^{+\infty}f(u,v)dvdu$ 密度函数：$f_X(x)=\int_{-\infty}^{+\infty}f(x,v)dv$

$F_Y(y)=\int_{-\infty}^{y}\int_{-\infty}^{+\infty}f(u,v)dudv$ $\qquad f_Y(y)=\int_{-\infty}^{+\infty}f(u,y)du$

③条件概率密度

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)},-\infty<y<+\infty,\quad f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)},-\infty<x<+\infty$$

3、随机变量的独立性

随机变量 X、Y 相互独立 $\Leftrightarrow F(x,y)=F_X(x)F_Y(y)$，

离散型：$p_{ij}=p_{i.}p_{.j}$，连续型：$f(x,y)=f_X(x)f_Y(y)$

4、二维随机变量和函数的分布(卷积公式)

离散型：$P(Z=z_k)=\sum\limits_{x_i+y_j=z_k}P(X=x_i,Y=y_j)$ 注意部分可加性

连续型：$f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx=\int_{-\infty}^{+\infty}f(z-y,y)dy$

四、随机变量的数字特征

1、数学期望

①定义：离散型 $E(X)=\sum_{k=1}^{+\infty} x_k p_k$，连续型 $E(X)=\int_{-\infty}^{+\infty} xf(x)dx$

②性质：$E(C)=C$，$E[E(X)]=E(X)$，$E(CX)=CE(X)$，$E(X\pm Y)=E(X)\pm E(Y)$

$E(aX\pm b)=aE(X)\pm b$，当 X、Y 相互独立时：$E(XY)=E(X)E(Y)$（正对逆错）

2、方差

①定义：$D(X)=E[(X-E(X))^2]=E(X^2)-E^2(X)$

②性质：$D(C)=0$，$D(aX\pm b)=a^2D(X)$，$D(X\pm Y)=D(X)+D(Y)\pm 2Cov(X,Y)$

当 X、Y 相互独立时：$D(X\pm Y)=D(X)+D(Y)$

3、协方差与相关系数

①协方差：$Cov(X,Y)=E(XY)-E(X)E(Y)$，当 X、Y 相互独立时：$Cov(X,Y)=0$

②相关系数：$\rho_{XY}=\dfrac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$，当 X、Y 相互独立时：$\rho_{XY}=0$（X, Y 不相关）

③协方差和相关系数的性质：$Cov(X,X)=D(X)$，$Cov(X,Y)=Cov(Y,X)$

$Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$，$Cov(aX+c,bY+d)=abCov(X,Y)$

Cov(x,a)=0(a 为常数)，$D(aX\pm bY)=a^2D(X)+b^2D(Y)\pm 2abCov(X,Y)$

4、常见随机变量分布的数学期望和方差

分布	数学期望 EX	方差 DX
0-1 分布 $b(1,p)$	p	$p(1-p)$
二项分布 $b(n,p)$	np	$np(1-p)$
泊松分布 $P(\lambda)$	λ	λ
均匀分布 $U(a,b)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
正态分布 $N(\mu,\sigma^2)$	μ	σ^2
指数分布 $e(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

五、大数定律与中心极限定理

1、切比雪夫不等式

若 $E(X)=\mu, D(X)=\sigma^2$，对于任意 $\varepsilon>0$ 有 $P\{|X-E(X)|\geq\varepsilon\}\leq\frac{D(X)}{\varepsilon^2}$

2、大数定律(普通班不重要)：

①切比雪夫大数定律：若 $X_1 \cdots X_n$ 相互独立，

$E(X_i)=\mu_i, D(X_i)=\sigma_i^2$ 且 $\sigma_i^2\leq C$，则：$\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n}E(X_i), (n\to\infty)$

②伯努利大数定律：设 n_A 是 n 次独立试验中事件 A 发生的次数，p 是事件 A 在每次试验中发生的概率，则 $\forall\varepsilon>0$，有：$\lim_{n\to\infty}P\left(\left|\frac{n_A}{n}-p\right|<\varepsilon\right)=1$

③辛钦大数定律：若 $X_1,\cdots,X_n$ 独立同分布，且 $E(X_i)=\mu$，则 $\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow[n\to\infty]{P} \mu$

3、★中心极限定理

①列维—林德伯格中心极限定理：独立同分布的随机变量 $X_i(i=1,2,\cdots)$，均值为 μ，方差为 $\sigma^2>0$，当 n 充分大时有：$Y_n=(\sum_{k=1}^{n} X_k-n\mu)\Big/\sqrt{n}\sigma \xrightarrow{\sim} N(0,1)$

②棣莫弗—拉普拉斯中心极限定理：随机变量 $X\sim B(n,p)$，则对任意 x 有：

$$\lim_{n\to\infty} P\{\frac{X-np}{\sqrt{np(1-p)}}\le x\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x)$$

③近似计算：$$P(a\le\sum_{k=1}^{n} X_k\le b)\approx\Phi(\frac{b-n\mu}{\sqrt{n}\sigma})-\Phi(\frac{a-n\mu}{\sqrt{n}\sigma})$$

六、数理统计的基本概念

1、总体和样本的分布函数

设总体 X~F(x)，则样本的联合分布函数 $F(x_1,x_2\cdots x_n)=\prod_{k=1}^{n}F(x_k)$

2、统计量

样本均值：$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$，样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-n\overline{X}^2)$

样本标准差：$S=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}$，样本 k 阶原点距：$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k, k=1,2\cdots$

样本 k 阶中心距：$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k, k=1,2,3\cdots$

3、三大抽样分布

(1) χ^2 分布：设随机变量 X~B(0, 1) $(i=1,2,\cdots,n)$ 且相互独立，则称统计量

$\chi^2=X_1^2+X_2^2+\cdots X_n^2$ 服从自由度为 n 的 χ^2 分布，记为 $\chi^2\sim\chi^2(n)$

性质：① $E[\chi^2(n)]=n, D[\chi^2(n)]=2n$ ②设 $X\sim\chi^2(m), Y\sim\chi^2(n)$ 且相互独立，则

$X+Y\sim\chi^2(m+n)$

(2) t 分布：设随机变量 $X\sim N(0,1), Y\sim\chi^2(n)$，且 X 与 Y 独立，则称统计量：

$T=\dfrac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布，记为 $T\sim t(n)$

性质：① $E(T)=0\ (n>1), D(T)=\dfrac{n}{n-2}\ (n>2)$ ② $\lim\limits_{n\to\infty} f_n(x)=\varphi(x)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

(3) F 分布：设随机变量 $X\sim\chi^2(m), Y\sim\chi^2(n)$，且 X 与 Y 独立，则称统计量