南京审计大学
NANJING AUDIT UNIVERSITY

概率统计学习讲座

数学学院　陆伟东

一、随机事件

1. 事件间的关系

事件的包含、互不相容事件、对立事件、事件的独立性

2. 随机事件的运算

$$A-B=A\cdot\overline{B} \qquad \overline{A+B}=\overline{A}\cdot\overline{B} \qquad \overline{A\cdot B}=\overline{A}+\overline{B}$$

二、随机事件的概率

1. 古典概型与几何概型

(1) 古典概型

(2) 几何概型

2. 概率的基本公式

(1) 加法公式 $P(A+B)=P(A)+P(B)-P(AB)$

$P(A+B)=P(A)+P(B)$ (A、B互不相容)

(2) 减法公式 $P(A-B)=P(A)-P(AB)$

$P(A-B)=P(A)-P(B)$ ($A\supset B$)

(3) 乘法公式 $P(AB)=P(A)P(B|A)$

$P(AB)=P(A)P(B)$ (A、B相互独立)

(4) 条件概率公式 $P(B|A)=\dfrac{P(AB)}{P(A)}$

(5) 逆事件概率公式 $P(\overline{A})=1-P(A)$

例1：设 A、B 为两个随机事件，$P(A)=0.92, P(B)=0.93$，$P(B\mid\overline{A})=0.85$，求：

(1) $P(\overline{A}B)$　(2) $P(AB)$　(3) $P(A+B)$

$\boxed{=0.068}$　$\boxed{=0.862}$　$\boxed{=0.988}$

例2：甲、乙、丙三人分别向同一目标射击，设甲命中的概率为0.8，乙命中的概率为0.7，丙命中的概率为0.6，求：目标被命中的概率．$\boxed{=0.976}$

3. 全概率公式与贝叶斯公式

(1) 全概率公式

$$P(B)=\sum_{i=1}^{n}P(A_i)P(B|A_i)$$

其中 $A_1,A_2,\cdots,A_n$ 构成一个完备事件组，且 $P(A_i)>0$

(2) 贝叶斯公式

$$P(A_i|B)=\frac{P(A_iB)}{P(B)}=\frac{P(A_i)P(B|A_i)}{\sum_j P(A_j)P(B|A_j)}$$

例3： 已知一批产品中有90%是合格品，检查产品质量时，一个合格品被误判为次品的概率为0.05，一个次品被误判为合格品的概率为0.04，求：

(1) 任意抽查一个产品，它被判为合格品的概率；$\boxed{=0.859}$

(2) 一个经检查被判为合格的产品确实是合格品的概率. $\boxed{=0.995}$

三、随机变量的分布与数字特征

1. 离散型随机变量

概率分布表：

X	x_1	x_2	$\cdots$	x_i	$\cdots$
P	p_1	p_2	$\cdots$	p_i	$\cdots$

2. 连续型随机变量

概率密度函数：$f(x)$

3. 随机变量函数的分布

分布函数法

例4：设连续型随机变量 X 的概率密度函数为

$$f(x)=\begin{cases} ax & 0<x<2 \\ 0 & \text{其他} \end{cases}$$

求：(1) 常数 a $\boxed{=1/2}$

(2) 分布函数 $F(x)$ $\boxed{=\begin{cases} 0 & x\le 0 \\ \frac{1}{4}x^2 & 0<x<2 \\ 1 & x\ge 2 \end{cases}}$

(3) $P\{1\le X\le \sqrt{2}\}$ $\boxed{=1/4}$

(4) EX $\boxed{=4/3}$

(5) DX $\boxed{=2/9}$

(6) $Y=2X+1$的概率密度函数 $\boxed{f_Y(y)=\begin{cases} \frac{y-1}{8} & 1<y<5 \\ 0 & \text{其他} \end{cases}}$

四、常用的随机变量分布

1. 常用的离散型分布

(1) $0-1$分布

X	0	1
P	$1-p$	p

$EX = p \qquad DX = p(1-p)$

(2) 二项分布 $X \sim b(n,p)$

$$P\{X=k\}=C_n^k p^k (1-p)^{n-k} \qquad k=0,1,2,\cdots,n$$

$$EX = np \qquad DX = np(1-p)$$

(3) 泊松分布 $X \sim P(\lambda)$

$$P\{X=k\}=\frac{\lambda^k}{k!}e^{-\lambda} \quad \lambda>0,\ k=0,1,2,\cdots$$

$$EX=\lambda \qquad DX=\lambda$$

2. 常用的连续型分布

(1) 均匀分布 $X \sim U[a,b]$

$$f(x)=\begin{cases}\dfrac{1}{b-a} & a \le x \le b \\ 0 & \text{其他}\end{cases} \qquad EX=\frac{a+b}{2} \quad DX=\frac{(b-a)^2}{12}$$

(2) 指数分布 $X \sim e(\lambda)$

$$f(x)=\begin{cases}\lambda e^{-\lambda x} & x \ge 0 \\ 0 & x<0\end{cases} \quad (\lambda>0) \quad EX=\frac{1}{\lambda} \quad DX=\frac{1}{\lambda^2}$$

(3) 正态分布 $X \sim N(\mu,\sigma^2)$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad EX=\mu \qquad DX=\sigma^2$$

例5：设顾客在某银行的窗口等待服务的时间 X (以分计) 服从参数为 $\lambda=\dfrac{1}{5}$ 的指数分布，某顾客在窗口等待服务，若超过 10 分钟，他就离开，求：

(1) 该顾客未等到服务而离开的概率；$\boxed{=e^{-2}}$

(2) 若该顾客一个月要到银行 4 次，求他至少有一次未等到服务就离开的概率. $\boxed{=1-(1-e^{-2})^4}$

五、二维随机向量的分布与数字特征

1. 离散型随机向量

(1) 联合概率分布：$P\{X=x_i, Y=y_j\}=p_{ij}$ ，$i,j=1,2,\cdots$

(2) 边缘概率分布：$p_i^X=\sum_j p_{ij} \qquad i=1,2,\cdots$

$$p_j^Y=\sum_i p_{ij} \qquad j=1,2,\cdots$$

(3) $Z=g(X,Y)$ 的概率分布：

$$P\{Z=z_k\}=P\{g(X,Y)=z_k\}$$

$$=\sum_{g(x_i,y_j)=z_k} P\{X=x_i, Y=y_j\},\ k=1,2,\cdots$$

(4) $E[g(X,Y)]=\sum_{i,j} g(x_i,y_j)p_{ij}$

(5) $\operatorname{cov}(X,Y)$ 与 $\rho_{X,Y}$

例6：袋中有四张卡片，分别写有数字1,2,3,4，用有放回的方式先后从袋中抽取两张卡片，用 X,Y 分别表示取到的卡片上的最小和最大数字，求：

(1) (X,Y) 的联合分布；

X \ Y	1	2	3	4
1	1/16	1/8	1/8	1/8
2	0	1/16	1/8	1/8
3	0	0	1/16	1/8
4	0	0	0	1/16

(2) 求 X,Y 的协方差

$$\operatorname{cov}(X,Y)=\frac{25}{64}$$

2. 连续型随机向量

(1) 联合密度函数：$f(x,y)$

(2) 边缘密度函数：$f_X(x)=\int_{-\infty}^{+\infty} f(x,y)\mathrm{d}y$

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

(3) X 与 Y 相互独立 $\Leftrightarrow f(x,y)=f_X(x)f_Y(y)$

(4) $P\{(X,Y)\in D\}=\iint\limits_D f(x,y)\,\mathrm{d}x\mathrm{d}y$

(5) $E[g(X,Y)]=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} g(x,y)f(x,y)\mathrm{d}x\mathrm{d}y$

例7：设随机向量(X,Y)的联合密度函数为

$$f(x,y)=\begin{cases}Axy^2 & 0<x<2,0<y<1\\ 0 & \text{其他}\end{cases}$$

求：(1) 常数 A；$\boxed{=\dfrac{3}{2}}$

(2) 求 X,Y 的边缘密度函数，并判断 X,Y 是否独立；

$$\boxed{f_X(x)=\begin{cases}\dfrac{1}{2}x, & 0\le x\le 2\\ 0, & \text{其他}\end{cases},\quad f_Y(y)=\begin{cases}3y^2, & 0\le y\le 1\\ 0, & \text{其他}\end{cases},X,Y\text{独立}}$$

(3) 计算(X,Y)落在以$(0,0),(0,2),(2,1)$为顶点的三角形中的概率

$\boxed{=0.6}$

3. 关于数字特征的一些重要结论

(1) $E(X \pm Y) = EX \pm EY$

(2) X, Y 独立时，$E(XY) = EXEY$

(3) $D(X \pm Y) = DX + DY \pm 2\operatorname{cov}(X, Y)$

(4) X, Y 独立时，$D(X \pm Y) = DX + DY$

(5) X, Y 独立时, $\operatorname{cov}(X, Y) = 0$

(6) X, Y 独立时, $\rho_{X,Y} = 0$ (不相关)

例8: 已知随机变量 X, Y 不相关，则下列不一定成立的是 (C)

(A) $D(X+Y) = DX + DY$

(B) $E(XY) = EXEY$

(C) X, Y 独立

(D) $E(X+Y) = EX + EY$

例9：已知随机变量 X,Y 相互独立，$X \sim b(16,0.5)$，Y 服从参数为 9 的泊松分布，则 $D(X-2Y)=($ **B** $)$

$(A)\ -32$ $(B)\ 40$ $(C)\ -14$ $(D)\ 22$

例10：已知随机变量 $X \sim b(16,0.5)$，Y 服从参数为 9 的泊松分布，且 X 与 Y 的相关系数为0.3，求 $\mathrm{cov}(X,Y)$ $\boxed{=1.8}$

六、中心极限定理

1. 林德伯格—勒维中心极限定理

设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一列独立同分布的随机变量，

且 $E\xi_i=\mu$, $D\xi_i=\sigma^2>0$, $i=1,2,\cdots$，则有

$$P\left\{\frac{\sum_{i=1}^{n}\xi_i-n\mu}{\sqrt{n}\,\sigma}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t$$

2. 棣莫弗—拉普拉斯中心极限定理

设 $X_n\sim b(n,p)$, $0<p<1$，则

$$\lim_{n\to\infty}P\left\{\frac{X_n-np}{\sqrt{npq}}\le x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\mathrm{d}t$$

例11：一包装工平均3分钟完成一项包装，假设他完成一件包装所用时间独立且服从指数分布，试求完成100件包装所需时间在5到6小时的概率. $\boxed{= 0.47725}$

例12：某系统由100个部件组成，运行期间每个部件是否损坏是相互独立的，损坏的概率均为0.1，如果有85个及以上的部件完好，则系统可以正常工作，试求系统正常工作的概率.

$\boxed{= 0.95254}$

七、统计量及抽样分布

1. 统计量

2. 常用的统计分布

(1) χ^2 分布 $X_1, X_2, \cdots, X_n$ 相互独立，且 $X_i \sim N(0,1)$,

则 $X = X_1^2 + X_2^2 + \cdots + X_n^2 \sim \chi^2(n)$

(2) F 分布 $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$，且 X 和 Y 相互独立，

则 $\dfrac{X/m}{Y/n} \sim F(m,n)$

(3) t 分布 $X \sim N(0,1)$, $Y \sim \chi^2(n)$，且 X 和 Y 相互独立，

则 $\dfrac{X}{\sqrt{Y/n}} \sim t(n)$

例13： 设总体 $X \sim N(0,9)$, $X_1, X_2, \cdots, X_8$ 是从总体中抽取的一个简单随机样本，试证明：$Y = \dfrac{X_1 + X_2 + X_3 + X_4}{\sqrt{X_5^2 + X_6^2 + X_7^2 + X_8^2}} \sim t(4)$

3. 抽样分布

设总体 $X \sim N(\mu, \sigma^2)$, $(X_1, X_2, \cdots, X_n)$ 是其容量为 n 的一个样本, $\overline{X}$ 与 S^2 分别为此样本的样本均值与样本方差，则有

(1) $\overline{X} \sim N(\mu, \dfrac{\sigma^2}{n})$;

(2) $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$;

(3) $\overline{X}$ 与 S^2 相互独立.

八、点估计

1. 点估计的评价标准

(1) 无偏性　(2) 有效性　(3) 相合性

2. 矩估计

3. 最大似然估计

例 14：设总体的密度函数为 $f(x)=\begin{cases}(\alpha+1)x^{\alpha} & 0<x<1\\ 0 & \text{其他}\end{cases}$ $(\alpha>-1)$，

$(X_1,X_2,\cdots,X_n)$ 是取自总体 X 的样本，求：

(1) α 的矩估计 $\boxed{\hat{\alpha}=\dfrac{1-2\overline{X}}{\overline{X}-1}}$

(2) α 的最大似然估计 $\boxed{\hat{\alpha}=-1-\dfrac{n}{\sum\limits_{i=1}^{n}\ln X_i}}$

九、区间估计

1. 方差 σ^2 已知，均值 μ 的置信区间 $\left(\overline{X}-u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right)$

2. 方差 σ^2 未知，均值 μ 的置信区间

$$\left(\overline{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\overline{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right)$$

3. 方差 σ^2 的置信区间 $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$

例15：假设某产品的零件长度 X 服从正态分布，今随机抽取16个配件，测得平均长度为3.05毫米，标准差为0.16毫米，试求 μ 的95%置信区间．

$\boxed{(2.96.3.14)}$

十、假设检验

1. 假设检验的基本思想：小概率原理

2. 检验的显著性水平与两类错误

3. 单正态总体的参数假设检验

(1) 方差 σ^2 已知，均值 μ 的检验

选取统计量 $U=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$

(2) 方差 σ^2 未知，均值 μ 的检验

选取统计量 $T=\dfrac{\overline{X}-\mu_0}{S/\sqrt{n}}$

(3) 方差 σ^2 的检验

选取统计量 $\chi^2=\dfrac{(n-1)S^2}{\sigma_0^2}$

例16：某洗衣粉包装机在正常工作情况下，每袋标准重量为1000克，标准差 σ 不能超过15克，假设每袋洗衣粉的净重服从正态分布，某人为检查机器工作是否正常，从已装好的袋中随机抽取10袋，测其净重(克)为：

$$1020, 1030, 968, 994, 1014, 998, 976, 982, 950, 1048$$

问：这天机器工作是否正常？ $(\alpha = 0.05)$

(1) $H_0: \mu = 1000$ 接受 H_0

(2) $H_0: \sigma^2 \le 15^2$ 拒绝 H_0