应用数理统计——第一章

1.1 概率基础

一、离散型随机变量的五大分布

分布类型	分布形式	分布律	期望	方差
两点分布（0-1 分布）	$X\sim B(1,p)$	$P\{X=k\}=p^k(1-p)^{1-k},k=0,1$	p	$p(1-p)$
二项分布	$X\sim B(n,p)$	$P\{X=k\}=C_n^k p^k(1-p)^{n-k},k=0,\dots,n$	np	$np(1-p)$
泊松分布	$X\sim P(\lambda)$	$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!},k=0,\dots,n$	λ	λ
超几何分布	$X\sim H(N,M,n)$	$P\{X=k\}=\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$, $k=l_1,l_1+1,\dots,l_2$, $l_1=\max\{0,n-(N-M)\}$, $l_2=\min\{0,M\}$	$\frac{nM}{N}$	--
几何分布	$X\sim G(p)$	$P\{X=k\}=(1-p)^{k-1}p,k=1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$

二、连续型随机变量的三大分布

分布类型	分布形式	概率密度	期望	方差
均匀分布	$X\sim U(a,b)$	$f(x)=\begin{cases}\frac{1}{b-a} & ,a<x<b\\ 0 & ,\text{其他}\end{cases}$	$\frac{a+b}{2}$	$\frac{(a-b)^2}{12}$
指数分布	$X\sim E(\lambda)$	$f(x)=\begin{cases}\lambda e^{-\lambda x} & ,x>0\\ 0 & ,x\le 0\end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态分布	$X\sim N(\mu,\sigma^2)$	$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

三、数字特征——期望、方差、协方差与相关系数

数字特征	定义	计算方法		性质
期望	取值乘概率求和	离散型	$E(X)=\sum_{i=1}^{n} x_i P_i$	$E(c)=c$
				$a<x<b \Rightarrow a<E(x)<b$
				$E(X+Y)=E(X)+E(Y)$
		连续型	$E(X)=\int_{-\infty}^{+\infty} xf(x)dx$	$E(cX)=cE(X)$
				独立：$E(XY)=E(X)E(Y)$
方差	$E[(X-E(X))^2]$	$E(X^2)-[E(X)]^2$		$D(c)=0$
				$D(X)\geq 0$
				$D(cX)=c^2D(X)$
				$D(X+Y)=D(X)+D(Y)+2Cov(X,Y)$
				独立：$D(X+Y)=D(X)+D(Y)$
协方差	$E[(X-E(X))(Y-E(Y))]$	$E(XY)-E(X)E(Y)$		$Cov(X,Y)=Cov(Y,X)$
				$Cov(X,c)=0$
				$Cov(aX,bY)=abCov(Y,X)$
				$Cov(X+Z,Y)=Cov(X,Y)+Cov(Z,Y)$
相关系数	$\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$	$\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$		$\rho(X,Y)=\rho(Y,X)$
				$\lvert\rho(X,Y)\rvert\leq 1$

1.2 数统基础

一、样本

从总体中取出的一个或具有代表性的部分个体，称为从总体中得到的一个样本。

二、统计量

1.充分统计量

利用因子分解定理，将似然函数 $f(x,\theta)$ 分解为两个因式的乘积：$f(x,\theta)=K(T(x),\theta)h(x)$。其中 $h(x)$ 仅与样本有关，$K(T(x),\theta)$ 与样本和参数均有关。则，$T(x)$ 是一个充分统计量。

2.完备统计量

若统计量 T 满足 $E_\theta g(T)=0 \Leftrightarrow P_\theta\{\mathrm{g(T)=0}\}=1$，则 T 为完备统计量。

3.顺序统计量

对于样本 $X_1,...,X_n$，对应观察值记为 $x_1,...,x_n$；按照样本观察值的大小关系排序：$x_{(1)}\le x_{(2)}\le...\le x_{(n)}$，相应的样本 $X_{(1)}\le X_{(2)}\le...\le X_{(n)}$，称为顺序统计量。

（1） 顺序统计量是充分统计量；
（2） 样本在顺序统计量中的位置称为“秩”。

三、数理统计五大分布

1.正态分布

（1）正态总体、样本和样本均值的分布

总体：$X\sim N(\mu,\sigma^2)$　　样本：$X_i\sim N(\mu,\sigma^2)$　　样本均值：$\bar{X}\sim N(\mu,\dfrac{\sigma^2}{n})$

（2）中心标准化

$X\sim N(\mu,\sigma^2)\xrightarrow{\text{中心标准化}}\dfrac{X-\mu}{\sigma}\sim N(0,1)$：（随机变量-期望）/标准差。

（3）正态总体的抽样分布定理

$X_1,X_2,...,X_n$ 为来自 $X\sim N(\mu,\sigma^2)$ 总体时，存在如下结论：

① $\dfrac{\sqrt{n}(\bar{X}-\mu)}{\sigma}\sim N(0,1)$　　② $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

③ $\bar{X}$ 与 S^2 独立　　　　④ $\dfrac{\sqrt{n}(\bar{X}-\mu)}{S} \sim t(n-1)$

（4）正态总体的联合分布

对于线性变换 $Y=AX$，若 $X \sim N(\vec{\theta}，\Sigma)$，则 $Y \sim N(A\vec{\theta}, A\Sigma A^T)$。

（5）标准正态分布的性质

①定义式：$P\{X<\alpha\}=\Phi(\alpha)$

②对称性：$\Phi(-\alpha)=1-\Phi(\alpha)$（负数）

$\Phi(|\alpha|)=2\Phi(\alpha)-1$（绝对值）

2. χ^2 分布

（1）定义：$K^2=\sum_{i=1}^{n} X_i^2, X_i \sim N(0,1) \Rightarrow K^2 \sim \chi^2(n)$（平方和分布）

（2）密度函数：$K_n=\dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}$

（3）期望：$E(K^2)=n$；方差：$D(K^2)=2n$

3. t 分布

（1）定义：$T=\dfrac{X}{\sqrt{Y/n}}, X \sim N(0,1), Y \sim \chi^2(n) \Rightarrow T \sim t(n)$（$\chi^2$ 开根比分布）

（2）密度函数：$T_n=\dfrac{\Gamma(\frac{n+1}{2})}{\sqrt{\pi n}\Gamma(\frac{n}{2})}(1+\dfrac{x^2}{n})^{-\frac{n+1}{2}}$

（3）期望：$E(T)=0(n\ge 2)$；方差：$D(T)=\dfrac{n}{n-2}(n\ge 3)$

4. F 分布

（1）定义：$F=\dfrac{X/m}{Y/n}, X \sim \chi^2(m), Y \sim \chi^2(n) \Rightarrow F \sim F(m,n)$（$\chi^2$ 比分布）

（2）密度函数：$f_{m,n}(X)=\dfrac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{m}{2})\Gamma(\frac{n}{2})} m^{\frac{m}{2}} n^{\frac{n}{2}} \dfrac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}$

（3）期望：$E(F)=\dfrac{n}{n-2}(n\ge 3)$

（4）性质 1：$T \sim t(n)$，则$T^2 \sim F(1,n)$。

（5）性质 2：$F \sim F(m,n)$，则$\frac{1}{F} \sim F(n,m)$。

5.Γ分布

（1）密度函数：$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, x > 0. \Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx, \alpha > 0$

（2）$\Gamma(1,\lambda)$为指数分布，$\Gamma(\frac{n}{2},\frac{1}{2})$为$\chi^2$分布.

（3）期望：$E(X) = \frac{\alpha}{\lambda}$；方差：$D(X) = \frac{\alpha}{\lambda^2}$

（4）Γ分布具有可加性

（5）$X \sim \Gamma(\alpha,\lambda)$，则$cX \sim \Gamma(\alpha,\lambda / c)$

1.3 题型总结

一、正态总体的样本容量问题

1.题型描述： 给出样本容量或求解样本容量大小的题目。

2.解题策略：

（1）求解样本均值的分布 $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$；

（2）根据题中要求建立等式或不等式关系；

（3）对样本均值进行中心标准化求解 $\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} \sim N(0,1)$。

（4）利用标准正态性质并查表。

3.注意事项：

（1）关注样本间的独立性问题；
（2）关注隐含性质条件。

4.题解：

例 1：设总体 $\xi \sim N(a, 0.5)$，若要以 0.997 的概率保证偏差 $|\bar{\xi}-a|<0.1$，试问<u>样本容量</u>n<u>应取多大</u>？

解：

（1） 求样本均值的分布：$\bar{\xi} \sim N(a, \frac{0.5}{n})$

（2） 根据题中要求建立不等式关系：$P\{|\bar{\xi}-a|<0.1\} \ge 0.997$

（3） 对样本均值进行中心标准化：$P\{\frac{\sqrt{n}|\bar{\xi}-a|}{\sqrt{0.5}} < \frac{0.1\sqrt{n}}{\sqrt{0.5}}\} \ge 0.997$

（4） 利用标准正态性质化简：$P\{\frac{\sqrt{n}|\bar{\xi}-a|}{\sqrt{0.5}} < \frac{0.1\sqrt{n}}{\sqrt{0.5}}\} = 2\Phi(\frac{0.1\sqrt{n}}{\sqrt{0.5}}) - 1 \ge 0.997$

（5） 查表：$n \ge 441.045$

例 2：设 $\xi_1, \xi_2, \ldots \xi_n$ 为总体 $\xi \sim N(a, 4)$ 的一个样本，$\bar{\xi}$ 为样本均值，试问<u>样本容量</u>n<u>至少应取多大</u>才能使：

（1）$E[|\bar{\xi}-a|^2] \le 0.1$　（2）$E[|\bar{\xi}-a|] \le 0.1$　（3）$P\{|\bar{\xi}-a| \le 0.1\} \ge 0.95$

解：

（1） $E[|\bar{\xi}-a|^2]\le 0.1$

①求样本均值的分布：$\bar{\xi}\sim N(a,\frac{4}{n})$

②*注意到隐含条件*：$E[|\bar{\xi}-a|^2]$表示$\bar{\xi}$的方差$D(\bar{\xi})=\frac{4}{n}$

③令$D(\bar{\xi})=\frac{4}{n}\le 0.1$得$n\ge 40$。

（2） $E[|\bar{\xi}-a|]\le 0.1$

①求样本均值的分布：$\bar{\xi}\sim N(a,\frac{4}{n})$

②*注意到隐含条件*：$E[|\bar{\xi}-a|]=\int_{-\infty}^{+\infty}|x|\frac{1}{\sqrt{2\pi}\sqrt{\frac{4}{n}}}e^{-\frac{x^2}{2\times\frac{4}{n}}}dx\le 0.1$

③利用期望的定义式积分求解得：$n\ge 254.65$

（3） $P\{|\bar{\xi}-a|\le 0.1\}\ge 0.95$

①求样本均值的分布：$\bar{\xi}\sim N(a,\frac{4}{n})$

②根据题中要求建立不等式关系：$P\{|\bar{\xi}-a|\le 0.1\}\ge 0.95$

③对样本均值进行中心标准化：$P\{\frac{\sqrt{n}\,|\bar{\xi}-a|}{2}\le\frac{0.1\sqrt{n}}{2}\}\ge 0.95$

④利用标准正态性质化简：$P\{\frac{\sqrt{n}\,|\bar{\xi}-a|}{2}\le\frac{0.1\sqrt{n}}{2}\}=2\Phi(\frac{0.1\sqrt{n}}{2})-1\ge 0.95$

⑤查表：$n\ge 1537$

例 3：设$\underline{\xi_1,\xi_2,\ldots\xi_{10}}$为总体$\xi\sim N(0,0.3^2)$的样本，求$P\{\frac{\xi_4^2}{(\xi_1-\bar{\xi}_3)^2}<10.73\}$。

解：

（1） 求样本均值的分布：$\bar{\xi}\sim N(0,\frac{0.3^2}{10})$

（2） *注意到$\bar{\xi}_3$与ξ_1不独立*，因此，此题属于独立性问题中非独立情况。

（3） $\xi_4\sim N(0,0.3^2),\xi_1-\bar{\xi}_3\sim N(0,\frac{10-1}{10}\times 0.3^2),$

（4） 对样本均值进行中心标准化：$P\{\frac{\xi_4^2}{0.3^2}/\frac{(\xi_1-\bar{\xi}_3)^2}{\frac{3-1}{3}\times 0.3^2}<10.73\times\frac{3}{2}\}$

（5） 查表：$P\{\frac{\xi_4^2}{0.3^2}/\frac{(\xi_1-\bar{\xi}_3)^2}{\frac{3-1}{3}\times 0.3^2}<10.73\times\frac{3}{2}\}$=0.95

二、顺序统计量问题

1.题型描述：存在顺序统计量的问题

2.解题策略：

（1）“大于最小、小于最大”表示“至少”→找对立事件
（2）“大于最大、小于最小”表示“全部”→求概率的 n 次方

3.题解：

例 1：设总体$\xi\sim N(12,4)$，$\underline{\xi_1,\xi_2,...\xi_5}$为$\xi$的一个样本，试求概率：

（1）$P\{\bar{\xi}>13\}$ （2）$\underline{P\{\xi_{(1)}<10\}}$ （3）$\underline{P\{\xi_{(5)}>15\}}$

解：

（1）$P\{\bar{\xi}>13\}$

①求样本均值的分布：$\bar{\xi}\sim N(12,0.8)$

②根据题中要求建立不等式关系：$P\{\bar{\xi}>13\}$

③对样本均值进行中心标准化：$P\{\frac{\bar{\xi}-12}{\sqrt{0.8}}>\frac{13-12}{\sqrt{0.8}}\}$

④利用标准正态性质化简：$P\{\frac{\bar{\xi}-12}{\sqrt{0.8}}>\frac{13-12}{\sqrt{0.8}}\}=1-\Phi(\frac{13-12}{\sqrt{0.8}})$

⑤查表：$P\{\frac{\bar{\xi}-12}{\sqrt{0.8}}>\frac{13-12}{\sqrt{0.8}}\}=0.1314$

（2）$P\{\xi_{(1)}<10\}$

①此题为顺序统计量中“大于最小”情况，表示“至少”含义

②构造对立事件“全部大于等于 10”：$P\{\xi_{(1)}<10\}=1-[P\{\xi\geq 10\}]^5$

③利用中心标准化并查表可得：$P\{\xi_{(1)}<10\}=1-[P\{\xi\geq 10\}]^5$=0.579

（3） $P\{\xi_{(5)} > 15\}$

①此题为顺序统计量中“小于最大”情况，表示“至少”含义

②构造对立事件“全部小于等于 15”：$P\{\xi_{(5)} > 15\} = 1 - [P\{\xi \le 15\}]^{5}$

③利用中心标准化并查表可得：$P\{\xi_{(5)} > 15\} = 1 - [P\{\xi \le 15\}]^5$=0.292

例 2：设电子元件寿命（时数）ξ 服从参数为 $\lambda = 0.0015$ 的指数分布，今测 6 个元件并记录下它们的失效时间，试问：

（1） 至 800 小时时没有一个元件失效的概率是多少？

（2） 至 3000 小时时所有元件都失效的概率是多少？

1-解：

①此题为顺序统计量中“小于最小”情况，表示“全部”含义

②求解概率的 6 次方如下：$P\{\xi_{(1)} > 800\} = [P\{\xi > 800\}]^6 = e^{-7.2}$

2-解：

①此题为顺序统计量中“大于最大”情况，表示“全部”含义

②求解概率的 6 次方如下：$P\{\xi_{(6)} < 3000\} = [P\{\xi < 3000\}]^6 = (1 - e^{-4.5})^6$

三、正态总体的线性变换问题

1.题型描述：与正态总体的线性变换相关的题目。

2.解题策略：

（1）写出线性变换矩阵 A

（2）计算线性变换后的均值向量、协方差矩阵：$\vec{\theta}_1 = A\theta, \Sigma_1 = A\Sigma A^T$

（3）计算线性变换后的方差为：$D(AX) = A\theta A^T$

3.题解：

例：设 $\xi^T = (\xi_1, \xi_2, \xi_3)$ 且 $Var(\xi) = \begin{pmatrix} 5 & 2 & 3 \\ 2 & 3 & 0 \\ 3 & 0 & 2 \end{pmatrix}$。

（1） 求 $\xi_1 - 2\xi_2 + \xi_3$ 的方差

（2） 求 $\eta = (\eta_1, \eta_2)^T$ 的协方差矩阵，其中 $\eta_1 = \xi_1 + \xi_2, \eta_2 = \xi_1 + \xi_2 + \xi_3$。

解：

（1） 线性变换矩阵为 $A=\begin{pmatrix}1 & -2 & 1\end{pmatrix}$

$$D(\xi_1-2\xi_2+\xi_3)=\begin{pmatrix}1 & -2 & 1\end{pmatrix}\begin{pmatrix}5 & 2 & 3\\ 2 & 3 & 0\\ 3 & 0 & 2\end{pmatrix}\begin{pmatrix}1\\ -2\\ 1\end{pmatrix}=17$$

（2） 线性变换矩阵为 $A=\begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 1\end{pmatrix}$

$$Var(\eta)=\begin{pmatrix}1 & 1 & 0\\ 1 & 1 & 1\end{pmatrix}\begin{pmatrix}5 & 2 & 3\\ 2 & 3 & 0\\ 3 & 0 & 2\end{pmatrix}\begin{pmatrix}1 & 1\\ 1 & 1\\ 0 & 1\end{pmatrix}=\begin{pmatrix}12 & 15\\ 15 & 20\end{pmatrix}$$

四、独立性问题

1.题型描述：存在多个样本时需要优先考虑独立性，当独立性起作用时的问题为独立性问题。

2.解题策略：

（1）协方差为 0 即独立；

（2）独立多维正态的线性变换满足 $AB^T=0$ 时，两种线性变换独立；

（3）常见的两类情况：

① X_{n+1} 与 $\bar{X}$ 独立， $X_{n+1}-\bar{X}\sim N(0,\dfrac{n+1}{n}\sigma^2)$

② X_n 与 $\bar{X}$ 不独立， $X_n-\bar{X}\sim N(0,\dfrac{n-1}{n}\sigma^2)$

五、证明题

- **常用公式：**

1. 样本与总体的关系：

$$\sum_{i=1}^{n}E(X_i)=nE(X),E(X_i)=E(X)$$

$$\sum_{i=1}^{n}D(X_i)=nD(X),D(X_i)=D(X)$$

2. 样本均值与总体的关系：

$$E(\bar{X})=E(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n}E(\sum_{i=1}^{n}X_i)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)$$

$$D(\bar{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}D(\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X)$$

3. 总体方差与样本方差：

$$D(X)=E(X^2)-[E(X)]^2=E[(X-E(X))^2]=\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

$$S^2(X)=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$