Basics of Probs and Stats

期末复习 (2)

梁雪吟

地球科学与资源学院
liangxueyin@cugb.edu.cn

Spring 2022

Outline

1. 随机变量的数学特征

数学期望

- 对任意两个随机变量 X, Y，如果其数学期望均存在，则:
  $E(X+Y)=EX+EY$
- 进一步地，如果 X, Y 独立，那么 E(XY) 存在，且:
  $E(XY)=EXEY$
- 设随机向量 (X, Y) 的函数 $Z=g(X,Y)$ 的数学期望存在，我们一般通过以下的两种来计算数学期望 E[Z]:
  - 求出随机变量 Z 的分布，再使用公式 $E[Z]=zf(z)dz$。其中 $\int_{\mathcal{R}} f(z)dz$ 在这里是 Z 的密度函数。当 (X, Y) 是离散型随机向量时，将积分换成求和，密度函数换成对应的概率分布即可。
  - 直接使用公式:
    $E[Z]=E[g(X,Y)]=\int_{\mathcal{R}}\int_{\mathcal{R}} g(x,y)f(x,y)dxdy$
    其中 $f(x,y)$ 是 (X, Y) 的联合密度函数，当 (X,Y) 是离散型随机向量时，将积分换成求和，密度函数换成对应的概率分布即可。

协方差

设 (X, Y) 为二维随机向量，EX 和 EY 均存在，如果 $E[(X-EX)(Y-EY)]$ 存在，则称之为随机变量 X 和 Y 之间的协方差，记作 $Cov(X, Y)$（理解：X 和 Y 的线性关系，若 X、Y 同方向变动，$Cov(X, Y) > 0$）

和协方差有关的性质:

- $cov(X, Y) = EXY - EXEY$
- $cov(X, X) = DX$
- $cov(X, Y) = cov(Y, X)$
- $cov(aX, bY) = abcov(X, Y)$
- $cov(X_1 + X_2, Y) = cov(X_1, Y) + cov(X_2, Y)$
- $cov(C, X) = 0$，其中 C 为一常数。
- 如果 X 和 Y 独立，则 $cov(X, Y) = 0$（反之不成立）
- $D(X + Y) = DX + DY + 2cov(X, Y)$，当 X 与 Y 独立时，
  $D(X + Y) = DX + DY$

相关系数

将协方差标准化我们便可以得到相关系数（也就是去掉了量纲的协方差）：

$$\rho_{X,Y} = \frac{\mathrm{cov}(X, Y)}{\sqrt{DX}\sqrt{DY}}$$

Cauchy—Schwarz 不等式告诉我们：

$$|\rho_{X,Y}| \leq 1$$

练习

设随机向量 (X, Y) 的概率分布为:

X \ Y	0	1
0	0.1	0.15
1	0.25	0.2
2	0.15	0.15

Table 1: X,Y 的概率分布

求 $Z = sin[\frac{\pi}{2}(X + Y)]$ 的数学期望

解答

$$
\begin{aligned}
E[Z] &= E[sin[\frac{\pi}{2}(X+Y)]] \\
&= \sum_{i=0}^{2}\sum_{j=0}^{1} sin[\frac{\pi}{2}(i+j)]P(X=i, Y=j) \\
&= 0.1 sin\, 0 + 0.15 sin\, \frac{\pi}{2} + 0.25 sin\, \frac{\pi}{2} + 0.2 sin\, \pi + 0.15 sin\, \pi \\
&+ 0.15 sin\, \frac{3\pi}{2} \\
&= 0.25
\end{aligned}
$$

设 X 服从参数 $\lambda = 1$ 的指数分布，求 $E(X + e^{-2x})$

设 X 服从参数 $\lambda=1$ 的指数分布，求 $E(X+e^{-2x})$

解答

由于 $E(X+e^{-2x})=E(X)+E(e^{-2x})$，根据题目可知:

$E(X)=1, E(e^{-2x})=\int_{-\infty}^{\infty} e^{-2x}f(x)\,dx$

其中 $f(x)$ 为 X 的概率密度, 由已知 $f(x)=\begin{cases} \mathrm{e}^{-x}, & x>0, \\ 0, & x\leqslant 0, \end{cases}$

$$E\left(\mathrm{e}^{-2X}\right)=\int_{0}^{+\infty} \mathrm{e}^{-2x}\cdot\mathrm{e}^{-x}\,\mathrm{d}x=\frac{1}{3}$$

由此可得:

$$E\left(X+\mathrm{e}^{-2X}\right)=1+\frac{1}{3}=\frac{4}{3}$$

已知 (X, Y) 为一个二维随机向量, $X_1 = X + Y$, $X_2 = X - Y$, (X_1, X_2) 的密度函数为:$\phi(x_1, x_2) = \frac{1}{2\sqrt{3}\pi} e^{-\frac{1}{2}\left[\frac{(x_1-4)^2}{3} + (x_2-2)^2\right]}$ 请求出: X 和 Y 的密度函数以及 $E[X_1, X_2]$

解答

$$\phi\left(x_1, x_2\right) = \frac{1}{2\sqrt{3}\pi} e^{-\frac{1}{2}\left[\frac{(x_1-4)^2}{3}+(x_2-2)^2\right]}$$
$$= \frac{1}{\sqrt{2\cdot 3\cdot \pi}} e^{-\frac{1}{2}\frac{(x_1-4)^2}{3}} \times \frac{1}{\sqrt{2\cdot \pi}} e^{\frac{(x_2-2)^2}{2}}$$

因此 X_1,X_2 独立，且 $X_1 \sim N(4,3)$，$X_2 \sim N(2,1)$，所以

$E[XY] = E[X]E[Y] = 8$，注意到 $X = \frac{X_1+X_2}{2}$，$Y = \frac{X_1-X_2}{2}$，所以，

$X \sim N(3,1)$，$Y \sim N(1,1)$

设随机变量 X 和 Y 独立、且都服从参数为 λ 的泊松分布，令 $U = 2X + Y$, $V = 2X - Y$，求 U 和 V 的相关系数.

设随机变量 X 和 Y 独立、且都服从参数为 λ 的泊松分布，令 $U = 2X + Y$, $V = 2X - Y$，求 U 和 V 的相关系数.

解答

$D(U) = D(2X + Y) = 4DX + DY = 5\lambda$

$D(V) = D(2X - Y) = 4DX + DY = 5\lambda$ 所以

$$\begin{aligned} cov(U, V) &= cov(2X + Y, 2X - Y) \\ &= cov(2X, 2X) + cov(Y, 2X) + cov(2X, -Y) + cov(Y, -Y) \\ &= 4DX - DY = 15\lambda \end{aligned}$$

$$\rho_{U,V} = \frac{Cov(U, V)}{\sqrt{D(U)D(V)}} = \frac{3}{5}$$

2. 大数定律与中心极限定理

练习

设随机变量 X，Y 独立，$X \sim U(-3,3)$，Y 的密度为

$$f_y(Y) = \begin{cases} \dfrac{3y^2}{16} & -2 < y < 2 \\ 0 & others \end{cases} \tag{1}$$

根据切比雪夫不等式，有 $P\{|X - Y| < 3\} \geq \quad ?$

设随机变量 X，Y 独立，$X \sim U(-3,3)$，Y 的密度为

$$f_y(Y) = \begin{cases} \dfrac{3y^2}{16} & -2 < y < 2 \\ 0 & others \end{cases} \tag{2}$$

根据切比雪夫不等式，有 $P\{|X - Y| < 3\} \geq$ ？

解答

$E(Y) = \int_{-2}^{2} \frac{3y^3}{16} dy = 0$

$E(Y^2) = \int_{-2}^{2} \frac{4y^4}{16} dy = \frac{12}{5}$

$E(X) = 0,\ D(X) = 3,\ E(Y) = 0,\ D(Y) = \frac{12}{5}$

则 $E(X - Y) = 0, D(X - Y) = D(X) + D(Y) = \frac{27}{5}$，所以，

$P\{|X - Y| < 3\} = P(|(X - Y) - E(X - Y)| < 3) \geq 1 - \frac{D(X-Y)}{9} = \frac{2}{5}$

设随机变量 $X_1, X_2, ..., X_{100}$，相互独立在区间 $[-1,1]$ 上服从均匀分布，则由中心极限定理 $P(\sum_{i=1}^{100} X_i \leq \frac{10}{\sqrt{3}}) \approx \quad ?$

设随机变量 $X_1, X_2, ..., X_{100}$，相互独立在区间 $[-1,1]$ 上服从均匀分布，则由中心极限定理 $P(\sum\limits_{i=1}^{100} X_i \le \frac{10}{\sqrt{3}}) \approx \quad ?$

解答

$U = \sum\limits_{i=1}^{100} X_i$，则 $E(U) = 0$，$D(U) = 100 \times \frac{4}{12} = \frac{100}{3}$

则 $P(\sum\limits_{i=1}^{100} X_i \le \frac{10}{\sqrt{3}}) = P(\frac{\sum\limits_{i=1}^{100} X_i}{\frac{10}{\sqrt{3}}} \le 1) \approx \Phi(1) = 0.8413$

将一均匀骰子连续扔六次，所出现的点数之和为 X，用切比雪夫不等式估计 $P(14 < X < 28) \geq \quad$ ？

将一均匀骰子连续扔六次，所出现的点数之和为 X，用切比雪夫不等式估计 $P(14 < X < 28) \geq \quad ?$

解答

$$X_i \sim \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix} (i = 1, 2, 3, 4, 5, 6)$$

$E(X_i) = \sum_{k=1}^{6} \frac{1}{6} \times k = \frac{7}{2}, E(X_i^2) = \sum_{k=1}^{6} \frac{1}{6} \times k^2 = \frac{91}{6}$

$D(X_i) = \frac{35}{12}$

则 $E(X) = 6 \times \frac{7}{21} = 21, D(X) = 6 \times \frac{35}{12} = \frac{35}{2}$ 由切比雪夫不等式，有

$P(14 < X < 28) = P(|X - E(X)| < 7) \geq 1 - \frac{D(X)}{7^2} = \frac{9}{14}$

设某概率统计考试共 100 道 4 选 1 的单选题，每题 1 分. 若某人答题全凭运气，求此人最终得分超过 35 分的概率.

设某概率统计考试共 100 道 4 选 1 的单选题，每题 1 分. 若某人答题全凭运气，求此人最终得分超过 35 分的概率.

解答

此人答题的事件服从二项分布 $X \sim B(100, 0.25)$, 所以，

$E(X) = 25, D(X) = \frac{75}{4}$

有中心极限定理可得：

$\frac{X-25}{\sqrt{\frac{75}{4}}} \sim N(0,1)$

$P(X > 35) = P(\frac{X-25}{\sqrt{\frac{75}{4}}} > \frac{35-25}{\sqrt{\frac{75}{4}}}) \approx 1 - \Phi(\frac{4}{\sqrt{3}}) = 1\%$

电话公司有 300 台分机，没台分机有 6% 的时间处于与外线通话的状态，设每台分机是否处于通话状态相互独立，用中心及限定理估计至少要安装多少条外线才能保证每台分机使用外线不必等候的概率不低于 0.95?

解答

$$X_i = \begin{cases} 1, & \text{第}i\text{ 台分机使用外线,} \\ 0, & \text{第}i\text{ 台分机不使用外线} \end{cases} \quad (i = 1, 2, \cdots, 300),$$

则，$X_i \sim \begin{pmatrix} 0 & 1 \\ 0.94 & 0.06 \end{pmatrix} (i = 1, 2, \cdots, 300)$，令 X 为表示需要使用外线的分机数，则 $X = \sum\limits_{i=1}^{300} X_i$，则

$E(X) = 300 \times 0.06 = 18$，$D(X) = 300 \times 0.0564 = 16.92$

设至少需要 n 条外线，由中心极限定理得：

$P(0 \leq X \leq n) \approx \Phi(\frac{n-18}{\sqrt{16.2}}) - \Phi(\frac{0-18}{\sqrt{16.2}}) \geq 0.95$

解得 $n \geq 24.8$，所以至少需要 25 条外线

3. 数理统计学基础

设 (X_1, X_2, X_3, X_4) 是总体 $N \sim (0, 2^2)$ 的简单随机样本，且
$Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2 \quad (a, b \neq 0)$，则当 a，b 分别为多少？
统计量 Y 服从自由度为多少的 χ^2 分布？

设 (X_1, X_2, X_3, X_4) 是总体 $N \sim (0, 2^2)$ 的简单随机样本，且

$Y = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2 \quad (a, b \neq 0)$，则当 a，b 分别为多少?

统计量 Y 服从自由度为多少的 χ^2 分布?

解答

由题意得，样本服从独立同分布，所以

$X_1 - 2X_2 \sim N(0, 5 \times 2^2) \quad \frac{X_1 - 2X_2}{2\sqrt{5}} \sim N(0, 1)$

$3X_3 - 4X_4 \sim N(0, 25 \times 2^2) \quad \frac{3X_3 - 4X_4}{10} \sim N(0, 1)$

所以，

$Y = (\frac{X_1 - 2X_2}{2\sqrt{5}})^2 + (\frac{3X_3 - 4X_4}{10})^2 = \frac{1}{20}(X_1 - 2X_2)^2 + \frac{1}{100}(3X_3 - 4X_4)^2 \sim \chi^2(2)$

所以，$a = \frac{1}{20}$，$b = \frac{1}{100}$，自由度为 2

设 (X_1, X_2, X_3, X_4) 是总体 $N \sim (1, \sigma^2) \quad (\sigma > 0)$ 的简单随机样本，则统计量 $Y = \frac{X_1 - X_2}{|X_3 + X_4 - 2|}$ 服从什么分布?

设 (X_1, X_2, X_3, X_4) 是总体 $N \sim (1, \sigma^2) \quad (\sigma > 0)$ 的简单随机样本，则统计量 $Y = \frac{X_1 - X_2}{|X_3 + X_4 - 2|}$ 服从什么分布?

解答

$X_1 - 2X_2 \sim N(0, 2\sigma^2)$， $\frac{X_1 - 2X_2}{\sqrt{2}\sigma} \sim N(0, 1)$

$X_3 + X_4 - 2 \sim N(0, 2\sigma^2)$， $\frac{X_3 + X_4 - 2}{\sqrt{2}\sigma} \sim N(0, 1)$， $(\frac{X_3 + X_4 - 2}{\sqrt{2}\sigma})^2 \sim \chi^2(1)$

且 $X_1 - 2X_2$， $X_3 + X_4 - 2$ 相互独立，所以

$$\frac{\frac{X_1 - 2X_2}{\sqrt{2}\sigma}}{\sqrt{\frac{(\frac{X_3 + X_4 - 2}{\sqrt{2}\sigma})^2}{1}}} = \frac{X_1 - X_2}{|X_3 + X_4 - 2|} \sim t(1)$$

设总体 $X(\mu_1, \sigma^2)$，总体 $Y(\mu_2, \sigma^2)$，若 $(X_1, X_2, X_3, ..., X_{n_1})$ 和 $(Y_1, Y_2, Y_3, ..., Y_{n_2})$ 分别为总体 X 和 Y 的简单随机样本，求

$$E[\frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2}{n_1 + n_2 - 2}]$$

解答

因为题目中要求总体样本方差的期望，所以我们首先考虑构造一个

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$，且 $E(\chi^2) = n-1$

$\frac{(n_1-1)S_1^2}{\sigma^2} = \frac{\sum\limits_{i=1}^{n_1}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2(n_1-1)$，$\frac{(n_2-1)S_2^2}{\sigma^2} = \frac{\sum\limits_{j=1}^{n_2}(Y_j-\bar{Y})^2}{\sigma^2} \sim \chi^2(n_2-1)$

有 $E[\frac{\sum\limits_{i=1}^{n_1}(X_i-\bar{X})^2}{\sigma^2}] = n_1-1$，$E[\frac{\sum\limits_{j=1}^{n_2}(Y_j-\bar{Y})^2}{\sigma^2}] = n_2-1$

所以

$$
\begin{aligned}
E[...] &= \frac{\sigma^2}{n_1+n_2-2}E[\frac{\sum\limits_{i=1}^{n_1}(X_i-\bar{X})^2}{\sigma^2} + \frac{\sum\limits_{j=1}^{n_2}(Y_j-\bar{Y})^2}{\sigma^2}] \\
&= \frac{\sigma^2}{n_1+n_2-2}(n_1-1+n_2-1) = \sigma^2
\end{aligned}
$$

设随机变量 X 服从 n 个自由度的 t 分布，定义 t_α 满足 $P[X \leq t_\alpha] = 1 - \alpha \quad (0 < \alpha < 1)$，若已知 $P(|X| > x) = b$，则 x 等于？（用上 α 分位点的形式表示）

设随机变量 X 服从 n 个自由度的 t 分布，定义 t_α 满足 $P(X \leq t_\alpha) = 1 - \alpha \quad (0 < \alpha < 1)$，若已知 $P(|X| > x) = b \quad (b > 0)$，则 x 等于？（用上 α 分位点的形式表示）

解答

根据 t 分布的对称性及 $(b > 0)$，可知 x>0。从而

$P(X \leq x) = 1 - P(X > x) = 1 - \frac{1}{2}P(|X| > x) = 1 - \frac{b}{2}$

因为 $P(X \leq t_\alpha) = 1 - \alpha$，可知 $x = t_{\frac{b}{2}}$

4. 参数估计

设 $(X_1, X_2, X_3, ..., X_n)$ 是来自总体的简单随机样本，已知总体 X 的概率密度为 $f(x;\theta) = \begin{cases} \frac{x}{\theta^2} \mathrm{e}^{-\frac{x^2}{2\theta^2}} & x > 0 \\ 0 & x \leqslant 0 \end{cases}$ $(\theta > 0)$ 则 θ 的最大似然估计量 $\hat{\theta}$ 是?

解答

似然函数为： $L(\theta) = \prod_{i=1}^{n} f(x_i;\theta) = \begin{cases} \frac{x_1\cdots x_n}{\theta^{2n}} \mathrm{e}^{-\frac{x^2+\cdots+x_i^2}{2\theta^2}} & x_1, x_2, \cdots, x_n > 0 \\ 0 & \text{其他} \end{cases}$

$\ln L(\theta) = \sum_{i=1}^{n} \ln x_i - 2n \ln\theta - \frac{x_1^2+x_2^2+\cdots+x_n^2}{2\theta^2}$

由 $\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta} = -2n\cdot\frac{1}{\theta} + \left(\sum_{i=1}^{n} x_i^2\right)\theta^{-3} = 0$ 解得 $\theta = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} x_i^2}$

所以 $\hat{\theta} = \sqrt{\frac{1}{2n}\sum_{i=1}^{n} X_i^2}$

设总体 X，Y 相互独立且服从 $N \sim (\mu, \sigma^2)$ 的分布，$(X_1, X_2, X_3, ..., X_{n_1})$ 与 $(Y_1, Y_2, Y_3, ..., Y_{n_2})$ 分别是来自总体的 X，Y 的简单随机样本，证明：

$$S^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2}{n_1 + n_2 - 2}$$

为参数 σ^2 的无偏估计

设总体 X，Y 相互独立且服从 $N \sim (\mu, \sigma^2)$ 的分布，$(X_1, X_2, X_3, ..., X_{n_1})$ 与 $(Y_1, Y_2, Y_3, ..., Y_{n_2})$ 分别是来自总体的 X，Y 的简单随机样本，证明：

$$S^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2 + \sum_{j=1}^{n_2}(Y_j - \bar{Y})^2}{n_1 + n_2 - 2}$$

为参数 σ^2 的无偏估计

解答

点这里看一下前面详细的证明过程

因为 $E(S^2) = \sigma^2$，所以无偏估计得证。

设总体 $X \sim U(\theta_1, \theta_2)$，$(X_1, X_2, X_3, ..., X_n)$ 是来自总体的简单随机样本，求 θ_1，θ_2 的矩估计和最大似然估计。

设总体 $X \sim U(\theta_1, \theta_2)$，$(X_1, X_2, X_3, ..., X_n)$ 是来自总体的简单随机样本，求 θ_1，θ_2 的矩估计和最大似然估计。

解答

矩估计如下:

$E(X) = \frac{\theta_1+\theta_2}{2}$，$D(X) = \frac{(\theta_2-\theta_1)^2}{12}$

$$\begin{cases} E(X) = \overline{X}, \\ D(X) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2, \end{cases} \Rightarrow \begin{cases} \theta_1 + \theta_2 = 2\bar{X} \\ \theta_2 - \theta_1 = 2\sqrt{3B_2} \end{cases}$$

$$\Rightarrow \begin{cases} \hat{\theta}_1 = \bar{X} - \sqrt{3B_2} \\ \hat{\theta}_2 = \bar{X} + \sqrt{3B_2} \end{cases}$$

解答

最大似然估计如下:

$$f(x;\theta_1,\theta_2)=\begin{cases}\frac{1}{\theta_2-\theta_1}, & \theta_1<x<\theta_2\\ 0, & \text{其他}.\end{cases}$$

$$L\left(x_1,x_2,\cdots,x_n;\theta_1,\theta_2\right)=\begin{cases}\frac{1}{(\theta_2-\theta_1)^n}, & \theta_1<x_i<\theta_2(i=1,2,\cdots,n)\\ 0, & \text{其他}.\end{cases}$$

$$\ln L\left(\theta_1,\theta_2\right)=-n\ln\left(\theta_2-\theta_1\right),\frac{\partial}{\partial\theta_1}\ln L\left(\theta_1,\theta_2\right)=\frac{n}{\theta_2-\theta_1}>0,$$

$$\frac{\partial}{\partial\theta_2}\ln L\left(\theta_1,\theta_2\right)=-\frac{n}{\theta_2-\theta_1}<0$$

因为 $\theta_1\leqslant\min\limits_{1\leqslant i\leqslant n}\{x_i\},\theta_2\geqslant\max\limits_{1\leqslant i\leqslant n}\{x_i\}$，$lnL(\theta_1,\theta_2)$ 是 θ_1 的单调增函数，是 θ_2 的单调减函数，所以，$\hat{\theta}_1=\min\limits_{1\leqslant i\leqslant n}\{X_i\}$，$\hat{\theta}_2=\max\limits_{1\leqslant i\leqslant n}\{X_i\}$

5. 区间估计

单个正态总体参数的区间估计

- 单个正态总体均值的区间估计
  - 方差 σ 已知，求均值 μ 的区间估计，取 $U = \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$
    - ★ 置信区间为：$(\bar{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$
  - 方差 σ 未知，求均值 μ 的区间估计，取 $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$
    - ★ 置信区间为：$(\bar{X} - t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}})$
- 单个正态总体方差和标准差的区间估计
  - 均值 μ 已知，求方差 σ 的区间估计，取 $\frac{X_i-\mu}{\sigma} \sim N(0,1)$
    - ★ 方差 σ^2 置信区间为：

$$\left(\frac{\sum\limits_{i=1}^{n} (X_i-\mu)^2}{\chi^2_{\frac{a}{2}}(n)}, \frac{\sum\limits_{i=1}^{n} (X_i-\mu)^2}{\chi^2_{1-\frac{a}{2}}(n)} \right) = \left(\frac{(n-1)S^2}{\chi^2_{\frac{a}{2}}(n)}, \frac{n-1)S^2}{\chi^2_{1-\frac{a}{2}}(n)} \right)$$

    - ★ 标准差 σ 置信区间为：$\left(\frac{\sqrt{n-1}S}{\sqrt{\chi^2_{\frac{a}{2}}(n-1)}}, \frac{\sqrt{n-1}S}{\sqrt{\chi^2_{1-\frac{a}{2}}(n-1)}} \right)$

两个正态总体参数的区间估计

- 两个正态总体均值的区间估计
  - σ_1^2 和 σ_2^2 已知，求 $\mu_1 - \mu_2$ 的区间估计，取：

    $\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$ 或 $\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1)$

    - ★ 置信区间为：$\left(\bar{X} - \bar{Y} \pm z_{a/2}\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}\right)$
    - ★ 当 n>50 时，样本方差满足无偏估计，置信区间为：

      $\left(\bar{X} - \bar{Y} \pm z_{a/2}\sqrt{S_1^2/n_1 + S_2^2/n_2}\right)$
  - 当 $\sigma_1^2 = \sigma_2^2 = \sigma^2$，但$\sigma^2$ 为未知 时，
    - ★ $\mu_1 - \mu_2$ 的置信度为 $1 - \alpha$ 的置信区间为:

      $\left(\bar{X} - \bar{Y} \pm S_w t_{a/2}\left(n_1 + n_2 - 2\right)\sqrt{1/n_1 + 1/n_2}\right)$

      其中 $S_w^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}, S_w = \sqrt{S_w^2}$

两个正态总体参数的区间估计

- 两个正态总体方差的区间估计
  - $\color{red}{\mu_1}$，$\color{red}{\mu_2}$ 未知，求 $\frac{\sigma_1^2}{\sigma_2^2}$ 的区间估计，取：$\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$
    - ★ $\frac{\sigma_1^2}{\sigma_2^2}$ 置信区间为：$\left(\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{a/2}(n_1-1,n_2-1)}, \frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}\right)$
    - ★ $\frac{\sigma_1}{\sigma_2}$ 置信区间为：$\left(\sqrt{\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{a/2}(n_1-1,n_2-1)}}, \sqrt{\frac{S_1^2}{S_2^2}\cdot\frac{1}{F_{1-\alpha/2}(n_1-1,n_2-1)}}\right)$

非正态总体的区间估计（了解）

- 思路：因为我们已经知道了正态总体的区间估计，我们又学过中心极限定理，样本量足够大的时候样本均值满足中心极限定理——依分布收敛到正态分布。所以中心极限定理（标准化参数）是此类型的桥梁。
- 假设总体 X 的均值 μ 与方差 σ^2 存在, $X_1, X_2, \cdots, X_n$ 为来自总体 X 的样本, $\bar{X}$ 与 S^2 为样本均值和样本方差, 取 $T = \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}}$
  - 均值 μ 的置信度为 $1-\alpha$ 的近似置信区间为:

$$\left(\bar{X} - \frac{S}{\sqrt{n}} z_{\alpha/2}, \bar{X} + \frac{S}{\sqrt{n}} z_{a/2}\right)$$

设总体 $X \sim N(\mu, \sigma^2)$，从总体中抽出容量为 36 的一个样本，样本均值为 3.5，样本方差为 4.

（1）若 $\sigma^2 = 1$，求 μ 置信度为 95% 的置信区间；

（2）若 σ^2 未知，求 μ 置信度为 95% 的置信区间；

已知 $\Phi(1.96) = 0.975, t_{0.025}(35) = 2.03$

设总体 $X \sim N(\mu, \sigma^2)$，从总体中抽出容量为 36 的一个样本，样本均值为 3.5，样本方差为 4.

（1）若 $\sigma^2 = 1$，求 μ 置信度为 95% 的置信区间；

已知 $\Phi(1.96) = 0.975, t_{0.025}(35) = 2.03$

解答（1）

由于 $\sigma^2 = 1$ 已知，可知 μ 的置信区间为：$(\bar{X} - u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}})$

由条件得 $1 - \alpha = 0.95, \alpha = 0.05, n = 36$，且 $\Phi(u_{\frac{0.05}{2}}) - \Phi(-u_{\frac{0.05}{2}}) = 0.95$

所以，$\Phi(u_{\frac{0.05}{2}}) = \frac{1+0.95}{2} = 0.975$，故 $u_{\frac{0.05}{2}} = 1.96$，可得置信区间为：

$\left(3.5 - \frac{1}{6} \times 1.96, 3.5 + \frac{1}{6} \times 1.96\right) = (3.17, 3.83)$

设总体 $X \sim N(\mu,\sigma^2)$，从总体中抽出容量为 36 的一个样本，样本均值为 3.5，样本方差为 4.

（2）若 σ^2 未知，求 μ 置信度为 95% 的置信区间；

已知 $\Phi(1.96)=0.975, t_{0.025}(35)=2.03$

解答（2）

由于 σ^2 未知，用 S^2 替换 σ^2，可知 μ 的置信区间为：

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

由条件得 $1-\alpha=0.95,\alpha=0.05,n=36,s=2,\bar{x}=3.5$,

且 $t_{\frac{0.05}{2}}(35)=t_{0.025}(35)=2.03,$

可得置信区间为：

$\left(3.5-\frac{2}{6}\times 2.03, 3.5+\frac{2}{6}\times 2.03\right)=(2.82,4.18)$

从甲、乙两厂生产的蓄电池产品中，分别抽出一批产品, 测得其电容量为:

甲厂:144,141,138,142,141,138,143,137

乙厂：142,143,139,140,138,141,140,138,142,136

设两厂的蓄电池的电容量分别服从正态分布,

即:$X \sim N(\mu_1, \sigma_1^2) \quad Y \sim N(\mu_2, \sigma_2^2)$

求：电容量的均值差 $\mu_1 - \mu_2$ 的置信度为 0.95 的置信区间 (设 $\sigma_1^2 = \sigma_2^2$)

解答

由已知有：$\bar{X}=140.5, \bar{Y}=139.9, S_1^2=2.563, S_2^2=2.183$

且 σ^2 未知

$\mu_1-\mu_2$ 的置信区间：$\left(\bar{X}-\bar{Y} \pm S_w t_{a/2}(n_1+n_2-2)\sqrt{1/n_1+1/n_2}\right)$

由 $n_1=8, n_2=10$ 可知：

$\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}=0.474, S_w=\sqrt{\frac{7\times 2.563^2+9\times 2.183^2}{16}}=2.357$

查表可得：$t_{\frac{\alpha}{2}}(n_1+n_2-2)$

所以所求置信区间为 $(-1.768, 2.968)$

Hint：我们可以看出置信区间包含 0，所以其实这两个厂的电池容量（总体）均值没有很大差别

一些小提示

- 几个特殊分布的概率密度函数、期望和方差考前尽量默写一遍。
- 基本公式考前记得复习一遍（e.g 离散型、连续性期望方差怎么求）
- 计算积分的时候找准上下限，区域一定要画对！
- 注意书里需要独立前提的定理、定义！
- 只要是几大分布的证明一定要检查自由度！（e.g 你的无偏性证明不出来，有一种可能是自由度取错了）
- 如果有时间，复习一遍自己作业中的错题，总结错误原因。
- 某些填空题考的比较细节，需要大家考前仔细读一遍书，没有经常强调的定义也要有个印象（e.g 边缘分布函数定义）
- 基本上第 1 章到第 6 章中每一章都会出一道计算大题。

最后祝大家考试顺利!

Thank You!

June 16, 2024