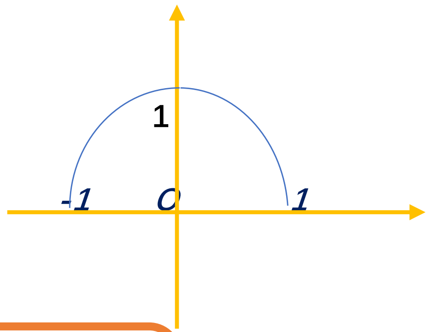
概率论与数理统计易错和格式分享

例题 1：设(X, Y)在区域D：$\{(x,y)|0 \le y \le 1-x^2\}$上服从**均匀分布**，求关于 X、Y 的边缘密度函数$f_X(x)$和$f_Y(y)$，判断独立性。



解得：$f(x,y) = \begin{cases} 3/4, & 0 \le y \le 1-x^2 \\ 0, & else \end{cases}$

$$f_X(x) = \int_{-\infty}^{+\infty} f(x,y)dy = \begin{cases} \int_0^{1-x^2} \frac{3}{4} dy = 3(1-x^2)/4 & -1 < x < 1, \\ 0, & else \end{cases}$$

分段函数

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x,y)dx = \begin{cases} \frac{3\sqrt{1-y}}{2}, & 0 < y < 1, \\ 0, & else \end{cases}$$(同上做法)

建议画图，重点注意边缘密度的计算格式

因为$f(x,y) \neq f_X(x) \times f_Y(y)$，所以不独立

例题 2：设随机变量$X_1, X_2, \dots, X_{100}$相互独立，且均服从相同的指数分布，概率密度函数为$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}, & \mathrm{x} > 0 \\ 0, & 其它 \end{cases}$,利用中心极限定理估计概率$P\{\sum_{i=1}^{100} X_i < 240\}$.

（结果用$\phi(\cdot)$表示）

解：由题意得

$$\mathrm{E}(X_i) = 2, \mathrm{D}(X_i) = 4, \mathrm{i} = 1,2,\dots,100$$

$$P\{\sum_{i=1}^{100} X_i < 240\} = P\left\{\frac{\sum_{i=1}^{100} X_i - 100\mathrm{E}(X_i)}{\sqrt{100\mathrm{D}(X_i)}} < \frac{240-100\mathrm{E}(X_i)}{\sqrt{100\mathrm{D}(X_i)}}\right\}$$

$$= P\left\{\frac{\sum_{i=1}^{100} X_i - 100\mathrm{E}(X_i)}{\sqrt{100\mathrm{D}(X_i)}} < \frac{40}{20}\right\}$$

$$\approx \phi(0.5)$$

约等号

例题 3：设总体$X \sim N(\mu,\sigma^2)$，μ,σ^2均未知，$x_1,x_2,\dots,x_n$为来自总体X的样值。求μ,σ^2的**最大似然估计值。**

解：由题意得，最大似然函数

$$L(\mu,\sigma^2)=\prod_{i=1}^{n} f(x_i;\mu,\sigma^2)$$

$$\boxed{=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}},\ x_i\in R}$$ 足标 i 别忘了

$$lnL(\mu,\sigma^2)=\sum_{i=1}^{n} ln\sqrt{2\pi}-\sum_{i=1}^{n} ln\sigma-\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\begin{cases}\frac{\partial lnL(\mu,\sigma^2)}{\partial\mu}=0\\ \frac{\partial lnL(\mu,\sigma^2)}{\partial\sigma^2}=0\end{cases}. \qquad \begin{cases}\hat{\mu}=\bar{x}\\ \widehat{\sigma^2}=\frac{n-1}{n}S^2\end{cases}$$

例题 4：设总体X的概率密度函数为$f(x)=\begin{cases}\frac{1}{2\theta}, & 0<x<\theta\\ \frac{1}{2(1-\theta)}, & \theta\le x<1\\ 0, & 其它\end{cases}$，其中参数$\theta(0<\theta<1)$未知，$X_1,X_2,\dots,X_n$为来自总体$X$的简单随机样本，求θ的**矩估计量**。

解：$\boxed{\text{step1 因为只有一个未知参数，所以 } A_1=\mu_1=E(X)}$ 几个未知数，几个方程，一般三步

Step2　又$A_1=\bar{X}$　　$\mu_1=\int_{-\infty}^{+\infty} xf(x)dx$

$$=\int_0^{\theta}\frac{1}{2\theta}xdx+\int_{\theta}^{1}\frac{1}{2(1-\theta)}xdx$$

$$=\frac{1+2\theta}{4}$$

Step3 $\frac{1+2\theta}{4}=\bar{X}\longrightarrow\hat{\theta}=\frac{4\bar{X}-1}{2}$ （估计量大写，估计值小写，别忘了带尖角）

例题 4：某公司利用两条自动化流水线灌装矿泉水，现从流水线上分别随机抽取样本$X_1,X_2,\dots,X_{18}$和$Y_1,Y_2,\dots,Y_{18}$，测得每瓶所装矿泉水的体积（单位：ml）。计算得$\bar{x}$ =501.2，$\bar{y}=499.8$，$s_1^2=3.8,$，$s_2^2=4.2$。设这两条流水线所装的矿泉

水的体积X，Y分别服从$N(\mu_1, \sigma^2)$和$N(\mu_2, \sigma^2)$。求$\mu_1-\mu_2$的置信水平为 0.95 的置信区间。（$t_{0.025}(36)$=2.028，$t_{0.025}(34)$=2.032，**数据保留两位小数)**

解：**step1:因为$\boldsymbol{\sigma_1^2}=\sigma_2^2=\sigma^2$**未知，**所以关于$\boldsymbol{\mu_1-\mu_2}$的置信水平为 0.95 的置信区间为**$\left(\bar{X}-\bar{Y}\pm S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\frac{\alpha}{2}}(n_1+n_2-2)\right)$，其中$S_\omega^2=\frac{(n_1-1)S_1^2+(n_2-2)S_2^2}{n_1+n_2-2}$

Step2:又$\bar{x}=501.2, \bar{y}=499.8, S_1^2=3.8, S_2^2=4.2, n_1=n_2=18, \alpha=0.05,$

$t_{0.025}(34)=2.032$ $S_\omega^2=4$ ，

Step3:代入得 $\left(\bar{X}-\bar{Y}\pm S_\omega\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}t_{\frac{\alpha}{2}}(n_1+n_2-2)\right)=$ （0.05，2.75）

例题 4: 某手表厂生产的男表表壳在正常情况下，其直径(单位:mm)服从正态分布 $N(20, 1)$。在某天的生产过程中，随机抽查 4 只表壳，测得直径分别为：19.5 19.8 20.0 20.5，经计算得$\bar{x}=19.95$. 问在$\alpha=0.05$显著性水平下，这天生产的表壳直径的均值是否正常？ （$z_{0.025}=1.96$）

解：**方法 1（拒绝域法）**

由题意得，设$H_0: \mu=\mu_0=20$ $H_1: \mu\neq\mu_0$ **（先假设再检验）**

因为σ^2 =1 已知，选取检验统计量$Z=\frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$，

拒绝域为：$|z|=\left|\frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geq z_{\alpha/2}$

$\bar{x}=19.95$, n=4, $\alpha=0.05$，$z_{0.025}=1.96$

数据代入得：$|z|=0.1<z_{\alpha/2}=1.96$ (没有落在拒绝域)

所以原假设为真，认为这天生产的直接的均值是正常的。

条件——>选取统计量——>拒绝域——>数据代入——>判断——>决策

方法 2（置信区间法）

由题意得，设$H_0: \mu=\mu_0=20$ $H_1: \mu\neq\mu_0$ （先假设再检验）

因为σ^2 =1 已知，关于μ的置信水平为 1-α = 0.95 的置信区间为：$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$

又 $\bar{x}$ = 19.95, n=4, $\alpha=0.05$, $z_{0.025}$ = 1.96

数据代入得：$\left(\bar{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\ \bar{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$ = (18.97，20.93)

因为$\mu_0=20\in$（18.97，20.93），所以原假设为真，认为这天生产的直接的均值是正常的。

条件——>置信区间——>数据代入——>判断——>决策

例 5 某厂对废水进行处理，要求某种有害物质的浓度不超过 19（毫克/立升）。抽样检查得到 10 个数据，其样本均值$\bar{x}=19.5$（毫克/立升），样本方差$s^2=1.25$（毫克/立升）。设该种有害物质的浓度服从正态分布，问在显著性水平$\alpha=0.05$下能否认为处理后的废水符合标准？

解： 设该种有害物质的浓度为$X\sim N(\mu,\sigma^2)$。

方法 1（拒绝域法） 根据题意，设

$$H_0: \mu\le\mu_0=19 \qquad H_1: \mu>\mu_0 \qquad \text{（右边检验）}$$

因为σ^2未知，所以选取统计量 $\mathbf{t}=\frac{\bar{X}-\mu_0}{S/\sqrt{n}}$

拒绝域为：$t\ge t_\alpha(n-1)$

又 $n=10, \bar{x}=19.5, \mu_0=19$，$s^2=1.25$，$\alpha=0.05$，$t_{0.05}(9)=1.833$代入得

$\mathrm{t}=1.414<t_{0.05}(9)=1.833$（没有落在拒绝域）

所以 H_0为真，可以认为处理后符合标准。

方法 1（置信区间法） 根据题意，设

$$H_0: \mu\le\mu_0=19 \qquad H_1: \mu>\mu_0 \qquad \text{（右边检验）}$$

因为σ^2未知，所以关于μ的置信水平为 1-α=0.95 的含有单侧置信下限的置信区间为$\left(\bar{x}-\frac{s}{\sqrt{n}}t_\alpha(9),+\infty\right)$，

又 $n=10, \bar{x}=19.5, \mu_0=19$，$s^2=1.25$，$\alpha=0.05$，$t_{0.05}(9)=1.833$

代入得 ：$\left(\bar{x}-\frac{s}{\sqrt{n}}t_\alpha(9),+\infty\right)=$ （$18.85,+\infty$）

因为$\mu_0=19\in(18.85,+\infty)$，所以 H_0为真，可以认为处理后符合标准