阿术的概率论与数理统计 Akyuu's Probability & Mathematical Statistic

衔接部分： §5 大数定律与中心极限定理

一、知识积累：

§5.1 切比雪夫不等式

1. $P\{|X-EX|\geq\varepsilon\}\leq\frac{DX}{\varepsilon^2}$ $\underline{\underline{\geq\varepsilon\leq}}$

§5.2 大数定律

2. 依概率收敛：$Y_1,Y_2\cdots Y_n$ 是一个随机变量序列，对于任意正数 ε 有

$$\lim_{n\to+\infty}P(|Y_n-a|<\varepsilon)=1$$

3. 切比雪夫大数定理， $X_1\cdots X_n$ 独立，$\exists c>0$，且 $DX_i\leq C$

$$\lim_{n\to+\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\frac{1}{n}\sum_{i=1}^{n}EX_i\right|<\varepsilon\right)=1$$

4. 辛钦大数定理（弱大数定理）

$X_1,X_2\cdots X_n\cdots$ 独立同分布 $EX_i=\mu$. 且则对 $\forall\varepsilon>0$，有 $\lim_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right)=1$

$\frac{1}{n}\sum_{i=1}^{n}X_i\xrightarrow{P}\mu$.

5. 伯努利大数定理：

设 f_A 是 n 次独立重复试验中事件A的次数，p 是事件A在每次试验中发生概率. 则对于 $\forall\varepsilon>0$ 有：

$$\lim_{n\to\infty}P\left(\left|\frac{f_A}{n}-p\right|<\varepsilon\right)=1$$

§5.3 中心极限定理

6. 列维林德伯格中心极限定理

$X_1,\cdots X_n\cdots$ 独立同分布 $EX_i=\mu$ $DX_i=\sigma^2$，则对于 $\forall x\in R$ 有

$$\lim_{n\to\infty}P\left(\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}\,\sigma}\leq x\right)=\Phi(x)$$

即：$\sum_{i=1}^{n}X_i\sim N(n\mu,n\sigma^2)$

7. 棣莫弗－拉普拉斯中心极限定理

$X_n\sim B(n,p)$，且 $X_1,X_2\cdots X_n$ 相互独立，则对于 $\forall x$ 有

$$\lim_{n\to\infty}P\left(\frac{\sum_{i=1}^{n}X_i-np}{\sqrt{np(1-p)}}\leq x\right)=\Phi(x)$$

二、例题典例

§5.4 切不等式之应用

1.设随机变量X的概……$f(x)=\begin{cases}\frac{1}{2}x^2e^{-x} & x>0\\ 0 & x\le 0\end{cases}$

请用切不等式估计概率 $P(1<X<5)\geq$.

解 $EX=\int_0^{\infty}\frac{1}{2}x^3e^{-x}dx=\frac{3!}{2}=3$

$EX^2=\int_0^{\infty}\frac{1}{2}x^4e^{-x}dx=\frac{4!}{2}=12$

$DX=EX^2-(EX)^2=3$

$P(1<X<5)=P(|X-3|<2)\geq 1-\frac{3}{2^2}=\frac{1}{4}$

§5.5 中心极限定理

例1:随机变量 $X_1, X_2 \cdots X_{100}$ 相互独立,且服从 $\lambda=\frac{1}{2}$ 的指数分布,则求 $P(\sum_{i=1}^{100}X_i<240)$.

$\mu=\frac{1}{\lambda}=2$ 且 $6^2=\frac{1}{\lambda^2}=4$. 则有: $\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{n}6}\le x=\Phi(x)$

即 $\sum_{i=1}^{n}X_i\sim N(200,400)$ $X<240\Rightarrow\frac{240-200}{20}=2$.

$\therefore P(\sum_{i=1}^{100}X_i<240)=\Phi(2)$.

例:设 $X_1\cdots X_n$ 独立,且服从Poisson分布 $P(2)$,则求 $\lim_{n\to\infty}(\frac{\sum_{i=1}^{n}X_i-2n}{\sqrt{n}}<2)$

解: $P(2)$: $\lambda=2$, $\mu=2$, $6=2$

则 $\lim_{n\to\infty}\frac{\sum_{i=1}^{n}X_i-n\cdot 2}{\sqrt{n}\cdot\sqrt{2}}<x=\Phi(x)$ $\lim_{n\to\infty}\frac{\sum_{i=1}^{n}X_i-2n}{\sqrt{n}\cdot\sqrt{2}}<\sqrt{2}\Rightarrow\Phi(\sqrt{2})$

3

阿求的概率论与数理统计　Akyuu's Probability& Mathematic Statistic

MS部分　§6　样本及抽样分布

一、知识储备

§6.1 基本概念

1、总体：试验全部可能之观察值

2、个体：每一个可能观察值

3、样本：设X是具有分布函数F的随机变量，若$X_1 \cdots X_n$是具有同一分布函数F的，相互独立随机变量，则称$X_1 \cdots X_n$为分布函数F得到的容量为n的简单随机样本

§6.2 统计量

4、统计量 设$X_1 \cdots X_n$是来自总体X的一个样本，$g(X_1 \cdots X_n)$为$X_1 \cdots X_n$的函数，若g中不含未知参数，则称$g(X_1 \cdots X_n)$是一统计量

本质上为一维随机变量

5、常见统计量

① 样本均值：$\bar{X}=\frac{1}{n}\sum_{i}^{n} X_i \xrightarrow{P} \mu$，$\xrightarrow{P} \sigma^2$

② 样本方差：$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$；样本标准差：$S=\sqrt{S^2} \xrightarrow{P} \sigma$

③ k阶原点矩　$A_k=\frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} EX^k$

④ k阶中心矩　$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$

§6.3 抽样分布

6、χ^2分布

设$X_1, X_2 \cdots X_n$相互独立同分布于$N(0,1)$，令$Z=X_1^2+X_2^2+\cdots+X_n^2$

则称$Z \sim \chi^2(n)$

有：(1) $E\chi^2(n)=n$，$D\chi^2(n)=2n$

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

7、t—分布：

$X\sim N(0,1)$，$Y\sim\chi^2(n)$，X、Y相互独立，令$Z=\frac{X}{\sqrt{Y/n}}$　称$Z\sim t(n)$

性质：① 类似于标准正态分布

② $EZ=0 \quad DZ=\frac{n}{n-2}$

③ n越大，t分布越高

不可加，对称

8. F-分布：$X \sim \chi^2(m)$，$Y \sim \chi^2(n)$ 且X、Y相互独立，令 $F=\frac{\frac{X}{m}}{\frac{Y}{n}}$ 称 $F \sim F(m,n)$

性质：$F \sim F(m,n)$ $\quad \frac{1}{F} \sim F(n,m)$

$t^2 \sim F(1,n)$

§6.4 分布分位点

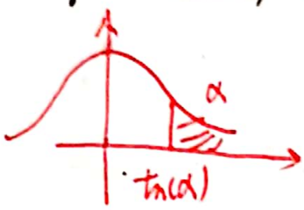
9. χ^2分布分位点：

$P(\chi^2 > \chi^2_\alpha(n)) = \alpha$

$\chi^2_\alpha(n)$

t分布分位点：

$P(t > t_\alpha(n)) = \alpha$

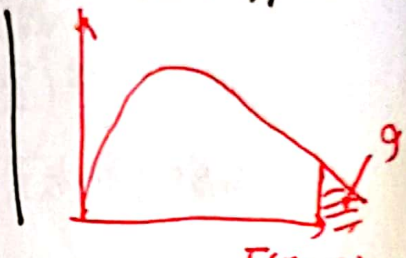


$t_\alpha(n)$

$t_{1-\alpha}(n) = -t_\alpha(n)$

F分布分位点：

$P(F > F_\alpha(m,n)) = \alpha$



$F_\alpha(m,n)$

$F_{1-\frac{\alpha}{2}}(m,n) = \frac{1}{F_{\frac{\alpha}{2}}(n,m)}$

§6.5 正态总体下常见抽样分布

10. 定理一 $X \sim N(\mu, \sigma^2)$ $X_1 \cdots X_n$

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(n)$

$\chi^2(n) \sim \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$

定理二：$\frac{(n-1)S^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$

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

定理三：$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$

定理四：$\frac{S_1^2}{S_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2} \sim F(n_1-1, n_2-1)$

二、重要题型

§6.6 t分布

设$X_1 \cdots X_n$来自正态分布$N(\mu,\sigma^2)$的一组简单随机样本，$\bar{X}$和S^2分别为…，设$X_{n+1} \sim N(\mu,\sigma^2)$与…独立，则 $\frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}$ 服从

$X_{n+1} \sim N(\mu, \sigma^2)$ $\quad \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ $\quad X_{n+1}-\bar{X} \sim N(0, \frac{n+1}{n}\sigma^2)$

$\therefore \frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}} \sim N(0,1)$ 则 $\frac{X_{n+1}-\bar{X}}{S}\sqrt{\frac{n}{n+1}}$ $\quad \frac{\frac{X_{n+1}-\bar{X}}{\sqrt{\frac{n+1}{n}\sigma^2}}}{\sqrt{\frac{\frac{(n-1)S^2}{\sigma^2}}{n-1}}} \sim t(n-1)$

解决方法：$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim t(n-1)$ 之变形

§6.7 χ^2分布

解决方法：① 将正态分布捆绑一起进行分析

② $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1) \sim \sum_{i=1}^{n}\left(\frac{X_i-\bar{X}}{\sigma}\right)^2$

例：$X \sim N(1,4)$，$X_1 \cdots X_n$为来自总体X之一样本，$\bar{X}$为样本均值，则

$E\left(\sum_{i=1}^{n}(X_i-\bar{X})^2\right)=(\quad)$

$\frac{1}{\sigma^2}\sum(X_i-\bar{X})^2 \sim \chi^2(n-1) \to \sum(X_i-\bar{X})^2 \sim \underbrace{\sigma^2\chi^2(n-1)}_{4(n-1)}$

§6.8 F-分布：

解决方法：① $t^2 \sim F(1,n) \to \overset{t(n)}{\cancel{t^2(n)}}$.

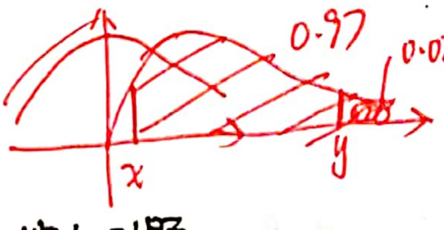
② $F(m,n)=\frac{\frac{\chi^2(m)}{m}}{\frac{\chi^2(n)}{n}}$

例：$T \sim t(15)$，则$T^2 \sim$

Ans: $F(1,15)$

§6.9 分位点理论

例：$X \sim \chi^2(2)$ $\quad P(x<X<y)=0.95$ $\quad P(X>y)=0.02$，请用分位点标记x, y.



$x = \chi^2_{0.97}(2)$

$y = \chi^2_{0.02}(2)$.

三、综合例题

1. ~~判~~ 判断 T&F

① $\dfrac{\sqrt{n-1}\,X_n}{\sqrt{\sum_{i=1}^{n-1} X_i^2}} \sim t(n-1)$

正确, $\dfrac{X_n}{\sqrt{\dfrac{\sum X_i}{n-1}}} \sim t(n-1)$

② $\dfrac{(\frac{n}{2}-1)\sum_{i=1}^{2} X_i^2}{\sum_{i=3}^{n} X_i^2} \sim F(2,n-2)$

原式: $\dfrac{\chi^2(2)}{\chi^2(n-2)} \cdot \dfrac{n-2}{2} \sim F(2,n-2)$ 正确

③ $\dfrac{\overline{X}-\mu}{S}\sqrt{n-1} \sim t(n-1)$

$\overline{X}-\mu \sim N(0,\frac{\sigma^2}{n})$

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

S: $\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2} \sim \chi^2(n-1)$ X

$X_i-\overline{X} \sim N(0,\frac{n+1}{n}\sigma^2)$

2. 求参数

① $Y_1=X_1+X_2$ $Y_2=X_1-X_2$. 当 $c=$___ 时 $\dfrac{c(Y_1-2\mu)}{|Y_2|} \sim t(\quad)$

$\dfrac{Y_1-2\mu}{\sqrt{2}\sigma} \sim N(0,1)$ $\dfrac{Y_2^2}{2\sigma^2} \sim \chi^2(1)$ $\dfrac{\frac{(Y_1-2\mu)}{\sqrt{2}\sigma}}{\sqrt{\frac{\frac{Y_2^2}{2\sigma^2}}{1}}} \sim t(1)$

② $(X_1 \cdots X_{10}) \sim N(0,4)$ $\overline{X}$ 均值, S^2 方差: $c=$____ $c\left(\frac{\overline{X}}{S}\right)^2$ 服从F分布

$\frac{2\sigma^4}{n-1}$

$\overline{X} \sim N(0,\frac{2}{5})$ $\dfrac{\overline{X}^2}{0.4} \sim \chi^2(1)$ $\dfrac{(10-1)S^2}{\sigma^2} \sim \chi^2(9)$

$\dfrac{\frac{\overline{X}}{0.4}}{\frac{(10-1)S^2}{4\times 9}} = \dfrac{\frac{4}{0.4}\overline{X}^2}{S^2} = 10\left(\frac{\overline{X}}{S}\right)^2$ $\therefore c=10$

③ $X_1 \cdots X_8 \sim N(0,4)$, $Y=\dfrac{c(X_1-X_2)}{\sqrt{\sum_{i=3}^{8} X_i^2}} \sim t(\quad)$ $P\{|\overline{X}| \leq \frac{\sqrt{2}}{2}\}=$____

$\dfrac{X_1-X_2}{2\sqrt{2}} \sim N(0,1)$ $\dfrac{X_i}{2} \sim N(0,1)$ $\sum_{i=3}^{8}\left(\frac{X_i}{2}\right)^2 \sim \chi^2(6)$ $\therefore \dfrac{\sum_{i=3}^{8} X_i^2}{24} \sim \chi^2(6)$

$\therefore \dfrac{\frac{2\sqrt{2}}{2\sqrt{2}}}{\sqrt{24}}$ $\dfrac{\Phi(X)}{\chi^2(6)}$ $\dfrac{\quad}{2\sqrt{6}}$ 解得 $c=\sqrt{3}$

$P(|\overline{X}| \leq \frac{\sqrt{2}}{2}) = P\left(\left|\frac{\overline{X}-0}{\frac{\sqrt{2}}{2}}\right| \leq \frac{\frac{\sqrt{2}}{2}}{\frac{\sqrt{2}}{2}}=1\right)=2\Phi(1)-1$

3. $(X_1 \cdots X_{10}),(Y_1 \cdots Y_{15}) \sim N(0,1)$ $Z=\sum_{i=1}^{10}(X_i-\overline{X})^2+\sum_{i=1}^{15}(Y_i-\overline{Y})^2$

$\therefore \sum_{i=1}^{10}(X_i-\overline{X})^2 \sim 9\sigma^2=9$ 同理: 后面14 $14+9=23$

§7 参数估计

§7.1 点估计量评选原则

1、无偏性：$E\hat{\theta}=\theta$　2.有效性 $D(\hat{\theta}_1)<D(\hat{\theta}_2)$ θ_1比θ_2更有效

3、一致性：$\hat{\theta}\xrightarrow{P}\theta$、

§7.2 点估计

4.矩估计

①计算总体矩：$\mu_l=E(X^l)=\int_{-\infty}^{+\infty}x^l f\,dx$

②计算样本矩：$A_l=\frac{1}{n}\sum_{i=1}^{n}x_i^l$

③令 $A_l=E(X_i^l)$.

则有：$A_1=\mu_1$　$A_2=\mu_2\cdots A_k=\mu_k'$

有：$\hat{\theta}=\theta_i(A_1\cdots A_k)$ $i=1,2,3\cdots k$.

5.极大似然估计

①离散型：$L(x_1,x_2\cdots x_n;\theta_1\cdots\theta_k)=\prod_{i=1}^{n}P(X=x_i)$.

连续型：$L(x_1,x_2\cdots x_n;\theta_1\cdots\theta_k)=\prod_{i=1}^{n}f(x_i,\theta_1\cdots\theta_k)$;

②求 $\ln L$.

③ $\frac{\partial}{\partial\theta_i}\ln L=0$. 解似然方程组.

若方程组无解，往往θ会取边界值，使得$L(\theta)$较大.
若有多解，$L(\theta)$较大的为优解

§7.3 区间估计

6.什么是区间估计

范围与可信程度

7.置信区间

①双侧置信区间：$P(\underline{\theta}<\theta<\overline{\theta})\geq1-\alpha$.

$\underline{\theta}$为置信下限，$\overline{\theta}$置信上限　$(\underline{\theta},\overline{\theta})$置信区间

②单侧置信区间：$P(\underline{\theta}<\theta)=1-\alpha$.

$P(\theta<\overline{\theta})=1-\alpha$.

8.基本原理.

总体$X\sim N(\mu,\sigma^2)$，σ^2已知，μ未知 设$X_1\cdots X_n$是来自X的样本，求μ置信水平为$1-\alpha$置信区间 枢轴量$W=\left(\frac{\bar X-\mu}{\sigma/\sqrt n}\right)\sim N(0,1)$

$$P(a<W<b)=P\left(a<\frac{\bar X-\mu}{\sigma/\sqrt n}<b\right)=1-\alpha \to P\left(\left|\frac{\bar X-\mu}{\sigma/\sqrt n}\right|<z_{\frac{\alpha}{2}}\right)=1-\alpha$$

$$P\left(\bar X-\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}}<\mu<\bar X+\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}}\right)=1-\alpha$$ 区间为$\left(\bar X-\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}},\bar X+\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}}\right)$

常见置信区间：

① σ^2已知，μ未知置信 $\left(\bar X-\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}},\bar X+\frac{\sigma}{\sqrt n}z_{\frac{\alpha}{2}}\right)$ 正态

② σ^2未知，μ …… $\left(\bar X-\frac{S}{\sqrt n}t_{\frac{\alpha}{2}}(n-1),\bar X+\frac{S}{\sqrt n}t_{\frac{\alpha}{2}}(n-1)\right)$ t

③ μ未知，σ^2…… $\left(\frac{(n-1)S^2}{\chi^2_{\frac{\alpha}{2}}(n-1)},\frac{(n-1)S^2}{\chi^2_{1-\frac{\alpha}{2}}(n-1)}\right)$ χ^2

④ μ已知，σ^2…… $\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\frac{\alpha}{2}}(n)},\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{1-\frac{\alpha}{2}}(n)}\right)$ χ^2

⑤ σ_1^2,σ_2^2已知，$\mu_1-\mu_2$ $\left(\bar X_1-\bar X_2\pm z_{\frac{\alpha}{2}}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right)$ 正态分布

后省略

二、重要题型

~~一、设~~ §7.4 连续型随机变量的参数估计

例：设总体X具有密度函数$f_X(x,\theta)=\frac{\theta^x e^{-\theta}}{x!}$且$0<x<\theta$，总体$Y$有密度函数$f(y,\lambda)=\frac{6y}{\lambda^3}(\lambda-y)$ $0<y<\lambda$，$X_1\cdots X_n$与$Y_1\cdots Y_n$为样本

(1)求θ的极大似然估计 (2)求λ的矩估计$\hat\lambda$ (3)求$\hat\lambda$的方差

解：(1) $L(\theta)=\prod_{i=1}^{n}e^{-\theta}\frac{\theta^{x_i}}{x_i!}=e^{-n\theta}\theta^{\sum_{i=1}^{n}x_i}\prod_{i=1}^{n}\frac{1}{x_i!}$

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

$\frac{\partial\ln(L(\theta))}{\partial\theta}=0\Rightarrow -n+\frac{\sum_{i=1}^{n}x_i}{\theta}=0\Rightarrow\hat\theta=\bar X$

(2) $E(Y)=\int_0^{\lambda}\frac{6y^2}{\lambda^3}(\lambda-y)=\frac{\lambda}{2}=\bar Y$ 则：$\lambda=2\bar Y$

(3) $D(\hat\lambda)=D(2\bar Y)=4D(\bar Y)=4D\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)=\frac{4}{n^2}D\left(\sum_{i=1}^{n}Y_i\right)=\frac{4}{n}DY$

$EY^2=\cancel{\int_{-\infty}^{+\infty}}\frac{3}{10}\lambda^2$. $DY=\frac{1}{20}\lambda^2$. $D(\hat\lambda)=\frac{\lambda^2}{5n}$

§7.5 参数估计之选择

例：设总体X概率密度 $f(x,\theta)=\begin{cases}\frac{3x^2}{\theta^3} & 0\le x\le\theta\\ 0 & 其他\end{cases}$

(1) 求 θ 的极大似然估计量 $\hat{\theta}$

(2) $\hat{\theta}$ 是无偏估计吗?

解: $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\frac{3^n\prod_{i=1}^{n} x_i^2}{\theta^{3n}}$

$\because \frac{\partial \ln L(\theta)}{\partial\theta}\ne 0$ 所以要使 $L(\theta)$ 越大,

$\therefore \theta=\max(x_i)$ 因为 $0<x<\theta$. θ 要大于X, 所以 ~~θ要大于~~ $\theta\ge\max(x_i)$

注意考虑范围

$P(\hat{\theta}<x)=P(X_1\le x)\cdots P(X_n\le x)$

$=\begin{cases}\left(\int_0^x \frac{3x^2}{\theta^3}dx\right)^n & 0\le x\le\theta\\ 0 & 其他\end{cases}$

$f_{\hat{\theta}}(x)=\begin{cases}\frac{3n x^{3n-1}}{\theta^{3n}} & 0\le x\le\theta\\ 0 & 其他\end{cases}$

$E=\int_0^{\theta}\frac{3n x^{3n}}{\theta^{3n}}dx=\frac{3n}{3n+1}\theta$ 有偏

§7.6 置信区间

例: $X\sim N(\mu,0.02^2)$ 25个 → $\bar{x}=1.05$ 则对 μ 置信水平为0.95置信区间为

______ ($\Phi(1.96)=0.975$ $\Phi(1.645)=0.95$)

解: 枢轴变量 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ 置信区间: $\left(\bar{X}-\frac{\sigma}{\sqrt{n}}u_{0.025},\ \bar{X}+\frac{\sigma}{\sqrt{n}}u_{0.025}\right)$

$=\left(1.05-\frac{0.02}{\sqrt{25}}\times1.96,\ 1.05+\frac{0.02}{\sqrt{25}}\times1.96\right)$

$=(1.042,\ 1.058)$