1 一维非均匀随机数的产生

1.1 连续非均匀随机数的产生

1.1.1 逆变换抽样法

Theorem. 设连续型随机变量 η 的分布函数 $F(x)$ 是连续且严格单调上升的分布函数, 其反函数 $F^{-1}(x)$ 存在, 则有
(1) 随机变量 $F(\eta)$ 服从 $(0,1)$ 上的均匀分布, 即 $F(\eta) \sim U(0,1)$
(2) 对于随机变量 $U \sim U(0,1)$, $F^{-1}(U)$的分布函数为 $F(x)$

Proof. $F(\eta)$ 的分布函数为:

$$P(F(\eta) \leq x) = \begin{cases} 1 & \text{if } x \geq 1 \\ P(\eta \leq F^{-1}(x)) = F(F^{-1}(x)) = x & \text{if } x \in [0,1] \\ 0 & \text{if } x \leq 0 \end{cases}$$

因此，$F(\eta) \sim U(0,1)$

$F^{-1}(U)$ 的分布函数为:

$$P\big(F^{-1}(U) \leq x\big) = P(U \leq F(x)) = F(x)$$

因此, $F^{-1}(U)$的分布函数为$F(x)$

Corollary. 已知连续型随机变量$\xi \sim G(x)$, 如果η的分布函数$F(x)$存在反函数, 则随机变量 $F^{-1}(G(\xi)) \sim F(x)$.

逆变换抽样法的步骤:
(1) 产生$U(0,1)$的随机数序列$\{u_i, i = 1, 2, ...\}$
(2) η的随机数序列为

$$\eta_i = F^{-1}(u_i), i = 1, 2, ...$$

逆变换抽样法的缺点: 有些分布的反函数不能用初等函数表出

1.1.2 舍选抽样法

Theroem. 设$f(\cdot), g(\cdot)$分别为概率密度函数, $h(\cdot)$为给定函数(不要求是概率密度函数), 若按下列步骤生成随机数η:
(1) 独立生成$X \sim f(x)$和$Y \sim g(y)$
(2) 如果X, Y满足$Y \leq h(X)$, 则令$\eta = X$并输出η, 否则返回上一步

令$G(y)$为随机变量Y的分布函数, 则η的概率密度函数为:

$$p(z) = \frac{f(z)G(h(z))}{\int_{-\infty}^{+\infty} f(y)G(h(y))\,\mathrm{d}y}$$

Proof. η的分布函数为:

$$
\begin{aligned}
P(\eta \le z) = P\{X \le z | Y \le h(X)\} &= \frac{P\{X \le z, Y \le h(X)\}}{P\{Y \le h(X)\}} \\
&= \frac{\int_{-\infty}^{z} \int_{-\infty}^{h(x)} f(x)g(y)\,\mathrm{d}y\,\mathrm{d}x}{\int_{-\infty}^{+\infty} \int_{-\infty}^{h(x)} f(x)g(y)\,\mathrm{d}y\,\mathrm{d}x} \\
&= \frac{\int_{-\infty}^{z} f(x)G(h(x))\,\mathrm{d}x}{\int_{-\infty}^{+\infty} f(x)G(h(x))\,\mathrm{d}x}
\end{aligned}
$$

因此, η的概率密度函数为

$$
p(z) = P'(z) = \frac{f(z)G(h(z))}{\int_{-\infty}^{+\infty} f(x)G(h(x))\,\mathrm{d}x}
$$

Corollary. 设随机变量η的概率密度函数为$p(z) \le M(z)$且$\int_{-\infty}^{+\infty} M(z)\,\mathrm{d}z = C < \infty$, 令概率密度函数$f(x) = \frac{M(x)}{C}$且$g(y)$是$U(0,1)$的概率密度函数,若按以下方法生成随机变量$\eta$, 则
(1) 独立生成$X \sim f(x)$和$Y \sim g(y)$
(2) 直到$Y \le \frac{P(X)}{M(X)}$时, 令$\eta = X$, 输出η

需满足
(1) $\forall x, p(x) \le M(x)$
(2) $f(x) = \frac{M(x)}{C}$的随机数容易生成

Corollary. 设随机变量η在有限区间$[a,b]$上取值, 且有概率密度函数$p(z)$并且$\sup_{z\in(a,b)} p(z) = M < \infty$, 则可以按下列方法生成随机变量$\eta$:
(1) 独立生成$X \sim U(a,b)$和$Y \sim U(0,1)$
(2) X, Y满足$Y \le \frac{p(X)}{M}$时, 令$\eta = X$, 输出η

Corollary. 设随机变量η的概率密度函数$p(z)$可表示为:

$$
p(z) = Lh(z)f(z)
$$

则可以按下列方法生成随机变量η:
(1) 独立生成$X \sim f(x)$和$Y \sim U(0,1)$
(2) 直到$Y \le h(X)$, 令$\eta = X$, 输出η

1.1.3 变换抽样法

Theorem. 设随机变量ξ有概率密度函数$f(x)$, 另有一函数$h(z)$严格单调, 其反函数记为$h^{-1}(z)$且导函数存在, 则$\eta = h(\xi)$是随机变量ξ的函数, 其概率密度函数为

$$
p(z) = f\big(h^{-1}(z)\big) \cdot \big|h^{-1}(z)'\big|
$$

Proof. η的分布函数为

$$
\begin{aligned}
P(\eta \le z) = P(h(\xi) \le z) &= P\big(\xi \le h^{-1}(z)\big) \\
&= \int_{-\infty}^{h^{-1}(z)} f(x)\,\mathrm{d}x
\end{aligned}
$$

故

$$p(z) = \frac{\mathrm{d}}{\mathrm{d}z} P(\eta \le z) = f\big(h^{-1}(z)\big)\big|h^{-1}(z)'\big|$$

Theorem. 设二维随机变量(ξ_1, ξ_2)的联合概率密度函数为$f(x_1, x_2)$. 令$\eta = g(\xi_1, \xi_2)$且存在唯一的逆变换$h(\cdot, \cdot)$, 其一阶偏导数存在, 则随机变量η的概率密度函数为:

$$p(y) = \int_{-\infty}^{+\infty} f(h(y, z), z)\left|\frac{\partial h}{\partial y}\right| \mathrm{d}z$$

Proof. 令

$$\begin{cases} \eta = g(\xi_1, \xi_2) \\ \zeta = \xi_2 \end{cases} \Longrightarrow \begin{cases} \xi_1 = h(\eta, \zeta) \\ \xi_2 = \zeta \end{cases}$$

则 Jacobi 行列式为

$$J = \begin{vmatrix} \frac{\partial h}{\partial \eta} & \frac{\partial h}{\partial \zeta} \\ 0 & 1 \end{vmatrix} = \frac{\partial h}{\partial \eta}$$

故(η, ζ)的联合概率密度函数为:

$$p(y, z) = f(h(y, z), z)|J| = f(h(y, z), z)\left|\frac{\partial h}{\partial y}\right|$$

故η的概率密度函数为:

$$p(y) = \int_{-\infty}^{+\infty} f(h(y, z), z)\left|\frac{\partial h}{\partial y}\right| \mathrm{d}z$$

1.1.4 复合抽样法

Theorem. 设随机变量η的分布函数和概率密度函数分别为$F(x)$和$f(x)$, 并且可以写成

$$F(x) = \sum_{j=1}^{K} p_j F_j(x) \text{或} f(x) = \sum_{j=1}^{K} p_j f_j(x)$$

其中$p_j \ge 0, \sum_{j=1}^{K} p_j = 1$, $F_j(x)$和$f_j(x)$分别是随机变量ξ_j的分布函数和概率密度函数. 复合抽样法的步骤为:

(1) 产生随机数$U \sim U(0,1)$, 如果$U \in [\sum_{j=1}^{l-1} p_j, \sum_{j=1}^{l} p_j)$, 则令$J = l$

(2) 产生分布函数为$F_J(x)$或概率密度函数为$f_J(x)$的随机变量, 记为η, 输出η

Proof. 由全概率公式

$$
\begin{aligned}
P(\eta \le x) &= \sum_{j=1}^{K} P(J=j)P(\eta \le x|J=j) \\
&= \sum_{j=1}^{K} p_j F_j(x)
\end{aligned}
$$

1.1.5 近似抽样法

1.1.5.a 利用中心极限定理近似抽样

Theorem. 设$\{\xi_1, \xi_2, ..., \xi_m\}$为独立同分布的随机变量序列，其数学期望$\mu$和方差$\sigma^2$存在，则$\sum_{i=1}^{m}\xi_i$的标准化形式为

$$
\eta_m = \frac{\sum_{i=1}^{m}\xi_i - m\mu}{\sigma\sqrt{m}}
$$

满足中心极限定理, 随机变量η_m的分布函数记为$F_m(x)$, 有

$$
\lim_{m\to\infty} F_m(x) = \Phi(x)
$$

1.1.5.b 对概率密度函数进行近似

若随机变量η的概率密度函数有:

$$
f(x) = \sum_{j=1}^{m} p_j f_j(x)
$$

则我们可以用简单且容易产生随机数的概率密度函数$\widetilde{f_j}(x)$近似$f_j(x)$, 然后用复合抽样法生成η.

Bulter 抽样法

(1) 将随机变量η的取值区间(a,b)划分为m个小区间，即$(x_{j-1}, x_j], j=1,2,...,m$，其中$x_0 = a, x_m = b$, 且有$F(x_j) - F(x_{j-1}) = \frac{1}{m}$

(2) 对概率密度函数$f(x)$进行分解, $f(x) = \sum_{j=1}^{m} p_j f_j(x), p_j = \frac{1}{m}$

$$
f_j(x) = mf(x)I\big(x \in (x_{j-1}, x_j]\big)
$$

(3) 在小区间$(x_{k-1}, x_k]$上用线性函数$\widetilde{f_k}(x)$作曲线$f_k(x)$的近似, 对于第j个小区间, 线性函数由$\big(x_{j-1}, mf(x_{j-1})\big)$和$\big(x_j, mf(x_j)\big)$两点确定

1.1.5.c 对分布函数进行近似

思想类似于概率密度函数, 对$F(x)$线性近似$\Longrightarrow$在小区间内均匀分布

1.1.5.d 经验分布抽样法

经验分布函数$F_n(x)$可以很好地近似分布函数$F(x)$

$$
F_n(x) = \sum_{i=1}^{n} I(X_i \le x)
$$

在小区间内, 服从均匀分布(思想类似于线性插值).

1.2 离散非均匀随机数的产生

1.2.1 逆变换法

Theorem. 设$\{x_i\}$为离散型随机变量ξ所有可能的取值, p_i是ξ取x_i的概率, 即

$$P(\xi = x_i) = p_i$$

ξ的分布函数为

$$F(x) = P(\xi \leq x) = \sum_{x_i \leq x} p_i$$

则按以下方法产生随机数ξ:
(1) 产生随机数$R \sim U(0,1)$
(2) 若$R \leq F(x_1)$, 则令$\xi = x_1$. 否则若$F(x_{i-1}) < R \leq F(x_i)$, 则令$\xi = x_i$, 输出$\xi$

ξ是分布为$F(x)$的随机数.

Proof. 由于

$$\begin{aligned} P(\xi = x_1) &= P(R \leq F(x_1)) = F(x_1) = p_1 \\ P(\xi = x_i) &= P(F(x_{i-1} < R \leq F(x_i))) = F(x_i) - F(x_{i-1}) = p_i \end{aligned}$$

因此$\xi \sim F(x)$.

Practice.

Rayleigh 分布具有密度函数

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}, \quad x \geq 0, \sigma > 0$$

Solution. 令$h(x) = \frac{x^2}{2\sigma^2}$, 则有 $f(x) = |h'(x)|e^{-h(x)}$. 因此, 若$X \sim \mathrm{Exp}(1)$, 则令$x = h(\xi)$, 即

$$\xi = h^{-1}(x) = \pm\sigma\sqrt{2x}$$

又因$\xi \geq 0$, 因此$\xi = \sigma\sqrt{2x}$, 故生成ξ的变换抽样法步骤如下:
(1) 生成$x \sim \mathrm{Exp}(1)$
(2) 令$\xi = \sigma\sqrt{2x}$, 输出ξ

2 随机向量随机数的抽样法

2.1 连续随机向量随机数的抽样法

2.1.1 变换抽样法

Theorem. 设随机向量(ξ_1, ξ_2)具有二维联合密度函数$f(x_1, x_2)$, 令

$$\begin{cases} y_1 = g_1(x_1, x_2) \\ y_2 = g_2(x_1, x_2) \end{cases}$$

且存在唯一的逆变换

$$\begin{cases} x_1 = h_1(y_1, y_2) \\ x_2 = h_2(y_1, y_2) \end{cases}$$

设h_1, h_2的一阶偏导数存在, 函数变换的 Jacobi 行列式为:

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} \neq 0$$

令

$$\begin{cases} \eta_1 = g_1(\xi_1, \xi_2) \\ \eta_2 = g_2(\xi_1, \xi_2) \end{cases}$$

则随机变量(η_1, η_2)的二维联合密度函数为:

$$p(y_1, y_2) = f(h_1(y_1, y_2), h_2(y_1, y_2))|J|$$

变换抽样法的步骤:
(1) 产生多维随机数$\left(\xi_1, \xi_2, ..., \xi_p\right) \sim f\left(x_1, x_2, ..., x_p\right)$
(2) 令$\eta_i = g_i\left(\xi_1, \xi_2, ..., \xi_p\right), i = 1, 2, ..., q$

2.1.2 条件分布法

$$f\left(x_1, x_2, ..., x_p\right) = f_1(x_1) f_2(x_2|x_1) ... f_p\left(x_p|x_1, x_2, ..., x_{p-1}\right)$$

条件分布法的步骤:
(1) 生成$x_1 \sim f_1(x_1)$
(2) 给定$\xi_1 = x_1$, 生成$x_2 \sim f_2(x_2|\xi_1 = x_1)$

$$\vdots$$

(p) 给定$\left(\xi_1, \xi_2, ..., \xi_{p-1}\right) = \left(x_1, x_2, ..., x_{p-1}\right)$, 产生随机数

$$x_p \sim f_p\left(x_p\middle|\left(\xi_1, \xi_2, ..., \xi_{p-1}\right) = \left(x_1, x_2, ..., x_{p-1}\right)\right)$$

得到密度函数为$f\left(x_1, x_2, ..., x_p\right)$的随机向量$\left(x_1, x_2, ..., x_p\right)$.

2.1.3 舍选抽样法

设随机向量$\left(\xi_1, \xi_2, ..., \xi_p\right)$在平行多面体$a_i \leq x_i \leq b_i, i = 1, 2, ..., p$上取值, 其联合密度函数的上界为:

$$f_0 = \sup_{a_i \leq x_i \leq b_i} f\left(x_1, x_2, ..., x_p\right) < \infty$$

则可以按以下方法生成随机向量:

(1) 从$U(0,1)$中独立产生随机数$U_0, U_1, ..., U_p$.

(2) 若$U_0, U_1, ..., U_p$满足$U_0 f_0 \le f\big((b_1 - a_1)U_1 + a_1, ..., (b_p - a_p)U_p + a_p\big)$, 则令 $\xi_1 = (b_1 - a_1)U_1 + a_1, ..., \xi_p = (b_p - a_p)U_p + a_p$

Proof. 如果$a_1 = a_2 = ... = a_p = 0, b_1 = b_2 = ... = b_p = 1$, 则根据上述抽样过程可知随机向量$(\xi_1, \xi_2, ..., \xi_p)$的分布函数为:

$$
\begin{aligned}
& P(\xi_1 \le x_1, \xi_2 \le x_2, ..., \xi_p \le x_p) \\
&= P\big(U_1 \le x_1, U_2 \le x_2, ..., U_p \le x_p | U_0 f_0 \le f(U_1, U_2, ..., U_p)\big) \\
&= \frac{P\big(U_1 \le x_1, U_2 \le x_2, ..., U_p \le x_p, U_0 f_0 \le f(U_1, U_2, ..., U_p)\big)}{P\big(U_0 f_0 \le f(U_1, U_2, ..., U_p)\big)} \\
&= \frac{\int_0^{x_1}\int_0^{x_2}...\int_0^{x_p}\int_0^{\frac{f(u_1,u_2,...,u_p)}{f_0}} \mathrm{d}u_0\,\mathrm{d}u_p ...\,\mathrm{d}u_2\,\mathrm{d}u_1}{\int_0^1\int_0^1...\int_0^1\int_0^{\frac{f(u_1,u_2,...,u_p)}{f_0}} \mathrm{d}u_0\,\mathrm{d}u_p ...\,\mathrm{d}u_2\,\mathrm{d}u_1} \\
&= \frac{\int_0^{x_1}\int_0^{x_2}...\int_0^{x_p} \frac{f(u_1,u_2,...,u_p)}{f_0}\,\mathrm{d}u_p ...\,\mathrm{d}u_2\,\mathrm{d}u_1}{\int_0^1\int_0^1...\int_0^1 \frac{f(u_1,u_2,...,u_p)}{f_0}\,\mathrm{d}u_p ...\,\mathrm{d}u_2\,\mathrm{d}u_1} \\
&= \int_0^{x_1}\int_0^{x_2}...\int_0^{x_p} f(u_1, u_2, ..., u_p)\,\mathrm{d}u_p ...\,\mathrm{d}u_2\,\mathrm{d}u_1
\end{aligned}
$$

2.2 离散随机向量随机数的抽样法

2.2.1 条件分布法

$$
P(X = x_i, Y = y_j) = p_{ij} \quad P(X = x_i) = p_{i\cdot} \quad P(Y = y_j) = p_{\cdot j}
$$

$$
P(X = x_i | Y = y_j) = \frac{p_{ij}}{p_{\cdot j}}
$$

3 参数估计数值计算

3.1 点估计数值计算

步骤:

(1) 从总体X中随机生成数据$X_{11}, X_{12}, ..., X_{1n}$. 基于数据得到$\theta$的估计$\hat{\theta} := \hat{\theta}(X_{11}, X_{12}, ..., X_{1n})$, 记为$\hat{\theta}^{(1)}$

(2) 重复上述步骤K次, 相应得到$\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ..., \hat{\theta}^{(K)}$

(3) 基于样本$\hat{\theta}^{(1)}, \hat{\theta}^{(2)}, ..., \hat{\theta}^{(K)}$的均值和方差估计总体$\hat{\theta}(X_1, X_2, ..., X_n)$的均值和方差, 即

$$\hat{E}\left(\hat{\theta}(X_1, X_2, ..., X_n)\right) = \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}^{(k)} := \hat{E}\left(\hat{\theta}\right)$$

$$\widehat{\text{Var}}\left(\hat{\theta}(X_1, X_2, ..., X_n)\right) = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\theta}^{(k)} - \hat{E}\left(\hat{\theta}\right)\right)^2$$

也可以用来估计 MSE, 即

$$\widehat{\text{MSE}} = \hat{E}\left(\hat{\theta} - \theta\right)^2$$

3.2 置信区间估计

Definition. 对感兴趣的参数 θ ,设 $X_1, X_2, ..., X_n$ 是来自该总体的一个样本, 对给定的 $\alpha(0 < \alpha < 1)$, 确定两个统计量

$$\hat{\theta}_L = \hat{\theta}_L(X_1, X_2, ..., X_n) \text{与} \hat{\theta}_U = \hat{\theta}_U(X_1, X_2, ..., X_n)$$

使得

$$P\left(\hat{\theta}_L \le \theta \le \hat{\theta}_U\right) = 1 - \alpha$$

则称随机区间 $\left[\hat{\theta}_L, \hat{\theta}_U\right]$ 是 θ 的置信水平为 $1-\alpha$ 的置信区间, $\left[\hat{\theta}_L, \hat{\theta}_U\right]$ 包含真值的概率(覆盖率)为

$$\text{Prob} = P_\theta\left(\hat{\theta}_L \le \theta \le \hat{\theta}_U\right)$$

对于置信区间的精确度, 一般的衡量标准是区间的平均值或平均长度

$$\text{Inter} = \left[E\left(\hat{\theta}_L\right), E\left(\hat{\theta}_U\right)\right] \text{或} E\left(\hat{\theta}_U\right) - E\left(\hat{\theta}_L\right)$$

数值模拟覆盖率和置信区间

(1) 从总体样本 X 中抽取随机样本 $X_1^{(1)}, X_2^{(1)}, ..., X_n^{(1)}$, 基于此样本计算

$$\hat{\theta}_L^{(1)} = \hat{\theta}_L\left(X_1^{(1)}, X_2^{(1)}, ..., X_n^{(1)}\right), \hat{\theta}_U^{(1)} = \hat{\theta}_U\left(X_1^{(1)}, X_2^{(1)}, ..., X_n^{(1)}\right)$$

(1) 重复上述步骤 K 次, 得到 $\hat{\theta}_L^{(1)}, \hat{\theta}_U^{(1)}, ..., \hat{\theta}_L^{(K)}, \hat{\theta}_U^{(K)}$

(2) 覆盖率 Prob 和置信区间的长度 Inter 的估计为:

$$\widehat{\text{Prob}} = \frac{1}{K}\sum_{k=1}^{K} I\left(\hat{\theta}_L^{(k)} \le \theta \le \hat{\theta}_U^{(k)}\right), \widehat{\text{Inter}} = \left[\frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_L^{(k)}, \frac{1}{K}\sum_{k=1}^{K}\hat{\theta}_U^{(k)}\right]$$

3.2.1 单总体置信区间估计

在此只列出枢轴量

3.2.1.a 总体均值的区间估计

当 σ 已知时:

$$\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{\sigma} \sim N(0,1)$$

当σ未知时:

$$\frac{\sqrt{n}\left(\overline{X}-\mu\right)}{S^2} \sim t(n-1)$$

3.2.1.b 总体方差的区间估计

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

3.2.1.c 总体比例的区间估计
当样本量足够大时, 可用渐进分布估计近似的置信区间. 若

$$X_1, X_2, ..., X_n \sim \mathrm{Bino}(1,p)$$

则

$$\frac{\sqrt{n}\left(\overline{X}-p\right)}{\sqrt{p(1-p)}} \xrightarrow{d} N(0,1)$$

3.2.2 两总体置信区间估计

3.2.2.a 两总体均值之差的区间估计
当σ_1^2, σ_2^2已知时:

$$U = \frac{\overline{X}_1-\overline{X}_2-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{m}}} \sim N(0,1)$$

当$\sigma_1^2=\sigma_2^2=\sigma^2$未知时:

$$T_{xy} = \frac{\left(\overline{X}-\overline{Y}\right)-(\mu_1-\mu_2)}{\sqrt{S_w\left(\frac{1}{n}+\frac{1}{m}\right)}} \sim t(n+m-2)$$

其中$S_w = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2+\sum_{i=1}^{m}\left(Y_i-\overline{Y}\right)^2}{n+m-2}$

3.2.2.b 两总体方差比的区间估计

$$F = \frac{\frac{S_x^2}{\sigma_1^2}}{\frac{S_y^2}{\sigma_2^2}} \sim F_{n-1,m-1}$$

其中$S_x^2 = \frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2}{n-1}, S_y^2 = \frac{\sum_{i=1}^{m}\left(Y_i-\overline{Y}\right)^2}{m-1}$

3.2.2.c 两总体比例之差的区间估计

若有

$$X_1, X_2, ..., X_n \sim \mathrm{Bino}(1, p_1)\text{且}Y_1, Y_2, ..., Y_m \sim \mathrm{Bino}(1, p_2)$$

则

$$\frac{\overline{X} - \overline{Y} - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n} + \frac{p_2(1-p_2)}{m}}} \xrightarrow{d} N(0,1)$$

4 假设检验数值计算

4.1 参数检验数值计算

用数值方法估计假设检验问题的第一类错误概率α以及检验的功效$1-\beta$.

在有显式表达式的情况下, 功效的检验:

$$\begin{aligned}
p &= P(\text{拒绝}H_0|H_1\text{成立}) \\
&= 1 - P\left(\left|\sqrt{n}\overline{X}\right| \le u_{1-\frac{\alpha}{2}}|H_1\text{成立}\right) \\
&= 1 - P\left(-u_{1-\frac{\alpha}{2}} - \sqrt{n}\mu \le \sqrt{n}\left(\overline{X} - \mu\right) \le u_{1-\frac{\alpha}{2}} - \sqrt{n}\mu|H_1\text{成立}\right) \\
&= 1 - \Phi\left(u_{1--\frac{\alpha}{2}} - \sqrt{n}\mu\right) + \Phi\left(-u_{1--\frac{\alpha}{2}} - \sqrt{n}\mu\right) \\
&= \Phi\left(-u_{1-\frac{\alpha}{2}} + \sqrt{n}\mu\right) + \Phi\left(-u_{1-\frac{\alpha}{2}} - \sqrt{n}\mu\right)
\end{aligned}$$

当然, 我们也可以用数值模拟的方法估计功效:

(1) 从备择假设$N(\mu, 1) \quad (\mu \ne 0)$中生成数据$X_{11}, X_{12}, ..., X_{1n}$, 样本均值为

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

相应的统计量为$\sqrt{n}\overline{X}^{(1)}$

(2) 重复上述步骤K次, 相应得到$\sqrt{n}\overline{X}^{(1)}, \sqrt{n}\overline{X}^{(2)}, ..., \sqrt{n}\overline{X}^{(K)}$

(3) 功效p的估计为

$$\hat{p} = \frac{1}{K}\sum_{i=1}^{K} I\left(\left|\sqrt{n}\overline{X}^{(i)}\right| > u_{1-\frac{\alpha}{2}}\right)$$

4.2 单样本的拟合优度检验

4.2.1 总体分布的卡方检验

随机样本为$r_1, r_2, ..., r_n$, 有假设检验:

$$H_0: r_1, r_2, ..., r_n \sim \text{某一分布} \quad H_1: \mathrm{others}$$

将分布的定义域划分成m份, 每份对应的概率为$\frac{1}{m}$, 我们定义n_i为随机样本中落入区间i的数目, 定义$\mu_i=\frac{n}{m}$为理论频数, 则有统计量

$$T=\sum_{i=1}^{m}\frac{(\mu_i-n_i)^2}{\mu_i}=\sum_{i=1}^{m}\left(\mu_i-\frac{n_i^2}{\mu_i}\right)\xrightarrow{L}\chi^2(m-1)$$

4.2.2 单样本 **K-S** 检验

为了检验样本的经验分布与理论分布(由原假设决定)是否一致. $X_1,X_2,...,X_n$为独立观测的样本, $X_{(1)}\le X_{(2)}\le...\le X_{(n)}$为次序样本, 经验分布函数为$F\left(X_{(i)}\right)=\frac{i}{n}$, 则经验分布函数和原假设下的分布函数的最大偏差为:

$$K_n=\max_{1\le i\le n}\left[\max\left(\left|F\left(X_{(i)}\right)-\frac{i-1}{n}\right|,\left|F\left(X_{(i)}\right)-\frac{i}{n}\right|\right)\right]$$

该统计量的分布函数为

$$P(K_n\le x)=1-2\sum_{j=1}^{\infty}(-1)^{j-1}\exp(-2nj^2x^2)$$

4.3 两样本的非参数检验

4.3.1 两样本的 **Mann-Whitney U** 检验

前提: 两样本的差异只在于位置参数

来自两总体的样本 $X_1,X_2,...,X_n$和 $Y_1,Y_2,...,Y_m$有对应的秩 $R_1,R_2,...,R_n$和$R_{n+1},R_{n+2},...,R_{n+m}$, 因此有

$$\begin{aligned}
W_x&=\sum_{i=1}^{n}R_i & W_y&=\sum_{i=n+1}^{n+m}R_i\\
E(W_x)&=\frac{n(n+m+1)}{2} & E(W_y)&=\frac{m(n+m+1)}{2}\\
\operatorname{Var}(W_x)&=mn\frac{n+m+1}{12} & \operatorname{Var}\left(W_y\right)&=mn\frac{n+m+1}{12}
\end{aligned}$$

又因

$$\begin{aligned}
W_x&=\sum_{i=1}^{n}\sum_{k=1}^{m}I(Y_k\le X_i)+\sum_{i=1}^{n}\sum_{j=1}^{n}I\left(X_j\le X_i\right)\\
&=\sum_{i=1}^{n}\sum_{k=1}^{m}I(Y_k\le X_i)+\frac{n(n+1)}{2}
\end{aligned}$$

令

$$W_{yx}=(mn)^{-1}\sum_{i=1}^{n}\sum_{k=1}^{m}I(Y_k\le X_i)\sim N\left(\frac{1}{2},\frac{n+m+1}{12mn}\right)$$

故

$$\frac{W_{yx}-\frac{1}{2}}{\sqrt{\frac{n+m+1}{12mn}}}\sim N(0,1)$$

4.3.2 两样本的 **K-S** 检验
假设检验如下:

$$H_0: F(x)=G(x)\quad H_1: F(x)\neq G(x)$$

此处不仅仅是位置参数不同, 让两总体的经验分布函数分别为:

$$F_n(x)=\sum_{i=1}^{n}\frac{I(X_i\leq x)}{n}\ \text{以及}\ G_m(x)=\sum_{i=1}^{m}\frac{I(Y_i\leq x)}{m}$$

因此类似单样本的 K-S 检验, 检验统计量为分布函数差距的最大值, 即

$$D_N=\max\left[\max_{1\leq i\leq n}(|F_n(X_i)-G_m(X_i)|),\max_{1\leq j\leq m}\left(\left|F_n\left(Y_j\right)-G_m\left(Y_j\right)\right|\right)\right]$$

分布函数同单样本情况下的分布函数.

4.3.3 独立性检验
相关系数的定义:

$$\rho=\frac{E(XY)-E(X)E(Y)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}=\frac{E\left(X-\overline{X}\right)\left(Y-\overline{Y}\right)}{\sqrt{\operatorname{Var}(X)\operatorname{Var}(Y)}}$$

Pearson 相关系数的定义如下:

$$r=\frac{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)\left(Y_i-\overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\sum_{i=1}^{n}\left(Y_i-\overline{Y}\right)^2}}$$

且有

$$t=r\sqrt{\frac{n-2}{1-r^2}}\sim t(n-2)$$

Spearman r_s秩相关系数的定义:

$$r_s=\frac{\sum_{i=1}^{n}\left(R_i-\overline{R}\right)\left(S_i-\overline{S}\right)}{\sqrt{\sum_{i=1}^{n}\left(R_i-\overline{R}\right)^2\sum_{i=1}^{n}\left(S_i-\overline{S}\right)^2}}=1-\frac{6\sum_{i=1}^{n}\left(R_i-S_i\right)^2}{n(n^2-1)}$$

且有

$$r_s\sqrt{n-1}\xrightarrow{L}N(0,1)$$

Kendall τ相关系数的定义为:

$$\tau = \binom{n}{2}^{-1} \sum_{1 \le i \le j \le n} \operatorname{sign}\big((X_i - Y_i)(X_j - Y_j)\big)$$

且有

$$\tau \sqrt{\frac{9n(n-1)}{2(2n+5)}} \xrightarrow{L} N(0,1)$$

5 重抽样方法

5.1 Bootstrap 估计的思想

Bootstrap 估计的步骤如下:

(1) 从样本$X_1, X_2, ..., X_n$中有放回地产生数据$X_{11}^*, X_{12}^*, ..., X_{1n}^*$, 得到$\theta$的估计, 记为 $\widehat{\theta}_1^* := \widehat{\theta}_1^*(X_{11}^*, X_{12}^*, ..., X_{1n}^*)$

(2) 重复上述步骤m次, 得到$\widehat{\theta}_1^*, \widehat{\theta}_2^*, ..., \widehat{\theta}_m^*$, 利用$\widehat{\theta}_1^*, \widehat{\theta}_2^*, ..., \widehat{\theta}_m^*$近似$\widehat{\theta}$的分布及其特征.

5.1.1 估计量的偏差的 Bootstrap 估计

估计量的偏差为$\operatorname{Bias}\big(\hat{\theta}\big) = E\big(\hat{\theta}\big) - \theta$, 由于$E\big(\hat{\theta}\big)$和$\theta$难以得到, 因此我们用$\hat{E}\big(\hat{\theta}\big)$以及$\hat{\theta}$进行估计, 由此得到偏差的估计:

$$\widehat{\operatorname{Bias}}\big(\hat{\theta}\big) = \hat{E}\big(\hat{\theta}\big) - \hat{\theta}$$

偏差的估计是无偏的吗? 以正态分布的方差为例:

$\operatorname{Bias}(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 = -\frac{1}{n}\sigma^2$, 而$E\Big(\widehat{\operatorname{Bias}}\big(\hat{\theta}\big)\Big) = \frac{1}{m}\sum_{i=1}^m E\big(\hat{\sigma}_i^{2^*}\big) - \hat{\sigma}^2$, 又因

$$\begin{aligned} E\big(\hat{\sigma}_1^{2^*} | X_1, X_2, ..., X_n\big) &= \frac{n-1}{n} \operatorname{Var}(X_{11}^* | X_1, X_2, ..., X_n) \\ &= \frac{n-1}{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}^2\right) \\ &= \frac{n-1}{n^2}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 \end{aligned}$$

故

$$\begin{aligned} &E\Big(\widehat{\operatorname{Bias}}(\hat{\sigma}^2) | X_1, X_2, ..., X_n\Big) \\ &= E\big(\hat{\sigma}_1^{2^*} | X_1, X_2, ..., X_n\big) - \hat{\sigma}^2 \\ &= \frac{n-1}{n^2}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 - \frac{1}{n}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 \\ &= -\frac{1}{n^2}\sum_{i=1}^n \left(X_i - \overline{X}\right)^2 = -\frac{1}{n}\hat{\sigma}^2 \end{aligned}$$

因此

$$E\left(\widehat{\text{Bias}}(\hat{\sigma}^2)\right) = E\left(E\left(\widehat{\text{Bias}}(\hat{\sigma}^2)|X_1, X_2, ..., X_n\right)\right) = -\frac{1}{n}E\left(\widehat{\sigma^2}\right)$$

因为σ^2的估计$\hat{\sigma}^2$不是无偏的, 因此偏差的估计也不是无偏的.

5.1.2 估计量方差的 Bootstrap 估计

$$\widehat{\text{Var}}\left(\hat{\theta}\right) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\theta}_i^* - \overline{\hat{\theta}}^*\right)^2$$

我们可以知道估计量方差的 Bootstrap 是渐进无偏的. 这里以正态分布的均值作为参数为例:

$$\text{Var}(\hat{\mu}) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right) = \frac{\sigma^2}{n}$$

正态分布的均值的方差估计为:

$$\widehat{\text{Var}}(\hat{\mu}) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{\mu}_i^* - \overline{\hat{\mu}}^*\right)^2 = \frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{*^2} - \overline{\hat{\mu}}^{*^2}$$

给定样本$X_1, X_2, ..., X_n$的情况下, 我们可以求得条件均值

$$\begin{aligned}
E\left(\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{*^2}|X_1, X_2, ..., X_n\right) &= E\left(\hat{\mu}_1^{*^2}|X_1, X_2, ..., X_n\right) \\
&= E\left(\left(\frac{1}{n}\sum_{i=1}^{n}X_{1i}^*\right)^2|X_1, X_2, ..., X_n\right) \\
&= \frac{1}{n}E\left(X_{11}^{*^2}|X_1, X_2, ..., X_n\right) + \frac{n-1}{n}(E(X_{11}^*|X_1, X_2, ..., X_n))^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{n}X_i^2 + \frac{n-1}{n}\overline{X}^2
\end{aligned}$$

$$\begin{aligned}
E\left(\overline{\hat{\mu}}^{*^2}|X_1, X_2, ..., X_n\right) &= E\left(\left(\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^*\right)^2|X_1, X_2, ..., X_n\right) \\
&= \frac{1}{m}E\left(\hat{\mu}_1^{*^2}|X_1, X_2, ..., X_n\right) + \frac{m-1}{m}\left(E\left(\hat{\mu}_1^*|X_1, X_2, ..., X_n\right)\right)^2 \\
&= \frac{1}{m}E\left(\hat{\mu}_1^{*^2}|X_1, X_2, ..., X_n\right) + \frac{m-1}{m}(E(X_{11}^*|X_1, X_2, ..., X_n))^2 \\
&= \frac{1}{m}E\left(\hat{\mu}_1^{*^2}|X_1, X_2, ..., X_n\right) + \frac{m-1}{m}\overline{X}^2
\end{aligned}$$

因此, 方差的估计的条件均值为:

$$E\left(\widehat{\operatorname{Var}}(\hat{\mu})|X_1,X_2,...,X_n\right)=E\left(\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i^{*^2}|X_1,X_2,...,X_n\right)-E\left(\overline{\hat{\mu}}^{*^2}|X_1,X_2,...,X_n\right)$$
$$=\frac{m-1}{m}\left(E\left(\hat{\mu}_1^{*^2}|X_1,X_2,...,X_n\right)-\overline{X}^2\right)$$
$$=\frac{m-1}{mn^2}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$$
$$=\frac{m-1}{mn}\hat{\sigma}^2$$

因此, 方差估计的期望为:

$$E\left(\widehat{\operatorname{Var}}(\hat{\mu})\right)=E\left(E\left(\widehat{\operatorname{Var}}(\hat{\mu})|X_1,X_2,...,X_n\right)\right)=\frac{m-1}{mn}E(\hat{\sigma}^2)=\frac{(m-1)(n-1)}{mn}\frac{1}{n}\sigma^2$$

因此正态分布期望方差的 Bootstrap 估计是有偏估计.

5.2 基于 Jackknife 法的估计

Jackknife 计算θ的具体步骤如下:

(1) 从观测样本$X_1,X_2,...,X_n$中去掉第i个数据X_i后的剩余样本, 定义为第i个 Jackknife 样本, 记为$X_{(-i)}=\left(X_1,...,X_{i-1},X_{i+1},...,X_n\right)$

(2) 基于第i个 Jackknife 样本$X_{(-i)}(i=1,2,...,n)$, 得到相应的估计

$$\hat{\theta}_{(-i)}=\hat{\theta}\left(X_{(-i)}\right)(i=1,2,...,n)$$

5.2.1 Bias 的 Jackknife 估计

$$\widehat{\operatorname{Bias}}\left(\hat{\theta}\right)=(n-1)\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{(-i)}-\hat{\theta}\right)$$

Bias 的 Jackknife 估计是无偏的, 证明如下:

假设参数为总体方差σ^2, 则有

$$\operatorname{Bias}(\hat{\sigma}^2)=E(\hat{\sigma}^2)-\sigma^2=\frac{n-1}{n}\sigma^2-\sigma^2=-\frac{1}{n}\sigma^2$$

同时, 我们有

$$E\left(\widehat{\operatorname{Bias}}(\hat{\sigma}^2)\right)=\frac{n-1}{n}E\left(\sum_{i=1}^{n}\left(\hat{\sigma}^2_{(-i)}-\hat{\sigma}^2\right)\right)=(n-1)E\left(\hat{\sigma}^2_{(-1)}-\hat{\sigma}^2\right)$$

又因

$$
\begin{aligned}
E\left(\hat{\sigma}^2_{(-1)} - \hat{\sigma}^2\right) &= E\left(\left(\hat{\sigma}^2_{(-1)} - \sigma^2\right) - \left(\hat{\sigma}^2 - \sigma^2\right)\right) \\
&= E\left(\hat{\sigma}^2_{(-1)} - \sigma^2\right) - E\left(\hat{\sigma}^2 - \sigma^2\right) \\
&= \mathrm{Bias}\left(\hat{\sigma}^2_{(-1)}\right) - \mathrm{Bias}\left(\hat{\sigma}^2\right) \\
&= -\frac{1}{n-1}\sigma^2 + \frac{1}{n}\sigma^2 \\
&= -\frac{1}{n(n-1)}\sigma^2 = \frac{\mathrm{Bias}\left(\hat{\sigma}^2\right)}{n-1}
\end{aligned}
$$

因此有

$$
E\left(\widehat{\mathrm{Bias}}\left(\hat{\sigma}^2\right)\right) = (n-1)\frac{\mathrm{Bias}\left(\hat{\sigma}^2\right)}{n-1} = \mathrm{Bias}\left(\hat{\sigma}^2\right)
$$

因此 Bias 的 Jackknife 估计是无偏的.

5.2.2 SE 的 Jackknife 估计

$$
\widehat{\mathrm{SE}}_{\mathrm{Jack}}\left(\hat{\theta}\right) = \left[\frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\theta}_{(-i)} - \hat{\theta}\right)^2\right]^{\frac{1}{2}}
$$

$\widehat{\mathrm{Var}}_{\mathrm{Jack}}\left(\hat{\theta}\right)$是$\hat{\theta}$的方差的无偏估计. 证明如下: 假设参数为总体均值$\hat{\mu}$, 则有

$$
\mathrm{Var}(\hat{\mu}) = \frac{\sigma^2}{n}
$$

同时我们有

$$
\begin{aligned}
\widehat{\mathrm{Var}}(\hat{\mu}) &= \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\mu}_{(-i)} - \hat{\mu}\right)^2 \\
&= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{n\overline{X} - X_i}{n-1} - \overline{X}\right)^2 \\
&= \frac{n-1}{n}\sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{n-1}\right)^2 \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2
\end{aligned}
$$

因此有

$$
E\left(\widehat{\mathrm{Var}}(\hat{\mu})\right) = \frac{\sigma^2}{n} = \mathrm{Var}(\hat{\mu})
$$

因此方差的 Jackknife 估计是无偏的.

5.3 Jackknife-after-Bootsrap 估计

步骤如下:

(1) 固定i, 从观测数据$X_{(i)} = (X_1, ..., X_{i-1}, X_{i+1}, ..., X_n)$中有放回地产生数据 $X^*_{(-i),1} = \left(X^*_{11}, X^*_{12}, ..., X^*_{1,n-1}\right)$, 基于此样本得到$\theta$的估计为$\hat{\theta}^*_{(-i),1}$

(2) 重复上述步骤m次, 得到$\hat{\theta}^*_{(-i),1}, \hat{\theta}^*_{(-i),2}, ..., \hat{\theta}^*_{(-i),m}$

(3) 计算$\overline{\hat{\theta}}^*_{(-i)} = \frac{1}{m}\sum_{j=1}^m \hat{\theta}^*_{(-i),j}$, 则有$\widehat{\mathrm{SE}}_{(-i)}\left(\hat{\theta}\right) = \left[\frac{1}{m}\sum_{i=1}^m \left(\hat{\theta}^*_{(-i),k} - \overline{\hat{\theta}}^*_{(-i)}\right)^2\right]^{\frac{1}{2}}$

(4) 对于$i = 1, 2, ..., n$, 重复(1) $\sim$ (3), 得到$\widehat{\mathrm{SE}}_{(-1)}\left(\hat{\theta}\right), \widehat{\mathrm{SE}}_{(-2)}\left(\hat{\theta}\right), ..., \widehat{\mathrm{SE}}_{(-n)}\left(\hat{\theta}\right)$, 则

$$\widehat{\mathrm{SE}}_{\mathrm{Jack}}\left(\widehat{\mathrm{SE}}\left(\hat{\theta}\right)\right) = \left[\frac{n-1}{n}\sum_{i=1}^n \left(\widehat{\mathrm{SE}}_{(-i)}\left(\hat{\theta}\right) - \overline{\widehat{\mathrm{SE}}}_{(\cdot)}\left(\hat{\theta}\right)\right)^2\right]^{\frac{1}{2}}$$

其中$\overline{\widehat{\mathrm{SE}}}_{(\cdot)} = \frac{1}{n}\sum_{i=1}^n \widehat{\mathrm{SE}}_{(-i)}\left(\hat{\theta}\right)$.

5.4 基于 Bootsrap 法的置信区间估计

5.4.1 标准正态 Bootstrap 置信区间

前提: 正态近似合理且$\hat{\theta}$无偏

$$Z = \frac{\hat{\theta} - E\left(\hat{\theta}\right)}{\mathrm{SE}\left(\hat{\theta}\right)} = \frac{\hat{\theta} - \theta}{\mathrm{SE}\left(\hat{\theta}\right)} \xrightarrow{L} N(0, 1)$$

因此θ的置信区间为

$$\hat{\theta} \pm u_{1-\frac{\alpha}{2}}\widehat{\mathrm{SE}}_B\left(\hat{\theta}\right)$$

5.4.2 基本的 Bootstrap 置信区间

计算$\hat{\theta}$的$\frac{\alpha}{2}$和$1 - \frac{\alpha}{2}$的 Bootstrap 百分位数

$$\left[\hat{\theta}^*_{(\lfloor m\frac{\alpha}{2}\rceil)}, \hat{\theta}^*_{(\lfloor m(1-\frac{\alpha}{2})\rceil)}\right]$$

5.4.3 Bootstrap 百分位数置信区间

$$P\left(L \leq \hat{\theta} - \theta \leq U\right) = 1 - \alpha$$

则

$$L = \hat{\theta}^*_{(\lfloor m\frac{\alpha}{2}\rceil)} - \hat{\theta} \text{ 以及 } U = \hat{\theta}^*_{(\lfloor m(1-\frac{\alpha}{2})\rceil)} - \hat{\theta}$$

因此θ的置信区间为

$$\left[2\hat{\theta} - \hat{\theta}^*_{(\lfloor m(1-\frac{\alpha}{2})\rceil)}, 2\hat{\theta} - \hat{\theta}^*_{(\lfloor m\frac{\alpha}{2}\rceil)}\right]$$

5.4.4 Bootstrap 的 t 置信区间方法

前提: $\hat{\theta}$是θ的无偏估计

$$t=\frac{\hat{\theta}-\theta}{\widehat{\mathrm{SE}}_B\left(\hat{\theta}\right)}$$

步骤:

(1) 从样本$(X_1,X_2,...,X_n)$中有放回地产生数据$(X_{11}^*,X_{12}^*,...,X_{1n}^*)$, 得到

$$\hat{\theta}_1^*=\hat{\theta}_1^*(X_{11}^*,X_{12}^*,...,X_{1n}^*)$$

(2) 从样本$(X_{11}*,X_{12}^*,...,X_{1n}^*)$中再次抽样$K$次, 得到$\hat{\theta}_1^*$的标准差$\widehat{\mathrm{SE}}_B\left(\hat{\theta}_1^*\right)$, 得到

$$t_1^*=\frac{\hat{\theta}_1^*-\hat{\theta}}{\widehat{\mathrm{SE}}_B\left(\hat{\theta}_1^*\right)}$$

(3) 重复上述步骤m次, 得到$t_1^*,t_2^*,...,t_m^*$, 于是得到置信区间为

$$\left[\hat{\theta}-t^*_{(\lfloor m(1-\frac{\alpha}{2})\rceil)}\widehat{\mathrm{SE}}_B\left(\hat{\theta}\right),\hat{\theta}-t^*_{(\lfloor m\frac{\alpha}{2}\rceil)}\widehat{\mathrm{SE}}_B\left(\hat{\theta}\right)\right]$$

Practice.

1 请证明逆变换抽样法的合理性并写出其流程
2 请证明舍选抽样法的合理性并写出其流程
3 请证明变换抽样法的合理性并写出其流程
4 请证明复合抽样法的合理性并写出其流程
5 请证明逆变换法的合理性并写出其流程
6 请证明随机向量的变换抽样法的合理性并写出其流程
7 请证明条件分布的合理性并写出其流程
8 请写出随机向量的舍选抽样法的流程
9 请证明离散型随机向量的条件分布法的合理性并写出其流程
10 请写出数值方法计算统计量$\hat{\theta}$的期望和方差的步骤
11 请写出数值方法求置信区间覆盖率$\widehat{\mathrm{Prob}}$和区间长度$\widehat{\mathrm{Inter}}$的流程
12 请写出数值方法求功效的估计$\hat{p}$的流程
13 请写出卡方检验的流程
14 请写出单样本 K-S 检验的流程(分布函数假设已知)
15 请写出 Mann-Whitney U 检验的流程
16 请写出两样本 K-S 检验的流程
17 请写出相关系数ρ以及r的计算公式
18 请写出 Bootstrap 方法的流程
19 Bootstrap 方法估计的偏差和方差是否是有偏的, 请证明之
20 请写出 Jackknife 方法的流程
21 请证明 Jackknife 方法估计的偏差和方差是无偏的
22 请写出 Jackknife-after-Bootstrap 的流程
23 请给出标准正态 Bootstrap 区间估计的流程, 并写明前提
24 请给出求基本的 Bootstrap 置信区间的流程
25 请给出求 Bootstrap 百分位数置信区间的流程

26 请给出求 Bootstrap 的 t 置信区间的流程, 并写明前提