第四章 区间估计 (interval estimation)

§4.1 区间估计的基本概念

点估计对未知参数θ作出了一种统计判断, 但这种判断的把握有多大,点估计在很多时候没有给出, 或者有时把握会很小.

参数估计有两种方案: 点估计、区间估计（两者互为补充）

一、区间估计

Definition

定义4.1.1 $\theta \in \Theta \subset \mathbb{R}$是待估参数. 设$\widehat{\theta}_L = \widehat{\theta}_L(\widetilde{X})$和$\widehat{\theta}_U = \widehat{\theta}_U(\widetilde{X})$是参数空间$\Theta$上取值的样本$\widetilde{X}$ 的两个统计量, 且满足$\widehat{\theta}_L(\widetilde{X}) \leq \widehat{\theta}_U(\widetilde{X})$. 则称随机区间$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$为参数$\theta$的区间估计(量) (interval estimation).

Example

设X_1, X_2, X_3, X_4是取自正态总体$N(\mu, 1)$的样本，则$[\overline{X}-1, \overline{X}+1]$是$\mu$的一个区间估计.

在点估计中，我们用$\overline{X}$估计μ，现在用$[\overline{X}-1, \overline{X}+1]$估计$\mu$，看上去点估计比区间估计精确，但有失必有得.

用$\overline{X}$估计μ时，估计正确的概率是$\mathsf{P}(\overline{X}=\mu)=0$.

而用区间$[\overline{X}-1,\overline{X}+1]$估计$\mu$时，估计正确的概率是

$$
\begin{aligned}
\mathsf{P}(\mu\in[\overline{X}-1,\overline{X}+1]) =& \mathsf{P}(-1\le \overline{X}-\mu\le 1)\\
=& \mathsf{P}\left(-2\le \frac{\overline{X}-\mu}{\sqrt{1/4}}\le 2\right)\\
=& \mathsf{P}\left(-2\le N(0,1)\le 2\right)=0.9544.
\end{aligned}
$$

二、区间估计的评价标准

如何评价一个区间估计的好坏?

常用的标准有两类:

(1) 置信度标准

(2) 精确度标准

1. 置信度

Definition

参数θ的区间估计$[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]$ 的涵盖概率(coverage probability)(也称置信度), 是指它涵盖待估参数θ的概率:

$$\mathsf{P}_\theta\left\{\theta\in[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]\right\}.$$

希望置信度越大越好.

一般说来,上述概率是依赖于θ的,如果一个区间估计对某个θ_1其置信度大，而对另一个θ_2其置信度小,那么此种区间估计的适应性差了一些,不能认为是一个好的区间估计.

假如对参数空间Θ中的任一个θ,其置信度都很大,那么此种区间估计就是一种好的区间估计.为此给出下面的定义.

Definition

定义4.1.2 参数θ的区间估计$[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]$ 的置信度在参数空间Θ上的下确界:

$$\inf_{\theta\in\Theta}\mathsf{P}_\theta\left\{\theta\in[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]\right\}$$

称为该区间估计的置信系数(confidence coefficient).

注意: 我们一般不讲“θ属于区间$[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]$的概率”，因为$\theta$不是随机的, 而区间是随机的.

显然，一个区间估计的置信系数是愈大愈好.

在实际中,为了计算置信度和置信系数,需要知道统计量的精确分布或者渐近分布.

Example

设$X_1, X_2, \cdots, X_n$是取自正态总体$N(\mu, \sigma^2)$的样本，其中μ和σ^2未知.考察总体均值μ的区间估计.

用样本均值$\overline{X}$和(无偏)样本方差${S_n^*}^2$(即,书中的S^2)可以给出总体均值μ的区间估计$[\overline{X} - kS_n^*/\sqrt{n}, \overline{X} + kS_n^*/\sqrt{n}]$(其中$k > 0$, $S_n^* = \sqrt{{S_n^*}^2}$). 则它的涵盖概率(即置信度)是

$$\begin{aligned}
p_k =& \mathsf{P}\left\{\overline{X} - kS_n^*/\sqrt{n} \le \mu \le \overline{X} + kS_n^*/\sqrt{n}\right\} \\
=& \mathsf{P}\left\{\left|\frac{\sqrt{n}(\overline{X} - \mu)}{S_n^*}\right| \le k\right\} \\
=& \mathsf{P}\left\{|t(n-1)| \le k\right\}.
\end{aligned}$$

上面的概率不依赖于未知参数μ和σ^2,故置信度就是置信系数.

在$n=20$时,取$k=1,2,3$,分别算出其区间的置信系数为:

$$
\begin{aligned}
p_1 =& \mathsf{P}\left\{\overline{X}-S_n^*/\sqrt{n}\le\mu\le\overline{X}+S_n^*/\sqrt{n}\right\}=0.6701,\\
p_2 =& \mathsf{P}\left\{\overline{X}-2S_n^*/\sqrt{n}\le\mu\le\overline{X}+2S_n^*/\sqrt{n}\right\}=0.9400,\\
p_3 =& \mathsf{P}\left\{\overline{X}-3S_n^*/\sqrt{n}\le\mu\le\overline{X}+3S_n^*/\sqrt{n}\right\}=0.9926.
\end{aligned}
$$

三个比较,第三个区间$[\overline{X}-3S_n^*/\sqrt{n},\overline{X}+3S_n^*/\sqrt{n}]$的置信系数最大.

2. 精确度

这类标准很多,这里仅介绍最常用的标准: 随机区间$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$的平均长度$\mathsf{E}_\theta\{\widehat{\theta}_U(\widetilde{X}) - \widehat{\theta}_L(\widetilde{X})\}$.

希望越精确越好,即随机区间的平均长度是越小越好.

在上例中随机区间$[\overline{X}-kS_n^*/\sqrt{n},\overline{X}+kS_n^*/\sqrt{n}]$的平均长度为

$$
\begin{aligned}
l_k =& \mathsf{E}\{2kS_n^*/\sqrt{n}\} = \frac{2k}{\sqrt{n(n-1)}}\sigma\mathsf{E}\left\{\sqrt{\sum_{i=1}^{n}\left(\frac{X_i-\overline{X}}{\sigma}\right)^2}\right\} \\
=& \frac{2k}{\sqrt{n(n-1)}}\sigma\mathsf{E}\{(\chi^2(n-1))^{1/2}\} \\
=& \frac{2k}{\sqrt{n(n-1)}}\sigma\mathsf{E}\left\{\sqrt{\Gamma(\frac{n-1}{2},\frac{1}{2})}\right\}.
\end{aligned}
$$

那么当样本容量n固定时,对于某个σ而言,

k越大, 区间平均长度越长,即精度越低;

k越大, 置信系数越大, 也即可靠性越高.

注意到：当n固定时，置信系数原则与精确度常常是相互制约的.

Neyman原则:
在保证置信系数达到指定要求的前提下，尽可能提高精确度.

三、置信区间

Definition

定义4.1.3 对给定的α $(0<\alpha<1)$，参数θ的置信系数不低于$1-\alpha$的区间估计$[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]$称为$\theta$的置信水平为$1-\alpha$的(双侧)置信区间(confidence interval)，即

$$\mathsf{P}_\theta\left\{\theta\in[\widehat{\theta}_L(\widetilde{X}),\widehat{\theta}_U(\widetilde{X})]\right\}\geq 1-\alpha,\quad \forall\theta\in\Theta.$$

此时,$\widehat{\theta}_L(\widetilde{X})$和$\widehat{\theta}_U(\widetilde{X})$分别称为$\theta$的置信水平为$1-\alpha$的(双侧)置信下限和置信上限.

如果有了一组样本观察值$\tilde{x}$，我们就可计算$\widehat{\theta}_L(\tilde{x})$和$\widehat{\theta}_U(\tilde{x})$的值，从而得到一个数值区间$[\widehat{\theta}_L(\tilde{x}), \widehat{\theta}_U(\tilde{x})]$，$[\widehat{\theta}_L(\tilde{x}), \widehat{\theta}_U(\tilde{x})]$称为是$\theta$的置信水平为$1-\alpha$的置信区间的一个实现，有时也简称为置信区间.

注意：我们不能写$\mathsf{P}_\theta\left\{\theta \in [\widehat{\theta}_L(\tilde{x}), \widehat{\theta}_U(\tilde{x})]\right\} \geq 1-\alpha$.

置信区间的含义(频率解释):

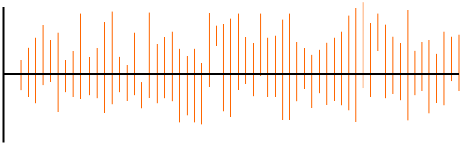
假如参数θ的一个置信水平为$1-\alpha$的置信区间$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$使用了很多次(譬如说$K$次), 对应的样本观察值为$\widetilde{x}^{(1)}, \widetilde{x}^{(2)}, \cdots, \widetilde{x}^{(K)}$, 那么在这些置信区间的实现$[\widehat{\theta}_L(\widetilde{x}^{(i)}), \widehat{\theta}_U(\widetilde{x}^{(i)})]$, $i=1,2,\cdots,K$, 中, 平均来讲大约至少有$100(1-\alpha)\%$的区间包含参数真值θ.

因此, 对于一个实现$[\widehat{\theta}_L(\widetilde{x}), \widehat{\theta}_U(\widetilde{x})]$, 其包含$\theta$大约有至少$100(1-\alpha)\%$的机会.

置信区间的含义：

若反复抽样多次，每个样本值确定一个区间$\left(\hat{\theta}_L, \hat{\theta}_U\right)$，每个这样的区间或者包含$\theta$的真值，或者不包含$\theta$的真值（见下图）

当$\alpha=0.05$，即置信水平为95%时，100个区间中大约95个包含μ值；
当$\alpha=0.01$，即置信水平为99%时，100个区间中有99个包含μ值.



从置信区间定义可知: 一个θ的置信水平为$1-\alpha$的置信区间, 其置信系数不小于其置信水平.

四、同等置信区间

Definition

定义 设$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$是参数$\theta$的一个区间估计,假定对给定的$\alpha(0 < \alpha < 1)$,有

$$\inf_{\theta\in\Theta} \mathsf{P}_\theta \left\{\theta \in [\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]\right\} = 1 - \alpha,$$

则称$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$是$\theta$的置信水平为$1 - \alpha$ 的同等(双侧) 置信区间.

五、置信限(单侧区间估计)

有时我们只关心参数不低于多少(例如:电视机的寿命)或不超过多少(例如:污染指数). 这时我们要用到置信限的概念.

Definition

定义4.1.4 对给定的α $(0<\alpha<1)$, 若参数θ 的区间估计$(-\infty, \widehat{\theta}_U(\widetilde{X})]$的置信系数不低于$1-\alpha$, 即

$$\mathsf{P}_\theta\left\{\theta \le \widehat{\theta}_U\right\} \ge 1-\alpha, \quad \forall \theta \in \Theta,$$

则称$\widehat{\theta}_U(\widetilde{X})$为$\theta$的置信水平为$1-\alpha$的单侧置信上限; $(-\infty, \widehat{\theta}_U(\widetilde{X})]$称为$\theta$的置信水平为$1-\alpha$的单侧置信区间.

Definition

定义4.1.4(续) 若参数θ 的区间估计$[\widehat{\theta}_L(\widetilde{X}), +\infty)$的置信系数不低于$1-\alpha$, 即

$$\mathsf{P}_\theta\left\{\theta \geq \widehat{\theta}_L\right\} \geq 1-\alpha, \quad \forall \theta \in \Theta,$$

则称$\widehat{\theta}_L(\widetilde{X})$为$\theta$的置信水平为$1-\alpha$的单侧置信下限; $[\widehat{\theta}_L(\widetilde{X}), +\infty)$称为$\theta$的置信水平为$1-\alpha$的单侧置信区间.

类似地,可以定义同等置信上限, 同等置信下限以及同等单侧置信区间.

下面给出一个引理，可知：单侧置信限与双侧置信限之间存在着一个简单的联系.

Lemma

设$\widehat{\theta}_L(\widetilde{X})$和$\widehat{\theta}_U(\widetilde{X})$分别是参数$\theta$的置信水平为$1-\alpha_1$ 和$1-\alpha_2$的单侧置信下、上限，且对任何样本$\widetilde{X}$,都有$\widehat{\theta}_L(\widetilde{X}) \le \widehat{\theta}_U(\widetilde{X})$.那么$[\widehat{\theta}_L(\widetilde{X}), \widehat{\theta}_U(\widetilde{X})]$是$\theta$的置信水平为$1-(\alpha_1+\alpha_2)$的双侧置信区间$(0<\alpha_1+\alpha_2<1)$.

证明： 下述三个事件是互不相容的,

$$\{\widehat{\theta}_L(\widetilde{X}) \le \theta \le \widehat{\theta}_U(\widetilde{X})\},\ \ \{\theta < \widehat{\theta}_L(\widetilde{X})\},\ \ \{\theta > \widehat{\theta}_U(\widetilde{X})\},$$

并且它们的并为"必然事件",所以

$$\begin{aligned}
&\mathsf{P}_\theta\left(\widehat{\theta}_L(\widetilde{X}) \le \theta \le \widehat{\theta}_U(\widetilde{X})\right)\\
=&1-\mathsf{P}_\theta\left(\theta < \widehat{\theta}_L(\widetilde{X})\right)-\mathsf{P}_\theta\left(\theta > \widehat{\theta}_U(\widetilde{X})\right)\\
=&1-\left\{1-\mathsf{P}_\theta\left(\theta \ge \widehat{\theta}_L(\widetilde{X})\right)\right\}-\\
&\qquad\left\{1-\mathsf{P}_\theta\left(\theta \le \widehat{\theta}_U(\widetilde{X})\right)\right\}\\
\ge&1-\alpha_1-\alpha_2.
\end{aligned}$$

六、置信域

Definition

设待估参数为$\theta = (\theta_1, \cdots, \theta_k) \in \Theta \subset \mathbb{R}^k$. 假设样本$\widetilde{X} = (X_1, \cdots, X_n)$的集合函数$S(\widetilde{X})$满足

- 对任一样本观察值$\widetilde{x}$, $S(\widetilde{x})$是Θ的一个子集;
- 对给定的α $(0 < \alpha < 1)$,

$$\mathsf{P}_\theta\left\{\theta \in S(\widetilde{X})\right\} \geq 1 - \alpha, \quad \forall \theta \in \Theta,$$

则称$S(\widetilde{X})$是θ的置信水平为$1-\alpha$的置信域(或置信集).

在实际中, 常常使用规则的几何图形作多维的置信域, 如: 长方形、球、椭球.

寻找区间估计的方法

§4.2 枢轴(变)量法 (Pivotal Quantity)

一、枢轴(变)量法

设样本$\widetilde{X}=(X_1, X_2, \cdots, X_n)$是取自总体$X$的一个样本. $\theta\in\Theta$是待估参数.

枢轴量法构造同等置信区间的步骤:

- 构造一个样本$\tilde{X}$和待估参数θ的函数$G(\tilde{X},\theta)$，要求G的分布不依赖于任何未知参数———一般称具有这种性质的函数$G(\tilde{X},\theta)$为枢轴量.
- 对给定的α $(0<\alpha<1)$，适当地选取两个常数c和d，使得

$$\mathsf{P}_\theta\left\{c\le G(\tilde{X},\theta)\le d\right\}=1-\alpha. \qquad (1)$$

- 若能将不等式$c\le G(\tilde{X},\theta)\le d$等价地转化为$\widehat{\theta}_L(\tilde{X})\le\theta\le\widehat{\theta}_U(\tilde{X})$，那么

$$\mathsf{P}_\theta\left\{\widehat{\theta}_L(\tilde{X})\le\theta\le\widehat{\theta}_U(\tilde{X})\right\}=1-\alpha.$$

从而$[\widehat{\theta}_L(\tilde{X}),\widehat{\theta}_U(\tilde{X})]$是$\theta$的置信水平为$1-\alpha$的同等置信区间.

注意点:

(i)用枢轴量法构造区间估计,关键在于构造枢轴量$G(\widetilde{X},\theta)$. 枢轴量常常在充分统计量或点估计的函数中寻找.

(ii) 枢轴量G要满足:

- G只是样本和待估参数θ的函数，它不含其它未知参数(注意它不是统计量);
- G的分布不依赖于任何未知参数. 即，一般选取的G，其分布是已知的;或在大样本情形下,G的渐近分布是已知的.

(iii)c和d的选取(一般满足(1)式的c和d很多):

- 使得置信区间的平均长度$\mathsf{E}_\theta\left[\widehat{\theta}_U(\widetilde{X})-\widehat{\theta}_L(\widetilde{X})\right]$最短; (最优置信区间)
- 不易做到以上这点时，常常取c和d使得

$$\mathsf{P}_\theta\left\{G(\widetilde{X},\theta)<c\right\}=\mathsf{P}_\theta\left\{G(\widetilde{X},\theta)>d\right\}=\alpha/2;$$

(等尾置信区间)

- 特别地,当G的分布对称时(如:$N(0,1)$, $t(2)$等),可取$d>0$使得

$$\mathsf{P}_\theta\left\{-d\le G(\widetilde{X},\theta)\le d\right\}=\mathsf{P}_\theta\left\{|G(\widetilde{X},\theta)|\le d\right\}=1-\alpha.$$

Example

设$X_1,\cdots,X_n$是取自均匀分布$U(0,\theta)$的一个样本. 对给定的α $(0<\alpha<1)$, 求θ的置信水平为$1-\alpha$的最优同等置信区间.

解: 分三步求θ的置信水平为$1-\alpha$的同等置信区间:

(I) 取θ 的MLE,即样本的最大次序统计量$X_{(n)}$,已知$X_{(n)}$还是参数θ的充分统计量. 而$X_{(n)}/\theta$的密度函数为

$$p(y;\theta)=ny^{n-1},\quad 0\le y\le 1,$$

它与参数θ无关, 故可取$X_{(n)}/\theta$为枢轴量.

(II)对给定的置信水平$1-\alpha$, 取c和d,$0\le c<d\le 1$, 使得

$$\mathsf{P}_\theta\{c\le X_{(n)}/\theta\le d)=d^n-c^n=1-\alpha.$$

(III) 利用不等式等价变形，可得θ的置信水平为$1-\alpha$的同等信区间$[X_{(n)}/d, X_{(n)}/c]$.

这样的区间很多. 我们找区间平均长度最短的那个. 区间的平均长度为$\mathsf{E}(X_{(n)})(1/c-1/d)$. 为使区间平均长度最短，只要使

$$\begin{aligned}\arg\min_{c,d}: &\quad 1/c-1/d\\ \text{subject to}: &\quad d^n-c^n=1-\alpha,\quad 0<c<d\le 1.\end{aligned}$$

解得$d=1$和$c=\sqrt[n]{\alpha}$. 所以$[X_{(n)}, X_{(n)}/\sqrt[n]{\alpha}]$是$\theta$的置信水平为$1-\alpha$的最优同等置信区间.

Example

设某产品的寿命X服从指数分布，其密度函数为

$$p(x;\theta)=\theta^{-1}e^{-x/\theta},\quad x>0,$$

$\theta>0$是该产品的平均寿命. 设$X_1,X_2,\cdots,X_n$是随机抽检的n个产品的寿命.求θ的单侧置信下限.

解： 注意到$T=\sum_{i=1}^n X_i$是θ的充分统计量. 且因为$X_i\sim \mathsf{E}(1/\theta)=\Gamma(1,1/\theta)$,且$X_i$之间相互独立，$i=1,2,\cdots,n$. 所以

$$T\sim\Gamma(n,1/\theta),$$

利用Gamma分布的伸缩性，可知

$$2T/\theta \sim \Gamma(n, 1/2) = \chi^2(2n).$$

即$2T/\theta$的分布不依赖于未知参数θ，故可取它做枢轴量.因为

$$\mathsf{P}_\theta\{2T/\theta \le \chi^2_\alpha(2n)\} = 1 - \alpha,$$

其中$\chi^2_\alpha(2n)$是$\chi^2(2n)$分布的上α分位数. 因此

$$\mathsf{P}_\theta\{2T/\chi^2_\alpha(2n) \le \theta\} = 1 - \alpha,$$

即平均寿命θ的置信水平为$1-\alpha$的单侧同等置信下限为

$$\widehat{\theta}_L = \frac{2T}{\chi^2_\alpha(2n)}.$$

二、正态总体参数的置信区间

(一)单个正态总体

设$X_1, X_2, \cdots, X_n$是取自正态总体$N(\mu, \sigma^2)$的一个样本. 记

$$\overline{X} = \sum_{i=1}^{n} X_i/n, \quad Q^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2,$$

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2, S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

1、正态总体均值μ的置信区间

(A) σ^2已知时

$\overline{X}$是充分统计量, $\overline{X} \sim N(\mu, \sigma^2/n)$.取枢轴量为

$$G(\widetilde{X}, \mu) = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

对于给定的$\alpha(0 < \alpha < 1)$,找c和d使得

$$\mathsf{P}_\mu\{c \le \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \le d\} = 1 - \alpha.$$

此时有

$$\mathsf{P}_\mu\{\overline{X} - d\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} - c\frac{\sigma}{\sqrt{n}}\} = 1 - \alpha.$$

那么μ的置信水平为$1-\alpha$的同等置信区间为$[\overline{X}-d\frac{\sigma}{\sqrt{n}}, \overline{X}-c\frac{\sigma}{\sqrt{n}}]$，其区间长度为$(d-c)\frac{\sigma}{\sqrt{n}}$. 由于正态分布是单峰对称的，为使得区间平均长度最短，只要取$d=-c=u_{\alpha/2}$，其中$u_{\alpha/2}$为标准正态分布的上侧$\alpha/2$分位数.所以

$$[\overline{X}-u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{X}+u_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$$

是μ的置信水平为$1-\alpha$的最优置信区间.区间长度为

$$l_n = 2u_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

当α确定且σ已知的情况下,要提高精确度(即减小区间平均长度),只有增加样本容量.

当α与n给定时，σ越大，l_n也越大.

当σ与n给定时，α越小，则$u_{\alpha/2}$的值越大，从而l_n也越大，即精确度越低.

(B) σ^2未知时

在Part (A)的$G = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 中用无偏样本方差S^2 代替未知的总体方差σ^2得

$$T = \frac{\overline{X}-\mu}{S/\sqrt{n}} = \frac{\sqrt{n(n-1)}(\overline{X}-\mu)}{Q} \sim t(n-1).$$

取它做枢轴量. 由于t分布的单峰对称性,只要取$c > 0$, 使得

$$\mathsf{P}_\mu\{|T| \le c\} = 1-\alpha,$$

就可使得所得的置信区间的平均长度为最短.这样的c为$c = t_{\alpha/2}(n-1)$,其中$t_{\alpha/2}(n-1)$为自由度为$(n-1)$ 的t分布的上侧$\alpha/2$分位数.故而μ的置信水平为$1-\alpha$的最优置信区间为

$$[\overline{X} - t_{\alpha/2}(n-1)\cdot S/\sqrt{n}, \overline{X} + t_{\alpha/2}(n-1)\cdot S/\sqrt{n}].$$

此时，区间的平均长度为

$$
\begin{aligned}
&\mathsf{E}(\overline{X}+t_{\alpha/2}(n-1)\cdot S/\sqrt{n}-(\overline{X}-t_{\alpha/2}(n-1)\cdot S/\sqrt{n})) \\
=&\frac{2t_{\alpha/2}(n-1)}{\sqrt{n}}\mathsf{E}S \\
=&\frac{2t_{\alpha/2}(n-1)}{\sqrt{n}}\mathsf{E}(\sqrt{\frac{(n-1)S^2}{\sigma^2}})\cdot\frac{\sigma}{\sqrt{n-1}} \\
=&\frac{2t_{\alpha/2}(n-1)}{\sqrt{n}}\mathsf{E}(\sqrt{\chi^2(n-1)})\cdot\frac{\sigma}{\sqrt{n-1}}
\end{aligned}
$$

Example

设某种植物的高度X (cm)服从正态分布$N(\mu,\sigma^2)$，随机选取36棵，其平均高度为15 cm. 就以下两种情形，求μ的置信水平为95%的双侧置信区间.

(1) $\sigma^2=16$;　(2)σ^2未知，$s^2=16$.

解: 现$\sigma=4$已知，估计μ，此时取$\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$作为枢轴量. 那么

$$\mathsf{P}\left\{-u_{\alpha/2}\le\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le u_{\alpha/2}\right\}=1-\alpha,$$

故可得

$$\mathsf{P}\left\{\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\le\mu\le\overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right\}=1-\alpha.$$

即μ的置信水平为$1-\alpha$的置信区间为

$$\left[\overline{X}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}, \overline{X}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}\right].$$

而$\alpha=0.05$, 查表得$u_{\alpha/2}=u_{0.025}=1.96$, 代入数据得

$$\overline{x}-\frac{\sigma}{\sqrt{n}}u_{\alpha/2}=15-\frac{1.96\times 4}{\sqrt{36}}=13.693, \quad \overline{x}+\frac{\sigma}{\sqrt{n}}u_{\alpha/2}=15+\frac{1.96\times 4}{\sqrt{36}}=16.307,$$

即μ的置信水平为95%的置信区间为$[13.693, \quad 16.307]$.

(2) σ^2未知，估计μ，此时取$\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$ 作为枢轴量．那么

$$\mathsf{P}\left\{\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1) \le \mu \le \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right\} = 1-\alpha,$$

即μ的置信水平为$1-\alpha$的置信区间为

$$\left[\overline{X}-\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1),\ \ \overline{X}+\frac{S}{\sqrt{n}}t_{\alpha/2}(n-1)\right].$$

而$\alpha = 0.05$，$n = 36$，查表得$t_{\alpha/2}(n-1) = t_{0.025}(35) = 2.0301$，代入数据得

$$15 - \frac{2.0301 \times 4}{\sqrt{36}} = 13.647, \quad 15 + \frac{2.0301 \times 4}{\sqrt{36}} = 16.353,$$

即μ的置信水平为95%的置信区间为$[13.647,\quad 16.353]$.

对比两种情形下的置信区间可以看到:

1. σ^2已知时, μ的置信水平为95%的置信区间(的实现)是$(13.693, \quad 16.307)$, 区间长度为2.614;
2. σ^2未知时, μ的置信水平为95%的置信区间(的实现)是$(13.647, \quad 16.353)$, 区间长度为2.706.

第二种情形对μ的区间估计的精度较低, 但在实际中更为常见, 因为多数时候σ^2是未知的.

2、正态总体方差σ^2(标准差σ)的置信区间

(A) μ已知时

此时, $\sum_{i=1}^n (X_i-\mu)^2$是σ^2的充分统计量, 取枢轴量为

$$\sum_{i=1}^n (X_i-\mu)^2/\sigma^2 \sim \chi^2(n).$$

由于

$$\mathsf{P}_{\sigma^2}\left\{\chi^2_{1-\alpha/2}(n) \le \sum_{i=1}^n (X_i-\mu)^2/\sigma^2 \le \chi^2_{\alpha/2}(n)\right\} = 1-\alpha.$$

σ^2的置信水平为$1-\alpha$的同等置信区间为

$$[\sum_{i=1}^n (X_i-\mu)^2/\chi^2_{\alpha/2}(n), \sum_{i=1}^n (X_i-\mu)^2/\chi^2_{1-\alpha/2}(n)].$$

(σ的置信水平为$1-\alpha$的同等置信区间为

$$\left[\sqrt{\sum_{i=1}^{n}(X_i-\mu)^2/\chi^2_{\alpha/2}(n)},\ \sqrt{\sum_{i=1}^{n}(X_i-\mu)^2/\chi^2_{1-\alpha/2}(n)}\right]).$$

注 χ^2分布不是对称分布，找平均长度最短的置信区间不容易．我们取上侧$1-\alpha/2$分位数$\chi^2_{1-\alpha/2}$和上侧$\alpha/2$分位数$\chi^2_{\alpha/2}$来构造置信区间,上面得到的置信区间其实是等尾置信区间,而不是最优置信区间.

(B) μ未知时

此时，$(\overline{X}, Q^2)$是(μ, σ^2)的充分统计量，取枢轴量为

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

由于

$$\mathsf{P}_{\sigma^2}\left\{\chi^2_{1-\alpha/2}(n-1) \le Q^2/\sigma^2 \le \chi^2_{\alpha/2}(n-1)\right\} = 1-\alpha.$$

σ^2的置信水平为$1-\alpha$的同等置信区间为

$$[Q^2/\chi^2_{\alpha/2}(n-1), Q^2/\chi^2_{1-\alpha/2}(n-1)].$$

(σ的置信水平为$1-\alpha$的同等置信区间为

$$[Q/\sqrt{\chi^2_{\alpha/2}(n-1)}, Q/\sqrt{\chi^2_{1-\alpha/2}(n-1)}].)$$

Example

为测得某种溶液中的甲醛浓度，取样测得4个独立测定值的平均值$\bar{x} = 8.4\%$，样本标准差$s = 0.03\%$，并设测量值近似服从正态分布,求总体方差σ^2和标准差σ的置信水平为95%的置信区间.

解： 取枢轴量为

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

可得σ^2的置信水平为$1-\alpha$的同等置信区间为

$$[(n-1)S^2/\chi^2_{\alpha/2}(n-1), (n-1)S^2/\chi^2_{1-\alpha/2}(n-1)].$$

现$n-1=3$, $\alpha/2=0.025$, $1-\alpha/2=0.975$, 查表得到$\chi^2_{0.025}(3)=9.348$, $\chi^2_{0.975}(3)=0.216$, $s^2=0.0009$, 因此σ^2 的置信水平为95%的置信区间为

$$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right]=[0.00029, 0.0125],$$

σ 的置信水平为95%的置信区间为

$$\left[\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}(n-1)}\right)^{1/2}, \left(\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}(n-1)}\right)^{1/2}\right]=[0.017, 0.112].$$

3、(μ,σ^2)的置信域

已知$(\overline{X},Q^2)$是(μ,σ^2)的充分统计量，而

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1),$$
$$\frac{(n-1)S^2}{\sigma^2}=\frac{Q^2}{\sigma^2}\sim\chi^2(n-1),$$

且两者独立，取$(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}},\frac{Q^2}{\sigma^2})$作枢轴量.

对给定的置信水平$1-\alpha$，取三个常数c，d_1，d_2使得

$$\mathsf{P}_{\mu,\sigma^2}\left\{\frac{|\overline{X}-\mu|}{\sigma/\sqrt{n}}\le c\right\}=\sqrt{1-\alpha}\hat{=}1-\alpha_1,$$
$$\mathsf{P}_{\mu,\sigma^2}\left\{d_1\le\frac{Q^2}{\sigma^2}\le d_2\right\}=\sqrt{1-\alpha}=1-\alpha_1.$$

只要取$c=u_{\alpha_1/2}$，$d_1=\chi^2_{1-\alpha_1/2}(n-1)$，$d_2=\chi^2_{\alpha_1/2}(n-1)$即可.

由$\overline{X}$与Q^2的独立性, 有

$$\mathsf{P}_{\mu,\sigma^2}\left\{(\overline{X}-\mu)^2 \le c^2\sigma^2/n,\ \ Q^2/d_2 \le \sigma^2 \le Q^2/d_1\right\} = 1-\alpha.$$

所以(μ,σ^2)的置信水平为$1-\alpha$的置信域为

$$\left\{(\mu,\sigma^2):\ \ (\overline{X}-\mu)^2 \le c^2\sigma^2/n,\ \ Q^2/d_2 \le \sigma^2 \le Q^2/d_1\right\}.$$

(二) 两个正态总体

$X \sim N(\mu_1, \sigma_1^2)$与$Y \sim N(\mu_2, \sigma_2^2)$是两个相互独立的正态总体.

样本$\widetilde{X} = (X_1, X_2, \cdots, X_m)$和$\widetilde{Y} = (Y_1, Y_2, \cdots, Y_n)$分别取自两个总体中的样本$(m \geq 2,\ n \geq 2)$.那么

样本$X_1, X_2, \cdots, X_m$, i.i.d. $\sim N(\mu_1, \sigma_1^2)$;

样本$Y_1, Y_2, \cdots, Y_n$, i.i.d. $\sim N(\mu_2, \sigma_2^2)$,

且两样本相互独立.

这里有四个参数, μ_1, μ_2, σ_1^2, σ_2^2, 对应地有四个统计量: $\overline{X}$, $\overline{Y}$,
$Q_1^2 = \sum_{i=1}^m (X_i - \overline{X})^2$ (或${S_X^*}^2 = Q_1^2/(m-1)$), $Q_2^2 = \sum_{i=1}^n (Y_i - \overline{Y})^2$ (或${S_Y^*}^2 = Q_2^2/(n-1)$).

这四个统计量相互独立,且有

$$\overline{X} \sim N(\mu_1, \sigma_1^2/m),$$
$$\overline{Y} \sim N(\mu_2, \sigma_2^2/n),$$
$$Q_1^2/\sigma_1^2 \sim \chi^2(m-1),$$
$$Q_2^2/\sigma_2^2 \sim \chi^2(n-1).$$

若记$d = \overline{Y} - \overline{X}$,还有

$$d - (\mu_2 - \mu_1) \sim N(0, \sigma_1^2/m + \sigma_2/n).$$

1、构作$\delta = \mu_2 - \mu_1$的置信区间(Behrens-Fisher问题)

(A) σ_1^2 和σ_2^2都已知的情形

取枢轴量为

$$G = \frac{d - \delta}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} \sim N(0, 1). \qquad *$$

从而δ的置信水平为$1 - \alpha$的最优置信区间为

$$[d - u_{\alpha/2}\sqrt{\sigma_1^2/m + \sigma_2^2/n}, d + u_{\alpha/2}\sqrt{\sigma_1^2/m + \sigma_2^2/n}].$$

(B)σ_1^2 和σ_2^2未知的情形

此时,自然的想法是用它们的无偏估计${S_X^*}^2$和${S_Y^*}^2$代替，得

$$G' = \frac{d-\delta}{\sqrt{{S_X^*}^2/m + {S_Y^*}^2/n}}.$$

遗憾的是，一般情况下，G'的分布依赖于未知参数，它不能成为枢轴量.

下面就几种特殊情形进行分析.

(a) $\sigma_1^2=\sigma_2^2\hat{=}\sigma^2$的情形

这时，$S^{*2}_{X+Y}=(Q_1^2+Q_2^2)/(m+n-2)$是$\sigma^2$的无偏估计. 用它代替$(*)$中的$\sigma_1^2$和$\sigma_2^2$ 得

$$t=\frac{d-\delta}{\sqrt{S^{*2}_{X+Y}/m+S^{*2}_{X+Y}/n}}=\sqrt{\frac{mn(m+n-2)}{m+n}}\frac{d-\delta}{\sqrt{Q_1^2+Q_2^2}}\sim t(m+n-2),$$

取它作为枢轴量.从而δ的置信水平为$1-\alpha$的同等置信区间上、下限为

$$d\pm t_{\alpha/2}(m+n-2)\cdot\sqrt{S^{*2}_{X+Y}(1/m+1/n)}.$$

即，$d\pm t_{\alpha/2}(m+n-2)\cdot\sqrt{\frac{1}{m}+\frac{1}{n}}\sqrt{\frac{Q_1^2+Q_2^2}{m+n-2}}$.

事实上，为了求得σ^2的好的估计，可以考察$\widetilde{X}$和$\widetilde{Y}$的联合pdf，

$$
\begin{aligned}
&\frac{1}{(2\pi\sigma_1^2)^{m/2}(2\pi\sigma_2^2)^{n/2}}\exp\left\{-\frac{\sum_{i=1}^m(X_i-\mu_1)^2}{2\sigma_1^2}-\frac{\sum_{j=1}^n(Y_j-\mu_2)^2}{2\sigma_2^2}\right\}\\
=&C(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2)\exp\left\{-\frac{\sum_{i=1}^m X_i^2}{2\sigma_1^2}-\frac{\sum_{j=1}^n Y_j^2}{2\sigma_2^2}+\frac{m\mu_1}{\sigma_1^2}\overline{X}+\frac{n\mu_2}{\sigma_2^2}\overline{Y}\right\}\\
=&C(\mu_1,\mu_2,\sigma^2)\exp\left\{-\frac{\sum_{i=1}^m X_i^2+\sum_{j=1}^n Y_j^2}{2\sigma^2}+\frac{m\mu_1}{\sigma^2}\overline{X}+\frac{n\mu_2}{\sigma^2}\overline{Y}\right\}
\end{aligned}
$$

由指数型分布族的性质知$(\overline{X},\overline{Y},\sum_{i=1}^m X_i^2+\sum_{j=1}^n Y_j^2)$是充分完备统计量. 从而$(\overline{X},\overline{Y},\sum_{i=1}^m(X_i-\overline{X})^2+\sum_{j=1}^n(Y_j-\overline{Y})^2)$ 也是(μ_1,μ_2,σ^2)的充分完备统计量, 那么$(\overline{X},\overline{Y},S^{*2}_{X+Y})$也是$(\mu_1,\mu_2,\sigma^2)$的充分完备统计量.

(b) σ_1^2和σ_2^2未知，但$\sigma_2^2/\sigma_1^2=\lambda$ 已知的情形

构造枢轴量为

$$t=\sqrt{\frac{mn(m+n-2)}{m\lambda+n}}\cdot\frac{d-\delta}{\sqrt{Q_1^2+Q_2^2/\lambda}}\sim t(m+n-2),$$

从而δ的置信水平为$1-\alpha$的同等置信区间的上、下限为

$$d\pm t_{\alpha/2}(m+n-2)\cdot\sqrt{\frac{m\lambda+n}{mn(m+n-2)}}\sqrt{Q_1^2+Q_2^2/\lambda}.$$

或: 这时

$$\frac{d-\delta}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}}=\frac{d-\delta}{\sqrt{\sigma_1^2(1/m+\lambda/n)}}.$$

而$S_X^{*\,2}$, $S_Y^{*\,2}/\lambda$都是σ_1^2 的无偏估计. 将它们综合起来,取

$$\widehat{\sigma_1^2}=\frac{(m-1)S_X^{*\,2}+(n-1)S_Y^{*\,2}/\lambda}{m+n-2}=\frac{Q_1^2+Q_2^2/\lambda}{m+n-2}$$

作为σ_1^2的无偏估计, 得

$$t=\frac{d-\delta}{\sqrt{\widehat{\sigma_1^2}(1/m+\lambda/n)}}=\frac{d-\delta}{\sqrt{\frac{Q_1^2+Q_2^2/\lambda}{m+n-2}}\cdot\sqrt{1/m+\lambda/n}}\sim t(m+n-2).$$

这样得δ的置信水平为$1-\alpha$的同等置信区间的上、下限为

$$d\pm t_{\alpha/2}(m+n-2)\cdot\sqrt{\widehat{\sigma_1^2}(1/m+\lambda/n)}.$$

(c) σ_1^2和σ_2^2未知，但$m=n$已知的情形

如果出现(a)和(b)的情形，只需将$m=n$处理即可.否则，就按如下方式进行估计.

令

$$Z_i = Y_i - X_i \sim N(\delta, \sigma_1^2+\sigma_2^2), \quad i=1,2,\cdots,n.$$

相应地有

$$\overline{Z} = \overline{Y} - \overline{X} \sim N(\delta, (\sigma_1^2+\sigma_2^2)/n),$$
$$\frac{(n-1){S_Z^*}^2}{\sigma_1^2+\sigma_2^2} = \frac{\sum_{i=1}^n (Z_i-\overline{Z})^2}{\sigma_1^2+\sigma_2^2} \sim \chi^2(n-1).$$

若记$Q_3^2=\sum_{i=1}^n(Z_i-\overline{Z})^2$,则有

$$\frac{\sqrt{n(n-1)}(\overline{Z}-\delta)}{Q_3}=\frac{\sqrt{n}(\overline{Z}-\delta)}{\sqrt{Q_3^2/(n-1)}}=\frac{\overline{Z}-\delta}{S_Z^*/\sqrt{n}}\sim t(n-1),$$

取之为枢轴量, 所以δ的置信水平为$1-\alpha$的同等置信区间为

$$[\overline{Z}-t_{\alpha/2}(n-1)\cdot Q_3/\sqrt{n(n-1)},\overline{Z}+t_{\alpha/2}(n-1)\cdot Q_3/\sqrt{n(n-1)}].$$

注:此法也适用于”成对数据”.

(d) 当m和n都充分大时

可用大样本方法.由于

$$\frac{d-\delta}{\sqrt{\sigma_1^2/m+\sigma_2^2/n}} \sim N(0,1).$$

又

$${S_X^*}^2 \xrightarrow{P} \sigma_1^2, \quad {S_Y^*}^2 \xrightarrow{P} \sigma_2^2.$$

可证

$$T=\frac{d-\delta}{\sqrt{{S_X^*}^2/m+{S_Y^*}^2/n}} \xrightarrow{D} N(0,1)$$

取之为枢轴量.因此当m和n都充分大时，δ的置信系数近似于$1-\alpha$的置信区间的上、下限为

$$d \pm u_{\alpha/2} \cdot \sqrt{{S_X^*}^2/m+{S_Y^*}^2/n}.$$

(e) 一般情形的近似解

令

$$T=\frac{d-\delta}{\sqrt{{S_X^*}^2/m+{S_Y^*}^2/n}}.$$

在一般情形下，T已不服从t分布，但近似服从$t(l)$分布，其中

$$l=\frac{\{{s_X^*}^2/m+{s_Y^*}^2/n\}^2}{\frac{{s_X^*}^4}{m^2(m-1)}+\frac{{s_Y^*}^4}{n^2(n-1)}}.$$

$\delta=\mu_2-\mu_1$的置信系数近似于$1-\alpha$的置信区间为

$$d\pm t_{\alpha/2}(l)\sqrt{{S_X^*}^2/m+{S_Y^*}^2/n}.$$

说明: 实际操作中，若采样得$\widetilde{x}, \widetilde{y}$，可得区间估计的一个实现$[\widehat{\delta}_L(\widetilde{x}, \widetilde{y}), \widehat{\delta}_U(\widetilde{x}, \widetilde{y})]$，若

(1)此区间中含有0,则大致可以认为μ_1与μ_2差不多;

(2)此区间的下限大于0,则大致可以认为μ_2比μ_1大;

(3)此区间的上限小于0,则大致可以认为μ_2比μ_1小.

2、方差比σ_1^2/σ_2^2的置信区间

(A) μ_1和μ_2都已知时

此时，$\sum_{i=1}^m (X_i-\mu_1)^2$是$\sigma_1^2$的充分统计量，且$\frac{1}{\sigma_1^2}\sum_{i=1}^m (X_i-\mu_1)^2 \sim \chi^2(m)$；同理,$\sum_{j=1}^n (Y_j-\mu_2)^2$是$\sigma_2^2$的充分统计量，且$\frac{1}{\sigma_2^2}\sum_{j=1}^m (Y_j-\mu_2)^2 \sim \chi^2(n)$;且相互独立．故取枢轴量为

$$\frac{\frac{1}{\sigma_1^2}\sum_{i=1}^m (X_i-\mu_1)^2/m}{\frac{1}{\sigma_2^2}\sum_{j=1}^m (Y_j-\mu_2)^2/n} \sim F(m,n).$$

(B) μ_1和μ_2都未知时

由于$Q_1^2/\sigma_1^2 \sim \chi^2(m-1)$, $Q_2^2/\sigma_2^2 \sim \chi^2(n-1)$,且相互独立. 构造枢轴量为

$$F = \frac{{S_X^*}^2/{S_Y^*}^2}{\sigma_1^2/\sigma_2^2} = \frac{Q_1^2/[\sigma_1^2(m-1)]}{Q_2^2/[\sigma_2^2(n-1)]} \sim F(m-1, n-1).$$

所以

$$\mathsf{P}\left\{F_{1-\alpha/2}(m-1, n-1) \le \frac{{S_X^*}^2/{S_Y^*}^2}{\sigma_1^2/\sigma_2^2} \le F_{\alpha/2}(m-1, n-1)\right\} = 1-\alpha.$$

得σ_1^2/σ_2^2 的置信水平为$1-\alpha$的同等置信区间为

$$\left[\frac{{S_X^*}^2}{{S_Y^*}^2}\frac{1}{F_{\alpha/2}(m-1, n-1)}, \quad \frac{{S_X^*}^2}{{S_Y^*}^2}\frac{1}{F_{1-\alpha/2}(m-1, n-1)}\right].$$

注意:

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

故σ_1^2/σ_2^2 的置信水平为$1-\alpha$的同等置信区间也可写为

$$\left[\frac{{S_X^*}^2}{{S_Y^*}^2}F_{1-\alpha/2}(n-1,m-1),\quad \frac{{S_X^*}^2}{{S_Y^*}^2}F_{\alpha/2}(n-1,m-1)\right].$$

说明: 实际操作中,若采样得$\tilde{x},\tilde{y}$, 可得区间估计的一个实现$[\widehat{\sigma_1^2/\sigma_2^2}_L(\tilde{x},\tilde{y}),\widehat{\sigma_1^2/\sigma_2^2}_U(\tilde{x},\tilde{y})]$, 若

(1)此区间中含有1,则大致可以认为σ_1^2与σ_2^2差不多;

(2)此区间的下限大于1,则大致可以认为σ_1^2比σ_2^2大;

(3)此区间的上限小于1,则大致可以认为σ_1^2比σ_2^2小.

Example

例 两台机床生产同一型号的滚珠，从甲机床生产的滚珠中随机抽取了8个，从乙机床生产的滚珠中随机抽取了9个，测得这些滚珠的直径(单位：mm)如下：

i	1	2	3	4	5	6	7	8	9
x_i(甲机床)	15.0	14.8	15.2	15.4	14.9	15.1	15.2	14.8	
y_i(乙机床)	15.2	15.0	14.8	15.1	14.6	14.8	15.1	14.5	15.0

设两机床生产的滚珠直径分别为X, Y，且$X \sim N(\mu_1, \sigma_1)$, $Y \sim N(\mu_2, \sigma_2)$.

(1) 若$\sigma_1 = 0.18$, $\sigma_2 = 0.24$，求$\mu_1 - \mu_2$的置信水平为0.90的最优同等置信区间；

(2) 若$\sigma_1 = \sigma_2$未知，求$\mu_1 - \mu_2$的置信水平为0.90的最优同等置信区间；

(3) 若μ_1, μ_2未知，求σ_1^2/σ_2^2的置信水平为0.90的等尾置信区间.

解 (1) 当$\sigma_1 = 0.18$, $\sigma_2 = 0.24$时，取枢轴量为

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

则$\mu_1 - \mu_2$的置信水平为$1-\alpha$的同等置信区间为

$$\left[\overline{X} - \overline{Y} - u_{\alpha/2}\sqrt{\sigma_1^2/m + \sigma_2^2/n},\ \ \overline{X} - \overline{Y} + u_{\alpha/2}\sqrt{\sigma_1^2/m + \sigma_2^2/n}\right].$$

现$\alpha = 0.1$, $u_{0.05} = 1.645$, 结合样本计算可得所求区间为$[-0.018,\ \ 0.318]$.

注 此区间中包含了0，这也意味着两机床生产的滚珠直径的期望之间没有显著差异.

(2) 当$\sigma_1=\sigma_2$未知时，估计$\mu_1-\mu_2$，取枢轴量为

$$T=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{S_W\sqrt{1/m+1/n}}\sim t(m+n-2),$$

其中$S_W^2=\frac{1}{m+n-2}\{(m-1)S_X^2+(n-1)S_Y^2\}$，$S_W=\sqrt{S_W^2}$. 则$\mu_1-\mu_2$的置信水平为$1-\alpha$的同等置信区间为

$$\Big[\overline{X}-\overline{Y}-t_{\alpha/2}(m+n-2)S_W\sqrt{1/m+1/n},\ \ \overline{X}-\overline{Y}+t_{\alpha/2}(m+n-2)S_W\sqrt{1/m+$$

现$\alpha=0.1$, $m=8,n=9$, $t_{0.05}(15)=1.753$, 结合样本计算可得所求区间为$[-0.044,\ \ 0.344]$.

注　此区间中包含了0, 这也意味着两机床生产的滚珠直径的期望之间没有显著差异.

(3) 当μ_1, μ_2未知时，估计σ_1^2/σ_2^2，取枢轴量为

$$F = \frac{S_X^2/{S_Y}^2}{\sigma_1^2/\sigma_2^2} \sim F(m-1, n-1).$$

则σ_1^2/σ_2^2的置信水平为0.90的等尾置信区间为

$$\left[\frac{{S_X}^2}{{S_Y}^2}\frac{1}{F_{\alpha/2}(m-1, n-1)}, \quad \frac{{S_X}^2}{{S_Y}^2}\frac{1}{F_{1-\alpha/2}(m-1, n-1)}\right].$$

现$\alpha = 0.1$, $m = 8, n = 9$, $F_{0.05}(7, 8) = 3.50$,
$F_{0.95}(7, 8) = \frac{1}{F_{0.05}}(8, 7) = \frac{1}{3.73} = 0.268$, 结合样本计算可得所求区间为$[0.227, \quad 2.965]$.

注 此区间中包含了1，这也意味着两机床生产的滚珠直径的方差之间没有显著差异.

（三）非正态总体的近似

1、大样本方法

当样本容量充分大时, 可用渐近分布来构造近似的置信区间.下面用例子来说明此种方法.

Example

总体X,分布未知,方差存在且为1, 设$X_1, X_2, \cdots, X_{100}$是来自该总体的一组样本,且$\overline{x}=5$, 求$X$的期望$\mu$的置信系数近似为$1-\alpha$的近似置信区间.

解: 注意到期望μ的矩法估计为$\overline{X}$, 且

$$\frac{\overline{X}-\mu}{1/\sqrt{n}} \text{ 近似服从} N(0,1).$$

所以可取之为枢轴量. 从而

$$\mathsf{P}(|\sqrt{n}(\overline{X}-\mu)| \le u_{\alpha/2}) \approx 1-\alpha.$$

所以μ的置信系数近似为95%的近似等尾置信区间为$\left[\overline{X}-u_{\alpha/2}/\sqrt{n}, \overline{X}+u_{\alpha/2}/\sqrt{n}\right]$, 其实现为

$$\left[\overline{x}-u_{\alpha/2}/\sqrt{n}, \overline{x}+u_{\alpha/2}/\sqrt{n}\right] = [5-0.196, 5+0.196].$$

Example

设$X_1, X_2, \cdots, X_n$是来自二点分布$B(1,p)$的一个样本，当样本容量充分大时，求p的置信系数近似为$1-\alpha$的近似置信区间.

解: 注意到$\widehat{p}=\overline{X}$是p的点估计量，又是其充分统计量，而由中心极限定理(CLT)知

$$G=\frac{\widehat{p}-p}{\sqrt{pq}/\sqrt{n}}=\frac{\sum_{i=1}^{n}(X_i-p)}{\sqrt{npq}}\xrightarrow{D}N(0,1),\quad 当\quad n\to\infty,$$

其中$q=1-p$. 即由CLT知，G的渐近分布是标准正态分布,与未知参数无关，所以可取它做枢轴量.

取$\lambda = u_{\alpha/2}$, 就有

$$\mathsf{P}_p\left\{-\lambda \le \frac{\widehat{p}-p}{\sqrt{pq}/\sqrt{n}} \le \lambda\right\}$$
$$\approx \mathsf{P}\left\{-\lambda \le N(0,1) \le \lambda\right\} = 1-\alpha, \quad \text{当} \quad n \to \infty.$$

上式中关于p的不等式等价于

$$(\widehat{p}-p)^2 \le \lambda^2 p(1-p)/n,$$
$$p^2(n+\lambda^2) - p(2n\widehat{p}+\lambda^2) + n\widehat{p}^2 \le 0,$$
$$\widehat{p}_L \le p \le \widehat{p}_U,$$

其中

$$\widehat{p}_{L,U} = \frac{n}{n+\lambda^2}\left[\widehat{p} + \frac{\lambda^2}{2n} \mp \lambda\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n} + \frac{\lambda^2}{4n^2}}\right], \quad \lambda = u_{\alpha/2}.$$

所以

$$\mathsf{P}_p\{\widehat{p}_L \le p \le \widehat{p}_U\} \approx 1-\alpha.$$

即，区间$[\widehat{p}_L, \widehat{p}_U]$是$p$置信系数近似为$1-\alpha$的置信区间.

由于

$$G_2 = \frac{\widehat{p}-p}{\sqrt{\widehat{p}(1-\widehat{p})/n}} \xrightarrow{D} N(0,1), \quad \text{当} \quad n \to \infty.$$

也可取G_2作为枢轴量，由它构造的近似置信区间是

$$[\widehat{p} - u_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}, \quad \widehat{p} + u_{\alpha/2}\sqrt{\widehat{p}(1-\widehat{p})/n}].$$

仅与前者相差一个阶为n^{-1}的项.

2、自助法(Bootstrap)

设$\widetilde{X}=(X_1,\cdots,X_n)$是取自总体$X\sim p(x;\theta)$的样本，$\widehat{\theta}=\widehat{\theta}(\widetilde{X})$是$\theta$的点估计量(或充分统计量)。

如果$\widehat{\theta}-\theta$ 的分布$F_{\widehat{\theta}-\theta}$已知，那么我们可以通过这一个分布求得$c$和$d$使得

$$\mathsf{P}\left(c\le\widehat{\theta}-\theta\le d\right)=1-\alpha.$$

当$F_{\widehat{\theta}-\theta}$未知时，可以求其近似分布。注意到$\widehat{\theta}$的分布是由总体分布确定的，而对于总体分布，我们用给定样本后的经验分布来代替。

即, 对于给定的样本观察值$\widetilde{x}=(x_1,x_2,\cdots,x_n)$, 把样本的经验分布函数当做总体分布函数(的近似), 即分布函数为

$$F_n(x)=\frac{\#\{i: x_i<x\}}{n}.$$

从这个总体中再抽出一个i.i.d.样本$\widetilde{X}^*=(X_1^*,\cdots,X_m^*)$ (再抽样). 按照构造$\widehat{\theta}$同样的方法, 构造一个新估计量$\widehat{\theta}^*=\widehat{\theta}(\widetilde{X}^*)$.

这样$\widehat{\theta}^*=\widehat{\theta}(\widetilde{X}^*)$的分布是$\widehat{\theta}=\widehat{\theta}(\widetilde{X})$ 的近似. 而$\widehat{\theta}$本身是θ的近似.

所以我们可以用, 给定样本观察值$\widetilde{X}=\widetilde{x}$ 后, $\widehat{\theta}^*-\widehat{\theta}(\widetilde{x})$ 的分布近似$\widehat{\theta}-\theta$的分布.

由于在再抽样中, 总体分布是已知的, 所以$\widehat{\theta}^* - \widehat{\theta}(\widetilde{x})$的分布$F^*$是可以求出的.
由这个分布出发,可以得到$\widehat{\theta}_1$和$\widehat{\theta}_2$使得

$$\mathsf{P}^*\left(\widehat{\theta}_1 \le \widehat{\theta}^* - \widehat{\theta}(\widetilde{x}) \le \widehat{\theta}_2\right) = 1 - \alpha. \tag{$*$}$$

对每一个$\widetilde{x}$,都有一对$\widehat{\theta}_1$和$\widehat{\theta}_2$. $\widehat{\theta}_1$和$\widehat{\theta}_2$实际上是$\widetilde{x}$的函数: $\widehat{\theta}_1(\widetilde{x})$, $\widehat{\theta}_2(\widetilde{x})$.
以$\widehat{\theta} - \theta$代替(*)中的$\widehat{\theta}^* - \widehat{\theta}(\widetilde{x})$,就得到区间

$$[\widehat{\theta}(\widetilde{X}) - \widehat{\theta}_2(\widetilde{X}), \widehat{\theta}(\widetilde{X}) - \widehat{\theta}_1(\widetilde{X})].$$

这个区间就是θ的置信水平近似为$1-\alpha$的置信区间, 称为bootstrap置信区间.

而

$$[\widehat{\theta}(\widetilde{x})-\widehat{\theta}_2(\widetilde{x}),\widehat{\theta}(\widetilde{x})-\widehat{\theta}_1(\widetilde{x})]$$

是一个实现. 在实际问题中, 我们可以从一个样本观察值得到一个实现.
在具体问题中,我们可用Monte Carlo方法计算$\widehat{\theta}^*-\widehat{\theta}$的分布$F^*$和$\widehat{\theta}_1(\widetilde{x})$, $\widehat{\theta}_2(\widetilde{x})$.
操作步骤为:

- 用计算机重复再抽样过程, 产生r个$\widehat{\theta}^*$ 的实现$\widehat{\theta}_1^*,\widehat{\theta}_2^*,\cdots,\widehat{\theta}_r^*$;
- $\widehat{\theta}_1^*-\widehat{\theta},\widehat{\theta}_2^*-\widehat{\theta},\cdots,\widehat{\theta}_r^*-\widehat{\theta}$ 从小到大排列,
- 在上述数据中, 取上$1-\alpha/2$分位数作为$\widehat{\theta}_1(\widetilde{x})$, 取上$\alpha/2$分位数作为$\widehat{\theta}_2(\widetilde{x})$.