第 六 章

参 数 估 计

Estimation of Parameters

参数

(1) 分布中所含未知参数；

(2) 未知参数的函数；

(3) 分布的各种特征数。

一般常用θ 表示参数，参数θ 所有可能取值组成的集合称为参数空间，常用 Θ 表示。

问题：

设θ表示未知参数，$X_1,\ldots,X_n$为来自总体的样本。

(1) 如何给出 θ的估计，比如构造估计量$\overline{\theta}=\overline{\theta}(X_1,\ldots,X_n)$;

(2) 如何评价估计量的好坏（评价标准）。

第一节：点估计（Point Estimation）

参数估计问题就是根据样本对未知参数作出估计。

参数估计的形式有两种：

- **点估计：**给出参数的估计量
- **区间估计：**给出参数的估计范围及其包含参数的可信度

第一节：点估计

问题的提法：

设总体X 的分布函数为$F(x,\theta)$,形式已知，含未知参数θ,

有样本$X_1,\ldots,X_n$，及观测值$x_1,\ldots,x_n$.

构造估计量$\overline{\theta}=\overline{\theta}(X_1,\ldots,X_n)$作为$\theta$的估计量;

用观测值$\overline{\theta}(x_1,\ldots,x_n)$作为$\theta$的估计值。

- **点估计的两种方法：**
  - **矩估计法（Method of Moments)**
  - **最大似然估计（Maximum Likelihood Estimate)**

第一节：点估计

- A **point estimate** of a parameter:

  --- **a single number** that can be regarded as a sensible value for the parameter.

  --- obtained by selecting a **suitable statistic** and computing its value from the given sample data.

- The selected statistic is called the **point estimator** of the parameter.

一、矩估计法 (Method of Moments)

替换原理：用样本的 AAA 替换总体的 AAA.

(1) 用样本矩 替换 总体矩(矩：原点距或中心距)；

(2) 用样本矩的函数 替换 总体矩的函数；

(3) 用经验分布函数 替换 总体的分布函数；

(4) 用样本的分位数替换总体的分位数；

(5) 用事件发生的频率 替换 事件发生的概率；

....

"替换"得到方程，解之。

思想

- From weak LLN, for iid variables, a sample moment converges to the population moment：

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k = E(X^K).$$

- Equate sample moments to population moments:

  Sample moments = Population moments

$$\begin{cases} A_1 = \mu_1(\theta), \\ \cdots \\ A_k = \mu_k(\theta). \end{cases}$$

- The point estimate is found by solving the above equation(s) for the parameters.

步骤：设 X 为离散型**RV**, $P(X=x)=p(x;\theta_1,\ldots,\theta_k)$;

或为连续型**RV**, $f(x)=f(x;\theta_1,\ldots,\theta_k)$; 共有 k 个参数。

Step 1：计算总体直到**k**阶矩：$\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k$ (用未知参数表示);

$$\boldsymbol{\mu}_j = E(X^j) = b_j(\theta_1,\ldots,\theta_k);$$

Step 2：解出未知参数，用总体矩表示 $\begin{cases}\theta_1 = g_1(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k),\\ \ldots\\ \theta_k = g_k(\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_k).\end{cases}$

Step 3：用样本矩替换总体矩：$\begin{cases}\overline{\theta_1} = g_1(A_1,\ldots,A_k),\\ \ldots\\ \overline{\theta_k} = g_k(A_1,\ldots,A_k).\end{cases}$

则 $\overline{\theta_1},\ldots,\overline{\theta_k}$ 为未知参数 $\theta_1,\ldots,\theta_k$ 的矩估计量。

若有 $\theta_1,\ldots,\theta_k$ 的函数 $\eta=h(\theta_1,\ldots,\theta_k)$, 则 $\overline{\eta}=h(\overline{\theta_1},\ldots,\overline{\theta_k})$ 为 η 的矩估计量。

理论依据

设$X_1,\ldots,X_n$是来自该总体X的样本,

样本k阶原点矩：$A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^{\,k}$.

定理：设总体X的k阶矩存在：$E(X^k)=\mu_k$,

则：(1) $A_k\xrightarrow{P}\mu_k$.　(用辛钦大数定律)

(2) $g(A_1,\ldots,A_k)\xrightarrow{P}g(\mu_1,\ldots,\mu_k)$ (g为连续函数)。

矩法估计的实质是用经验分布函数去替换总体分布函数，其理论基础是格里纹科定理 --- 经验分布函数是总体分布函数的一致良好近似。

矩估计量可能不唯一，尽量采用低阶矩。

例：设总体$X \sim U[0,\theta]$. 求θ的矩估计量。

解：**Step 1**：计算总体一阶矩：μ_1(用未知参数表示);

$$\mu_1 = E(X) = \frac{\theta}{2};$$

Step 2：反解出未知参数，用总体矩表示

$$\theta = 2\mu_1.$$

Step 3：用样本矩替换总体矩：

$$\overline{\theta} = 2\mathbf{A}_1 = 2\overline{X}.$$

则θ的矩估计量为$\overline{\theta} = 2\mathbf{A}_1 = 2\overline{X}.$

例：设总体$X \sim U[-\theta,\theta]$. 求θ的矩估计量。

解：**Step 1**：计算总体一阶矩：μ_1(用未知参数表示);

$$\mu_1 = E(X) = 0; \quad \text{(与未知参数无关!)}$$

计算总体二阶矩：μ_2(用未知参数表示);

$$\mu_2 = E(X^2) = \theta^2/3;$$

Step 2：反解出未知参数，用总体矩表示

$$\theta = \sqrt{3\mu_2}.$$

Step 3：用样本矩替换总体矩：

$$\overline{\theta} = \sqrt{3A_2}$$

则θ的矩估计量为$\overline{\theta} = \sqrt{3A_2}$.

例：设总体X为指数分布$f(x,\lambda)=\lambda\cdot e^{-\lambda x}, x>0$. 求$\lambda$的矩估计量。

解：**Step 1:** 计算总体一阶矩：μ_1(用未知参数表示)

$$\mu_1 = E(X) = \frac{1}{\lambda};$$

原点矩

Step 2: 反解出未知参数，用总体矩表示

$$\lambda = \frac{1}{\mu_1}.$$

Step 3: 用样本矩替换总体矩：

$$\overline{\lambda} = \frac{1}{A_1} = \frac{1}{\overline{X}}$$

则λ的矩估计量为$\overline{\lambda} = \dfrac{1}{\overline{X}}$.

另解：**Step 1:** 计算总体二阶(中心)矩：

$$D(X) = 1/\lambda^2;$$

Step 2: 反解出未知参数，用总体矩表示

$$\lambda = 1/\sqrt{D(X)}.$$

Step 3: 用样本矩替换总体矩：$\overline{\lambda} = 1/S$

则λ的矩估计量为$\overline{\lambda} = 1/S$.

例：设总体$X\sim U[a,b]$. 求a,b的矩估计量。

解：**Step 1:** 计算总体矩(原点或中心矩,用未知参数表示)

$$E(X)=\frac{a+b}{2},\qquad D(X)=\frac{(b-a)^2}{12}.$$

$$\Rightarrow\begin{cases}a+b=2E(X)\\ b-a=\sqrt{12D(X)}\end{cases}$$

Step 2: 反解出未知参数，用总体矩表示

$$\Rightarrow\begin{cases}a=E(X)-\sqrt{3D(X)}\\ b=E(X)+\sqrt{3D(X)}\end{cases}$$

原点矩　　中心矩

Step 3: 用样本矩替换总体矩：$\begin{cases}\overline{a}=A_1-\sqrt{3}\,S,\\ \overline{b}=A_1+\sqrt{3}\,S.\end{cases}$

例：

设总体X的均值μ及方差均σ^2存在，但未知，求μ,σ^2的矩估计量。

解：**Step 1:** 计算总体一阶、二阶原点矩：(用未知参数表示)；

$$\mu_1 = E(X) = \mu,$$

$$\mu_2 = E(X^2) = \sigma^2 + \mu^2.$$

Step 2: 反解出未知参数，用总体矩表示

$$\Rightarrow \begin{cases} \mu = \mu_1, \\ \sigma^2 = \underbrace{\mu_2}_{\text{原点矩}} - \underbrace{\mu_1^2}_{\text{原点矩}}. \end{cases}$$

Step 3: 用样本矩替换总体矩：
$$\begin{cases} \overline{\mu} = A_1 = \overline{X}, \\ \overline{\sigma^2} = A_2 - A_1^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2. \end{cases}$$

分别是：
样本均值
样本方差

注：结果与总体分布无关。

例:

设总体X服从自由度为m的χ^2分布,m未知,求m的矩估计量。

解:

Step 1: 计算总体的一阶矩(用未知参数表示):

$$\mu_1 = E(X) = m.$$

Step 2: 反解出未知参数，用总体矩表示

$$\Rightarrow m = \mu_1.$$

Step 3: 用样本矩替换总体矩:

$$\bar{\mu} = A_1 = \bar{X}.$$

例：甲乙两人独立地校对某书样稿。校完后，甲发现a个错字，乙发现b个错字，共同发现的错字有c个。试用矩估计法估计：

(1) 该书样稿总错字数；（2）未被发现的错字数。

解：(1)设该书样稿总错字数为d. 设甲、乙识别错字的概率分别为P_1和P_2.由独立性，同一错字能被甲、乙同时识别的概率为

$$P = P_1 \cdot P_2,$$

由频率替换概率的思想:

$$\Rightarrow \begin{cases} P_1 \approx \overline{P_1} = a/d, \\ P_2 \approx \overline{P_2} = b/d, \\ P \approx \overline{P} = c/d. \end{cases}$$

近似有：$\overline{P} = \overline{P_1} \cdot \overline{P_2} \Rightarrow \dfrac{c}{d} = \dfrac{a}{d} \cdot \dfrac{b}{d} \Rightarrow \bar{\mathbf{d}} = \dfrac{ab}{c}.$

（2）未被发现的错字数的估计是总错字数的估计减去

甲、乙发现的错字数,即为 $\bar{\mathbf{d}} - (a+b-c) = \dfrac{ab}{c} - a - b + c.$

二、最大似然估计

Maximum Likelihood Estimate **(1821 Gauss, 1922 Fisher)**

最大似然思想：

有两个射手甲和乙，命中率分别为 0.9 和 0.1. 现在他们中的一个向目标射击了一发，结果命中了。是谁射击的？

分析：如果是甲，则 $P(\text{命中}) = 0.90$;

如果是乙，则 $P(\text{命中}) = 0.10$.

所以，"最像"是甲命中的。

或者：

$$P(\text{甲} \mid \text{命中}) = \frac{P(\text{命中} \mid \text{甲}) \cdot P(\text{甲})}{P(\text{命中} \mid \text{甲}) \cdot P(\text{甲}) + P(\text{命中} \mid \text{乙}) \cdot P(\text{乙})} = 0.9.$$

最大似然思想

一般，事件 A 发生的概率与某参数 $\theta\in\Theta$ 有关，

参数 θ 取值不同，则概率 $P(A)$ 也不同。

事件 A 发生的概率应记为 $P(A|\theta)$.

若A发生了，则认为此时的 θ 值应是在所有可能参数 Θ 中使事件A发生的概率 $P(A|\theta)$ 达到最大(最像)的那一个。

最大似然思想

- This method has an intuitive appeal to it, since it asks **the question**:

  For what parameter value is the observed data most likely to have arisen?

- **The maximum likelihood principle**

  Given a dataset, choose the parameter(s) in such a way that the data are **most likely**.

例:

设有来自Poisson总体 P(λ)的样本观测值:

$X_1=4, X_2=7, X_3=2$. 哪个 λ 最可能?

$$P(X_1=4, X_2=7, X_3=2)=\frac{\lambda^4}{4!}e^{-\lambda}\cdot\frac{\lambda^7}{7!}e^{-\lambda}\cdot\frac{\lambda^2}{2!}e^{-\lambda}$$

$$=\frac{\lambda^{13}}{4!7!2!}e^{-3\lambda}.$$

选取λ,使上述概率最大。

$$\log\left(\frac{\lambda^{13}}{4!7!2!}e^{-3\lambda}\right)=13\log(\lambda)-3\lambda-\log(4!7!2!).$$

求导,令之为0,得$\lambda=13/3$.

二阶导<0, 确为最大值点。

1. 总体为离散型

设$P(X=x)=p(x;\theta),\theta\in\Theta$(参数空间，即参数可能取值全体)。

$X_1,\ldots,X_n$为样本，$x_1,\ldots,x_n$为观测值。

则样本$(X_1,\ldots,X_n)$取到观测值$(x_1,\ldots,x_n)$的概率为：

$$P(X_1=x_1,\cdots,\ X_n=x_n)=\prod_{i=1}^{n}\ p(x_i;\theta).$$

定义：$L(\theta):=\prod_{i=1}^{n}\ p(x_i;\theta)$，称$L(\theta)$为样本的似然函数。

不同的参数θ，$L(\theta)$不同。

问题：参数θ取什么值时，概率$P(X_1=x_1,\cdots,\ X_n=x_n)$最大？

即哪个θ最可能导致$\{X_1=x_1,\cdots,\ X_n=x_n\}$的发生？

最大似然思想：选择$\theta\in\Theta$，使$L(\theta)$达到最大。

最大似然估计

Maximum Likelihood Estimate, MLE

定义：如果$\overline{\theta}=\overline{\theta}(x_1,\dots,x_n)$满足

$$L(\overline{\theta}):=\max_{\theta\in\Theta}L(\theta),$$

则称$\overline{\theta}$为θ的最大似然估计(**MLE**),

$\overline{\theta}(x_1,\dots,x_n)$为**MLE**估计值，

$\overline{\theta}(X_1,\dots,X_n)$为**MLE**估计量。

2. 总体为连续型

设总体X的密度：$f(x;\theta),\theta\in\Theta$（参数空间）。

$X_1,\ldots,X_n$为样本，$x_1,\ldots,x_n$为观测值。

则$(X_1,\ldots,X_n)$的联合密度为：$\prod_{i=1}^{n} f(x_i;\theta)$.

随机向量$(X_1,\ldots,X_n)$落在观测点$(x_1,\ldots,x_n)$的邻域(不是一点)

的概率：

$$\prod_{i=1}^{n} f(x_i;\theta)\cdot\prod_{i=1}^{n} dx_i$$ (概率微元，后者与θ无关)

这个概率与θ有关。

最大似然思想：选择$\theta\in\Theta$，使上述概率达到达到最大。

2. 总体为连续型

定义：$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)$ 为样本的似然函数。

如果 $\bar{\theta}=\bar{\theta}(x_1,\ldots,x_n)$满足

$$L(\bar{\theta}):=\max_{\theta\in\Theta} L(\theta),$$

则称 $\bar{\theta}$ 为θ的最大似然估计(**MLE**),

$\bar{\theta}(x_1,\ldots,x_n)$为**MLE**估计值，

$\bar{\theta}(X_1,\ldots,X_n)$为**MLE**估计量。

注记：

似然函数 vs 联合分布密度函数

不同的视角:

似然函数形式上等于样本的联合分布密度函数,

但似然函数是未知参数θ的函数(其中的x_i为确定的观测值)

在联合分布中，x_i为变量，θ为已知参数(固定)。

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta).$$

注记：

- 可把参数 θ 和样本 $X_1,\ldots,X_n$ 分别看成“原因”和“结果”。
- 定了 θ 的值，就完全确定了样本分布，也就定下了得到种种结果的可能性的大小。
- 反过来，当有了结果（样本）时，我们问：当参数 θ 取各种不同的值(原因)时，导出这个结果的可能性有多大？这导出“似然”函数的概念。“似然”---“看起来像”。
- 在已得样本的情况下，不能唯一决定 θ 。
  取哪个值去估计 θ 呢？--- 取那个“看起来最像”的值，

即，使得“似然”函数达到最大的那个 θ 值。

3. 求MLE的方法

Step 1: 写出似然函数：$L(\theta)=\begin{cases}\prod_{i=1}^{n} p(x_i;\theta), & \text{总体为离散型，}\\ \prod_{i=1}^{n} f(x_i;\theta), & \text{总体为连续型。}\end{cases}$

Step 2: 求似然函数$L(\theta)$的最大值点。

方法1：如$p(x;\theta)$，$f(x;\theta)$关于θ可微，

令：$\dfrac{dL(\theta)}{d\theta}=0$，解出$\theta$，再讨论它是否为最大值点。

方法2：对数似然函数$\ln L(\theta)=\sum \ln f(x_i;\theta)$.

$L(\theta)$与$\ln L(\theta)$在同一θ处取得最大值(因$\ln x$单调).

令：$\dfrac{d\ln L(\theta)}{d\theta}=0$，解出$\theta$，再讨论它是否为最大值点。

方法3：其它方法(观察法)，求$\bar{\theta}$，满足 $L(\bar{\theta}):=\max\limits_{\theta\in\Theta} L(\theta)$.

3. 求MLE的方法

推广到多个参数的情形:

Step 1: 写出似然函数: $L(\theta)=\begin{cases}\prod_{i=1}^{n} p(x_i;\theta_1,\ldots,\theta_k), \text{总体为离散型,}\\ \prod_{i=1}^{n} f(x_i;\theta_1,\ldots,\theta_k), \text{总体为连续型。}\end{cases}$

Step 2: 求似然函数$L(\theta_1,\ldots,\theta_k)$或对数似然函数$\ln L(\theta_1,\ldots,\theta_k)$的最大值点。

令: $\begin{cases}\dfrac{\partial L(\theta_1,\ldots,\theta_k)}{\partial\theta_1}=0,\\ \vdots\\ \dfrac{\partial L(\theta_1,\ldots,\theta_k)}{\partial\theta_k}=0,\end{cases}$ 或 $\begin{cases}\dfrac{\partial \ln L(\theta_1,\ldots,\theta_k)}{\partial\theta_1}=0,\\ \vdots\\ \dfrac{\partial \ln L(\theta_1,\ldots,\theta_k)}{\partial\theta_k}=0,\end{cases}$

解出$(\theta_1,\ldots,\theta_k)$,再讨论它是否为最大值点。

需要判断对应的**Hessian**矩阵是否为“非正定”

4. MLE举例

例：产品质量X分为合格(X=0)与不合格(X=1)，不合格率为p

(未知)。抽得样本：$x_1, x_2, \ldots, x_n$. 求 p 的 MLE.

解：总体的分布：$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$. 即 $P(X=x) = p^x(1-p)^{1-x}, x = 0,1.$

似然函数 $L(p) = P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n)$

$$= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

对数似然函数：$\ln L(p) = \left(\sum x_i\right)\ln p + \left(n - \sum x_i\right)\ln(1-p).$

令 $\dfrac{d\ln L(p)}{dp} = \left(\sum x_i\right)\dfrac{1}{p} - \left(n - \sum x_i\right)\dfrac{1}{1-p} = 0,$

解得 $\overline{p} = \dfrac{1}{n}\sum x_i$，又 $\ln L(p)$ 关于 p 的二阶导数 $< 0.$

$\Rightarrow$ MLE估计值为 $\overline{p} = \dfrac{1}{n}\sum x_i$，MLE估计量为 $\overline{p} = \dfrac{1}{n}\sum X_i$. (与矩估计相同)

例：

设总体：$X \sim \begin{pmatrix} a & b & c \\ \theta^2 & 2\theta(1-\theta) & (1-\theta)^2 \end{pmatrix}$，$\theta$为参数。

现做了n次试验，观测到a,b,c分别出现n_1,n_2,n_3次。求θ的MLE.

解：似然函数

$$\underline{L(\theta)=P(X_1=x_1,\dots,X_n=x_n)}=P(X_1=x_1)\cdots P(X_n=x_n)$$

$$=C\,[\theta^2]^{n_1}\cdot[2\theta(1-\theta)]^{n_2}\cdot(1-\theta)^{2n_3}=C\,2^{n_2}\theta^{2n_1+n_2}\cdot(1-\theta)^{n_2+2n_3}.$$

此处 C 为多项系数，与参数theta无关

对数似然函数：

$$\ln L(\theta)=(2n_1+n_2)\ln\theta+(n_2+2n_3)\ln(1-\theta)+\ln C+n_2\ln 2.$$

令 $$\frac{d\ln L(\theta)}{d\theta}=(2n_1+n_2)\frac{1}{\theta}-(n_2+2n_3)\frac{1}{1-\theta}=0,$$

解得 $\bar{\theta}=\dfrac{2n_1+n_2}{2n}$，又$\ln L(\theta)$关于$\theta$的二阶导数$<0$.

$\Rightarrow$ MLE估计值为$\bar{\theta}=\dfrac{2n_1+n_2}{2n}$.

例

设总体 $X\sim P(\lambda)$, λ为参数。设有观测值$x_1,\ldots,x_n$,

求λ的MLE.

解：总体的分布列 $f(k;\lambda)=\dfrac{\lambda^k}{k!}e^{-\lambda}, k=0,1,\ldots$

似然函数 $L(\lambda)=\dfrac{\lambda^{x_1}}{x_1!}e^{-\lambda}\cdots\dfrac{\lambda^{x_n}}{x_n!}e^{-\lambda}=\dfrac{\lambda^{x_1+\cdots+x_n}}{x_1!\cdots x_n!}e^{-n\lambda}.$

对数似然函数：

$$\ln L(\lambda)=(x_1+\cdots+x_n)\ln(\lambda)-n\lambda-\ln(x_1!\cdots x_n!).$$

令 $\dfrac{\ln L(\lambda)}{d\lambda}=(x_1+\cdots+x_n)\cdot\dfrac{1}{\lambda}-n$, 解得 $\overline{\lambda}=\dfrac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}.$

再计算二阶导数，易知为负，上述解为MLE.

例

设总体 $X \sim N(\mu,\sigma^2)$, (μ,σ^2)为参数.设有观测值$x_1,\dots,x_n$,求μ,σ^2的**MLE**.

解：总体的密度$f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$.

似然函数

$$L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{(x_i-\mu)^2}{2\sigma^2}\right)=(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

对数似然函数：$\ln L(\mu,\sigma^2)=-\dfrac{1}{2\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2}\ln(\sigma^2)-\dfrac{n}{2}\ln(2\pi)$.

令$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)=0,\\[2ex] \dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=\dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2\sigma^2}=0,\end{cases}$ 解得$\begin{cases}\overline{\mu}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}x_i=\bar{x},\\[2ex] \overline{\sigma^2}=\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i-\bar{x})^2.\end{cases}$

再计算**Hessian**矩阵，可判定为非正定,上述解为**MLE** (与矩估计相同).

$\boxed{\theta=\sigma^2}$

注：并非在所有场合均可求导

设总体$X \sim U[a,b]$, a,b未知，$x_1,\ldots,x_n$为观测值，求a,b的**MLE**.

解：记$x_{(1)}=\min\{x_1,\ldots,x_n\}$, $x_{(n)}=\max\{x_1,\ldots,x_n\}$.

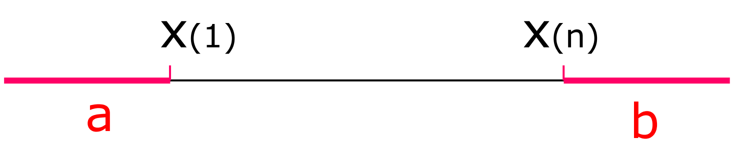
总体X的密度 $f(x)=\dfrac{1}{b-a}\cdot \mathbf{I}_{\{a\le x\le b\}}$.

似然函数：$L(a,b)=\prod_{i=1}^{n}\dfrac{1}{b-a}\cdot \mathbf{I}_{\{a\le x_i\le b\}}=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_{(1)}, b\ge x_{(n)},\\ 0, & \text{其它。}\end{cases}$

对满足$a\le x_{(1)}, b\ge x_{(n)}$的所有$a,b$,均有

作为a, b的函数

$$L(a,b)=\frac{1}{(b-a)^n}\le \frac{1}{(x_{(n)}-x_{(1)})^n},$$

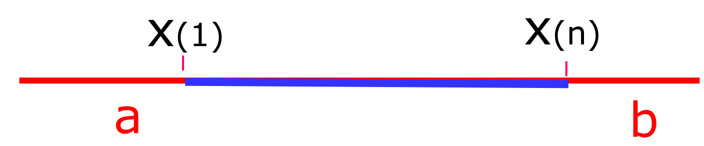


在$a=x_{(1)}, b=x_{(n)}$时，等号成立，即$L(a,b)$达最大。

故a,b的**MLE**为$\bar{a}=x_{(1)}=\min\{x_1,\ldots,x_n\}$, $\bar{b}=x_{(n)}=\max\{x_1,\ldots,x_n\}$.

注：并非在所有场合均可求导

似然函数：$L(a,b)=\prod_{i=1}^{n}\frac{1}{b-a}\cdot \mathbf{I}_{\{a\le x_i\le b\}}=\begin{cases}\frac{1}{(b-a)^n}, & a\le x_{(1)}, b\ge x_{(n)},\\ 0, & 其它。\end{cases}$



对满足$a\le x_{(1)}, b\ge x_{(n)}$的所有$a,b$,

$\ln L(a,b)=-n\ln(b-a).$

$$\frac{\partial \ln L(a,b)}{\partial a}=n/(b-a)>0, \quad \frac{\partial \ln L(a,b)}{\partial b}=-n/(b-a)<0.$$

表明$\ln L(a,b)$对a单调上升，a最大时最好，$\bar{a}=x_{(1)}=\min\{x_1,\ldots,x_n\}.$

$\ln L(a,b)$对b单调下降，b最小时最好，$\bar{b}=x_{(n)}=\max\{x_1,\ldots,x_n\}.$

故a,b的**MLE**为$\bar{a}=x_{(1)}=\min\{x_1,\ldots,x_n\}, \bar{b}=x_{(n)}=\max\{x_1,\ldots,x_n\}.$

5. MLE的不变性

设 $\bar{\theta}$ 是 θ 的MLE,即 $\mathbf{L}(\bar{\theta}) = \max\limits_{\theta\in\Theta} L(\theta)$.

又设 $\eta = g(\theta)$ (一一对应的函数)，则 $\bar{\eta} := g(\bar{\theta})$ 是 η 的MLE.

因关于 η 的似然函数 $L^*(\eta)$ 可由关于 θ 的似然函数 $L(\theta)$ 通过代人 $\theta = g^{-1}(\eta)$ 得到:

$$L^*(\eta) = L(g^{-1}(\eta)) = L(\theta),$$

$$\Rightarrow L^*(\bar{\eta}) = L(g^{-1}(\bar{\eta})) = L(\bar{\theta}),$$

因 $\bar{\theta}$ 使 $L(\theta)$ 达最大,所以 $\bar{\eta} := g(\bar{\theta})$ 使 $L^*(\eta)$ 达最大。

$\boxed{\text{不变性使一些复杂结构的参数的极大似然估计容易得到}}$

例

设总体 $X \sim N(\mu,\sigma^2)$, (μ,σ^2)为参数。

设有观测值$x_1,\ldots,x_n$, 已求出μ,σ^2的MLE是：

$$\begin{cases} \overline{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}, \\ \overline{\sigma^2} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = (s^*)^2. \end{cases}$$

$\Rightarrow$ 标准差σ的MLE是：$\overline{\sigma} = s^*$.（标准差$=\sqrt{\text{方差}}$）

概率$\eta := P(X < 2) = \Phi\left(\dfrac{2-\mu}{\sigma}\right)$的MLE是$\overline{\eta} = \Phi\left(\dfrac{2-\overline{x}}{s^*}\right)$.

第二节、点估计的评价标准

- Two different methods (method of moments, MLE) of finding estimators for population parameters have been introduced.
- We have seen that it is possible to have several estimators for the same parameter.
- Should we use the method of moments estimator, the MLE, or an estimator obtained through some other methods?

一、相合性

- 点估计是一个统计量，是RV，在样本量一定的条件下，不可能要求它完全等同于参数的真值。
- 随着样本量的不断增大，经验分布函数逼近真实分布函数，因此可以要求估计量随着样本量的不断增大而逼近参数真值，这就是相合性。
- 相合性是对估计的一个最基本要求。如果一个估计量，在样本量不断增大时，它都不能近似被估参数到任意指定的精度，那么这个估计是值得怀疑的。

一、相合性

大意：估计量随着样本容量的增大而“逼近”真参数。

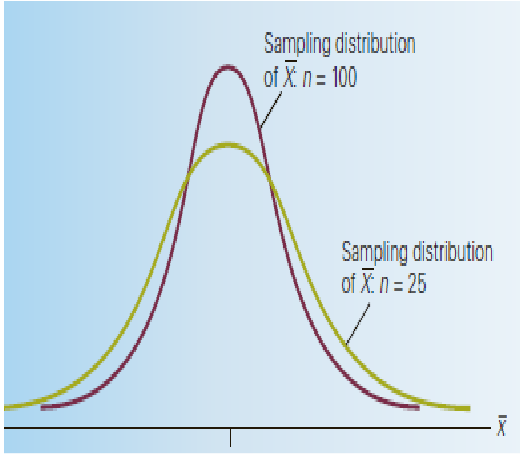
定义：设$\bar{\theta}_n=\bar{\theta}_n(X_1,\dots,X_n)$为参数$\theta$的估计量，

若 $\bar{\theta}_n \xrightarrow{P} \theta$,

即$\forall\varepsilon>0$, $\lim\limits_{n\to\infty}P(|\bar{\theta}_n-\theta|\ge\varepsilon)=0$,

或 $\lim\limits_{n\to\infty}P(|\bar{\theta}_n-\theta|<\varepsilon)=1.$

则称$\bar{\theta}_n$为θ的相合估计。



证明相合性可用：各种大数定律或依概率收敛的性质。

例(续)

设$X_1,\ldots,X_n$是来自该总体X的样本,样本k阶原点矩为

$A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$. 设总体$X$的$k$阶矩存在：$E(X^k)=\mu_k$,

则:(1) $A_k\xrightarrow{P}\mu_k$.　(用辛钦大数定律)

(2) $g(A_1,\ldots,A_k)\xrightarrow{P}g(\mu_1,\ldots,\mu_k)$ (g为连续函数)。

表明：A_k为μ_k的相合估计,

$g(A_1,\ldots,A_k)$为$g(\mu_1,\ldots,\mu_k)$的相合估计。

例：

特别，样本方差

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}X_i{}^2-\overline{X}^2\xrightarrow{P}\mu_2-\mu_1^2=\sigma^2.$$

$$\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=\left(\frac{n}{n-1}\right)\cdot\left(\frac{1}{n}\sum_{i=1}^{n}X_i{}^2-\overline{X}^2\right)\xrightarrow{P}\sigma^2.$$

（依概率）趋于1

即两种样本方差均为总体方差的相合估计。

相合性的判定

定理：设$\bar{\theta}_n=\bar{\theta}_n(X_1,\ldots,X_n)$为参数$\theta$的一个相合估计量，

$\eta=g(\theta)$是θ的连续函数，

则 $\bar{\eta}_n:=g(\bar{\theta}_n)$为$\eta$的相合估计。

证：因 $\bar{\theta}_n\xrightarrow{P}\theta$，又$\eta=g(\theta)$是$\theta$的连续函数，

由依概率收敛的性质

$$g(\bar{\theta}_n)\xrightarrow{P}g(\theta).$$

即$\bar{\eta}_n=g(\bar{\theta}_n)$为$\eta$的相合估计。

注：此处θ可为向量。

矩估计一般都是样本矩的连续函数，由该定理，一般具有相合性

相合性的判定定理:

设$\overline{\theta}_n=\overline{\theta}_n(X_1,\ldots,X_n)$为参数$\theta$的一个估计量,

若$\lim\limits_{n\to+\infty}E(\overline{\theta}_n)=\theta$, $\lim\limits_{n\to+\infty}D(\overline{\theta}_n)=0$, 则$\overline{\theta}_n$为$\theta$的相合估计。

证：注意到 $\overline{\theta}_n-\theta=[\overline{\theta}_n-E(\overline{\theta}_n)]+[E(\overline{\theta}_n)-\theta]$.

由**Chebyshev**不等式有$P(|\overline{\theta}_n-E(\overline{\theta}_n)|>\varepsilon)\le\dfrac{D(\overline{\theta}_n)}{\varepsilon^2}$.

因$\lim\limits_{n\to+\infty}D(\overline{\theta}_n)=0$, 所以$\overline{\theta}_n\xrightarrow{P}E(\overline{\theta}_n)$, 即$\overline{\theta}_n-E(\overline{\theta}_n)\xrightarrow{P}0$.

又$\lim\limits_{n\to+\infty}E(\overline{\theta}_n)=\theta$, 所以$E(\overline{\theta}_n)\xrightarrow{P}\theta$, 即$E(\overline{\theta}_n)-\theta\xrightarrow{P}0$.

(因n充分大时，必$|E(\overline{\theta}_n)-\theta|<\varepsilon\Rightarrow P(|E(\overline{\theta}_n)-\theta|>\varepsilon)=0$.)

$$\overline{\theta}_n-\theta=[\overline{\theta}_n-E(\overline{\theta}_n)]+[E(\overline{\theta}_n)-\theta]\xrightarrow{P}0$$

则$\overline{\theta}_n$为θ的相合估计。

二、无偏性

定义：

设$\overline{\theta}_n=\overline{\theta}_n(X_1,\ldots,X_n)$为参数$\theta$的估计量，

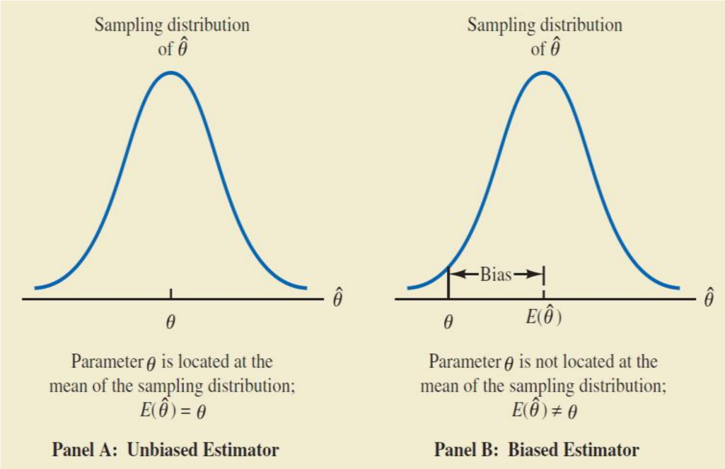
若 $E(\overline{\theta}_n)=\theta \quad (\forall\theta\in\Theta)$

则称$\overline{\theta}_n$为θ的无偏估计。否则称为有偏估计。

注：$E(\overline{\theta}_n)-\theta$ 称为“系统误差”

无偏性就是无系统误差。

MLEs or moment estimators are not, in general, unbiased.



例：设总体X的均值μ及方差均σ^2存在，但未知，

$$\mu,\sigma^2\text{的矩估计量}:\begin{cases}\overline{\mu}=A_1=\overline{X},\\ \overline{\sigma^2}=A_2-A_1^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=(S^*)^2.\end{cases}$$

显然，$E(\overline{\mu})=E(\overline{X})=\mu \;\Rightarrow\; \overline{\mu}$为$\mu$ 的无偏估计量；

又$E(S^2)=E\left(\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)=\sigma^2$ (已证)

即：S^2是σ^2的无偏估计量。

但，$E[(S^*)^2]=E\left(\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\right)$

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

$\Rightarrow (S^*)^2$不是σ^2的无偏估计量(是渐近无偏估计量).

注记：

无偏估计不具有不变性。

设 $\overline{\theta}$ 是 θ 的无偏估计，又设 $\eta = g(\theta)$,

则 $\overline{\eta} := g(\overline{\theta})$ 一般不是 η 的无偏估计。

如：设 $\overline{\theta}$ 是 θ 的无偏估计，且 $D(\overline{\theta}) > 0$,

则 $(\overline{\theta})^2$ 不是 θ^2 的无偏估计。

因为 $E(\overline{\theta}) = \theta$,

所以 $E[(\overline{\theta})^2] = D(\overline{\theta}) + [E(\overline{\theta})]^2 = D(\overline{\theta}) + \theta^2 > \theta^2$.

三、有效性

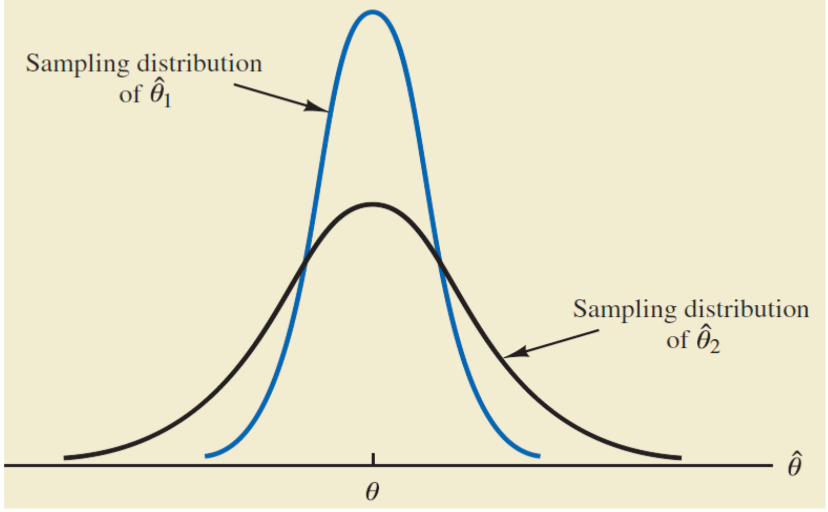
定义：

设 $\overline{\theta}_1,\overline{\theta}_2$ 为参数θ的两个无偏估计，

若 $\forall\theta\in\Theta$，有 $D(\overline{\theta}_1)\le D(\overline{\theta}_2)$，

且至少有一个$\theta\in\Theta$，使上述不等号严格成立，

则称$\overline{\theta}_1$比$\overline{\theta}_2$更有效。



例：

设总体X的均值μ及方差均σ^2存在，但未知。

作估计量：$\overline{\mu}_1 = X_1,\ \overline{\mu}_2 = \overline{X}.$

则 $E(\overline{\mu}_1) = E(X_1) = \mu,$

$$E(\overline{\mu}_2) = E(\overline{X}) = \mu.$$

即$\overline{\mu}_1,\ \overline{\mu}_2$均为$\mu$的无偏估计量。

但$D(\overline{\mu}_1) = D(X_1) = \sigma^2,$

$$D(\overline{\mu}_2) = D(\overline{X}) = \sigma^2 / n,$$

这表明用全部数据的平均估计总体均值要比只使用部分数据更有效。

所以，当$n > 1$时，$\overline{\mu}_2$比$\overline{\mu}_1$更有效。

例： 设总体$X \sim U[0,\theta]$, θ未知，$X_1,\dots,X_n$为样本。

(1) 求θ的MLE $\overline{\theta}_{\mathbf{MLE}}$.

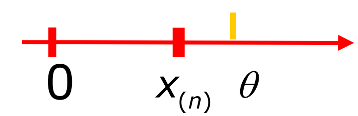
---

(2) 修正$\overline{\theta}_{\mathbf{MLE}}$，使之成为无偏估计，记为$\overline{\theta}_{修}$.

(3)比较$\overline{\theta}_{修}$与θ的矩估计(记为$\overline{\theta}_{矩}$)的有效性。

解:(1)记 $x_{(n)}=\max\{x_1,\dots,x_n\}$. 总体$X$ 的密度 $f(x)=\frac{1}{\theta}\cdot \mathbf{I}_{\{0\le x\le\theta\}}$.

似然函数：$L(\theta)=\prod_{i=1}^{n}\frac{1}{\theta}\cdot \mathbf{I}_{\{0\le x_i\le\theta\}}=\frac{1}{\theta^n}\cdot \mathbf{I}_{\{\theta\ge x_{(n)}\}}$.



对满足$\theta\ge x_{(n)}$的所有θ,均有 $L(\theta)=\frac{1}{\theta^n}\le\frac{1}{(x_{(n)})^n}$,

在$\theta=x_{(n)}$时,等号成立.

故θ的MLE为$\overline{\theta}_{\mathbf{MLE}}=X_{(n)}=\max\{X_1,\dots,X_n\}$.

例(续)

(2) $\bar{\theta}_{\mathrm{MLE}} =: Z$ 的分布函数为$F_{\mathrm{Max}}(z) = (z/\theta)^n, 0 \le z \le \theta,$

其密度函数 $f_{\mathrm{Max}}(z) = n\, z^{n-1}/\theta^n, 0 \le z \le \theta.$

$\Rightarrow E(\bar{\theta}_{\mathrm{MLE}}) = \int_0^\theta z\ f_{\mathrm{Max}}(z)\,dz = \dfrac{n}{n+1}\theta$ (即$\bar{\theta}_{\mathrm{MLE}}$有偏)。

令 $\bar{\theta}_{修} = \dfrac{n+1}{n}\ \bar{\theta}_{\mathrm{MLE}}$

$\Rightarrow E(\bar{\theta}_{修}) = \dfrac{n+1}{n} E(\bar{\theta}_{\mathrm{MLE}}) = \theta$,即$\bar{\theta}_{修}$ 无偏。

例(续)

解:(3).θ 的矩估计为 $\bar{\theta}_{矩}=2\bar{X}$.

设总体 $X\sim U[0,\theta]$. 求 θ 的矩估计量。

解：Step 1：计算总体一阶矩：μ_1(用未知参数表示);

$$\mu_1=E(X)=\frac{\theta}{2};$$

Step 2：反解出未知参数，用总体矩表示 $\theta=2\mu_1$.

Step 3：用样本矩替换总体矩：$\bar{\theta}=2A_1=2\bar{X}$.

则 θ 的矩估计量为 $\bar{\theta}=2A_1=2\bar{X}$.

$$\Rightarrow E(\bar{\theta}_{矩})=E(2\bar{X})=2\cdot\frac{\theta}{2}=\theta.$$ （无偏）

$$D(\bar{\theta}_{矩})=4D(\bar{X})=4\cdot D(X)/n=4\cdot\theta^2/(12n)=\theta^2/(3n).\ (*)$$

因 $\bar{\theta}_{\mathrm{MLE}}$ 的密度函数 $f_{\mathrm{Max}}(z)=nz^{n-1}/\theta^n, 0\le z\le\theta$.

$$\Rightarrow D(\bar{\theta}_{\mathrm{MLE}})=\frac{n}{(n+1)^2(n+2)}\theta^2$$

因 $\bar{\theta}_{修}=\frac{n+1}{n}\bar{\theta}_{\mathrm{MLE}}$

$$\Rightarrow D(\bar{\theta}_{修})=\left(\frac{n+1}{n}\right)^2 D(\bar{\theta}_{\mathrm{MLE}})=\frac{1}{n(n+2)}\theta^2\ (**).$$

比较(*)与(**),当 $n>1$ 时,$3n<n(n+2)\Rightarrow\bar{\theta}_{修}$ 比 $\bar{\theta}_{矩}$ 更有效。

注记：上例中矩估计量和MLE的相合性

(1) 矩估计$\overline{\theta}_{矩}$是θ的相合估计.

证：由大数律$\overline{X} \xrightarrow{P} \dfrac{\theta}{2}$(独立同分布,方差有限),

所以$\overline{\theta}_{矩} = 2\overline{X} \xrightarrow{P} \theta$.

或者由Chebyshev不等式：

$$0 < \mathbf{P}(|\overline{\theta}_{矩} - \theta| \geq \varepsilon) \leq \frac{\mathbf{D}(\overline{\theta}_{矩})}{\varepsilon^2} = \frac{\theta^2}{3n\varepsilon^2} \to 0.$$

注记：上例中矩估计量和MLE的相合性

(2) 最大似然估计$\overline{\theta}_{\text{MLE}}$是$\theta$的相合估计.

证：$E(\overline{\theta}_{\text{MLE}}) = \dfrac{n}{n+1}\theta \to \theta,$

$$\text{且}\,\mathbf{D}(\overline{\theta}_{\text{MLE}}) = \frac{n\theta^2}{(n+1)^2(n+2)} \to 0 \ (n \to +\infty),$$

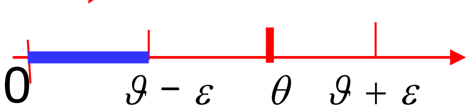
由以前的定理,$\overline{\theta}_{\text{MLE}}$是$\theta$的相合估计。（相合性判定定理）

$$\text{或者：}\ 0 < \mathbf{P}(|\overline{\theta}_{\text{MLE}} - \theta| \geq \varepsilon) = \int_{|z-\theta|\geq\varepsilon} n\frac{z^{n-1}}{\theta^n}dz$$

$$= \int_0^{\theta-\varepsilon} n\frac{z^{n-1}}{\theta^n}dz = \frac{(\theta-\varepsilon)^n}{\theta^n} \to 0.$$

$(z \leq \theta)$

Theta 的epsilon 邻域之外积分

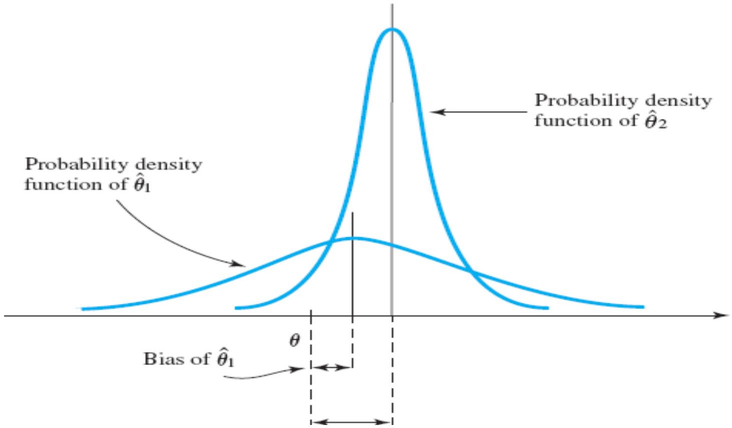


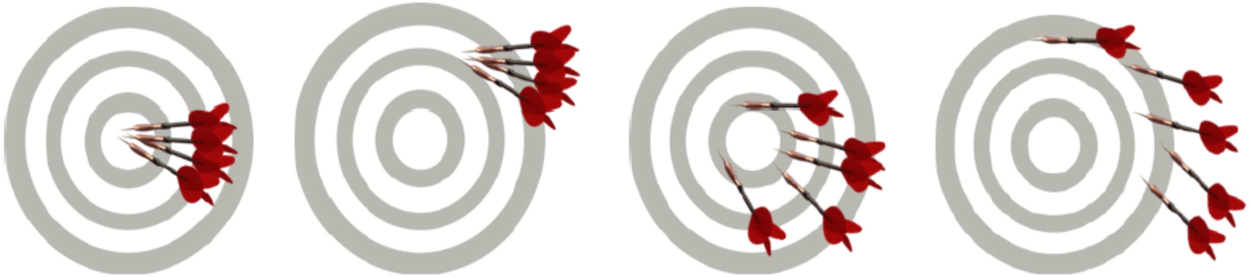
四、均方误差

Two point estimates may have:

- different expectations and
- different variances.

How to compare them?





Bias low Var low (a)

Bias high Var low (b)

Bias low Var high (c)

Bias high Var high (d)

定义：均方误差

设 $\overline{\theta}$ 为参数 θ 的估计量，定义

$$\mathbf{MSE}(\overline{\theta}) := E(\overline{\theta}-\theta)^2$$ 为均方误差。

点估计值与参数真值的距离平方的期望

均方误差的分解：

$$\mathbf{MSE}(\overline{\theta}) = E(\overline{\theta}-\theta)^2 = E\{[\overline{\theta}-E(\overline{\theta})]+[E(\overline{\theta})-\theta]\}^2$$

$$= D(\overline{\theta})+[E(\overline{\theta})-\theta]^2+2E\{[\overline{\theta}-E(\overline{\theta})]\cdot[E(\overline{\theta})-\theta]\}$$

$$= D(\overline{\theta})+[\underbrace{E(\overline{\theta})-\theta}_{\text{系统偏差}}]^2 = \text{方差}+\text{系统偏差的平方。}$$

注记：

对无偏估计，第二项为0，**MSE**准则就是方差(有效性)准则。

对有偏估计，应使**MSE**越小越好。

均方误差是评价点估计的最一般的标准。

无偏估计 不一定 比有偏估计更优。

例： 设总体 $X \sim U[0,\theta]$.

$\bar{\theta}_{\text{MLE}} = X_{(n)} = \max\{X_1,\ldots,X_n\}$是$\theta$的**MLE,** 是有偏的。

令$\bar{\theta}_{修} = \dfrac{n+1}{n} X_{(n)} \Rightarrow E(\bar{\theta}_{修}) = \dfrac{n+1}{n} E(X_{(n)}) = \theta$,即$\bar{\theta}_{修}$无偏。

$$\mathbf{MSE}(\bar{\theta}_{修}) = D(\bar{\theta}_{修}) = \frac{1}{n(n+2)}\theta^2.$$

（上例已计算）

另作形如$\bar{\theta}_a = a\,X_{(n)}$的估计量。选取$a$使**MSE**最小。

$$\mathbf{MSE}(\bar{\theta}_a) = \mathbf{D}(\bar{\theta}_a) + [\mathbf{E}(\bar{\theta}_a) - \theta]^2$$

$$= a^2 \frac{n}{(n+1)^2(n+2)}\theta^2 + \left(\frac{na}{n+1} - 1\right)^2 \theta^2,$$

当$a = \dfrac{n+2}{n+1}$时,$\bar{\theta}_a = a\,X_{(n)}$的**MSE**最小。

注：$E(X_{(n)}) = \dfrac{n}{(n+1)}\theta$

$D(X_{(n)}) = \dfrac{n\theta^2}{(n+1)^2(n+2)}$

在均方误差意义下，有些有偏估计优于无偏估计。

例: 设总体$X \sim N(\mu,\sigma^2),\mu,\sigma^2$ 未知。考虑σ^2 的形如

$T = c\sum_{i=1}^{n}(X_i - \overline{X})^2$的估计量.确定$c$,使对应的**MSE**最小.

注：$c = 1/(n-1)$时为无偏;$c = 1/n$为**MLE**.

解：$\chi^2 := \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 \sim \chi^2(n-1) \Rightarrow T = c\sigma^2\chi^2,$

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

$$E(T) = c\sigma^2 E(\chi^2) = c\sigma^2(n-1),$$

$$D(T) = c^2\sigma^4 D(\chi^2) = c^2\sigma^4 2(n-1).$$

$$\mathbf{MSE(T)} = D(T) + [E(T) - \sigma^2]^2$$

$$= c^2\sigma^4 2(n-1) + [c\sigma^2(n-1) - \sigma^2]^2$$

$$= \sigma^4[(n^2-1)c^2 - 2(n-1)c + 1].$$

当$c = 1/(n+1)$时，**MSE**最小。

四、均方误差（续）

定义：对参数估计问题，设有一个估计类，

称 $\bar{\theta}$ 是该估计类中的一致最小均方误差估计，

如果对该估计类中其它任意的 θ^*，在参数

空间上都有

$$\mathrm{MSE}(\bar{\theta}) \le \mathrm{MSE}(\theta^*).$$

注意：

一致最小均方误差估计通常在某个确定的估计类中进行。如果对估计类不加限制，一致最小均方误差估计一般不存在。

五、一致最小方差无偏估计 了解

在无偏估计类中，均方误差 = 方差。

按均方误差准则，在无偏估计类中，方差愈小愈先。

定义：对参数估计问题，设 $\bar{\theta}$ 是参数 θ 的一个无偏估计，如果对其它任意的 θ 的无偏估计 θ^*，在参数空间上都有

$$D(\bar{\theta}) \le D(\theta^*),$$

则称 $\bar{\theta}$ 是 θ 的一致最小方差无偏估计 (**UMVUE**)

Uniformly Minimum Variance Unbiased Estimate.

第三节、区间估计

- **点估计：**

  对未知参数，适当构造统计量作为其估计（精度不知）。

- **区间估计：**

  给出一个范围(区间)，希望它包含参数真值的可信度较高

Point estimate

Fishing with a spear

Confidence interval

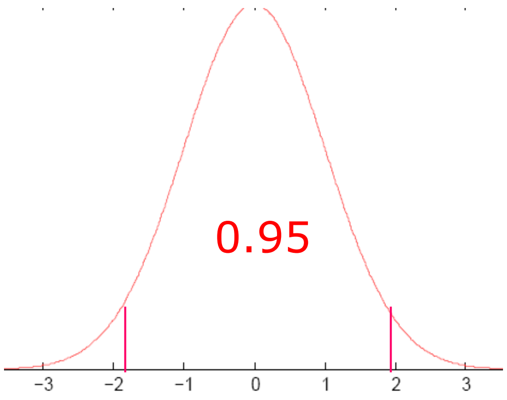
Fishing with a net

例：

设总体X 服从正态分布$N(\mu,\sigma^2)$. σ^2已知，μ未知。

均值μ的MLE：$\overline{\mu}=\overline{X}$. 点估计，精确度未知

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



$$\Rightarrow \quad P\left(-u_{0.975}\le\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\le u_{0.975}\right)=0.95;\ (*)$$

其中，$u_{0.975}$ --- $N(0,1)$的下侧**0.975**分位数。

即：$\mathbf{P}\{\overline{X}-1.96\sigma/\sqrt{n}\le\mu\le\overline{X}+1.96\sigma/\sqrt{n}\}=\mathbf{0.95}$.

区间($\overline{X}-1.96\sigma/\sqrt{n}$, $\overline{X}+1.96\sigma/\sqrt{n}$）为随机区间，

区间端点均为统计量。该区间包含真值的可信度为95%.

一、置信区间

设θ是总体X的一个参数$(\theta \in \Theta)$, $X_1, \ldots, X_n$为样本，

对给定的$\alpha\ (0 < \alpha < 1)$, 若有两个统计量

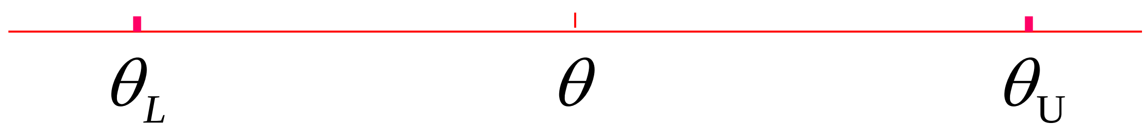
$\theta_L = \theta_L(X_1, \ldots, X_n)$和$\theta_U = \theta_U(X_1, \ldots, X_n)$, 对$\theta \in \Theta$，满足

$$P_\theta(\theta_L \le \theta \le \theta_U) \ge 1 - \alpha,$$

则称随机区间$[\theta_L, \theta_U]$为参数θ的置信水平为$1-\alpha$的置信区间；

θ_L, 和θ_U分别称为(双侧)置信下限和置信上限；

$1-\alpha$称为置信水平。



同等置信区间：要求$P_\theta(\theta_L \le \theta \le \theta_U) = 1 - \alpha.$

同等置信区间是把给定的置信水平$1-\alpha$ 用足, 在总体为连续分布时常可实现。

注意：

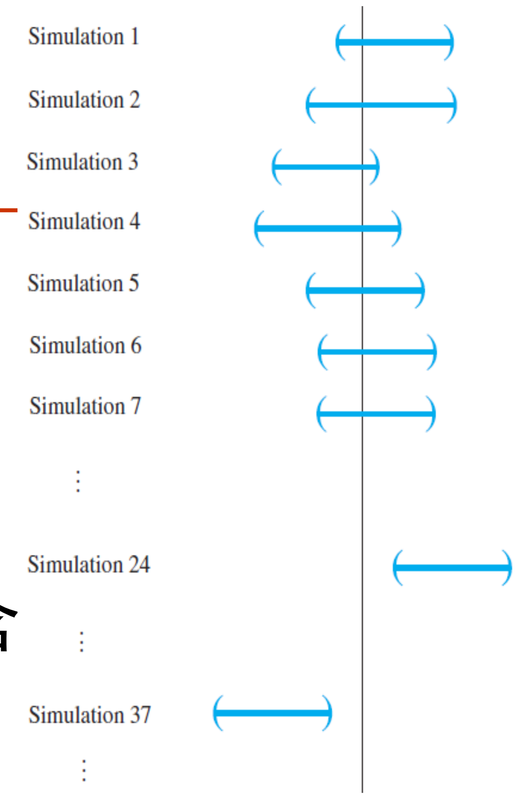
不能说，参数θ落在区间$[\theta_L, \theta_U]$的概率为$1-\alpha$.

θ为定数，区间$[\theta_L, \theta_U]$的两端点为随机变量。

随机区间

- When we have a 95% CI of (2, 4), we say: we are 95% confident that the interval (2, 4) contains the true population mean μ.
- But we cannot say: 95% probability that the true population mean μ is in (2, 4). (此处没有RV)

置信区间的频率解释



设θ是总体X的一个参数($\theta \in \Theta$),

重复抽样：$x_1^k, \ldots, x_n^k$, $k = 1, \ldots, M$.

每组样本观测值确定一个区间$[\theta_L^k, \theta_U^k]$

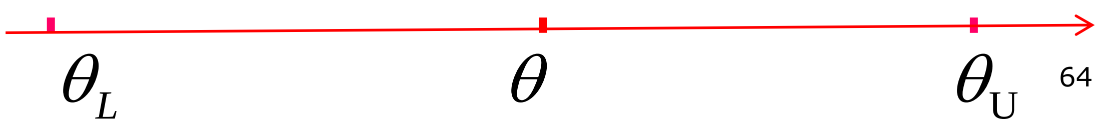
这样的区间或者包含真值θ，或者不包含

注意到：$P(\theta_L \le \theta \le \theta_U) = (1-\alpha)$,

由Bernoulli大数律，含真值θ的区间约占$100(1-\alpha)\%$.

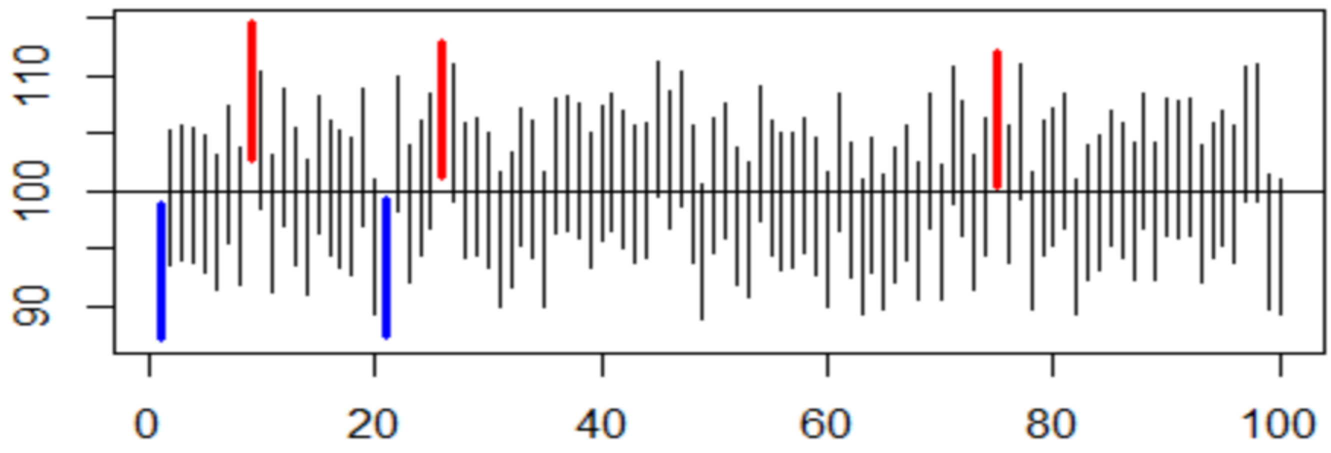
如：$\alpha = 0.01$, $M = 1000$:

含真值θ的区间约有$M \cdot 99\% = 990$.



置信区间的频率解释

- 100人，每人从正态N(100,18)生成容量为36的样本。
- 每人从自己的样本中构造总体均值的95%的置信区间。
- 画图展示100个CI.



区间估计的优良性准则

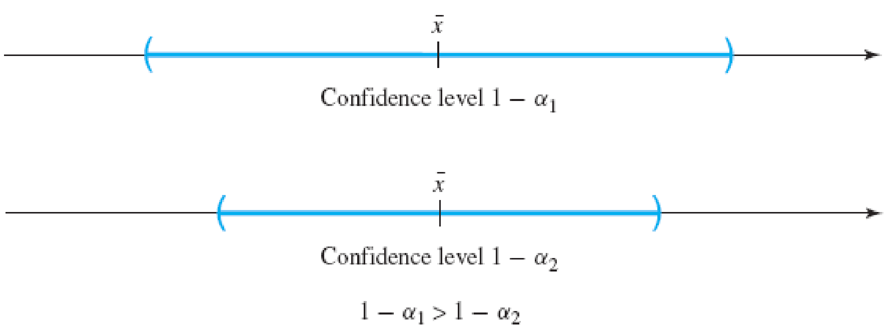
- **可靠性**：区间包含未知参数的概率有多大？
- **精　度**：区间的长度。

若把可靠度提高，则区间长度增加。

若把区间的长度减小，可靠度也变小。

样本量一定时，两者如何协调？

一般，给定置信水平，
以保证有一定的可靠度，
在此前提下，尽量选择精度
更高的区间估计。



二、置信区间构造方法

枢轴量法：

Step 1. 构造样本和θ的函数$G=G(X_1,\ldots,X_n;\theta)$，

使G的分布已知且不含未知参数θ（枢轴量）；

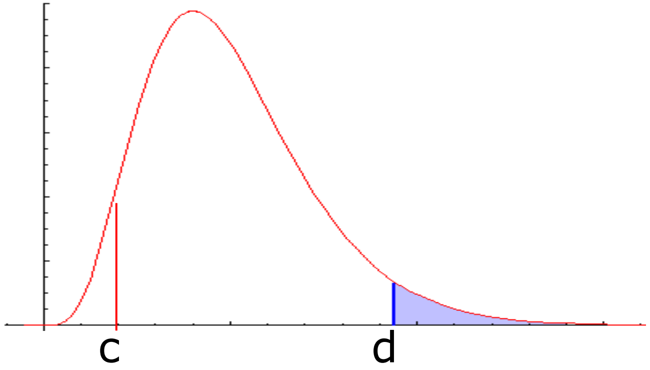
Step 2. 对给定的置信水平$1-\alpha$，选取常数c,d，使

$P(c\le G(X_1,\ldots,X_n;\theta)\le d)=1-\alpha$；

Step 3. 从$c\le G\le d$得到等价不等式：

$\theta_L\le\theta\le\theta_U$，

则 $P(\theta_L\le\theta\le\theta_U)=1-\alpha.$



则$[\theta_L,\theta_U]$为θ的置信水平为$1-\alpha$的(同等)置信区间。

注记：

满足置信度要求的 c 与 d 通常不唯一。

若有可能，应选置信区间平均长度达到最短者。

当G 的分布为对称分布时容易实现。

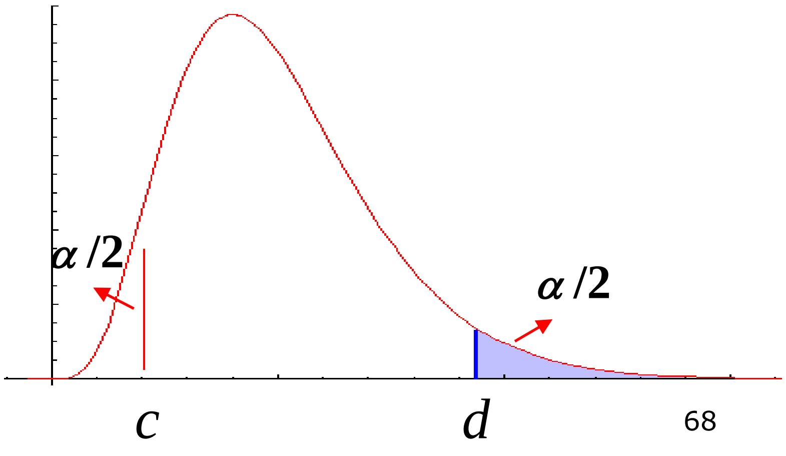
一般情况下，选使平均长度最短的c和d往往很难实现。

常选择 c 与 d，使得两个尾部概率各为 $\alpha/2$，即

$$P(G < c) = P(G > d) = \alpha/2.$$

这样的置信区间称为等尾置信区间

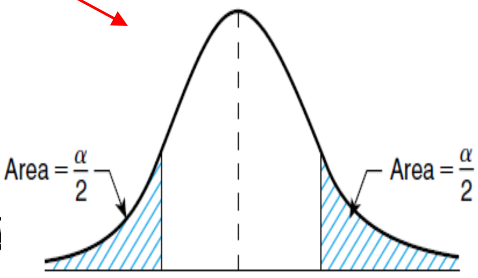
在G 的分布为偏态分布场合常采用



三、单个正态总体参数的置信区间

1. 均值的置信区间（方差已知的情形）

Step 1. 从均值μ的点估计出发：作$G := \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$



Step 2. 对给定的置信水平$1-\alpha$，选取常数c, d，使

$$P(c \le G \le d) = 1-\alpha;\ (*)$$

可取$d=-c=u_{1-\alpha/2}$ --- $N(0,1)$的下侧$1-\alpha/2$分位数

Step 3. 从$c \le G \le d$ 即 $c \le \dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le d$ 得等价不等式：

$$\overline{X}-u_{1-\alpha/2}\sigma/\sqrt{n} \le \mu \le \overline{X}+u_{1-\alpha/2}\sigma/\sqrt{n},$$

则 $P(\overline{X}-u_{1-\alpha/2}\sigma/\sqrt{n} \le \mu \le \overline{X}+u_{1-\alpha/2}\sigma/\sqrt{n}) = 1-\alpha.$

$\Rightarrow \mu$的置信水平为$1-\alpha$的(同等)置信区间为

$$[\overline{X}-u_{1-\alpha/2}\sigma/\sqrt{n},\ \overline{X}+u_{1-\alpha/2}\sigma/\sqrt{n}]，\text{常表示为：}\overline{X}\pm u_{1-\alpha/2}\underline{\sigma/\sqrt{n}}.$$

Point estimate $\pm$ (Critical value)(Standard error)

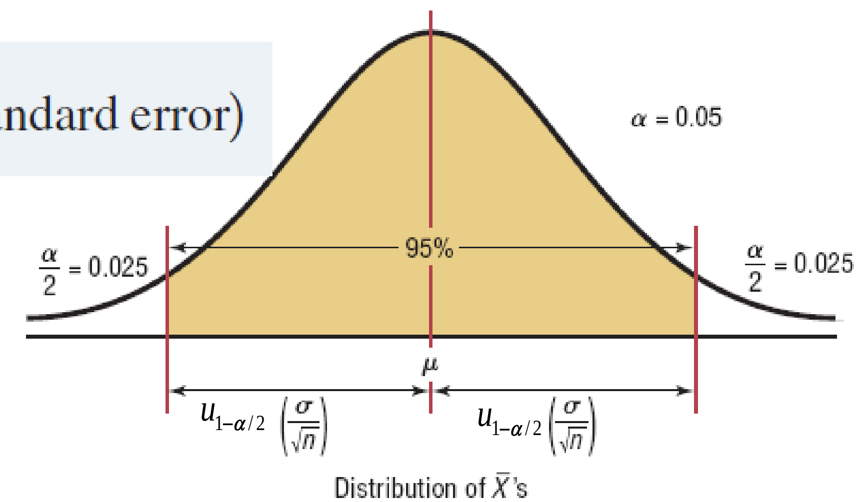
standard error of $\overline{X}$

Four commonly used confidence level

$1-\alpha$	α	$\alpha/2$	$u_{1-\alpha/2}$
.90	.10	.05	1.645
.95	.05	.025	1.96
.98	.02	.01	2.33
.99	.01	.005	2.575

Point estimate $\pm$ (Critical value)(Standard error)

CI General Format



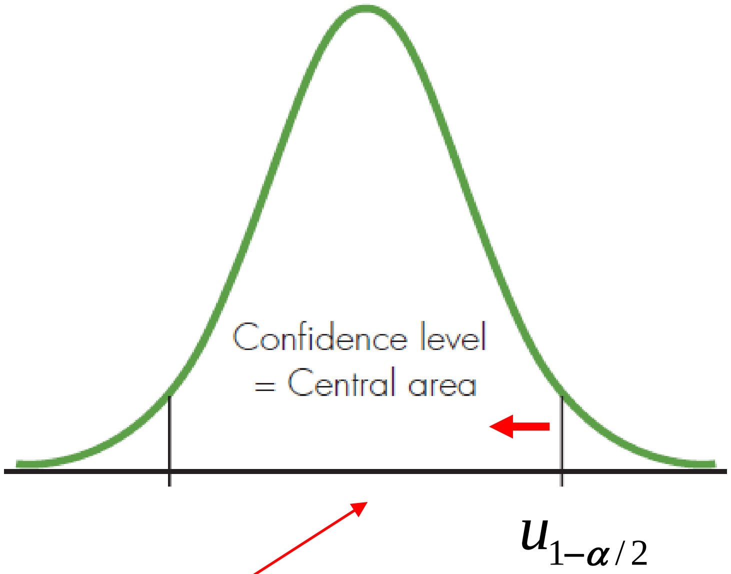
置信水平和样本容量的影响

Margin of error：the half-length of CI. It is the maximum likely difference between point estimate and the actual value of the parameter.

- Conf Interval: $\bar{X} \pm u_{1-\sigma/2}\dfrac{\sigma}{\sqrt{n}}$,

Margin of Error $E = u_{1-\sigma/2}\dfrac{\sigma}{\sqrt{n}}$,

Sample size $n = \dfrac{\left(u_{1-\sigma/2}\right)^2\sigma^2}{E^2}$

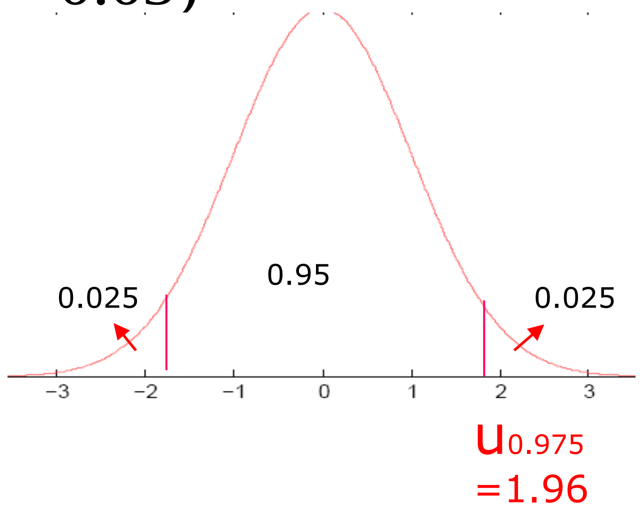


- 置信水平减小时（分位点减），置信区间长度缩小。
- 样本量增大时，置信区间长度缩小（长度与样本量的平方根成反比）。

例：

例如：对给定的置信水平 $1-\alpha=0.95\ (\alpha=0.05)$

$$P\left(c \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le d\right)=0.95\quad(*)$$



可取 $d=-c=u_{1-\alpha/2}=u_{0.975}=1.96$

(对称取点时，可使区间长度最短)

(*)等价于：

$$P(\overline{X}-1.96\sigma/\sqrt{n} \le \mu \le \overline{X}+1.96\sigma/\sqrt{n})=1-\alpha.$$

则 μ 的置信水平为0.95的(同等)置信区间为

$$[\overline{X}-1.96\sigma/\sqrt{n},\ \overline{X}+1.96\sigma/\sqrt{n}].$$

$\overline{X}-1.96\sigma/\sqrt{n}$ $\quad$ $\overline{X}$ $\quad$ $\overline{X}+1.96\sigma/\sqrt{n}$

例：

用天平秤某物体的重量9次，得平均值为 15.4(g).
已知天平秤量结果为正态分布，其标准差为0.1克。
试求该物体重量的 0.95 置信区间。

解： 此处$1-\alpha=0.95$，$\alpha=0.05$，查表知$u_{0.975}=1.96$，
于是该物体重量μ的0.95置信区间为

$$[\bar{x}-1.96\sigma/\sqrt{n},\ \bar{x}+1.96\sigma/\sqrt{n}].$$

从而该物体重量的0.95置信区间为

[15.3347，15.4653].

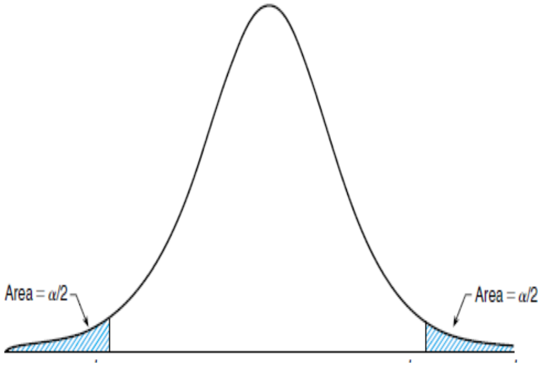
2. 均值的置信区间（方差未知的情形）

用t分布代替正态分布

Step 1. 从均值μ的点估计出发：作$G=\dfrac{\overline{X}-\mu}{S/\sqrt{n}}\sim t(n-1)$

Step 2. 对给定的置信水平$1-\alpha$，选取常数c,d,使

$$P\left(c\le \frac{\overline{X}-\mu}{S/\sqrt{n}}\le d\right)=1-\alpha;\ (*)$$



可取$d=-c=t_{1-\alpha/2}(n-1)$,即$t(n-1)$的下侧$1-\alpha/2$分位数。

$t_{1-\alpha/2}(n-1)$

Step 3. $(*)$等价于

$$P(\overline{X}-t_{1-\alpha/2}(n-1)S/\sqrt{n}\le\mu\le\overline{X}+t_{1-\alpha/2}(n-1)S/\sqrt{n})=1-\alpha.$$

$\Rightarrow$ 则μ的置信水平为$1-\alpha$的(同等)置信区间为

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

与方差已知的情形类似

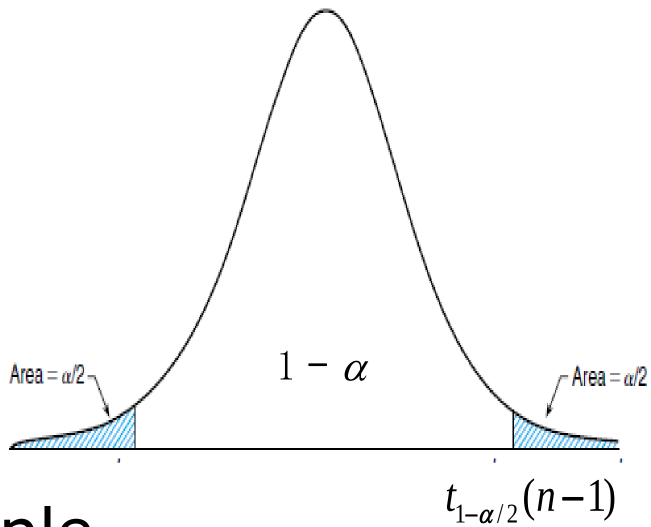
Remark: The Effect of Sample Size

置信区间：

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

区间长度：

$$L = \frac{2\, t_{1-\alpha/2}(n-1)S}{\sqrt{n}} \propto \frac{1}{\sqrt{n}}.$$



A *fourfold* increase in the sample size reduces the CI length by *half*.

3. 方差的置信区间（均值未知）

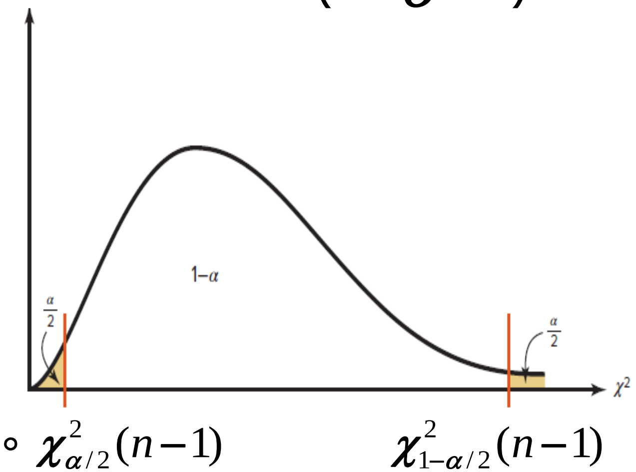
Step 1. 作 $G=\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$.　注：当$\mu$已知时作$G=\sum\left(\dfrac{X_i-\mu}{\sigma}\right)^2$

Step 2. 对给定的置信水平$1-\alpha$，选取常数c,d,使

$$P\left(c\le\frac{(n-1)S^2}{\sigma^2}\le d\right)=1-\alpha;\ (*)$$

可取 $d=\chi^2_{1\text{-}\alpha/2}(n-1)$,$c=\chi^2_{\alpha/2}(n-1)$,

分别为$\chi^2(n-1)$分布的下侧$1-\alpha/2$，$\alpha/2$分位数。



Step 3. (*)等价于

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

σ^2的置信水平为$1-\alpha$的(同等)置信区间为$\left[\dfrac{(n-1)S^2}{\chi^2_{1\text{-}\alpha/2}(n-1)},\ \dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}\right]$.

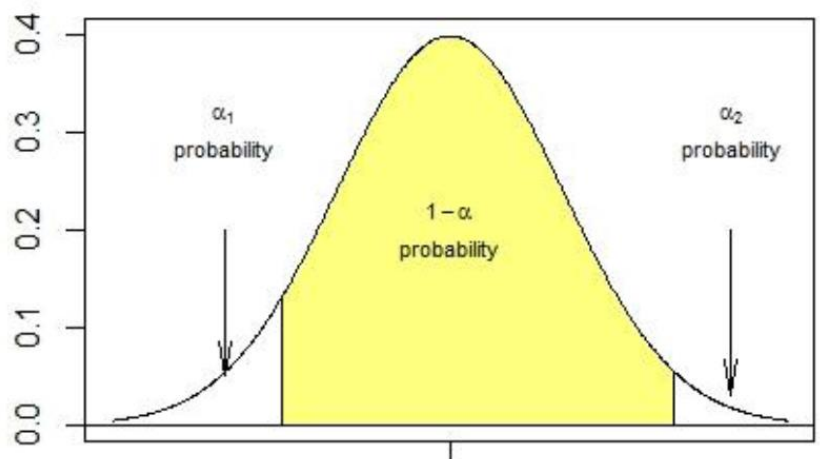
四、单侧置信区间

均值的对称的CI：

$$[\overline{X}+u_{\alpha/2}\sigma/\sqrt{n},\overline{X}+u_{1-\alpha/2}\sigma/\sqrt{n}].$$

亦可非对称的CI：

$$[\overline{X}-u_{\alpha_1}\sigma/\sqrt{n},\overline{X}+u_{1-\alpha_2}\sigma/\sqrt{n}],\alpha_1+\alpha_2=1.$$



常对均值的上界或下界感兴趣：单侧置信区间

四、单侧置信区间

实际问题中, 关心产品寿命的“下限”(标志质量，越长越好) 杂质（毒性）含量的“上限”（越小越好）。

设θ是总体X的一个参数$(\theta \in \Theta)$, $X_1,\dots,X_n$为样本，

对给定的$\alpha\ (0<\alpha<1)$, 若有统计量$\theta_L=\theta_L(X_1,\dots,X_n)$,

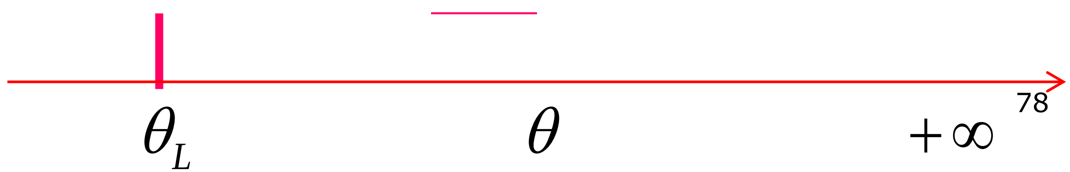
满足

$$P_\theta(\theta_L \le \theta) \ge 1-\alpha,\ \theta \in \Theta$$

则θ_L为参数θ的置信水平为$1-\alpha$的(单侧)置信下限。

若 $$P_\theta(\theta_L \le \theta) = 1-\alpha,\ \theta \in \Theta$$

则θ_L为参数θ的置信水平为$1-\alpha$的同等置信下限。



四、单侧置信区间

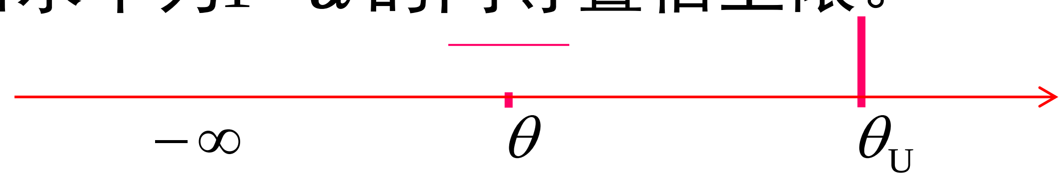
设θ是总体X的一个参数$(\theta \in \Theta)$, $X_1,\ldots,X_n$为样本，对给定的$\alpha\ (0<\alpha<1)$, 若有统计量$\theta_U=\theta_U(X_1,\ldots,X_n)$, 满足

$$P_\theta(\theta \le \theta_U) \ge 1-\alpha,\ \theta \in \Theta$$

则θ_U为参数θ的置信水平为$1-\alpha$的(单侧)置信上限。

若 $P_\theta(\theta \le \theta_U) = 1-\alpha,\ \theta \in \Theta$

则θ_U为参数θ的置信水平为$1-\alpha$的同等置信上限。



单侧置信限是(双侧)置信区间的特殊情形。
寻求(双侧)置信区间的方法可以用来寻找单侧置信限。

例： 设总体$X \sim N(\mu,\sigma^2)$，μ,σ^2均未知。

求：(1) μ的置信水平为$1-\alpha$的单侧置信下限；

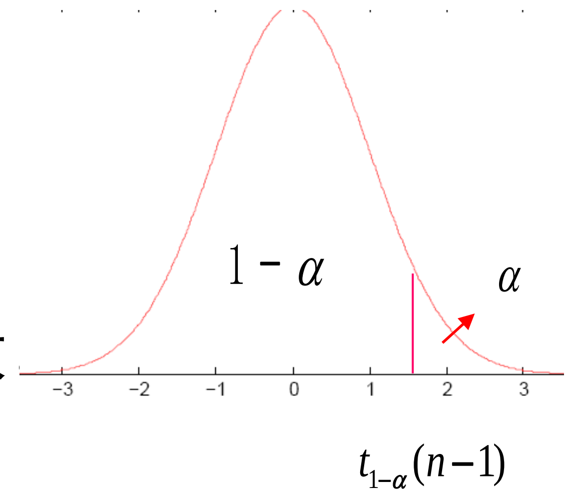
(2) σ^2的置信水平为$1-\alpha$的单侧置信上限。

Step 1. 从均值μ的点估计出发：作$G=\dfrac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$

Step 2. 对给定的置信水平$1-\alpha$，有

$$P\left(\frac{\overline{X}-\mu}{S/\sqrt{n}} \le t_{1-\alpha}(n-1)\right)=1-\alpha;\ (*)$$

$t_{1-\alpha}(n-1)$ 为$t(n-1)$的下侧$1-\alpha$分位数



Step 3. (*)等价于

下限

$$P(\mu \ge \overline{X}-t_{1-\alpha}(n-1)S/\sqrt{n})=1-\alpha.$$

$\Rightarrow$ μ的置信水平为$1-\alpha$的(同等)置信下限为$\overline{X}-t_{1-\alpha}(n-1)S/\sqrt{n}$.

例： 设总体$X \sim N(\mu,\sigma^2)$，μ,σ^2均未知。

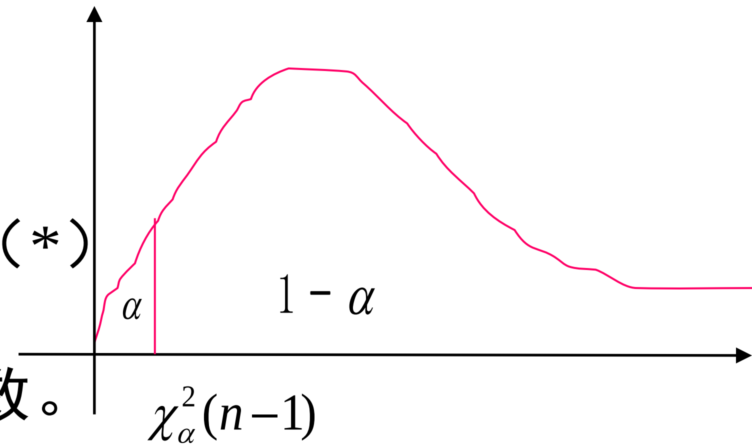
求：(1) μ的置信水平为$1-\alpha$的单侧置信下限；

(2) σ^2的置信水平为$1-\alpha$的单侧置信上限。

Step 1. 作$G=\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

Step 2. 对给定的置信水平$1-\alpha$,

$$P\left(\frac{(n-1)S^2}{\sigma^2} \geq \chi^2_\alpha(n-1)\right)=1-\alpha; \quad (*)$$



$\chi^2_\alpha(n-1)$为$\chi^2(n-1)$分布的下侧α分位数。

Step 3. (*)等价于

上限

$$P\left(\sigma^2 \leq \frac{(n-1)S^2}{\chi^2_\alpha(n-1)}\right)=1-\alpha.$$

则σ^2的置信水平为$1-\alpha$的(同等)置信上限为$\dfrac{(n-1)S^2}{\chi^2_\alpha(n-1)}$.

四、单侧置信区间

- 均值的下侧、上侧置信区间（置信水平 $1-\alpha$，方差已知）

下侧： $[\overline{X}+u_{\alpha}\dfrac{\sigma}{\sqrt{n}},+\infty)$. **上侧：** $(-\infty,\overline{X}+u_{1-\alpha}\dfrac{\sigma}{\sqrt{n}}]$.

- 方差的下侧、上侧置信区间：

下侧： $\left(\dfrac{(n-1)S^2}{\chi^2_{1-\alpha}(n-1)},\ +\infty\right)$. **上侧：** $\left(0,\dfrac{(n-1)S^2}{\chi^2_{\alpha}(n-1)}\right)$.

五、大样本置信区间

- Consider a binomial experiment with n trials and success probability p, number of successes is X:

$$X \sim B(n, p).$$

- The parameter is population proportion p.
- We compute sample proportion x/n ---
  the proportion of trials resulting in success.

五、大样本置信区间

Let p = binomial probability of success.

Let $\bar{p}$ = sample proportion or proportion of success.

If X ~ B(n, p), then

$$\bar{p} = \bar{X} = X/n$$

has **approximately normal distribution**

$$\bar{p} = \bar{X} \sim N(p, p(1-p)/n).$$

(np and $n(1-p)$ must each be at least 5)

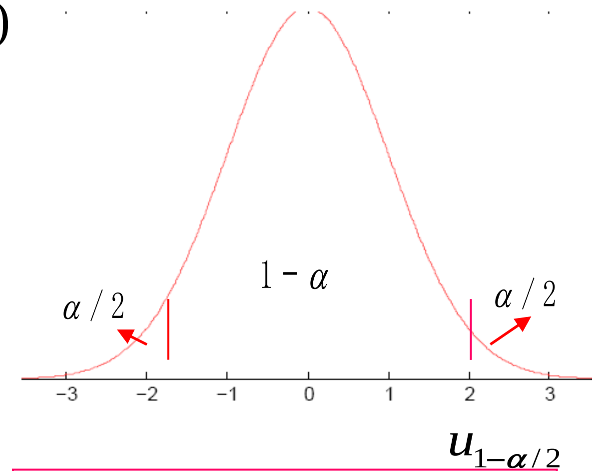
考虑 0-1 分布的总体X， P(X=1)=p, P(X=0)=1-p，p未知，样本 X₁， ···， Xₙ.

Step 1. 作$G=\dfrac{n\overline{X}-np}{\sqrt{np(1-p)}}\overset{\text{近似}}{\sim}N(0,1)$ (中心极限定理)

Step 2. 对给定的置信水平$1-\alpha$，选取常数$u_{1-\alpha/2}$，使

$$P\left(\left|\frac{n\overline{X}-np}{\sqrt{np(1-p)}}\right|\le u_{1-\alpha/2}\right)\approx 1-\alpha;\ (*)$$

$u_{1-\alpha/2}$ --- $N(0,1)$的下侧$1-\alpha/2$分位数。



Step 3. (*)式等价于 $P(ap^2+bp+c\le 0)\approx 1-\alpha,$

二次三项式，开口向上

其中，$a=n+u^2_{1-\sigma/2}$，$b=-(2n\overline{X}+u^2_{1-\sigma/2})$，$c=n\overline{X}^2$.

即 $P(p_L\le p\le p_U)\approx 1-\alpha$, $p_L=\dfrac{1}{2a}(-b-\sqrt{b^2-4ac})$, $p_U=\dfrac{1}{2a}(-b+\sqrt{b^2-4ac})$.

则p的置信水平为$1-\alpha$的置信区间近似为$[p_L,p_U]$

五、大样本置信区间

上述置信区间可近似化为：

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

对比正态总体情形，均值的置信区间：

$$\left[\overline{X}-u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}},\quad \overline{X}+u_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}}\right].$$

(相当于 σ^2 用 $\overline{X}(1-\overline{X})$ 代替)

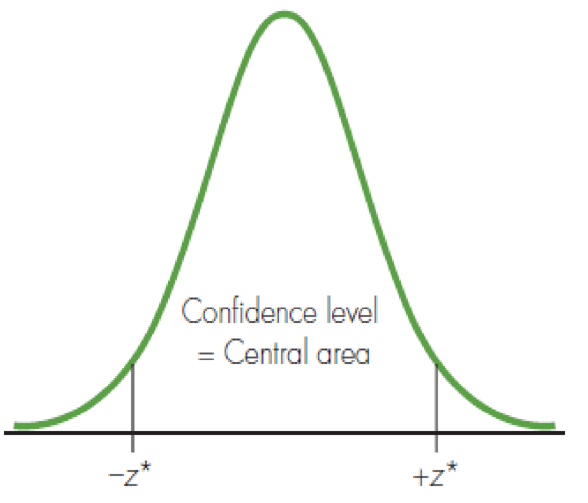
Point estimate $\pm$ (Critical value)(Standard error)

例：

调查某地100座桥梁，发现有 33 座有缺陷。求该地桥梁有缺陷比例的 90% 的置信区间。

解：置信区间

$$\left(\hat{p}-z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}},\ \hat{p}+z^{*}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$



其中，$z^*=1.645$ （0. 95下侧分位数），

$$\hat{p}=\frac{33}{100}=0.33,\quad n=100$$

置信区间(0.33-0.077, 0.33+0.077)，

即 (0.253，0.407)

例：

调查收视率p，为使p的$1-\alpha$的置信区间长度不超过d_0，应调查多少用户？

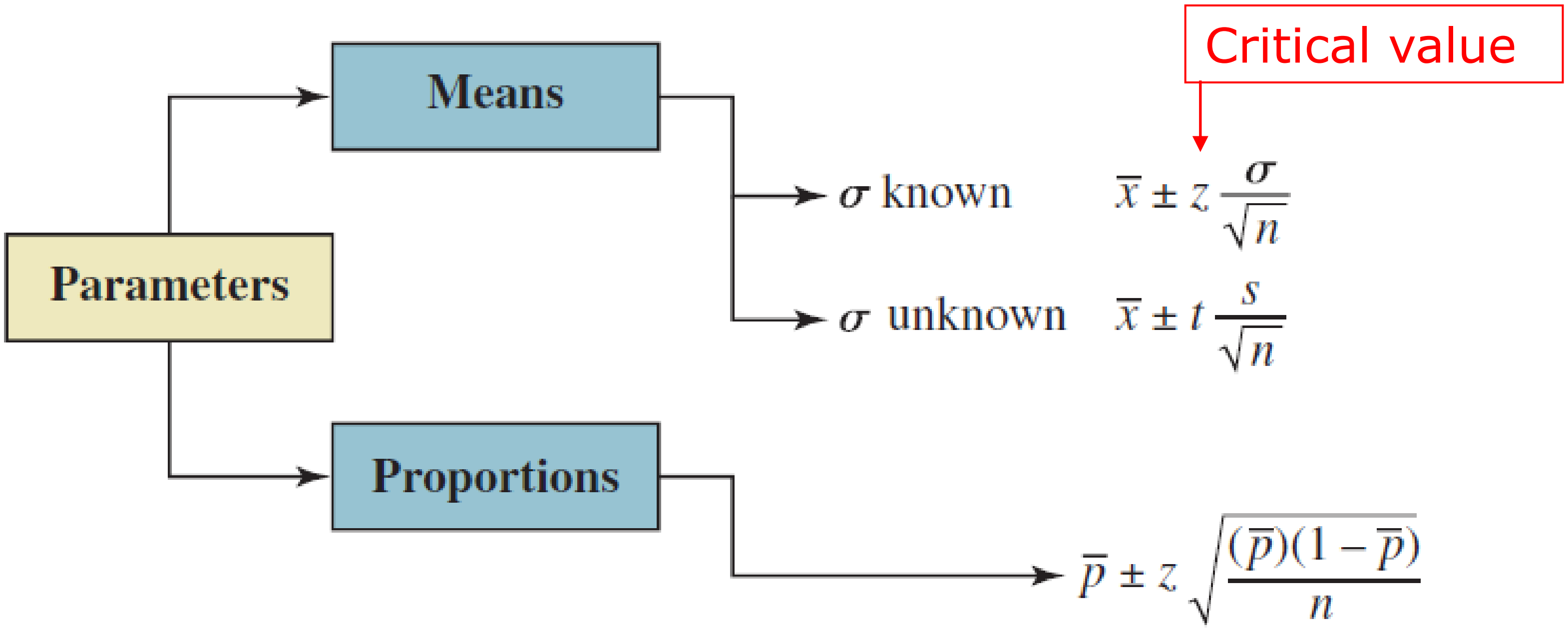
解：置信区间的长度为: $d := 2u_{1-\alpha/2}\sqrt{\dfrac{\overline{X}(1-\overline{X})}{n}}$.

易知 $d \le u_{1-\alpha/2}\sqrt{1/n}$ （因 $\overline{X}(1-\overline{X}) \le 1/4$）

要使置信区间的长度不超过d_0，只要

$$u_{1-\alpha/2}\sqrt{1/n} \le d_0 \Rightarrow n \ge \left(\frac{u_{1-\alpha/2}}{d_0}\right)^2.$$

Comparison



General formula:

$$\sigma_{\bar{p}} \approx \sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$$

$$\boxed{\text{Point estimate} \pm (\text{Critical value})(\text{Standard error})}$$

CI for Poisson mean

For large n, we can use the z-interval $(\bar{X} + z_{\alpha/2} \cdot se(\bar{X}), \bar{X} + z_{1-\alpha/2} \cdot se(\bar{X}))$ as a 1-α CI for λ.

Since Poisson variance is λ, we use $\hat{\lambda} = \bar{X}$ instead of sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ to estimate the population variance. Hence $se(\bar{X}) = \sqrt{\frac{\bar{X}}{n}}$.

Therefore the two-sided 1-α CI for λ is $(\bar{X} + z_{\alpha/2}\sqrt{\frac{\bar{X}}{n}}, \bar{X} + z_{1-\alpha/2}\sqrt{\frac{\bar{X}}{n}})$.

The one-sided 1-α CI for λ is $(\bar{X} + z_{\alpha}\sqrt{\frac{\bar{X}}{n}}, \infty)$ or $(-\infty, \bar{X} + z_{1-\alpha}\sqrt{\frac{\bar{X}}{n}})$.

Point estimate $\pm$ (Critical value)(Standard error)

The End of Chapter 6