ch9：区间估计

9.1 引言

第 7 章我们讨论过参数 θ 的点估计，那里的推断是猜测一个单个值作为 θ 的值. 这一章我们讨论区间估计及更一般的集合估计. 集合估计问题中的推断就是陈述"$\theta \in C$"，其中 $C \subset \Theta$ 并且 $C=C(\boldsymbol{X})$ 是一个由观测数据 $\boldsymbol{X}=\boldsymbol{x}$ 的值决定的集合. 如果 θ 是实值的，则我们通常更喜欢集合估计 C 是一个区间. 区间估计将是这章的主题.

一些定义

区间估计量

定义 9.1.1 一个实值参数 θ 的区间估计是样本的任意一对函数 $L(x_1, \cdots, x_n)$ 和 $U(x_1, \cdots, x_n)$，对于所有的 $\boldsymbol{x} \in \mathcal{X}$ 满足 $L(\boldsymbol{x}) \leqslant U(\boldsymbol{x})$. 如果观测到样本 $\boldsymbol{X}=\boldsymbol{x}$，就做出推断 $L(\boldsymbol{x}) \leqslant \theta \leqslant U(\boldsymbol{x})$. 随机区间 $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ 叫做区间估计量 (interval estimator).

单侧区间

我们将按照过去的习惯，把 θ 的一个基于随机样本 $\boldsymbol{X}=(X_1, \cdots, X_n)$ 的区间估计量写成 $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ 而实现值的区间写成 $[L(\boldsymbol{x}), U(\boldsymbol{x})]$. 虽然在多数情况我们处理的是 L 和 U 的有限值，但是有时兴趣在单侧区间估计上. 例如，若 $L(\boldsymbol{x})=-\infty$，我们就有单侧的区间 $(-\infty, U(\boldsymbol{x})]$ 并断言"$\theta \leqslant U(\boldsymbol{x})$"而不提及下界. 类似. 我们也可取 $U(\boldsymbol{x})=\infty$ 而得到一个单侧区间 $[L(\boldsymbol{x}), \infty)$.

虽然定义提及的是一个闭区间 $[L(\boldsymbol{x}), U(\boldsymbol{x})]$，但是有的时候更自然地使用开区间 $(L(\boldsymbol{x}), U(\boldsymbol{x}))$，甚或一个半开半闭的区间，就像上段中那样. 虽然优先使用的是闭区间，但是我们将根据手边的特定问题，哪种更合适就使用哪种.

例子

例 9.1.2（区间估计量） 设 X_1，X_2，X_3，X_4 是来自 $\mathrm{n}(\mu, 1)$ 的样本，$[\bar{X}-1, \bar{X}+1]$ 是 μ 的一个区间估计量. 这意味我们将断言 μ 在这个区间里. ||

在这里，自然要问及通过使用区间估计我们得到了什么. 过去我们用 $\bar{X}$ 估计 μ，而现在使用的估计量 $[\bar{X}-1, \bar{X}+1]$ 不如前者精确. 我们想必得到了什么！通过放弃估计值（或关于 μ 的断言）的某些精确性，我们得到关于这个断言之正确性的某些自信或保证.

例 9.1.3（例 9.1.2 续） 当我们用 $\overline{X}$ 估计 μ 时，恰好正确的概率，即 $P(\overline{X}=\mu)$，是 0. 但是若使用一个区间估计量，断言正确的概率就是正的. μ 被 $[\overline{X}-1, \overline{X}+1]$ 所覆盖的概率可以计算如下

$$\begin{aligned}P(\mu\in[\overline{X}-1,\overline{X}+1])&=P(\overline{X}-1\leqslant\mu\leqslant\overline{X}+1)\\&=P(-1\leqslant\overline{X}-\mu\leqslant 1)\\&=P\left(-2\leqslant\frac{\overline{X}-\mu}{\sqrt{1/4}}\leqslant 2\right)\\&=P(-2\leqslant Z\leqslant 2)\left(\frac{\overline{X}-\mu}{\sqrt{1/4}}\text{是标准正态的}\right)\\&=0.9544\end{aligned}$$

因此我们有超过 95％的机会用我们的区间估计量覆盖这个未知的参数. 一个点换成一个区间，牺牲估计值的某些精确性，却使我们对于断言是正确的更有信心.

使用区间估计而不使用点估计，目的在于对于捕获感兴趣的参数有某种保证. 这个保证的确定是用以下定义量化的.

区间覆盖概率

定义 9.1.4 对于一个对参数 θ 的区间估计量 $[L(\boldsymbol{X}), U(\boldsymbol{X})]$，$[L(\boldsymbol{X}), U(\boldsymbol{X})]$ 的覆盖概率（coverage probability）是指随机区间 $[L(\boldsymbol{X}), U(\boldsymbol{X})]$ 覆盖真实参数 θ 的概率. 在符号上它记作 P_θ（$\theta\in[L(\boldsymbol{X}), U(\boldsymbol{X})]$）或 P（$\theta\in[L(\boldsymbol{X}), U(\boldsymbol{X})]|\theta$）.

置信系数

定义 9.1.5 对于一个参数 θ 的区间估计量 $[L(\boldsymbol{X}), U(\boldsymbol{X})]$，$[L(\boldsymbol{X}), U(\boldsymbol{X})]$ 的置信系数（confidence coefficient）是指覆盖概率的下确界 $\inf_\theta P_\theta$（$\theta\in[L(\boldsymbol{X}), U(\boldsymbol{X})]$).

注意

1

从这些定义里我们认识到很多事情. 首先，一定牢记这个区间是随机的量，而参数不是，因此，当我们书写像 P_θ（$\theta\in[L(\boldsymbol{X}), U(\boldsymbol{X})]$）这样的概率陈述时，是针对 $\boldsymbol{X}$ 而非针对 θ 的. 换句话说，P_θ（$\theta\in[L(\boldsymbol{X}), U(\boldsymbol{X})]$）等同于 P_θ（$L(\boldsymbol{X})\leqslant\theta$，$U(\boldsymbol{X})\geqslant\theta$），而后者是一个关于 $\boldsymbol{X}$ 的陈述.

2

区间估计量，加之信心的一个量度（通常为置信系数），有时被称为一个置信区间（confidence interval）. 我们将经常把这个术语与区间估计量交替使用. 虽然我们主要关心置信区间，但是我们有时候也会处理更一般的集合. 当工作于一般情况，不能十分确信我们的集合的确切形式时，将使用置信集合（confidence set）这样的说法. 一个具有置信系数等于 $1-\alpha$ 的置信集合简称为一个 $1-\alpha$ 置信集合.

另外一个重点牵涉到覆盖概率与置信系数. 因为我们不知道 θ 的真值，所以我们只能保证一个覆盖的概率的下确界，即置信系数. 在某些情况这并不紧要，因为覆盖概率是 θ 的一常数函数. 而在其他情况，覆盖概率可能随 θ 不同而有很大变化.

例

例 9.1.6（均匀分布 尺度参数区间估计量） 设 X_1，…，X_n 是来自均匀分布总体 $U(0,\theta)$ 的随机样本并设 $Y=\max\{X_1,\cdots,X_n\}$. 我们对 θ 的区间估计感兴趣. 考虑两个候选估计量：$[aY, bY]$，$1\leqslant a<b$ 和 $[Y+c, Y+d]$，$0\leqslant c<d$，

- 求两个区间的置信系数

9.2 区间估计量的求法

9.2.1 反转一个检验统计量

例子

假设检验与区间估计有很强的对应关系. 事实上我们可以讲，一般每个置信集合对应一个检验，反之也对. 考虑下面的例子.

例 9.2.1（反转一个正态均值检验） 设 X_1，…，X_n 是 iid n (μ,σ^2) 的并考虑检验 $H_0:\mu=\mu_0$ 对 $H_1:\mu\neq\mu_0$. 对于固定的水平 α，一个合理的检验（事实上是最大功效无偏检验）具有拒绝区域 $\{\boldsymbol{x}: |\bar{x}-\mu_0|>z_{\alpha/2}\sigma/\sqrt{n}\}$. 注意到对于符合 $|\bar{x}-\mu_0|\leqslant z_{\alpha/2}\sigma/\sqrt{n}$ 或者等价地满足

$$\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\mu_0\leqslant\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

的样本点，H_0 则被接受.

因为这个检验具有真实水平 α，这意味着 $P(H_0$ 被拒绝 $|\mu=\mu_0)=\alpha$，或者换言之，$P(H_0$ 被接受 $|\mu=\mu_0)=1-\alpha$. 把它和上面接受区域描述结合起来，我们就能写出

$$P\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\mu_0\leqslant\overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\,\middle|\,\mu=\mu_0\right)=1-\alpha$$

但是这里对概率的陈述对于每一个 μ 都真. 因此陈述

$$P\left(\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)=1-\alpha$$

为真. 通过反转这个水平为 α 的检验的接受区域而获得的区间 $[\bar{x}-z_{\alpha/2}\sigma/\sqrt{n}, \bar{x}+z_{\alpha/2}\sigma/\sqrt{n}]$ 就是一个 $1-\alpha$ 置信区间. ||

置信集合与检验的对应

我们已经举例说明了置信集合与检验的对应，**样本空间**中使得 $H_0:\mu=\mu_0$ 被接受的集合由下式给出

$$A(\mu_0)=\left\{(x_1,\cdots,x_n):\mu_0-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\bar{x}\leqslant\mu_0+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

而置信区间是由**参数空间**中似乎可信的参数值构成的集合，由下式给出

$$C(x_1,\cdots,x_n)=\left\{\mu:\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

这两个集合通过等价关系

$$(x_1,\cdots,x_n)\in A(\mu_0)\Leftrightarrow \mu_0\in C(x_1,\cdots,x_n)$$

建立起联系.

关于双侧正态问题的检验与区间估计的对应画在了图 9.2.1 中. 此处也许更易看到检验和区间间的是同样的问题，不过是看问题的观点略有不同. 两个过程都寻找样本统计量与总体参数的一致. 假设检验是固定参数并询问什么样本值（接受区

域）与该固定值相符合．置信集合固定样本值并询问什么参数值（置信区间）使得这个样本值好像最有道理．

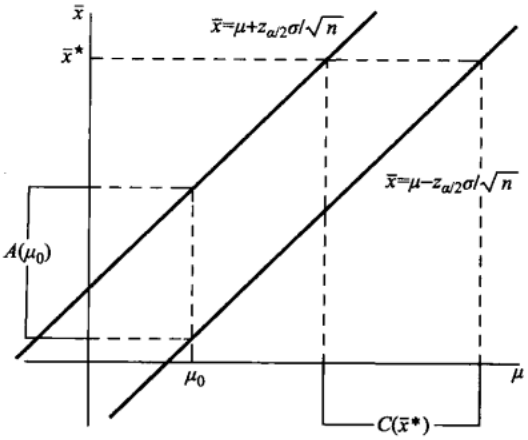

图 9.2.1　置信区间与检验的接受区域的关系．上面的线是 $\bar{x}=\mu+z_{\alpha/2}\sigma/\sqrt{n}$ 而下面的线是 $\bar{x}=\mu-z_{\alpha/2}\sigma/\sqrt{n}$.

检验的接受区域与置信集合之间的对应普遍成立．以下的定理给出这种对应的一个正式的说法．

定理9.2.2

定理 9.2.2　对每一个 $\theta_0\in\Theta$，设 $A(\theta_0)$ 是 $H_0: \theta=\theta_0$ 的一个水平为 α 的检验的接受区域．对每一个 $\boldsymbol{x}\in\mathcal{X}$，在参数空间里定义一个集合 $C(\boldsymbol{x})$

$$C(\boldsymbol{x})=\{\theta_0: \boldsymbol{x}\in A(\theta_0)\} \tag{9.2.1}$$

则随机集合 $C(\boldsymbol{x})$ 是一个 $1-\alpha$ 置信集合．反之，设 $C(\boldsymbol{x})$ 是一个 $1-\alpha$ 置信集合．对任意的 $\theta_0\in\Theta$，定义

$$A(\theta_0)=\{\boldsymbol{x}: \theta_0\in C(\boldsymbol{x})\}$$

则 $A(\theta_0)$ 是 $H_0: \theta=\theta_0$ 的一个水平为 α 的检验的接受区域．

证明　关于第一部分，因为 $A(\theta_0)$ 是一个水平为 α 的检验的接受区域，所以

$$P_{\theta_0}(\boldsymbol{X}\notin A(\theta_0))\leqslant\alpha \text{ 并且因此 } P_{\theta_0}(\boldsymbol{X}\in A(\theta_0))\geqslant 1-\alpha$$

由于 θ_0 是任意的，可以把 θ_0 改写成 θ．把上面的不等式与 (9.2.1) 合在一起，就证明了集合 $C(\boldsymbol{X})$ 的覆盖概率是

$$P_\theta(\theta\in C(\boldsymbol{X}))=P_\theta(\boldsymbol{X}\in A(\theta))\geqslant 1-\alpha$$

证明了 $C(\boldsymbol{X})$ 是一个 $1-\alpha$ 置信集合．

关于第二部分，对 $H_0: \theta=\theta_0$ 的以 $A(\theta_0)$ 作接受区域的检验，它犯第一类错误的概率是

$$P_{\theta_0}(\boldsymbol{X}\notin A(\theta_0))=P_{\theta_0}(\theta_0\notin C(\boldsymbol{X}))\leqslant\alpha$$

所以是一个水平为 α 的检验．

■

注意

虽然经常说是把一个检验反转获得一个置信集合，但是定理 9.2.2 表明我们实际上有一族检验，每一个值 $\theta_0\in\Theta$ 对应一个检验，置信集合是通过把这一族检验进行反转而得到的.

检验能够经转化得到置信集合且反之也对这个事实在理论上是有趣的，但是定理 9.2.2 在实际中有用的部分是它的第一部分. 构造一个水平为 α 的接受区域相对较容易，困难的是构造一个置信集合. 所以通过转化一个接受区域来获得一个置信集合的方法就相当有用. 我们所持有的获得检验的全部手法都能够立刻用来构造置信集合.

在定理 9.2.2 中我们提到原假设 $H_0:\theta=\theta_0$. 对接受区域的全部要求就是

$$P_{\theta_0}(\boldsymbol{X}\in A(\theta_0))\geqslant 1-\alpha$$

实际上，在通过反转检验构造一个置信集合时，我们在心中也会同时有了一个就像 $H_1:\theta\neq\theta_0$ 或 $H_1:\theta>\theta_0$ 的备择假设. 这个备择假设将规定 $A(\theta_0)$ 的合理形式，而 $A(\theta_0)$ 的形式将决定 $C(\boldsymbol{x})$ 的形状. 但是注意，我们非常小心地宁可使用集合这个词而不是区间这个词. 这是因为不能保证通过转化检验获得的置信集合是个区间. 不过多数情况下，单侧检验给出单侧区间，双侧检验给出双侧区间，奇怪形状的接受区域给出奇怪形状的置信集合. 后面的例子将展现这一点.

被反转的检验的性质也转而保留（有时适当修订）到置信集合上. 例如，转化无偏检验，将产生无偏的置信集合. 而且，更重要的是，因为我们知道在寻找一个好的检验时可以把注意力集中在充分统计量上，由此就可以推出，当我们在寻找一个好的置信集合时，也可以把注意力集中在充分统计量上.

在我们毫无直觉并且没有好的想法来组成一个合理的集合的情况之下，反转检验的方法确实最有帮助. 我们只需依靠通用的方法构建一个合理的检验.

例子

例 9.2.3（反转一个 LRT） 设我们希望求指数总体 EXPO（λ）的均值 λ 的一个置信区间. 我们可以通过反转 $H_0:\lambda=\lambda_0$ 对 $H_1:\lambda\neq\lambda_0$ 的一个水平为 α 的检验获得这样一个区间.

设我们取得一组随机样本 $X_1,\cdots,X_n$，则 LRT 检验统计量由下式给出

$$\frac{\frac{1}{\lambda_0^n}\mathrm{e}^{-\sum x_i/\lambda_0}}{\sup\limits_{\lambda}\frac{1}{\lambda^n}\mathrm{e}^{-\sum x_i/\lambda}}=\frac{\frac{1}{\lambda_0^n}\mathrm{e}^{-\sum x_i/\lambda_0}}{\frac{1}{(\sum x_i/n)^n}\mathrm{e}^{-n}}=\left(\frac{\sum x_i}{n\lambda_0}\right)^n\mathrm{e}^n\mathrm{e}^{-\sum x_i/\lambda_0}$$

对于固定的 λ_0，接受区域是

$$A(\lambda_0)=\left\{\boldsymbol{x}:\left(\frac{\sum x_i}{\lambda_0}\right)^n\mathrm{e}^{-\sum x_i/\lambda_0}\geqslant k^*\right\}\tag{9.2.2}$$

其中 k^* 是一个常数，它的选择满足 P_{λ_0}（$\boldsymbol{X}\in A$（λ_0））$=1-\alpha$（常数 e^n/n^n 被吸收进 k^*）. 这是如图 9.2.2 所示样本空间内的一个集合. 反转这个接受区域，就给出

$1-\alpha$ 置信集合

$$C(\boldsymbol{x})=\left\{\lambda:\left(\frac{\sum x_i}{\lambda}\right)^n \mathrm{e}^{-\sum x_i/\lambda}\geqslant k^*\right\}$$

这是如图 9.2.2 所示样本空间内的一个区间.

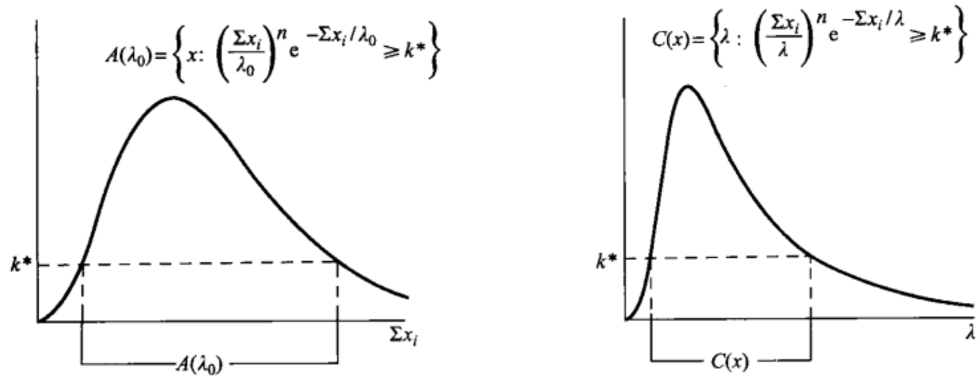

图 9.2.2　例 9.2.3 的接受区域与置信区间. 它的接受区域是

$$A(\lambda_0)=\left\{\boldsymbol{x}:\left(\sum_i x_i/\lambda_0\right)^n \mathrm{e}^{-\sum_i x_i/\lambda_0}\geqslant k^*\right\}$$

而置信区间是

$$C(\boldsymbol{x})=\left\{\lambda:\left(\sum_i x_i/\lambda\right)^n \mathrm{e}^{-\sum_i x_i/\lambda}\geqslant k^*\right\}$$

$C(\boldsymbol{x})$的定义表达式仅通过 $\sum x_i$ 依赖于 $\boldsymbol{x}$. 所以这个置信区间能够表示成如下形式

$$C\left(\sum x_i\right)=\left\{\lambda:L\left(\sum x_i\right)\leqslant\lambda\leqslant U\left(\sum x_i\right)\right\} \tag{9.2.3}$$

其中 L 和 U 是由集合（9.2.2）有 $1-\alpha$ 概率以及

$$\left[\frac{\sum x_i}{L\left(\sum x_i\right)}\right]^n \mathrm{e}^{-\sum x_i/L(\sum x_i)}=\left[\frac{\sum x_i}{U\left(\sum x_i\right)}\right]^n \mathrm{e}^{-\sum x_i/U(\sum x_i)} \tag{9.2.4}$$

这些限制条件决定的函数. 若我们设

$$\frac{\sum x_i}{L\left(\sum x_i\right)}=a \text{ 和 } \frac{\sum x_i}{U\left(\sum x_i\right)}=b \tag{9.2.5}$$

其中 $a>b$ 是常数，则式（9.2.4）变成

$$a^n \mathrm{e}^{-a}=b^n \mathrm{e}^{-b} \tag{9.2.6}$$

此方程易于数值求解. 为了解得详细一些，设 $n=2$ 并且注意 $\sum X_i/\lambda\sim\text{gamma}(2,1)$. 因此，由式（9.2.5），置信区间变成 $\left\{\lambda:\frac{1}{a}\sum x_i\leqslant\lambda\leqslant\frac{1}{b}\sum x_i\right\}$，其中 a 和 b 满足

$$P_\lambda\left(\frac{1}{a}\sum X_i \leqslant \lambda \leqslant \frac{1}{b}\sum X_i\right) = P\left\{b \leqslant \frac{\sum X_i}{\lambda} \leqslant a\right\} = 1-\alpha$$

而且，由式（9.2.6），$a^2 e^{-a} = b^2 e^{-b}$. 于是

$$P\left\{b \leqslant \frac{\sum X_i}{\lambda} \leqslant a\right\} = \int_b^a t e^{-t} \mathrm{d}t$$

$$= e^{-b}(b+1) - e^{-a}(a+1) \text{（分部积分）}$$

为了得到一个 90％置信区间，我们必须同时满足这个概率条件和限制条件. 计算到小数点后三位，我们得到 $a=5.480$，$b=0.441$，置信系数是 0.90006. 这样

$$P_\lambda\left(\frac{1}{5.480}\sum X_i \leqslant \lambda \leqslant \frac{1}{0.441}\sum X_i\right) = 0.90006$$

||

注意

通过反转检验 $H_0: \theta=\theta_0$ 对 $H_1: \theta\neq\theta_0$ 的 LRT（定义 8.2.1）得到的区域的形式为

$$\text{接受 } H_0 \text{ 如果} \frac{L(\theta_0|\boldsymbol{x})}{L(\hat{\theta}|\boldsymbol{x})} \geqslant k(\theta_0),$$

导致置信区域

(9.2.7) $$\{\theta: L(\theta|\boldsymbol{x}) \geqslant k'(\boldsymbol{x},\theta)\}$$

函数 k' 使得置信系数为 $1-\alpha$.

在某些情况（像正态分布和伽玛分布），函数 k' 将不依赖于 θ. 在这种情况，上述似然区域有一个令人愉快的解释，即它是由那些似然最高的 θ 的值组成的. 我们也将看到这种区间起因于频率论者（定理 9.3.2）和 Bayes 学派（推论 9.3.10）都有的最优化的考虑.

例

例 9.2.4（正态　单侧置信界）　设 X_1，…，X_n 是来自 N（μ，σ^2）总体的随机样本，考虑对 μ 构造一个 $1-\alpha$ 置信上界.

例子

例 9.2.5（二项　单侧置信界）　作为一个更为困难的单侧置信区间的例子，考虑给出一列伯努利试验成功概率 p 的一个 $1-\alpha$ 置信下界. 就是说，我们观测 X_1，…，X_n，其中 $X_i \sim \text{Bernoulli}(p)$，而我们想要求得一个形式为（$L(X_1, \cdots, X_n), 1]$ 的区间，其中 $P_p(p \in (L(X_1, \cdots, X_n), 1]) \geqslant 1-\alpha$.（我们将看到，获得的这个区间左边是开的.）因为我们想要一个给出置信下界的单侧区间，所以考虑反转来自关于

$$H_0: p=p_0 \text{ 对 } H_1: p>p_0$$

的检验的接受区域.

为了简化，我们知道可以用基于 $T=\sum_{i=1}^{n} X_i \sim \text{binomial}(n, p)$ 的检验，因为 T 关于 p 是充分的（参见杂录一节）. 因为二项分布具有单调似然比（见习题 8.25），根据 Karlin-Rubin 定理（定理 8.3.17），当 $T>k(p_0)$ 就拒绝 H_0 的检验是其真实水平上的 UMP 检验. 对于每个 p_0，我们想选择常数 $k(p_0)$（它可以是一个整数）以使我们持有一个水平为 α 的检验. 由于 T 的离散性，除去对于某些特定的 p_0 值，我们无法使检验的真实水平精确地是 α. 不过我们可以这样选择 $k(p_0)$，以使检验的真实水平尽可能靠近 α，而不比它大. 这样，$k(p_0)$ 就被定义为 0 和 n 之间的同时满足以下两个不等式的整数

$$\begin{aligned} &\sum_{y=0}^{k(p_0)} \binom{n}{y} p_0^y (1-p_0)^{n-y} \geqslant 1-\alpha \\ &\sum_{y=0}^{k(p_0)-1} \binom{n}{y} p_0^y (1-p_0)^{n-y} < 1-\alpha \end{aligned} \tag{9.2.8}$$

由于二项分布的 MLR 性质，对于任意的 $k=0, 1, \cdots, n$，

$$f(p_0 \mid k)=\sum_{y=0}^{k} \binom{n}{y} p_0^y (1-p_0)^{n-y}$$

是 p_0 的一个下降函数（见习题 8.26）. 当然地有 $f(0 \mid 0)=1$，所以 $k(0)=0$ 而且对于某一段区间上的 p_0 值 $f(p_0 \mid 0)$ 在 $1-\alpha$ 之上. 接着，有某点使得 $f(p_0 \mid 0)=1-\alpha$，对于大于这点的 p_0 的值，$f(p_0 \mid 0)<1-\alpha$. 于是在这点上，$k(p_0)$ 增加到 1. 这个模式继续下去. 这样，$k(p_0)$ 就是一个整数值阶梯函数，在 p_0 的一段范围内它是常数，然后跳到更大的一个整数. 因为 $k(p_0)$ 是 p_0 的非减的函数，

这样就给出置信下界（参见习题 9.5，它是关于置信上界的）. 求解式（9.2.8）中关于 $k(p_0)$ 的不等式组，就同时给出检验的接受区域和置信集合.

对每个 p_0，接受区域由 $A(p_0)=\{t: t\leqslant k(p_0)\}$ 给出，其中 $k(p_0)$ 满足式（9.2.8）. 对每个 t 的值，置信集合是 $C(t)=\{p_0: t\leqslant k(p_0)\}$. 然而这个集合现在的形式对于我们并不是很实用. 虽然它在形式上是正确的并且是一个 $1-\alpha$ 置信集合，但是它对于 p_0 是按照隐函数定义的而我们希望它被定义成对于 p_0 的显式.

因为 $k(p_0)$ 是非减的，所以对于一个给定的观测 $T=t$，在所有小于或等于某个值的 p_0 都满足 $k(p_0)<t$ 时，把这个值叫做 $k^{-1}(t)$. 在 $k^{-1}(t)$，$k(p_0)$ 上跳到 t 而对于所有的 $p_0>k^{-1}(t)$ 都有 $k(p_0)\geqslant t$.（注意在 $p_0=k^{-1}(t)$ 时，$f(p_0\mid t-1)=1-\alpha$. 所以式（9.2.8）对于 $k(p_0)=t-1$ 依然满足. 只有当 $p_0>k^{-1}(t)$ 时才有 $k(p_0)\geqslant t$.）这样，置信集合就是

$$C(t)=\{p_0: t\leqslant k(p_0)\}=\{p_0: p_0>k^{-1}(t)\} \tag{9.2.9}$$

从而我们就构造出了一个形式为 $C(T)=(k^{-1}(T), 1]$ 的 $1-\alpha$ 置信下界.

数 $k^{-1}(t)$ 可以定义为

$$k^{-1}(t)=\sup\left\{p: \sum_{y=0}^{t-1}\binom{n}{y}p^y(1-p)^{n-y}\geqslant 1-\alpha\right\} \tag{9.2.10}$$

因为 $k(p_0)$ 并不是一个一对一的函数，所以应认识到 $k^{-1}(t)$ 并不是 $k(p_0)$ 的一个真正的反函数. 但是，表达式（9.2.8）和（9.2.10）给予了我们关于 k 和 k^{-1} 的确切定义.

二项置信界问题是由 Clopper 和 Pearson（1934）首次论述的，他们得到的解答类似于双侧区间（参见习题 9.21），从此开辟了一条至今仍然活跃的研究线路. 参见杂录 9.5.2. ‖

9.2.2 枢轴量

我们在例 9.1.6 见到的两个置信区间在很多方面是有区别的. 一个重要区别就是区间 $[aY, bY]$ 的覆盖概率不依赖于参数 θ 的值，而 $[Y+c, Y+d]$ 则不然. 这是由于 $[aY, bY]$ 的覆盖概率能够经由量 Y/θ 来表示，而该随机变量其分布不依赖于参数，这个量就被称为枢轴量（pivotal quantity）或枢轴（pivot).

把枢轴量用于构造置信集合就导致了所谓枢轴推断（pivotal inference），这主要归功于 Barnard（1949，1980)，然而可以向上追溯到 Fisher（1930)，他使用了逆概率（inverse probability）术语. 与之密切相关的是 D. A. S. Fraser（1968，1979）的结构推断（structural inference）理论 . Berger 和 Wolpert（1984）对这些方法的长处与弱点做了有趣的讨论.

一些定义

定义 9.2.6 一个随机变量 $Q(\boldsymbol{X},\theta)=Q(X,\cdots,X_n,\theta)$ 是一个枢轴量或枢轴，如果 $Q(\boldsymbol{X},\theta)$ 的分布独立于所有的参数. 就是说，如果 $X\sim F(\boldsymbol{x}\mid\theta)$，则 $Q(\boldsymbol{X},\theta)$ 对于所有的 θ 值具有同样的分布.

函数 $Q(\boldsymbol{x},\theta)$ 通常会明显地同时包含参数与统计量，但是对任何集合 A，$P_\theta(Q(\boldsymbol{X},\theta)\in A)$ 不能依赖于 θ. 从枢轴构造置信集合的技术靠的是能求出一个枢轴与一个集合 A 使得集合 $\{\theta: Q(\boldsymbol{X},\theta)\in A\}$ 是 θ 的一个集估计.

例子

例 9.2.7（位置-尺度枢轴） 在位置和尺度情况里有很多枢轴量．我们在这里展示几个，更多的可在习题 9.8 找到．设 X，…，X_n 是来自一个指定的概率密度函数的随机样本，并设 $\overline{X}$ 和 S 是样本均值和样本标准差．为了证明表 9.2.1 中的量是枢轴，我们只需证明它们的概率密度函数与参数是无关的（细节在习题 9.9 里）．特别地，注意到当 X，…，X_n 是来自正态总体 n（μ，σ^2）的随机样本时，t 统计量（$\overline{X}-\mu$）/（$S/\sqrt{n}$）是一个枢轴，因为 t 分布不依赖于参数 μ 和 σ^2．

表 9.2.1 位置-尺度枢轴

pdf 的形式	pdf 的类型	枢轴量
$f(x-\mu)$	位置	$\overline{X}-\mu$
$\frac{1}{\sigma}f(\frac{x}{\sigma})$	尺度	$\frac{\overline{X}}{\sigma}$
$\frac{1}{\sigma}f(\frac{x-\mu}{\sigma})$	位置-尺度	$\frac{\overline{X}-\mu}{S}$

在 9.2.1 节使用检验反转法构造的那些区间当中，有些实际上是基于枢轴的（例 9.2.3 和例 9.2.4），有些则不是（例 9.2.5）．没有通用的求枢轴的策略，但是我们可以略微聪明一些而不是完全依靠猜测．例如，求出位置或尺度参数的枢轴就是相对容易的事情．一般讲，差是位置问题的枢轴而比（或乘积）是尺度问题的枢轴．

例子

例 9.2.8（伽玛分布 枢轴） 设 X，…，X_n 是指数分布 EXPO（λ）的 iid 样本，则 $T=\sum X_i$ 是关于 λ 的充分统计量并且 $T\sim$gamma（n，λ）．在伽玛分布的概率密度函数中 t 和 λ 以 t/λ 的形式一起出现并且，事实上 gamma（n，λ）的概率密度函数（$\Gamma(n)\ \lambda^n)^{-1}t^{n-1}e^{-t/\lambda}$是一个尺度族．这样，如果 $Q(T，\lambda)=2T/\lambda$，则

$$Q(T,\lambda)\sim \text{gamma}(n,\lambda(2/\lambda))=\text{gamma}(n,2)$$

它不依赖于 λ．所以，量 Q（T，λ）$=2T/\lambda$ 是一个枢轴，服从 gamma（n，2）分布，或者说 χ^2_{2n}分布．

有时我们能够通过观察概率密度函数的形式看出是否存在枢轴．在上例中，量 t/λ 出现在概率密度函数里并且它实际上就是一个枢轴．在正态概率密度函数中，有量（$\bar{x}-\mu$）/σ 出现并且这个量也是一个枢轴．一般，设一个统计量 T 的概率密

度函数 $f(t\mid\theta)$ 能够表示成如下形式

$$f(t|\theta)=g(Q(t,\theta))\left|\frac{\partial}{\partial t}Q(t,\theta)\right| \tag{9.2.11}$$

其中 g 是某个函数而 Q 是某个单调（对于每个 t，关于 θ 单调）函数. 则可以利用定理 2.1.5 证明 $Q(T,\theta)$ 是一个枢轴（见习题 9.10）.

一旦我们有了一个枢轴，我们怎样利用它来构造一个置信集合？实际上这就相当简单了. 如果 $Q(\boldsymbol{X},\theta)$ 是一个枢轴，则对于一个指定的 α 值，我们能够求出数 a 和 b，它们不依赖于 θ，满足

$$P_\theta(a\leqslant Q(\boldsymbol{X},\theta)\leqslant b)\geqslant 1-\alpha$$

则对于每个 $\theta_0\in\Theta$，

$$A(\theta_0)=\{\boldsymbol{x}: a\leqslant Q(\boldsymbol{x},\theta_0)\leqslant b\} \tag{9.2.12}$$

就是关于 $H_0:\theta=\theta_0$ 的一个水平为 α 的检验的接受区域. 我们将用检验反转法构造置信集合，但现在用枢轴指出了接受区域的具体形式. 利用定理 9.2.2，反转这些检验而得到

$$C(\boldsymbol{x})=\{\theta_0: a\leqslant Q(\boldsymbol{x},\theta_0)\leqslant b\} \tag{9.2.13}$$

并且 $C(\boldsymbol{X})$ 是关于 θ 的一个 $1-\alpha$ 置信集合. 如果 θ 是一个实值参数并且对于每个 $\boldsymbol{x}\in\mathcal{X}$，$Q(\boldsymbol{x},\theta)$ 是 θ 的一个单调函数，则 $C(\boldsymbol{x})$ 将是一个区间. 事实上，如果 $Q(\boldsymbol{x},\theta)$ 是 θ 的一个增函数，则 $C(\boldsymbol{x})$ 具有 $L(\boldsymbol{x},a)\leqslant\theta\leqslant U(\boldsymbol{x},b)$ 的形式. 如果 $Q(\boldsymbol{x},\theta)$ 是 θ 的一个减函数（这是典型的），则 $C(\boldsymbol{x})$ 具有 $L(\boldsymbol{x},b)\leqslant\theta\leqslant U(\boldsymbol{x},a)$ 的形式.

例

- 如果$X_1,X_2,\dots,X_n$是iid的$n(\mu,\sigma^2)$的，找出μ一个枢轴，并利用该枢轴考虑σ已知和未知时，μ的置信区间

9.2.3 枢轴化累积分布函数

在上节我们看到一个枢轴 Q 导致如式（9.2.13）形式的置信集合，即

$$C(\boldsymbol{x})=\{\theta_0: a\leqslant Q(\boldsymbol{x},\theta_0)\leqslant b\}$$

如果对于每个 $\boldsymbol{x}$，函数 $Q(\boldsymbol{x},\theta)$ 是 θ 的一个单调函数，则可保证置信集合 $C(\boldsymbol{x})$ 是一个区间. 迄今为止，我们见过的枢轴主要是用位置和尺度变换做成的，从而导致单调的函数 Q 并且因此得到置信区间.

在这一节我们来处理另一种枢轴，它是十分一般的并且附带较少的假定而能保证得到区间.

在拿不准的或者陌生的情况下，我们建议在可能的时候构造一个基于反转 LRT 的置信集合. 这样的一个集合虽然不能保证它是最优的，但是决不会很坏. 但是在某些情况下这个方法十分困难，这种困难是解析上的或计算上的；有时对接受区域的反转相当烦琐. 而若可以运用这一节的方法，计算就相当直接并且常会生成一个合理的集合.

定理9.2.12

定理 9.2.12（枢轴化一个连续型累积分布函数） 设 T 是一个以 $F_T(t\mid\theta)$ 为其累积分布函数的连续型统计量. 设 $\alpha_1+\alpha_2=\alpha$ 其中 $0<\alpha<1$ 是固定值. 假定对于每个 $t\in T$，函数 $\theta_L(t)$ 和 $\theta_U(t)$ 可以被如下定义.

i. 如果对于每个 t，$F_T(t\mid\theta)$ 都是 θ 的一个减函数，则由

$$F_T(t|\theta_U(t))=\alpha_1,F_T(t|\theta_L(t))=1-\alpha_2$$

定义 $\theta_L(t)$ 和 $\theta_U(t)$.

ii. 如果对于每个 t，$F_T(t\mid\theta)$ 都是 θ 的一个增函数，则由

$$F_T(t|\theta_U(t))=1-\alpha_2,F_T(t|\theta_L(t))=\alpha_1$$

定义 $\theta_L(t)$ 和 $\theta_U(t)$.

那么随机区间 $[\theta_L(T),\theta_U(T)]$ 是 θ 的一个 $1-\alpha$ 置信区间.

使用此方法需注意

1

我们注意到在缺少附加信息的情况，通常选择 $\alpha_1=\alpha_2=\alpha/2$. 虽然这也许不是最优选择（见定理 9.3.2），但是在大多数情况它肯定是一个合理的策略. 如果要得到一个单侧区间，选取 α_1 或 α_2 等于 0 即可.

在随机递增的情况，等式

$$F_T(t|\theta_U(t))=\alpha_1,\quad F_T(t|\theta_L(t))=1-\alpha_2 \tag{9.2.15}$$

也能经由统计量 T 的概率密度函数来表达. $\theta_U(t)$ 和 $\theta_L(t)$ 可以被定义为满足

$$\int_{-\infty}^{t} f_T(u\mid\theta_U(t))\mathrm{d}u=\alpha_1 \text{ 和} \int_{t}^{\infty} f_T(u\mid\theta_L(t))\mathrm{d}u=\alpha_2$$

对于随机递减的情况，则有一组类似等式成立.

2

使用这个方法有两点要注意. 首先，只有在统计量的值实际观测到才需要求解实际的方程组（9.2.15）. 如果观测到 $T=t_0$，则 θ 的实际置信区间就是 $[\theta_L(t_0),\theta_U(t_0)]$. 这样，我们只需要求解关于 $\theta_L(t_0)$ 和 $\theta_U(t_0)$ 的两个方程

$$\int_{-\infty}^{t_0} f_T(u\mid\theta_u(t_0))\mathrm{d}u=\alpha_1 \text{ 和} \int_{t_0}^{\infty} f_T(u\mid\theta_L(t_0))\mathrm{d}u=1-\alpha_2$$

其次，要认识到即使这些方程不能被解析解出，由于在证明我们有一个 $1-\alpha$ 置信区间的时候并未要求一个解析的解，所以我们只需用数值方法求解.

例子

例 9.2.13（位置指数区间） 这个方法可以用于获得位置指数概率密度函数的一个置信区间.（在习题 9.25 把此处的答案与似然法和枢轴法进行比较. 也见习题 9.41）

如果 X_1，…，X_n 是 iid 的，具有概率密度函数 $f(x|\mu)=\mathrm{e}^{-(x-\mu)}I_{[\mu,\infty)}(x)$，则 $Y=\min\{X_1, \cdots, X_n\}$ 关于 μ 是充分的并具有概率密度函数

$$f_Y(y|\mu)=n\mathrm{e}^{-n(y-\mu)}I_{[\mu,\infty)}(y)$$

固定 α 并定义 $\mu_L(y)$ 和 $\mu_U(y)$ 满足

$$\int_{\mu_U(y)}^{y} n\mathrm{e}^{-n(u-\mu_U(y))}\mathrm{d}u=\frac{\alpha}{2}, \int_{y}^{\infty} n\mathrm{e}^{-n(u-\mu_L(y))}\mathrm{d}u=\frac{\alpha}{2}$$

计算这些积分就可以得到方程组

$$1-\mathrm{e}^{-n(y-\mu_U(y))}=\frac{\alpha}{2}, \quad \mathrm{e}^{-n(y-\mu_L(y))}=\frac{\alpha}{2}$$

其解为

$$\mu_U(y)=y+\frac{1}{n}\log\left(1-\frac{\alpha}{2}\right), \quad \mu_L(y)=y+\frac{1}{n}\log\left(\frac{\alpha}{2}\right)$$

因此随机区间

$$C(Y)=\left\{\mu: Y+\frac{1}{n}\log\left(\frac{\alpha}{2}\right)\leqslant\mu\leqslant Y+\frac{1}{n}\log\left(1-\frac{\alpha}{2}\right)\right\}$$

是关于 μ 的一个 $1-\alpha$ 置信区间. ||

定理9.2.14

定理 9.2.14（枢轴化一个离散型累积分布函数） 设 T 是一个以 $F_T(t|\theta)=P(T\leqslant t|\theta)$ 为其累积分布函数的离散型统计量. 设 $\alpha_1+\alpha_2=\alpha$ 其中 $0<\alpha<1$ 是固定值. 假定对于每个 $t\in\mathcal{T}$，函数 $\theta_L(t)$ 和 $\theta_U(t)$ 可以被如下定义.

i. 如果对于每个 t，$F_T(t|\theta)$ 都是 θ 的一个减函数，则由

$$P(T\leqslant t|\theta_U(t))=\alpha_1, \quad P(T\geqslant t|\theta_L(t))=\alpha_2$$

定义 $\theta_L(t)$ 和 $\theta_U(t)$.

ii. 如果对于每个 t，$F_T(t|\theta)$ 都是 θ 的一个增函数，则由

$$P(T\geqslant t|\theta_U(t))=\alpha_2, \quad P(T\leqslant t|\theta_L(t))=\alpha_1$$

定义 $\theta_L(t)$ 和 $\theta_U(t)$.

那么随机区间 $[\theta_L(T), \theta_U(T)]$ 是关于 θ 的一个 $1-\alpha$ 置信区间.

例子

例 9.2.15（Poisson 区间估计量） 设 X，…，X_n 是来自参数为 λ 的 Poisson 总体的随机样本并且定义 $Y=\sum X_i$，则 Y 关于 λ 是充分的且 $Y\sim$ Poisson $(n\lambda)$. 应用上面的方法取 $\alpha_1=\alpha_2=\alpha/2$，如果观测到 $Y=y_0$，则求解关于 λ 的方程组

$$(9.2.16)\qquad \sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!}=\frac{\alpha}{2}\text{ 与 }\sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!}=\frac{\alpha}{2}$$

回忆来自例 3.3.1 的联系 Poisson 与伽玛分布族的恒等式. 把这个恒等式应用于式 (9.2.16) 里的和式，我们就可以写出（记住 y_0 是 Y 的观测到的值）

$$\frac{\alpha}{2}=\sum_{k=0}^{y_0} e^{-n\lambda}\frac{(n\lambda)^k}{k!}=P(Y\leqslant y_0\mid\lambda)=P(\chi^2_{2(y_0+1)}>2n\lambda)$$

其中 $\chi^2_{2(y_0+1)}$ 是一个自由度是 $2(y_0+1)$ 的 χ^2 分布随机变量. 这样，以上方程的解为

$$\lambda=\frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}$$

类似地，把这个恒等式应用于式 (9.2.16) 里的另一个方程就得到

$$\frac{\alpha}{2}=\sum_{k=y_0}^{\infty} e^{-n\lambda}\frac{(n\lambda)^k}{k!}=P(Y\geqslant y_0\mid\lambda)=P(\chi^2_{2y_0}<2n\lambda)$$

做一些代数计算，我们就得到关于 λ 的 $1-\alpha$ 置信区间为

$$(9.2.17)\qquad \left\{\lambda:\frac{1}{2n}\chi^2_{2y_0,1-\alpha/2}\leqslant\lambda\leqslant\frac{1}{2n}\chi^2_{2(y_0+1),\alpha/2}\right\}$$

（在 $y_0=0$ 我们定义 $\chi^2_{0,1-\alpha/2}=0$.）

这种区间首次由 Garwood (1936) 得到. 图 9.2.5 给出了覆盖概率的一张图. 注意这个图形相当地参差不齐. 跳跃出现在不同置信区间的端点，此处组成覆盖概率的求和项被加入或减掉（见习题 9.24）.

作为一个数值的例子，考虑 $n=10$ 并且观测到 $y_0=\sum x_i=6$. 一个关于 λ 的 90%置信区间是

$$\frac{1}{20}\chi^2_{12,0.95}\leqslant\lambda\leqslant\frac{1}{20}\chi^2_{14,0.05}$$

9.2.4 Bayes区间

到现在，在描述置信区间与参数的相互关系时，我们小心地讲区间覆盖参数而不讲参数在区间里. 这样做是有意的. 我们希望强调随机量是区间而不是参数. 因此，我们试图让行为动词应用于区间而不是参数.

在例 9.2.15 我们看到如果 $y_0=\sum_{i=1}^{10}x_i=6$，则 $0.262\leqslant\lambda\leqslant1.184$ 是一个关于 λ 的 90%置信区间. 这就引诱人们去说（并且很多试验者确实就说）“λ 在区间 [0.262,

1.184] 里的概率是 90%.”但是在经典统计学里这句话是无效的，因为参数 λ 被设想为固定的. 正式讲，[0.262，1.184] 是随机区间$\left[\frac{1}{2n}\chi^2_{2Y,0.95}\leqslant\lambda\leqslant\frac{1}{2n}\chi^2_{2(Y+1),0.05}\right]$的可能实现值之一，并且因为参数 λ 不能移动，所以 λ 在已实现区间 [0.262，1.184] 里的概率是 0 或 1. 当我们讲已实现区间 [0.262，1.184] 有 90%的覆盖机会的时候，只意味着我们知道随机区间的样本点中 90%覆盖真实参数.

与此形成对比的是，Bayes 体制允许我们讲 λ 在 [0.262，1.184] 里的概率，可以不是 0 或 1. 这是因为在 Bayes 模型中，λ 是一个有概率分布的随机变量. 所有 Bayes 的关于覆盖的断言都是关于参数的后验分布来说的.

Bayes 集合和经典的集合有相当不同的概率解释. 为了把它们区别开来，称 Bayes 估计集合为可信集合（credible sets）而不是置信集合.

这样，若 $\pi(\theta|\boldsymbol{x})$ 是给定 $\boldsymbol{X}=\boldsymbol{x}$ 条件下 θ 的后验分布，则对任意的集合 $A\subset\Theta$，A 的可信概率就是

$$P(\theta\in A|\boldsymbol{x})=\int_A\pi(\theta|\boldsymbol{x})\mathrm{d}\theta \tag{9.2.18}$$

而 A 是关于 θ 的一个可信集合. 如果 $\pi(\theta|\boldsymbol{x})$是一个概率质量函数，则我们把以上表达式中的积分换成求和.

注意，Bayes 可信集合的解释和构造都比经典的置信集合直截了当. 但是记住，没有什么是免费的. 解释和构造的省力伴随着要附加假定. Bayes 模型比经典模型要求更多的输入.

9.3 区间估计量的评价方法

我们现在已经见到许多导出置信集合的方法，而且实际上对于相同的问题我们能够导出不同的置信集合. 在这种情况下我们当然想选择一个最佳的. 因此，现在我们来考察一些旨在评价集合估计量所用的方法与标准.

集合估计量有两个互相对立竞争的量，就是尺寸和覆盖概率. 自然，我们希望我们的集合具有小的尺寸和大的覆盖概率，但是这样的集合通常难以构造.（显然，我们可以通过增加集合的尺寸获取大的覆盖概率. 区间（$-\infty$，∞）的覆盖概率是1!）在我们对于尺寸和覆盖概率来最优化一个集合之前，我们必须决定怎样度量这些量.

一个置信集合的覆盖概率除特殊情况之外是参数的一个函数，所以要考虑的不是一个而是无穷个值. 然而大多数情况我们将通过置信系数（confidence coefficient），即覆盖概率的下确界，去度量覆盖概率性能. 这是一种方式，但不是唯一的总括覆盖概率信息的可用方式.（例如，我们可以计算平均覆盖概率.）

当我们说到一个置信集合的尺寸，如果置信集合是一个区间，我们通常意指置信集合的长度. 如果这个集合不是一个区间，或者我们在处理一个多维集合，则长度一般改为体积.（也有这样的情况：其中长度以外的尺寸度量是自然的，特别是把同变性当作一种考虑的时候. 在 Schervish 1995 的第 6 章和 Berger 1985 的第 6 章论述了这个题目.）

9.3.1 尺寸和覆盖概率

- 对一个给定的覆盖概率求具有最短长度的置信区间

举个例子

例 9.3.1（最优化长度） 设 X，…，X_n 是 iid n（μ，σ^2）的，其中 σ 已知. 根据 9.2.2 节的方法以及事实上

$$Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$$

是一个具有标准正态分布的枢轴，所以任何满足

$$P(a\leqslant Z\leqslant b)=1-\alpha$$

的 a 和 b 将给出置信区间

$$\left\{\mu:\bar{x}-b\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\bar{x}-a\frac{\sigma}{\sqrt{n}}\right\}$$

哪种 a 和 b 的选择最优？更正式地讲，a 和 b 的什么选择将在保持 $1-\alpha$ 覆盖的情况下最小化置信区间的长度？注意到置信区间的长度等于（$b-a$)$\sigma/\sqrt{n}$，但是由于因子 $\sigma/\sqrt{n}$ 出现在每个区间的长度中，所以可以不考虑它而把长度的比较基于 $b-a$ 的值. 这样，我们想求得一对数 a 和 b 满足 $P(a\leqslant Z\leqslant b)=1-\alpha$ 并且 $b-a$ 最小.

在例 9.2.1 中我们取 $a=-z_{\alpha/2}$ 和 $b=z_{\alpha/2}$，但是没有提到最优性. 如果我们取 $1-\alpha=0.90$，则下表 9.3.1 中的任何一对数都给出了 90%区间：

表 9.3.1 三个 90%正态置信区间

a	b	概　率	$b-a$
-1.34	2.33	$P(Z<a)=0.09$，$P(Z>b)=0.01$	3.67
-1.44	1.96	$P(Z<a)=0.075$，$P(Z>b)=0.025$	3.40
-1.65	1.65	$P(Z<a)=0.05$，$P(Z>b)=0.05$	3.30

这个数值研究建议选择 $a=-1.65$ 和 $b=1.65$ 给出最优的区间，并且确实如此. 在这个情况下等分概率 α 是一个最优的策略.

注意

等分概率 α 的策略在上面例子中是最优的，但不总是最优的. 上例中等分 α 之所以成为最优是因为在 $-z_{\alpha/2}$ 和 $z_{\alpha/2}$ 处概率密度函数的高度相同. 我们现在证明一个定理，从而论证这个事实. 该定理可以在较一般的情况下使用，它只需要假定概率密度函数是单峰的（unimodal). 回忆单峰的定义：一个概率密度函数 $f(x)$ 是单峰的，如果存在 x^* 使得 $f(x)$ 在 $x\leqslant x^*$ 非减而 $f(x)$ 在 $x\geqslant x^*$ 非增.（这是一个相当弱的要求.）

定理9.3.2

定理 9.3.2 设 $f(x)$ 是一个单峰的概率密度函数. 如果区间 $[a, b]$ 满足

i. $\int_a^b f(x)\mathrm{d}x = 1-\alpha$,

ii. $f(a)=f(b)>0$,

iii. $a \leqslant x^* \leqslant b$,其中 x^* 是 $f(x)$ 的一个众数(mode),

则 $[a, b]$ 是所有满足(i)的区间中最短的.

证明 设 $[a', b']$ 是任意的一个使 $b'-a'<b-a$ 的区间. 我们将证明这蕴涵 $\int_{a'}^{b'} f(x)\mathrm{d}x < 1-\alpha$. 仅就 $a' \leqslant a$ 去证明结论,如果 $a<a'$,证明是类似的. 此外,需要考虑 $b' \leqslant a$ 和 $b'>a$ 两种情况.

如果 $b' \leqslant a$,则 $a' \leqslant b' \leqslant a \leqslant x^*$ 并且有

$$\begin{aligned}
\int_{a'}^{b'} f(x)\mathrm{d}x &\leqslant f(b')(b'-a') && (x \leqslant b' \leqslant x^* \Rightarrow f(x) \leqslant f(b')) \\
&\leqslant f(a)(b'-a') && (b' \leqslant a \leqslant x^* \Rightarrow f(b') \leqslant f(a)) \\
&< f(a)(b-a) && (b'-a' < b-a \text{ 而且 } f(a)>0) \\
&\leqslant \int_a^b f(x)\mathrm{d}x && \left(\begin{array}{l}\text{(i),(ii) 以及单峰性} \\ \Rightarrow f(x) \geqslant f(a) \text{ 对 } a \leqslant x \leqslant b\end{array}\right) \\
&= 1-\alpha && \text{由 (i)}
\end{aligned}$$

这就证完了第一种情况.

如果 $b'>a$,则 $a' \leqslant a < b' < b$,这是因为如果 $b' \geqslant b$,则就要 $b'-a' \geqslant b-a$. 在这种情况,我们可以把积分写成

$$\begin{aligned}
\int_{a'}^{b'} f(x)\mathrm{d}x &= \int_a^b f(x)\mathrm{d}x + \left[\int_{a'}^{a} f(x)\mathrm{d}x - \int_{b'}^{b} f(x)\mathrm{d}x\right] \\
&= 1-\alpha + \left[\int_{a'}^{a} f(x)\mathrm{d}x - \int_{b'}^{b} f(x)\mathrm{d}x\right]
\end{aligned}$$

如果我们能证出方括号里的表达式为负则定理就证出了. 现在利用 f 的单峰性,$a' \leqslant a < b' < b$ 的次序以及条件(ii),我们就有

$$\int_{a'}^{a} f(x)\mathrm{d}x \leqslant f(a)(a-a')$$

和

$$\int_{b'}^{b} f(x)\mathrm{d}x \geqslant f(b)(b-b')$$

于是

$$\begin{aligned}
\int_{a'}^{a} f(x)\mathrm{d}x - \int_{b'}^{b} f(x)\mathrm{d}x &\leqslant f(a)(a-a') - f(b)(b-b') \\
&= f(a)[(a-a')-(b-b')] \qquad (f(a)=f(b)) \\
&= f(a)[(b'-a')-(b-a)]
\end{aligned}$$

如果 $(b'-a')<(b-a)$ 且 $f(a)>0$ 则此式为负. ‖

9.3.2 与检验相关的最优性

- 置信集合的与检验相关的最优性性质并不直接涉及这个集合的尺寸,而是涉及到它覆盖假值的概率

覆盖假值的概率，或假值覆盖概率（probability of false coverage）间接地度量一个置信集合的尺寸. 直观地看，较小的集合覆盖较少的值，因此较少可能覆盖假值. 而且我们后面将看到一个连接尺寸与假值覆盖概率的方程.

我们首先考虑一般情况，这里 $\boldsymbol{X}\sim f(\boldsymbol{x}\mid\theta)$，并且通过反转接受区域 $A(\theta)$ 来构造一个对于 θ 的置信集合 $C(\boldsymbol{x})$. $C(\boldsymbol{x})$ 的覆盖概率，即真值覆盖概率是由 $P_\theta(\theta\in C(\boldsymbol{X}))$ 给出的 θ 的函数. 假值覆盖概率是 θ 和 θ' 的函数，它定义为当 θ 为真值时，覆盖 θ' 的概率

$$
\begin{aligned}
&P_\theta(\theta'\in C(\boldsymbol{X})),\ \theta\neq\theta',\ \text{若 } C(\boldsymbol{X})=[L(\boldsymbol{X}),U(\boldsymbol{X})]\\
&P_\theta(\theta'\in C(\boldsymbol{X})),\ \theta'<\theta,\ \text{若 } C(\boldsymbol{X})=[L(\boldsymbol{X}),\infty)\\
&P_\theta(\theta'\in C(\boldsymbol{X})),\ \theta'>\theta,\ \text{若 } C(\boldsymbol{X})=[-\infty,U(\boldsymbol{X})]
\end{aligned}
\tag{9.3.2}
$$

分别地对于单侧和双侧区间定义假值覆盖概率是有意义的. 例如，若我们有一个置信下界，就是说肯定 θ 比一个值大，于是覆盖假值只有在我们的区间覆盖了过小的 θ 值的情况下才发生. 类似的论证引导我们给出用于置信上界与双侧置信界的假值覆盖概率定义.

一个在一类 $1-\alpha$ 置信集合上最小化假值覆盖概率的 $1-\alpha$ 置信集合叫做一致最精确（uniformly most accurate，UMA）置信集合. 例如，我们可以考虑在形如 $[L(\boldsymbol{x}),\infty)$ 的集合中寻找一个 UMA 置信集合. 下面我们将要证明，UMA 置信集合是通过反转 UMP 检验的接受区域来构造的. 遗憾的是，虽然 UMA 置信集合是一个理想的集合，但是它仅在（就像做 UMP 检验）相当稀少的情况下才存在. 特别地，因为 UMP 检验一般是单侧的，所以 UMA 区间也是这样. 然而它们在理论上是优美的，从下面的定理我们就会看到 $H_0:\theta=\theta_0$ 对 $H_1:\theta>\theta_0$ 的一个 UMP 检验产生一个 UMA 置信下界.

定理9.3.5

定理 9.3.5 设 $\boldsymbol{X}\sim f(\boldsymbol{x}\mid\theta)$，其中 θ 是一个实值参数. 对于每个 $\theta_0\in\Theta$，设 $A^*(\theta_0)$是关于 $H_0:\theta=\theta_0$ 对 $H_1:\theta>\theta_0$ 的一个 UMP 水平 α 检验的接受区域. 设 $C^*(\boldsymbol{x})$是通过反转上述 UMP 接受区域所建立的 $1-\alpha$ 置信集合. 则对于任何其他的 $1-\alpha$ 置信集合 C，有

$$P_\theta(\theta'\in C^*(\boldsymbol{X}))\leqslant P_\theta(\theta'\in C(\boldsymbol{X}))\text{对于所有的 }\theta'<\theta\text{ 成立}$$

举个例子

例 9.3.6（UMA 置信界） 设 $X_1,\cdots,X_n$ 是 iid n (μ,σ^2) 的，其中 σ^2已知. 区间

$$C(\bar{x})=\left\{\mu:\mu\geqslant\bar{x}-z_\alpha\frac{\sigma}{\sqrt{n}}\right\}$$

是一个 $1-\alpha$ UMA 置信下界，这是因为它可以通过反转关于假设 $H_0:\mu=\mu_0$ 对 $H_1:\mu>\mu_0$ 的 UMP 检验获得.

更常见的双侧区间

$$C(\bar{x})=\left\{\mu:\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\leqslant\mu\leqslant\bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\}$$

不是 UMA 的，因为它是通过转化检验 $H_0:\mu=\mu_0$ 对 $H_1:\mu\neq\mu_0$ 的双侧接受区域得来的，而关于它的假设不存在 UMP 检验. ||

定义9.3.7

在检验问题当中，当考虑双侧检验时，我们发现无偏性性质既有说服力也很有用. 在置信区间问题中，应用相似的想法. 当我们处理双侧置信区间的时候，有理由把考虑的范围限制在无偏置信集合上面. 记得一个无偏检验就是它在备择假设的功效总比在原假设的功效大. 在读下面定义时请记住这点.

定义 9.3.7 一个 $1-\alpha$ 置信集合 $C(\boldsymbol{x})$ 是无偏的（unbiased），如果 $P_\theta(\theta'\in C(\boldsymbol{X}))\leqslant 1-\alpha$ 对于所有的 $\theta\neq\theta'$ 成立.

这样，对于一个无偏的置信集合，假值覆盖概率决不会大于最小的真值覆盖概率. 可以通过反转无偏检验得到无偏置信集合. 就是说，如果设 $A(\theta_0)$ 是关于 H_0：$\theta=\theta_0$ 对 H_1：$\theta\neq\theta_0$ 的一个无偏的水平 α 检验的接受区域，而 $C(\boldsymbol{x})$ 是由反转该接受区域得到的 $1-\alpha$ 置信集合，则 $C(\boldsymbol{x})$ 就是一个无偏的 $1-\alpha$ 置信集合（见习题 9.46）. ||

定理9.3.9

假值覆盖概率达到最小的集合也叫做 Neyman 最短的（Neyman-shortest）. 这个名称有长度的内涵，以下的定理表明这个称呼有些理由，该定理归于 Pratt (1961).

定理 9.3.9 (Pratt) 设 X 是一个实值随机变量，$X\sim f(x|\theta)$，其中 θ 是一个实值参数. 设 $C(x)=[L(x),U(x)]$ 是一个关于 θ 的置信区间. 如果 $L(x)$ 和 $U(x)$ 都是 x 的增函数，则对于任意的值 θ^*，有

$$E_{\theta^*}(\text{Length}[C(X)])=\int_{\theta\neq\theta^*} P_{\theta^*}(\theta\in C(X))\mathrm{d}\theta \tag{9.3.3}$$

定理 9.3.9 讲的是 C（X）的期望长度等于假值覆盖概率的求和（积分），积分域取遍参数的所有假值.

定理 9.3.9 表明一个置信区间的长度与它的假值覆盖概率有关. 在双侧的情况，这意味着极小化假值覆盖概率带有长度最优的某些保证. 然而在单侧情况，这样说就不行了. 这时，那种极小化假值覆盖概率的区间牵涉的参数仅是参数空间的一部分而长度最优可能得不到. Madansky（1962）给出一个例子，对于其中的 $1-\alpha$UMA 区间（单侧），能构造出一个更短的 $1-\alpha$ 区间（见习题 9.45）. 另外，Maatta 和 Casella（1987）证明了通过反转一个 UMP 检验得到的区间在使用其他合理的准则衡量时可能是次优的.

9.3.3 Bayes最优

获得具有指定覆盖概率的最小置信集合的目标也能利用 Bayes 准则达到．如果我们有一个后验分布 $\pi(\theta\mid\boldsymbol{x})$，即给定 $\boldsymbol{X}=\boldsymbol{x}$ 时 θ 的后验分布，而我们欲求集合 $C(\mathrm{x})$，满足

(i) $\int_{C(\boldsymbol{x})}\pi(\theta\mid\boldsymbol{x})\mathrm{d}\theta=1-\alpha$（原文中积分变元为 $\boldsymbol{x}$，译者注）

(ii) 尺寸$(C(x))\leqslant$尺寸$(C'(\boldsymbol{x}))$

对于任何满足 $\int_{C'(\boldsymbol{x})}\pi(\theta\mid\boldsymbol{x})\mathrm{d}\theta\geqslant1-\alpha$ 的集合 $C'(\boldsymbol{x})$ 成立．

如果用长度当作尺寸大小的测度，则我们可以应用定理 9.3.2 得到以下的结果．

推论 9.3.10 如果后验密度 $\pi(\theta\mid\boldsymbol{x})$ 是单峰的，则对于一个给定的 α 值，关于 θ 的最短可信区间由

$$\{\theta:\pi(\theta\mid\boldsymbol{x})\geqslant k\}\text{其中}\int_{\{\theta:\pi(\theta\mid\boldsymbol{x})\geqslant k\}}\pi(\theta\mid\boldsymbol{x})\mathrm{d}\theta=1-\alpha$$

给出．

推论 9.3.10 中给出的可信集合叫做最高后验密度（highest posterior density，HPD）区域，因为它由那些后验密度最高的参数的值所组成．注意 HPD 区域在形式上与似然区域是类似的．

9.3.4 损失函数最优

在前两节我们考察区间估计量的最优性时，首先要求它们有最小的覆盖概率，然后找寻最短的区间．但是可以在一个损失函数下把这些要求放在一块，然后用判决理论求得一个最佳估计量．在区间估计里，行为空间 $\mathcal{A}$ 将由参数空间的子集所组成，而更一般地，我们应当说“集合估计量”，因为一个最佳的法则未必是区间．然而出于实际应用上的考虑，我们主要去寻找区间估计量，幸运的是，很多最优解给出的是区间．

我们用 C（对于置信区间）表示 $\mathcal{A}$ 的元素，它具有的意义是行为 C 给出了区间估计“$\theta\in C$”．一个判决法则 $\delta(\boldsymbol{x})$ 就是指出对每个 $\boldsymbol{x}\in\mathcal{X}$，如果观测到 $\boldsymbol{X}=\boldsymbol{x}$，把哪

个集合 $C\in\mathcal{A}$ 作为 θ 的一个估计. 这样，我们将像以前那样使用记号 $C(x)$.

区间估计问题里的损失函数通常包括两个量：集合估计是否正确地包括 θ 真值的测度和集合估计尺寸的测度. 我们将主要考虑是区间的那种集合 C，所以尺寸的一个自然的测度就是 Length（C）$=C$ 的长度. 为表示正确性测度，一般使用

$$I_C(\theta)=\begin{cases}1 & \theta\in C\\ 0 & \theta\notin C\end{cases}$$

这就是说，如果估计正确，$I_C(\theta)=1$ 否则是 0. 实际 $I_C(\theta)$ 就是集合 C 的示性函数. 但是要认识到 C 是由数据 $\boldsymbol{X}$ 的值决定的一个随机集合.

损失函数应当反映出一个好估计具有小的 Length(C) 和大的 $I_C(\theta)$ 这样的事实. 下式是这样的一个损失函数

(9.3.4) $$L(\theta,C)=b\mathrm{Length}(C)-I_C(\theta)$$

其中 b 是一个正的常数，它反映了我们想给予两个准则的相对权重，由于两个量是非常不同的，因此这点考虑是必需的. 如果对于估计的正确性关心更多，则 b 就应当小，而如果更关心区间长度，则就应当使用大的 b.

与式（9.3.4）相关联的风险函数特别简单，由下式给出

$$\begin{aligned}R(\theta,C)&=bE_\theta\left[\mathrm{Length}(C(\boldsymbol{X}))\right]-E_\theta I_{C(\boldsymbol{X})}(\theta)\\&=bE_\theta\left[\mathrm{Length}(C(\boldsymbol{X}))\right]-P_\theta\left(I_{C(\boldsymbol{X})}(\theta)=1\right)\\&=bE_\theta\left[\mathrm{Length}(C)(\boldsymbol{X}))\right]-P_\theta\left(\theta\in C(\boldsymbol{X})\right)\end{aligned}$$

这个风险有两个成分，即区间的期望长度和区间估计量的覆盖概率. 该风险反映这样的事实，即我们希望期望长度小而同时覆盖概率高，这和过去几节一样. 但是，与过去先要求有最小覆盖概率然后最小化长度的做法不同的是，现在的风险指明了在两个量之间的权衡考虑. 也许一个有较小的覆盖概率的判决由于能大幅度减少长度而被采用.

通过变动损失（9.3.4）中的 b，我们可以变动区间估计量的尺寸与覆盖概率间的相对重要性，这在以前是做不到的. 作为说明当前这个设置的适应性的例子，考虑一些极端的情况. 如果 $b=0$，就是不考虑尺寸只考虑覆盖概率，于是估计量 $C=(-\infty,\infty)$ 是最佳判决法则，它的覆盖概率是 1. 类似，如果 $b=\infty$，则覆盖概率无关紧要，于是点集合是最佳的. 因此，判决法则的选择范围包含了所有可能的情形. 在下面的例子中，对于 b 的一个指定的有限范围，选择一个好法则等于利用风险函数去决定置信区间，而如果 b 在这个范围之外，最优判决法则就是一个点估计量.

习题

9.1

9.1 如果 $L(x)$ 和 $U(x)$ 满足 $P_\theta(L(X)\leqslant\theta)=1-\alpha_1$ 和 $P_\theta(U(X)\geqslant\theta)=1-\alpha_2$，并且对于所有的 x，$L(x)\leqslant U(x)$，证明：$P_\theta(L(X)\leqslant\theta\leqslant U(X))=1-\alpha_1-\alpha_2$.

9.4

9.4 设 $X_1,\cdots,X_n$ 是来自 n（0，σ_X^2）的随机样本，而 $Y_1,\cdots,Y_n$ 是来自 n(0，σ_Y^2)的随机样本，并且与诸 X 独立. 定义 $\lambda=\sigma_Y^2/\sigma_X^2$.

(a) 求 $H_0:\lambda=\lambda_0$ 对 $H_1:\lambda\neq\lambda_0$ 的水平 α 的 LRT.

(b) 用一个 F 分布的随机变量来表示（a）中 LRT 的拒绝区域.

(c) 求关于 λ 的一个 $1-\alpha$ 置信区间.

9.6

9.6 (a) 关于二项分布的参数 p，通过反转 $H_0: p=p_0$ 对 $H_1: p\neq p_0$ 的 LRT，推出它的一个置信区间.

9.7

9.7 (a) 基于来自 n（θ，$a\theta$）族的样本 X_1，…，X_n，通过反转假设 $H_0: a=a_0$ 对 $H_1: a\neq a_0$ 的 LRT，求 a 的 $1-\alpha$ 置信集合，其中 θ 未知.

(b) 对于相关联的分布族 n（θ，$a\theta^2$），可以问类似的问题. 如果 X_1，…，X_n 是 iid n（θ，$a\theta^2$）的，其中 θ 未知，基于反转关于 $H_0: a=a_0$ 对 $H_1: a\neq a_0$ 的 LRT，求 a 的 $1-\alpha$ 置信集合.

9.12

9.12 求：基于抽自一个 n(θ，θ) 总体、样本量为 n 的随机样本的一个枢轴量，其中 $\theta>0$. 利用这个枢轴量建立一个关于 θ 的 $1-\alpha$ 置信区间.

9.13

9.13 设 X 是来自 beta（θ，1）的一个单一的观测.

(a) 设 $Y=-(\log X)^{-1}$. 计算集合 $[y/2, y]$ 的置信系数.

(b) 求一个枢轴量并且利用它建立一个和（a）中区间具有相同置信系数的置信区间.

(c) 比较这两个置信区间.

9.16

9.16 设 X_1，…，X_n 是 iid n（θ，σ^2）的，其中 σ^2 已知．对于以下每个假设，

写出一个水平 α 检验的接受区域以及由反转它得到的 $1-\alpha$ 置信区间．

（a）H_0：$\theta=\theta_0$ 对 H_1：$\theta\neq\theta_0$

（b）H_0：$\theta\geqslant\theta_0$ 对 H_1：$\theta<\theta_0$

（c）H_0：$\theta\leqslant\theta_0$ 对 H_1：$\theta>\theta_0$

9.34

9.34 设 X_1，…，X_n 是来自 n(θ，σ^2）总体的随机样本．

（a）如果 σ^2 已知，求 n 的最小值以保证 μ 的一个 $1-\alpha$ 置信区间的长度不超过 $\sigma/4$．

（b）如果 σ^2 未知，求 n 的最小值以保证 μ 的一个 $1-\alpha$ 置信区间的长度以 90% 的概率不超过 $\sigma/4$．

9.36

9.36 设 X_1，…，X_n 是独立的，具有概率密度函数 $f_{X_i}(x|\theta)=\mathrm{e}^{i\theta-x}I_{[i\theta,\infty)}(x)$．证明 $T=\min\limits_i(X_i/i)$是关于 θ 的一个充分统计量．基于 T，求 θ 的一个具有最短长度的形式为［$T+a$，$T+b$］的 $1-\alpha$ 置信区间．

9.37

9.37 设 X_1，…，X_n 是 iid $U(0，\theta)$ 的．设 Y 是最大的次序统计量．证明Y/θ 是一个枢轴量并且证明区间

$$\left\{\theta：y\leqslant\theta\leqslant\frac{y}{\alpha^{1/n}}\right\}$$

是最短的 $1-\alpha$ 枢轴区间．