书籍摘要及习题解答：《数理统计学讲义（第 2 版）》

曾焰

版本 0.1，最后修订于 2017-03-06

摘要

陈家鼎等 [1] 的内容摘要及习题解答。

目录

1 绪论

1.1 数理统计学的研究对象

（1）若可以知道抽样产品的准确寿命，如何估计产品的寿命？例：某些电子设备（在使用寿命服从指数分布的假设下，第 16 页例 1.1 用最大似然法给出了估计）。

又若只能知道抽样产品有效还是无效，如何估计产品的寿命？例：导弹。

（2）以针织品断裂强力为对象，研究两个正态总体的假设检验问题：如何判断它们的均值是相等的？（第 116 页例 4.3 和第 119 页例 4.4）

（3）以两个刑侦组的破案率为对象，研究服从两点分布（伯努利分布）的两个总体的比较问题：如何判断它们的参数是相等的？（第 146 页例 6.3）

（4）在使用海洋重力仪探测海底石油时，观测信号 $x(t)$、干扰信号 $n(t)$ 和有用信号 $s(t)$ 满足关系：$x(t) = s(t) + n(t)$ $(0 \leq t \leq T)$。干扰信号 $n(t)$ 的强度是有用信号 $s(t)$ 的十万倍。问：如何从强干扰背景中把微弱信号 $s(t)$ 提取出来？

（5）以多元回归分析为工具，研究合金材料的最佳组成成分（第 235 页例 5.2）。

（6）利用列联表的独立性检验，探讨吸烟与患慢性支气管炎的关联（第 156 页例 7.4）。

1.2 数理统计学的基本概念

总体。总体的分布函数。总体的样本。随机抽样法：简单随机样本（通过有放回的逐次随机抽取或对总体进行多次独立的重复观测，得到独立同分布的观测值）、单纯随机样本（总体有限的情形下的无放回随机抽取）。统计推断主要有估计和假设检验两类。抽样分布。

1.3 数理统计学发展简史

略，参看陈希孺 [2]。

2 估计

2.1 参数估计的方法

2.1.1 最大似然法

最大似然估计是基于概率最大原则这一朴素的统计思想。数学上可以证明，在一定条件下，只要样本量 n 足够大，最大似然估计 $\hat{\theta_n}$ 与真值 θ 可相差任意小：

$$\lim_{n\to\infty} P\left(\|\hat{\theta}_n - \theta\| > \varepsilon\right) = 0 \tag{2.1}$$

或

$$P\left(\lim_{n\to\infty} \hat{\theta}_n = \theta\right) = 1. \tag{2.2}$$

若 (2.1) 成立，通常称 $\hat{\theta}_n$ 有**相合性**；若 (2.2) 成立，通常称 $\hat{\theta}_n$ 有**强相合性**。

（1）**两点分布**（也称**伯努利（Bernoulli）分布**）：两点分布的最大似然估计是均值。由强大数律知该最大似然估计有强相合性。

（2）**指数分布**：指数分布的最大似然估计是均值的倒数。由强大数律知该最大似然估计有强相合性。

（3）**正态分布**：均值 μ 和方差 δ 的最大似然估计分别是

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

和

$$\hat{\delta} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

由强大数律知它们有强相合性。

（4）**韦布尔（Weibull）分布**：最大似然估计没有解析解只有数值解。数学上可以证明，韦布尔分布的最大似然估计是强相合估计。

（5）**均匀分布**：$[a,b]$ 区间上的均匀分布的最大似然估计是 $\hat{a} = \min_{1\le i\le n} x_i, \hat{b} = \max_{1\le i\le n} x_i$。数学上可以证明，它们是强相合估计（通过计算可以证明 $\sum_{n=1}^{\infty} P\left(|\hat{b} - b| \ge \varepsilon\right) < \infty$ 和 $\sum_{n=1}^{\infty} P\left(|\hat{a} - a| \ge \varepsilon\right) < \infty$，然后用 Borel-Cantelli 引理）。

2.1.2 矩法

这是英国统计学家 K. Pearson 在 1894 年提出来的。设随机变量 X 的密度函数（或概率函数）是 $f(x,\theta)$，其中 $\theta = (\theta_1,\cdots,\theta_m)$ 是未知参数。若 X 的 k 原点距 $V_k = E[X^k]$ 存在，则 V_k 也是 θ_1，$\cdots$，θ_m 的函数，记为 $g_k(\theta_1,\cdots,\theta_m)$（$k = 1,2,\cdots$）。假定从方程组

$$\begin{cases} g_1(\theta_1,\cdots,\theta_m) = V_1 \\ g_2(\theta_1,\cdots,\theta_m) = V_2 \\ \cdots \\ g_m(\theta_1,\cdots,\theta_m) = V_m \end{cases}$$

可解出：

$$\begin{cases} \theta_1 = f_1(V_1,\cdots,V_m) \\ \theta_2 = f_2(V_1,\cdots,V_m) \\ \cdots \\ \theta_m = f_m(V_1,\cdots,V_m) \end{cases}$$

设 x_1，x_2，$\cdots$，x_n 是 X 的样本，用样本距

$$\tilde{V}_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$$

来估计 V_k，然后用

$$\tilde{\theta}_k = f_k(\tilde{V}_1,\cdots,\tilde{V}_m)$$

来估计 θ_k（$k = 1,2,\cdots,m$）。这种估计未知参数的方法叫做距法。矩法的根据是强大数律。

2.1.3 其他估计方法（以二战期间的序列号统计方法为例）

由于估计问题多种多样，需要人们针对各个具体问题的实际情况制定恰当的估计方法，以满足实际工作的需要。

问题：二战期间德国的每件装备都有一个序列号，表明它按该类装备制造出来的次序排在第几。假设到某天为止制造出的某型装备共有 N 件，而盟军通过各种手段（俘虏或截获邮件）知道了 n 件装备的序列号为 $k_1 < k_2 < \cdots < k_n$。问：如何估计该型装备的总数 N?

我们可把 N 看成某个随机向量的概率分布中的参数。从 1，2，$\cdots$，N 这 N 个数中随机抽取 n 个，共有 C_N^n 种结果。我们假定各种结果出现的机会均等，出现 $(k_1, k_2, \cdots, k_n)$ 的概率为 $1/C_N^n$。N 究竟比 k_n 大多少呢？自然想到应该利用 $k_1 < k_2 < \cdots < k_n$ 中相邻两个数的间隔来估计：

$$W_1 := k_n + \frac{1}{n}\sum_{i=2}^{n}(k_i - k_{i-1} - 1) = k_n + \frac{1}{n-1}(k_n - k_1) - 1.$$

容易看出，对任何 i（$1 \le i \le N - n + 1$）有

$$P(k_1 = i) = \frac{\text{剩下的后（}n-1\text{）个数的不同排列数目}}{C_N^n} = \frac{C_{N-i}^{n-1}}{C_N^n}$$

对任何 i（$n \le i \le N$）有

$$P(k_n = i) = \frac{\text{剩下的前（}n-1\text{）个数的不同排列数目}}{C_N^n} = \frac{C_{i-1}^{n-1}}{C_N^n}$$

由此可通过计算证明，用 W_1 估计 N 是没有系统偏差的：$E_N[W_1] = N$（这里 E_N 表示由 N 所决定的概率测度下的数学期望）。

估计量 W_1 还可以改进为

$$W_2 = k_n + \frac{1}{n}\sum_{i=1}^{n}(k_i - k_{i-1} - 1) = \frac{n+1}{n}k_n - 1 \;\; (k_0 := 0)$$

则 W_2 也是无偏的，而且可用初等方法证明其方差最小：设 $\psi(k_1, \cdots, k_n)$ 是 N 的任何估计，只要满足 $E_N[\psi(k_1, \cdots, k_n)] = N$（对一切 $N \ge n$），就一定成立

$$E_N[(\psi(k_1, \cdots, k_n) - N)^2] \ge E_N[(W_2 - N)^2]$$

而且当 $\psi(k_1, \cdots, k_n) \ne W_2$ 时，上式左端大于右端。

更进一步，可把 n 看作随机的。则上述结论在将叙述改为“条件方差”、“条件期望”、“条件概率”之后仍然成立。

2.2 估计的优良性标准

设 X 的密度函数是 $f(x, \theta)$，其中 $\theta = (\theta_1, \cdots, \theta_m) \in \Theta \subset \mathbb{R}^m$。设 $g(\theta)$ 是 θ 的函数，X_1，X_2，$\cdots$，X_n 是 X 的样本。所谓**$g(\theta)$ 的估计量**，是指样本的函数 $\varphi(X_1, \cdots, X_n)$。直观上看，$|\varphi(X_1, \cdots, X_n) - g(\theta)|$ 越小，φ 就越好。但 $\varphi(X_1, \cdots, X_n)$ 的值依赖于样本值，它本身是个随机变量，而 $g(\theta)$ 是未知的。所以衡量优良性的标准不很简单也不唯一。

定义 2.1. 设 $\varphi(X_1,\cdots,X_n)$ 是 $g(\theta)$ 的估计，称 $M_\theta(\varphi)=E_\theta[(\varphi(X_1,\cdots,X_n)-g(\theta))^2]$ 为 φ 的**均方误差**。

定义 2.2. 若 $\varphi_1(X_1,\cdots,X_n)$ 和 $\varphi_2(X_1,\cdots,X_n)$ 都是 $g(\theta)$ 的估计量，且 $M_\theta(\varphi_1)\le M_\theta(\varphi_2)$ $(\forall\theta\in\Theta)$，则称φ_1 **不次于** φ_2，此时若还有 θ_0 使得 $M_{\theta_0}(\varphi_1)<M_{\theta_0}(\varphi_2)$，则称$\varphi_1$ **比** φ_2 **有效**。

因为均方误差是 θ 的函数，两个估计量并不一定总能比较优劣。在绝大多数情况下，不存在均方误差最小的估计量。因此在确定估计的优良性标准时，往往不得不缩小范围，例如在所有无偏估计中找方差最小的。

定义 2.3. 称$\varphi(X_1,\cdots,X_n)$ **是** $g(\theta)$ **的无偏估计**，若

$$E_\theta[\varphi(X_1,\cdots,X_n)]=g(\theta),\ \forall\theta\in\Theta$$

例 1. 根据计算和强大数律可知，$\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$ 是 $Var_\theta(X)$ 的强相合的无偏估计。

例 2. 当 $g(\theta)=E_\theta[X]$ 时，$\bar{X}_n=\frac{1}{n}\sum_1^n X_i$ 是 $g(\theta)$ 的无偏估计且 $M_\theta(\bar{X}_n)=Var_\theta(\bar{X}_n)=\frac{1}{n}Var_\theta(X)$。只要 $0<Var_\theta(X)<\infty$，$\bar{X}_n$ 比 $\bar{X}_{n-1}$ 有效。

例 3. 当 $\sum_1^n\lambda_i=1$ 时，$\varphi_1=\frac{1}{n}\sum_1^n X_i$ 与 $\varphi_2=\sum_1^n\lambda_iX_i$ 都是由样本得到的 $E_\theta[X]$ 的无偏估计，且 φ_1 不次于 φ_2。若 λ_1，$\cdots$，λ_n 不全相等，则 φ_1 比 φ_2 有效。

定义 2.4. 设 $\varphi(X_1,\cdots,X_n)$ 是 $g(\theta)$ 的无偏估计，且对一切无偏估计 $\psi(X_1,\cdots,X_n)$ 均有 $M_\theta(\varphi)\le M_\theta(\psi)$ $(\forall\theta\in\Theta)$，则称 φ 是 $g(\theta)$ 的**（一致）最小方差无偏估计**。

注 1. “最小方差无偏估计”就是一种最优的估计量，它在很多情况下确实存在，当然也有不存在的时候。有时无偏性显得不合理。还有别的优良性标准，如贝叶斯标准，*minimax* 标准等等。

2.3 置信区间（区间估计）

2.4 分布函数与密度函数的估计

2.5 习题

▶ 1. 设 X 是服从集合分布

$$P(X=k)=p(1-p)^{k-1},\ k=1,2,\cdots$$

X_1，X_2，$\cdots$，X_n 是 X 的简单随机样本，试找出 p 的最大似然估计。

解答. 设样本是 k_1，k_2，$\cdots$，k_n，则似然函数是

$$L(k_1,\cdots,k_n;p)=\prod_{i=1}^n p(1-p)^{k_i-1}=p^n(1-p)^{\sum_{i=1}^n k_i-n}$$

取对数并对 p 求一阶和二阶导得

$$\frac{d}{dp}\ln(L)=\frac{n}{p}-\frac{\sum_{i=1}^n k_i-n}{1-p},\ \frac{d^2}{dp^2}\ln(L)=-\frac{n}{p^2}-\frac{\sum_{i=1}^n k_i-n}{(1-p)^2}<0$$

所以 p 的最大似然估计是 $\frac{n}{\sum_{i=1}^n k_i}$。 □

注 2. 简单计算可得 $E[X]=\frac{1}{p}$。所以根据强大数律，$\lim_{n\to\infty}\frac{n}{\sum_{i=1}^{n}k_i}=\frac{1}{E[X]}=p$。

▶ 2. 设 X 的分布密度是

$$f(x)=\frac{1}{2\sigma}e^{-\frac{1}{\sigma}|x|}\ (\sigma>0)$$

X_1，X_2，$\cdots$，X_n 是 X 的样本，试求 σ 的最大似然估计。

解答. 设样本是 x_1，x_2，$\cdots$，x_n，则似然函数是

$$L(x_1,\cdots,x_n;\sigma)=\frac{1}{(2\sigma)^n}e^{-\frac{1}{\sigma}\sum_{i=1}^{n}|x_i|}$$

取对数并对 σ 求导得

$$\frac{d}{d\sigma}\ln(L)=-\frac{n}{\sigma}+\frac{\sum_{i=1}^{n}|x_i|}{\sigma^2}$$

则 $\frac{d}{d\sigma}\ln(L)$ 在 $\frac{\sum_{i=1}^{n}|x_i|}{n}$ 处取值为 0，在其左侧大于 0，在其右侧小于 0。所以 σ 的最大似然估计是 $\frac{\sum_{i=1}^{n}|x_i|}{n}$。 □

▶ 3. 设 X_1，X_2，$\cdots$，X_n 是来自 $[\theta,\theta+1]$ 上均匀分布的样本，其中 $-\infty<\theta<+\infty$。试证明 θ 的最大似然估计不止一个。你能求出 θ 的全部最大似然估计吗？你能找出 θ 的无偏估计吗？

解答. 设样本是 x_1，x_2，$\cdots$，x_n，则似然函数是

$$L(x_1,\cdots,x_n;\theta)=\prod_{i=1}^{n}1_{[\theta,\theta+1]}(x_i)$$

对顺序统计量 $x_{(1)}$，$x_{(2)}$，$\cdots$，$x_{(n)}$，显然我们应有 $x_{(n)}-x_{(1)}\in[0,1]$。并且似然函数取得最大值 1，当且仅当 $\theta\le x_{(1)}\le x_{(n)}\le\theta+1$。换言之，似然函数取得最大值 1 当且仅当

$$x_{(n)}-1\le\theta\le x_{(1)}$$

这一条件给出了 θ 的全部最大似然估计。容易计算 $X_{(k)}$ 的概率密度是（参见 [3] 第三章 §7）

$$f_{(k)}(x;\theta)=\frac{n!}{(k-1)!(n-k)!}(x-\theta)^{k-1}(1+\theta-x)^{n-k}\cdot 1_{[\theta,\theta+1]}(x)$$

所以

$$\begin{aligned}
E[X_{(1)}] &= \int_{\theta}^{\theta+1}x\cdot n(1+\theta-x)^{n-1}dx\\
&= \int_{\theta}^{\theta+1}n\left[-(1+\theta-x)(1+\theta-x)^{n-1}+(1+\theta)(1+\theta-x)^{n-1}\right]dx\\
&= (1+\theta-x)|_{\theta}^{\theta+1}-(1+\theta)\left(1+\theta-x\right)|_{\theta}^{\theta+1}\\
&= \theta
\end{aligned}$$

故 $X_{(1)}$ 是一个 θ 的无偏估计。容易验证

$$E[X_{(n)}]=\int_{\theta}^{\theta+1}x\cdot n(x-\theta)^{n-1}dx=\frac{n}{n+1}+\theta$$

故 $X_{(n)}-1$ 不是一个 θ 的无偏估计。 □

▶ *4. 设随机变量 X 以均等机会按 $N(0,1)$ 分布取值和按 $N(\mu,\sigma^2)$ 分布取值（μ,σ^2 未知，$-\infty<\mu<\infty$，$\sigma^2>0$）。这时 X 的分布密度为这两个分布的密度的平均：

$$f(x;\mu,\sigma^2)=\frac{1}{2}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}+\frac{1}{2}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

设 X_1，X_2，$\cdots$，X_n 为此混合分布的简单随机样本，试证明 μ 和 σ^2 不存在最大似然估计。

注 3. 参见*Introduction to EM: Gaussian Mixture Models, by Matt Bonakdarpour.*

▶ 5. 在买面包作早点的男、女消费者中，男性购买者的比例 p 未知，但知道 $\frac{1}{2}\le p\le\frac{2}{3}$。设在 70 个购买者中发现 12 个是男性，58 个是女性，求 p 的最大似然估计。

如果对 p 无限制（$0\le p\le 1$），求 p 的最大似然估计。

解答. 如 §1 中的例子，设随机变量 X 根据购买者为男性或女性而取值 1 或 0，且 $p=P(X=1)=1-P(X=0)$。则 X 的概率函数是

$$P(X=x)=p^x(1-p)^{1-x},\ (x=0,1)$$

设 X 的样本是 x_1，x_2，$\cdots$，x_n，则似然函数的对数是

$$\ln L(x_1,\cdots,x_n;p)=n\bar{x}\ln p+n(1-\bar{x})\ln(1-p),\ \bar{x}=\frac{\sum_{i=1}^n x_i}{n}\in[0,1]$$

对 p 求一阶和二阶导后得

$$\frac{d}{dp}\ln L(x_1,\cdots,x_n;p)=n\cdot\frac{\bar{x}}{p}-n\cdot\frac{1-\bar{x}}{1-p},\ \frac{d^2}{dp^2}\ln L(x_1,\cdots,x_n;p)=-n\cdot\frac{\bar{x}}{p^2}-n\cdot\frac{1-\bar{x}}{(1-p)^2}\le 0$$

如果对 p 无限制，则 p 的最大似然估计是 $\bar{x}$。

若知道 $\frac{1}{2}\le p\le\frac{2}{3}$，我们注意到 $\lim_{p\downarrow 0}\ln L=-\infty$，$\lim_{p\uparrow 1}\ln L=-\infty$，

$$\frac{d}{dp}\ln L(x_1,\cdots,x_n;p)\begin{cases}>0,\ \text{if}\ p<\bar{x}\\ <0,\ \text{if}\ p>\bar{x}\end{cases}$$

所以 $L(p)$ 在 $(0,\bar{x})$ 单调上升，在 $(\bar{x},1)$ 单调下降。故 p 的最大似然估计是 $\bar{x}1_{[\frac{1}{2},\frac{2}{3}]}(\bar{x})+\frac{2}{3}1_{(\frac{2}{3},1)}(\bar{x})+\frac{1}{2}1_{(0,\frac{1}{2})}(\bar{x})$。 □

▶ 6. 设 X_1，X_2，$\cdots$，X_n 是来自下列两参数指数分布的样本：

$$f(x;\theta_1,\theta_2)=\begin{cases}\frac{1}{\theta_2}e^{-\frac{1}{\theta_2}(x-\theta_1)}, & x\ge\theta_1\\ 0, & x<\theta_1\end{cases}$$

其中 $\theta_1\in(-\infty,+\infty)$，$\theta_2\in(0,+\infty)$，试求出 θ_1 和 θ_2 的最大似然估计。

解答. 设样本是 x_1，x_2，$\cdots$，x_n，则似然函数是

$$L(x_1,\cdots,x_n;\theta_1,\theta_2)=\prod_{i=1}^n\frac{1}{\theta_2}e^{-\frac{x_i-\theta_1}{\theta_2}}\cdot 1_{(-\infty,x_i]}(\theta_1)$$

对于任何给定的 θ_2，$L(x_1,\cdots,x_n;\theta_1,\theta_2)$ 在 $(-\infty,x_{(1)}]$ 上是 θ_1 的非零递增函数，而在 $(x_{(1)},\infty)$ 上恒等于零。所以 θ_1 的最大似然估计是 $\theta_1^{mle}=x_{(1)}=\min_{1\le i\le n}x_i$。

对于取定的 $\theta_1^{mle}=x_{(1)}$，$L(x_1,\cdots,x_n;\theta_1^{mle},\theta_2)$ 是 θ_2 的光滑函数。取对数求导可得

$$\frac{d}{d\theta_2}\ln L=-\frac{n}{\theta_2}+\frac{n(\bar{x}-x_{(1)})}{\theta_2^2}=\frac{n}{\theta_2^2}\left[(\bar{x}-x_{(1)})-\theta_2\right]$$

可知 $\ln L$ 在 $\bar{x}-x_{(1)}$ 左侧单调上升，在 $\bar{x}-x_{(1)}$ 右侧单调下降。所以 θ_2 的最大似然估计是 $\bar{x}-x_{(1)}$。 □

▶ 7. 设 X_1，X_2，$\cdots$，X_n 是来自下列 Weibull 分布的样本：

$$F(x;\alpha,\eta)=\begin{cases}1-e^{-\left(\frac{x}{\eta}\right)^\alpha}, & x\ge 0\\ 0, & x<0\end{cases}$$

其中 $\alpha>0$，$\eta>0$，试证明 α，η 的最大似然估计存在。

证明. 上述两参数 Weibull 分布的密度函数为

$$f(x;\alpha,\eta)=\begin{cases}\frac{\alpha x^{\alpha-1}}{\eta^\alpha}e^{-\left(\frac{x}{\eta}\right)^\alpha}, & x\ge 0\\ 0, & x<0\end{cases}$$

设样本是 x_1，x_2，$\cdots$，x_n，则似然函数的对数是

$$\ln L(x_1,\cdots,x_n;\alpha,\eta)=\ln\left(\prod_{i=1}^n\frac{\alpha x_i^{\alpha-1}}{\eta^\alpha}e^{-\left(\frac{x_i}{\eta}\right)^\alpha}\right)=n\ln\alpha-n\alpha\ln\eta+(\alpha-1)\sum_{i=1}^n\ln x_i-\sum_{i=1}^n\left(\frac{x_i}{\eta}\right)^\alpha$$

$\ln L$ 对 η 求偏导，我们可得

$$\frac{\partial}{\partial\eta}\ln L=-\frac{n\alpha}{\eta}+\alpha\sum_{i=1}^n x_i^\alpha\eta^{-\alpha-1}=\frac{n\alpha}{\eta}\left(\frac{\sum_{i=1}^n x_i^\alpha/n}{\eta^\alpha}-1\right)$$

可见对任何固定的 α，$\ln L(x_1,\cdots,x_n;\alpha,\eta)$ 在 $\eta=\left(\frac{\sum_{i=1}^n x_i^\alpha}{n}\right)^{1/\alpha}$ 处取最大值。

$\ln L$ 对 α 求偏导，我们可得

$$\begin{aligned}\frac{\partial}{\partial\alpha}\ln L &= \frac{n}{\alpha}-n\ln\eta+\sum_{i=1}^n\ln x_i-\sum_{i=1}^n\left(\frac{x_i}{\eta}\right)^\alpha\ln\frac{x_i}{\eta}\\ &= \frac{n}{\alpha}-n\ln\eta+\sum_{i=1}^n\ln x_i-\frac{\sum_{i=1}^n x_i^\alpha\ln x_i-\sum_{i=1}^n x_i^\alpha\ln\eta}{\eta^\alpha}\end{aligned}$$

将 $\eta=\left(\frac{\sum_{i=1}^n x_i^\alpha}{n}\right)^{1/\alpha}$ 代入上述公式，可得

$$\frac{\partial}{\partial\alpha}\ln L=\frac{n}{\alpha}+\sum_{i=1}^n\ln x_i-n\frac{\sum_{i=1}^n x_i^\alpha\ln x_i}{\sum_{i=1}^n x_i^\alpha}=n\left[\frac{1}{\alpha}+\frac{\sum_{i=1}^n\ln x_i}{n}-\frac{\sum_{i=1}^n(x_i^\alpha\ln x_i)}{\sum_{i=1}^n x_i^\alpha}\right]$$

我们注意到

$$\lim_{\alpha\downarrow 0}\left[\frac{1}{\alpha}+\frac{\sum_{i=1}^n\ln x_i}{n}-\frac{\sum_{i=1}^n(x_i^\alpha\ln x_i)}{\sum_{i=1}^n x_i^\alpha}\right]=\lim_{\alpha\downarrow 0}\frac{1}{\alpha}=+\infty$$

以及（$x_{max} = \max_{1\le i\le n} x_i$）

$$
\begin{aligned}
&\lim_{\alpha\uparrow+\infty}\left[\frac{1}{\alpha}+\frac{\sum_{i=1}^n \ln x_i}{n}-\frac{\sum_{i=1}^n (x_i^\alpha \ln x_i)}{\sum_{i=1}^n x_i^\alpha}\right]\\
=\ &\lim_{\alpha\uparrow+\infty}\left[\frac{1}{\alpha}+\frac{\sum_{i=1}^n \ln x_i}{n}-\frac{\sum_{i=1}^n \left(\frac{x_i}{x_{max}}\right)^\alpha \ln x_i}{\sum_{i=1}^n \left(\frac{x_i}{x_{max}}\right)^\alpha}\right]\\
=\ &\frac{\sum_{i=1}^n \ln x_i}{n}-\ln x_{max}\cdot\frac{\#\{x_i = x_{max}: 1\le i\le n\}}{\#\{x_i = x_{max}: 1\le i\le n\}}\\
\le\ &0
\end{aligned}
$$

此处不等号中的等号成立，当且仅当所有的 x_i 都相等。这样的集合是个 $\mathbb{R}^n$ 中的零测集，所以我们可以认为

$$
\lim_{\alpha\uparrow+\infty}\left[\frac{1}{\alpha}+\frac{\sum_{i=1}^n \ln x_i}{n}-\frac{\sum_{i=1}^n (x_i^\alpha \ln x_i)}{\sum_{i=1}^n x_i^\alpha}\right]<0
$$

所以方程 $\frac{\partial}{\partial\alpha}\ln L = 0$ 至少有一个解。如能证明此方程有唯一解，则 $\ln L$ 在此解左侧单调上升（导数为正），在此解右侧单调下降（导数为负），可推知该唯一解是 α 的最大似然估计。

事实上，定义 $y_i = \ln x_i$ 以及 $w_i(\alpha) = \frac{x_i^\alpha}{\sum_{k=1}^n x_k^\alpha}$，我们有

$$
\frac{1}{n}\frac{\partial}{\partial\alpha}\ln L = \frac{1}{\alpha}+\frac{\sum_{i=1}^n y_i}{n}-\sum_{i=1}^n w_i(\alpha) y_i
$$

注意到

$$
w_i'(\alpha) = \frac{d}{d\alpha}\frac{e^{\alpha y_i}}{\sum_{k=1}^n e^{\alpha y_k}} = y_i w_i(\alpha) - e^{\alpha y_i}\frac{\sum_{k=1}^n e^{\alpha y_k}\cdot y_k}{(\sum_{k=1}^n e^{\alpha y_k})^2} = y_i w_i(\alpha) - w_i(\alpha)\sum_{k=1}^n w_k(\alpha) y_k
$$

故此

$$
\begin{aligned}
\frac{1}{n}\frac{\partial^2}{\partial\alpha^2}\ln L &= -\frac{1}{\alpha^2}-\sum_{i=1}^n\left(y_i w_i(\alpha) - w_i(\alpha)\sum_{k=1}^n w_k(\alpha) y_k\right) y_i\\
&= -\frac{1}{\alpha^2}-\sum_{i=1}^n w_i(\alpha) y_i^2 + \left(\sum_{i=1}^n w_i(\alpha) y_i\right)^2
\end{aligned}
$$

由柯西-施瓦茨不等式，

$$
\left(\sum_{i=1}^n w_i(\alpha) y_i\right)^2 \le \left(\sum_{i=1}^n w_i(\alpha) y_i^2\right)\left(\sum_{i=1}^n w_i(\alpha)\right) = \sum_{i=1}^n w_i(\alpha) y_i^2
$$

所以 $\frac{\partial^2}{\partial\alpha^2}\ln L < 0$，$\frac{\partial}{\partial\alpha}\ln L$ 单调下降。结合前面结果可知 $\frac{\partial}{\partial\alpha}\ln L = 0$ 有唯一解。

综合起来，可知 α 和 η 的最大似然估计存在。 □

▶ 8. 设 X_1，X_2，$\cdots$，X_n 是来自下列 Γ 分布的样本：

$$
f(x;\alpha,\beta) = \begin{cases}\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}, & x>0\\ 0, & x\le 0\end{cases}
$$

其中 $\alpha>0$，$\beta>0$，试证明 α，β 的最大似然估计存在。

证明. 设样本是 x_1，x_2，$\cdots$，x_n，则似然函数的对数是

$$\ln L(x_1, \cdots, x_n; \alpha, \beta) = n\alpha \ln \beta - n \ln \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^{n} \ln x_i - \beta \sum_{i=1}^{n} x_i$$

对 β 求一阶和二阶偏导可得

$$\frac{\partial}{\partial \beta} \ln L = \frac{n\alpha}{\beta} - n\bar{x}, \frac{\partial^2}{\partial \beta^2} \ln L = -\frac{n\alpha}{\beta^2} < 0$$

所以对于给定的 α，β 的最大似然估计是 $\beta_{mle} = \frac{\alpha}{\bar{x}}$。

似然函数的对数对 α 求一阶偏导并代入 $\beta_{mle} = \frac{\alpha}{\bar{x}}$ 可得

$$\frac{\partial}{\partial \alpha} \ln L = n \ln \alpha - n \ln \bar{x} - n\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \sum_{i=1}^{n} \ln x_i$$

我们分析 $\frac{\partial}{\partial \alpha} \ln L$ 在 $\alpha \downarrow 0$ 和 $\alpha \uparrow \infty$ 两种情形下的性质以决定上述方程是否有解。

首先，由分部积分可得

$$\Gamma(\alpha + 1) = \int_0^\infty z^\alpha e^{-z} dz = -\left. z^\alpha e^{-z} \right|_0^\infty + \int_0^\infty \alpha z^{\alpha-1} e^{-z} dz = \alpha \Gamma(\alpha)$$

所以

$$\frac{\Gamma'(\alpha+1)}{\Gamma(\alpha+1)} = \frac{\Gamma(\alpha) + \alpha\Gamma'(\alpha)}{\alpha\Gamma(\alpha)} = \frac{1}{\alpha} + \frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$$

故当 $\alpha \downarrow 0$ 时，

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} \sim -\frac{1}{\alpha} + \Gamma'(1)$$

其中 $\Gamma'(\alpha) = \int_0^\infty \ln z \cdot z^{\alpha-1} e^{-z} dz$ 对任何 $\alpha > 0$ 是有限数。所以在 α 趋近于 0 时，

$$\frac{\partial}{\partial \alpha} \ln L \sim n \ln \alpha + \frac{n}{\alpha} \to +\infty$$

其次，根据 Binet 对 Γ 函数的第二积分表示（参见 Whittaker and Watson [5] 第 251 页，12.32，例）：

$$\frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \ln \alpha - \frac{1}{2\alpha} - 2\int_0^\infty \frac{tdt}{(t^2 + \alpha^2)(e^{2\pi t} - 1)}$$

以及对数函数是凹函数这一事实，我们有当 $\alpha \uparrow \infty$ 时，

$$\begin{aligned} \frac{\partial}{\partial \alpha} \ln L &= n\left(\frac{1}{2\alpha} + 2\int_0^\infty \frac{tdt}{(t^2 + \alpha^2)(e^{2\pi t} - 1)}\right) + n\left(\frac{\sum_{i=1}^n \ln x_i}{n} - \ln \bar{x}\right) \\ &\to n\left(\frac{\sum_{i=1}^n \ln x_i}{n} - \ln \bar{x}\right) \le 0 \end{aligned}$$

其中等号成立当且仅当所有的 x_i 都相等。这样的集合是个 $\mathbb{R}^n$ 中的零测集，所以我们可以认为

$$\lim_{\alpha \uparrow \infty} \frac{\partial}{\partial \alpha} \ln L < 0$$

根据 $\frac{\partial}{\partial \alpha} \ln L(\alpha)$ 在零点和无穷处的渐进性质，我们可推知 $\frac{\partial}{\partial \alpha} \ln L = 0$ 必然有解。

最后，仍然根据 Binet 的 Γ 函数第二积分表示，我们可知 $\frac{\partial}{\partial \alpha} \ln L$ 是 α 的单调递减函数，所以 $\frac{\partial}{\partial \alpha} \ln L = 0$ 的解必然唯一。综合起来，我们可以推知 α 的最大似然估计存在。 □

▶ 9. 设 X_1，X_2，$\cdots$，X_n 是来自下列分布密度的样本：

$$f(x,\theta)=\begin{cases}\frac{\Gamma(\theta+1)}{\Gamma(\theta)\Gamma(1)}x^{\theta-1},\\ 0,\end{cases}$$

其中 $\theta\in(0,+\infty)$，试用矩阵法估计 θ。

解答. 我们注意到 $f(x,\theta)=\theta x^{\theta-1}1_{[0,1]}(x)$，所以

$$V_1=E[X]=\int_0^1 xf(x,\theta)dx=\int_0^1\theta x^\theta dx=\frac{\theta}{\theta+1}$$

于是对 θ 的距估计为

$$\tilde{\theta}=\frac{\tilde{V}_1}{1-\tilde{V}_1}=\frac{\bar{x}}{1-\bar{x}}$$

□

▶ 10. 设 X_1，X_2，$\cdots$，X_n 是来自分布密度为

$$f(x;c,\theta)=\frac{1}{2\theta}I_{[c-\theta,c+\theta]}(x)$$

的总体的样本，$-\infty<c<\infty$，$\theta>0$，试用距法估计 c 和 θ。

解答.

$$V_1=E[X]=\int_{c-\theta}^{c+\theta}\frac{xdx}{2\theta}=\frac{1}{4\theta}x^2\Big|_{c-\theta}^{c+\theta}=\frac{(c+\theta)^2-(c-\theta)^2}{4\theta}=c$$

$$V_2=E[X^2]=\int_{c-\theta}^{c+\theta}\frac{x^2dx}{2\theta}=\frac{x^3|_{c-\theta}^{c+\theta}}{6\theta}=c^2+\frac{\theta^2}{3}$$

所以 c 和 θ 的距估计分别是

$$\begin{cases}\tilde{c}=\tilde{V}_1\\ \tilde{\theta}=\sqrt{3(\tilde{V}_2-\tilde{V}_1^2)}\end{cases}$$

□

▶ 11. 设 X_1，X_2，$\cdots$，X_n 为 $N(\mu,\sigma^2)$ 分布的样本，参数 μ，σ^2 未知（$-\infty<\mu<+\infty$，$\sigma^2>0$），令

$$S_0^2=\frac{1}{n}\sum_{i=1}^n(X_i-\bar{X})^2,\ S_1^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2$$

证明对所有可能的 μ 和 σ^2 的值，作为 σ^2 的估计，S_0^2 的均方误差比 S_1^2 的均方误差都小，即 S_0^2 比 S_1^2 有效；但 S_1^2 是无偏估计而 S_0^2 不是；S_0^2 是最大似然估计。

证明. 由定理 2.1，可知 S_1^2 是无偏估计而 S_0^2 不是。为证 S_0^2 比 S_1^2 有效，我们注意到

$$M_0=E[(S_0^2-\sigma^2)^2]$$

同时，

$$M_1=E[(S_1^2-\sigma^2)^2]$$

为证 S_0^2 是最大似然估计， □

3 假设检验

4 回归分析与线性模型

4.1 引言

4.1.1 回归分析的两种观点

自变量 x_1，x_2，$\cdots$，x_p 有些情况下可以人为控制取值，有些情况下只能观测得到，由此产生了回归分析中的两种观点：

（1）把自变量 x_1，x_2，$\cdots$，x_p 看成非随机的普通变量，因变量 y 看成随机变量，这称为**第一类回归分析**。如自变量的取值是人为控制的，则这种观点比较自然。

（2）把自变量 x_1，x_2，$\cdots$，x_p 和因变量 y 都看成随机变量，这称为**第二类回归分析**。如果自变量是观测得到的，则这种观点比较自然。但实际工作中人们一般还是用第一类回归分析来处理。

我们将主要讨论第一类回归分析，关于第二类回归分析只作简单的讨论。我们将指出，在许多情形下，两类回归分析有相同的计算公式。

4.1.2 线性回归分析与线性模型

线性回归模型

$$y = b_0 + b_1x_1 + \cdots + b_px_p + e$$

的基本问题就是如何确定 b_0，b_1，$\cdots$，b_p 使得 e 在某种意义下最小。线性回归分析是回归分析里最重要的组成部分，因为实际当中不仅是经常遇到线性回归模型，而且许多非线性回归模型经过适当的变换可以化为线性回归模型。

4.2 一元线性回归

4.2.1 最小二乘法

定理 4.1. *一元线性回归* $y = a + bx + e$ *的最小二乘估计是*

$$\hat{a} = \bar{y} - \hat{b}\bar{x},\ \hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

4.2.2 线性关系的检验

要使用**回归值**$\hat{y} = \hat{a} + \hat{b}x$ 来进行预报和控制，需要判断回归值和真实值的差距 $\hat{y} - y$ 确实都比较小。而这是一个随机变量，判断一个随机变量是否比较小只能用假设检验的方法，称为**线性相关关系的检验**。为此，我们需要一个具有统计意义的平方和分解公式：

引理 4.1. *令* $\hat{y}_i = \hat{a} + \hat{b}x_i$ $(i = 1, 2, \cdots, n)$，*其中* $\hat{a}$，$\hat{b}$ *为最小二乘估计，则*

$$l_{yy} = Q + U \tag{4.1}$$

其中

$$l_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

叫 y 的**总离差平方和**，

$$Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

叫**残差平方和**，

$$U = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$$

叫**回归平方和**。

公式（4.1）可以解读为“真实值 y 的分散程度 $=$ 拟合误差 $+$ 拟合值 $\hat{y}$ 的分散程度”。我们用比值 U/Q 的大小来衡量线性相关关系的可信程度，这个比值越大，我们越有把握说两变量之间确实存在线性相关关系。

假设 e_1，e_2，$\cdots$，e_n 为独立同 $N(0,\sigma^2)$ 分布的随机变量（$\sigma^2 > 0$ 未知）。对假设 $H_0 : b = 0$ 的检验称作**相关性检验**。当 H_0 被否定时，称**回归方程是显著的**。可以证明

$$F = \frac{U}{Q/(n-2)}$$

在 H_0 下服从自由度为 $(1, n-2)$ 的 F 分布。取否定域 $W = \{F > \lambda\}$ 其中 λ 为 $F(1, n-2)$ 分布的 $1-\alpha$ 分位数，α 为检验水平。

4.2.3 相关系数

在实际工作中，要判断变量 y 与变量 x 之间的线性相关性，人们还常用 y 与 x 的**样本相关系数**

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

r 是把 y 和 x 看成随机变量时其相关系数的距估计。虽然这里我们把 x 作为非随机的普通变量，我们还是可以用估计量 r 来作为 y 与 x 的线性相关性的度量。事实上，

$$r^2 = \frac{U}{l_{yy}} = 1 - \frac{Q}{l_{yy}}$$

所以 r^2 越大说明 y 与 x 线性相关性越强。

r 与最小二乘估计 $\hat{b}$ 和 F 统计量还有如下关系：

$$r = \hat{b}\sqrt{\frac{l_{xx}}{l_{yy}}},\ F = \frac{r^2}{(1-r^2)/(n-2)}$$

故对 $H_0 : b = 0$ 的检验可通过 r 进行，但临界值仍根据上式用 F 分布来求。

N.B. 在用统计量 F 检验假设 $H_0 : b = 0$ 时，我们用到了假定“e_1，e_2，$\cdots$，e_n 为独立同 $N(0,\sigma^2)$ 分布的随机变量（$\sigma^2 > 0$ 未知）”。如何根据数据判别这个假设是否成立，又是一个检验问题。定义**残差**$\hat{e}_i = y_i - \hat{y}_i$ $(i = 1, \cdots, n)$ 和**学生化残差**

$$\gamma_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-p_i}}\ (i = 1, \cdots, n)$$

其中

$$\hat{\sigma} = \sqrt{\frac{Q}{n-2}},\ Q = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}\hat{e}_i^2,\ p_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{l_{xx}},\ l_{xx} = \sum_{k=1}^{n}(x_k - \bar{x})^2$$

数学上可以证明，如果上述假定（“e_1，e_2，$\cdots$，e_n 为独立同 $N(0,\sigma^2)$ 分布的随机变量（$\sigma^2 > 0$ 未知）”）成立，则 γ_1，$\cdots$，γ_n 近似地相互独立，且各 γ_i 近似地服从标准正态分布（韦博成等 [4] §3.1）。从而在 n 比较大时，γ_1，$\cdots$，γ_n 中大约有 $[0.95n]$ 个 γ_i 满足 $|\gamma_i| \le 2$。反之，若根据数据算出满足不等式 $|\gamma_i| \le 2$ 的 γ_i 的个数与 $[0.95n]$ 差别很大，则表明上述假定不成立，这时用统计量 F 检验 $H_0 : b = 0$ 就缺乏根据了。

4.2.4 预测和控制

当我们用 $\hat{y}_0 = \hat{a} + \hat{b}x_0$ 去预测 y_0 时，由于 y_0 是随机变量，故 $\hat{y}_0$ 不是第二章中对参数的估计，而是对随机变量的估计，我们也可以构造 y_0 的置信区间。如果我们假定“e_1，e_2，$\cdots$，e_n 为独立同 $N(0,\sigma^2)$ 分布的随机变量（$\sigma^2 > 0$ 未知）”，则

$$T = \frac{y_0 - \hat{y}_0}{\sqrt{dQ/(n-2)}}$$

服从 $n-2$ 个自由度的 t 分布，其中

$$d = 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}},\ \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

给定 $\alpha \in (0,1)$，从 t 分布表中可找到临界值 λ 满足

$$P(|T| \le \lambda) = 1 - \alpha$$

于是可得 y_0 的 $1-\alpha$ 水平置信区间

$$[\hat{y}_0 - \lambda\sqrt{dQ/(n-2)}, \hat{y}_0 + \lambda\sqrt{dQ/(n-2)}$$

当自变量 x_0 在横轴上变化时，置信区间的上、下限的相应变化形成双曲线。

若要以 $1-\alpha$ 的概率控制 y_0 取值在 $[A,B]$ 内，则 x_0 必须且只需满足不等式

$$\begin{cases} \hat{a} + \hat{b}x_0 - \lambda\sqrt{dQ/(n-2)} \ge A \\ \hat{a} + \hat{b}x_0 + \lambda\sqrt{dQ/(n-2)} \le B \end{cases}$$

解上述不等式可得 $x_0 \in [C_1, C2]$。

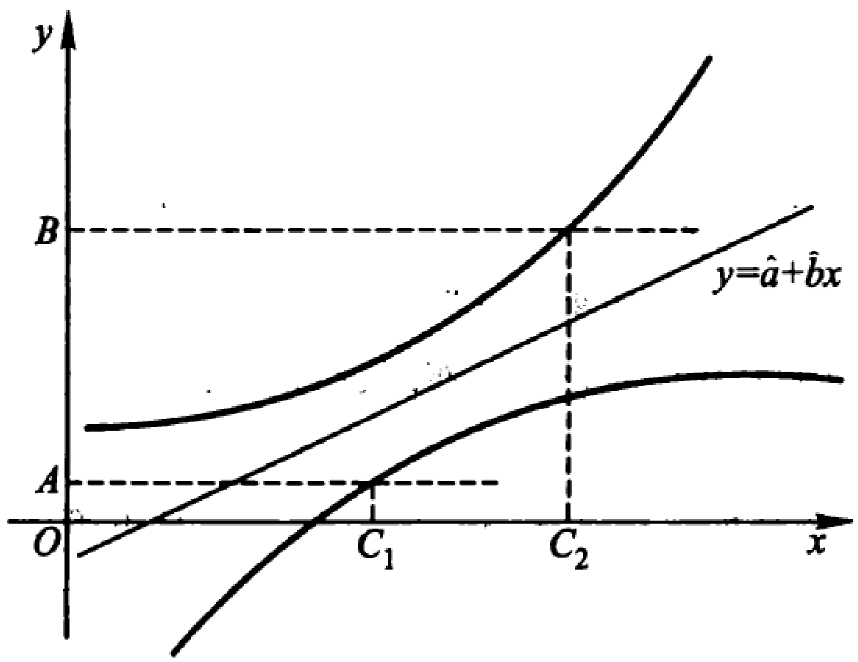

图 1: 预测和控制的置信区间

4.3 线性模型的参数估计

4.3.1 线性模型参数的最小二乘估计

使用矩阵和向量记号，我们记

$$Y=\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix},\ X=(x_{ij})_{n\times p},\ e=\begin{bmatrix}e_1\\e_2\\\vdots\\e_n\end{bmatrix},\ \beta=\begin{bmatrix}\beta_1\\\beta_2\\\vdots\\\beta_p\end{bmatrix}$$

则多元线性回归的一般形式是（称为**线性模型**）

$$Y=X\beta+e \tag{4.2}$$

其中 e 为 n 维随机向量，X 为 $n\times p$ 常数矩阵（其中第一列可以都是常数 1 以包括常数项非零的情形），β 为 p 维未知参数向量。对随机项 e 常做下列两种假定：

- **假定** A：$E[e]=0, Cov(e,e)=\sigma^2I$。其中 I 为单位阵，$Cov(\xi,\eta)$ 表示 ξ 与 η 的协方差阵。
- **假定** B：$Y\sim N(X\beta,\sigma^2I)$。

N.B.：线性模型研究只要不涉及假设检验和置信区间一般只用到假定 A。

定义 4.1. 对线性模型（*4.2*），称 $\hat{\beta}$ 是 β 的**最小二乘估计**，如果对一切 β，有

$$Q(\beta)\geq Q(\hat{\beta})$$

其中 $Q(\beta)=\|Y-X\beta\|^2$（这里的记号 $Q(\cdot)$ 与一元线性回归中残差平方和的记号 Q 是一致的）。

定理 4.2. （1）最小二乘估计 $\hat{\beta}$ 一定存在。

（2）$\hat{\beta}$ 是最小二乘估计的充要条件是 $\hat{\beta}$ 适合模型（4.2）的**正规方程**

$$X'X\beta = X'Y$$

（3）$Q(\hat{\beta}) = Y'Y - Y'X\hat{\beta}$

（4）若 X 满秩，则最小二乘估计唯一且

$$\hat{\beta} = (X'X)^{-1}X'Y,\ Q(\hat{\beta}) = Y'(I - X(X'X)^{-1}X')Y$$

定理 4.3. 对线性模型（4.2），设 X 满秩，且假定 A 成立，则对 β 的最小二乘估计 $\hat{\beta}$ 有下列结论：

（1）$E[\hat{\beta}] = \beta$（$\hat{\beta}$ 是 β 的无偏估计）。

（2）$Cov(\hat{\beta}, \hat{\beta}) = \sigma^2(X'X)^{-1}$

（3）$E[Q(\hat{\beta})] = (n-p)\sigma^2$（$\hat{\sigma}^2 = \frac{1}{n-p}Q(\hat{\beta}) = \frac{1}{n-p}\|Y - X\hat{\beta}\|$ 是 σ^2 的无偏估计）。

4.3.2 线性可估性

当 X 不是满秩的时候，单个 β_i 可能没有无偏估计，但多个 β_i 的线性组合却有可能有无偏估计。

定义 4.2. 称 $c'\beta = \sum_{i=1}^{n} c_i\beta_i$ 是**（线性）可估的**，若存在 Y 的线性函数 $a'Y = \sum_{i=1}^{n} a_i y_i$ 使得 $E[a'Y] = c'\beta$ （$\forall\beta \in \mathbb{R}^p$）。

注 4. 要理解上述定义中的“$\forall\beta \in \mathbb{R}^p$”，应将“可估性”看作是对向量 c 的要求，使得存在一个通用的 a，对任何被 β 参数化的线性模型 $Y = X\beta + \varepsilon$ 都有 $E[a'Y] = c'\beta$。

若 X 满秩，则 $\forall c \in \mathbb{R}^p$ 都有 $c'\beta$ 可估，且 $a' = c'(X'X)^{-1}X'$。下面的定理讨论了一般情形

定理 4.4. 对线性模型（4.2），设假定 A 成立，为了 $c'\beta$ 可估，必须且只须 c' 是 X 的行的线性组合，即 $c \in \mu(X')$。

定理 4.5 (高斯-马尔科夫). 对线性模型（4.2），设假定 A 成立，$\hat{\beta}$ 是 β 的最小二乘估计，那么如果 $c'\beta$ 线性可估，$c'\hat{\beta}$ 必为 $c'\beta$ 的唯一的最小方差线性无偏估计。

注 5. 如果 $c'\beta$ 可估，并且对线性模型（4.2）的假定 B 成立，则可以证明，$c'\hat{\beta}$ 是所有 $c'\beta$ 的无偏估计（不限于线性的）中方差最小的。

下面的定理把定理4.3的第三条结论推广到了 X 不满秩的情形。

定理 4.6. 对线性模型（4.2），设假定 A 成立，矩阵 X 的秩为 r，则

$$\hat{\sigma}^2 = \frac{Q(\hat{\beta})}{n-r}$$

是 σ^2 的无偏估计。

注 6. 如果假定 B 成立，可以进一步证明 $\hat{\sigma}^2$ 是 σ^2 的最小方差无偏估计。

4.3.3 带约束的线性模型的参数估计

约束最小二乘估计永远存在，因为优化的目标函数是个凸函数（二次型），而线性限制条件定义了一个凸集。它的存在性也不依赖于假定 A 或者假定 B。对于带约束的最小二乘估计，也可以建立相应的高斯-马尔科夫定理。

有两种方法可以求约束最小二乘估计：（1）消去多余参数法；（2）拉格朗日乘子法。

4.3.4 关于最小二乘估计的讨论

当 X 满秩时，所得的最小二乘估计是无偏的而且是所有线性无偏估计中方差最小的。但并不是在任何情况下，最小二乘估计都好。

（1）当行列式 $|X'X|$ 接近零时，最小二乘估计 $\hat{\beta}$ 很差，取值很不稳定，这是因为其方差矩阵 $Cov\ \hat{\beta},\hat{\beta}) = \sigma^2(X'X)^{-1}$。在这种情况下，一个重要的替代估计是**岭估计**：

$$\hat{\beta} = (X'X + \lambda I)^{-1}X'Y$$

其中 λ 是正参数。

（2）最小二乘估计可能受个别数据的较大影响。一个比最小二乘估计稳健的估计是**最小一乘估计**。

4.4 线性模型的假设检验

4.4.1 参数线性相关性的检验

本节恒假设假定 B 成立，即 $\varepsilon \sim N(0, \sigma^2 I)$。考虑检验问题

$$H_0 : H\beta = 0 \leftrightarrow H_a : H\beta \neq 0 \tag{4.3}$$

其中 H 为 $s \times p$ 矩阵。

我们用广义似然比来构造检验（4.3）的检验统计量。令

$$W = \mu(X) = \{\eta = X\beta : \beta \in \mathbb{R}^p\} \subset \mathbb{R}^n,\ W_0 = \{\eta = X\beta : H\beta = 0, \beta \in \mathbb{R}^p\} \subset W$$

记

$$\Theta = \{(\xi, \sigma^2) : \xi \in W, \sigma^2 > 0\},\ \Theta_0 = \{(\xi, \sigma^2) : \xi \in W_0, \sigma^2 > 0\}$$

在假定 B 之下，Y 的分布密度为

$$p(Y; \xi, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|Y - \xi\|^2\right\}$$

这也是参数 (ξ, σ^2) 的似然函数。不难看出，要使这个函数达到最大值，必须且只须选 ξ 使 $\|Y - \xi\|^2$ 最小，再选 σ^2 使表达式达到最大值。记

$$\hat{\xi} = \mathrm{Proj}_W Y,\ \hat{\xi}_0 = \mathrm{Proj}_{W_0} Y$$

易知 $p(Y, \xi, \sigma^2)$ 在 $\xi = \hat{\xi}$，$\sigma^2 = \frac{1}{n}\|Y - \hat{\xi}\|^2$ 时达到最大值 $\sup_{(\xi,\sigma^2)\in\Theta} p(Y, \xi, \sigma^2)$，在 $\xi = \hat{\xi}_0$，$\sigma^2 = \frac{1}{n}\|Y - \hat{\xi}_0\|^2$ 时达到最大值 $\sup_{(\xi,\sigma^2)\in\Theta_0} p(Y, \xi, \sigma^2)$。于是广义似然比

$$\lambda = \frac{\sup_{\Theta} p(Y, \xi, \sigma^2)}{\sup_{\Theta_0} p(Y, \xi, \sigma^2)} = \left[\frac{\|Y - \hat{\xi}_0\|^2}{\|Y - \hat{\xi}\|^2}\right]^{n/2} = \left[1 + \frac{\|\hat{\xi} - \hat{\xi}_0\|^2}{\|Y - \hat{\xi}\|^2}\right]^{n/2}$$

定理 4.7. 对于线性模型(*4.2*),设假定 B 成立,则在假设 H_0 成立时,统计量 F 的分布为 $F(r-q,n-r)$。

推论 4.1. 对模型 *(4.2)*，设假定 B 成立，$\hat{\beta}$ 为 β 的最小二乘估计，则 $X\hat{\beta}$ 与残差 $\hat{e}$ 独立，从而 $X\hat{\beta}$ 与残差平方和 Q 独立，且

$$Q/\sigma^2 \sim \chi^2(n-r)$$

4.4.2 参数线性组合的检验与置信区间

定理 4.8. 对模型（*4.2*），设假定 B 成立，$c'\beta$ 是 β 的可估线性组合，$\hat{\beta}$ 是 β 的最小二乘估计，Q 为残差平方和，则 $c'\hat{\beta}$ 与 Q 相互独立。

推论 4.2. 对模型（*4.2*），设假定 B 成立，如果 X 满秩，则 $\hat{\beta}$ 与 Q 独立，且

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}),\ Q/\sigma \sim \chi^2(n-p)$$

定理 4.9. 对模型（*4.2*），设假定 B 成立，并设 $c'\beta$ 可估（$c \neq 0$），则

$$\frac{c'(\hat{\beta}-\beta)}{\hat{\sigma}\|a\|} \sim t(n-r)$$

其中 a 为 c 的伴随元，r 为 X 的秩，$\hat{\sigma} = \sqrt{\frac{Q}{n-r}}$。特别的，若 X 满秩，则

$$\frac{c'(\hat{\beta}-\beta)}{\hat{\sigma}\sqrt{c'(X'X)^{-1}c}} \sim t(n-p)$$

此定理允许我们进行假设检验和构造置信区间。

4.5 回归分析

前面两节中建立的线性模型的一般理论可以用来研究回归分析问题。

4.5.1 回归参数估计

4.5.2 预测

4.5.3 控制

4.5.4 回归模型中的假设检验

4.5.5 回归模型的残差分析

如何判别“假定 B”是否成立?

在假定 B 之下，**残差**$\hat{e} = Y - X\hat{\beta}$ 服从正态分布，且

$$Cov(\hat{e},\hat{e}) = \sigma^2(I-P),\ E[\hat{e}_i] = 0,\ Var(\hat{e}_i) = \sigma^2(1-p_{ii})$$

其中 $P = X(X'X)^{-1}X'$。

令

$$\hat{\sigma} = \sqrt{\frac{Q}{n-p-1}},\ \gamma_i = \frac{\hat{e}_i}{\hat{\sigma}\sqrt{1-p_i i}},\ i = 1, \cdots, n$$

在假定 B 之下，只要 n 相当大，**学生化残差**γ_1，$\cdots$，γ_n 近似地相互独立同分布，且 γ_i 近似服从标准正态分布。

如果假定 B 成立，n 又比较大，则 γ_1，$\cdots$，γ_n 中大约有 $[0.95n]$ 个 γ_i 满足 $|\gamma_i| \le 2$。若根据实际数据算出的 γ_1，$\cdots$，γ_n 中有较多的 γ_i 不满足这个不等式，则应拒绝假定 B。

4.5.6 第二类回归简介

第二类回归与第一类回归的计算公式完全相同。

4.5.7 关于回归分析的几点注意事项

（1）对实际问题，要根据问题的物理背景适当地选择模型。如果没有理论的指导可以用不同的模型进行比较。一般地，可以比较模型的回归 F 统计量的值，值越大，说明模型越合适。同时还应该看一看残差对自变量的散布图（对一元回归而言）。

（2）使用回归方法要注意，变量的统计相关不等于变量之间的因果关系。

（3）下一节讲的回归自变量选择方法只解决了部分问题，在选择自变量的问题上还是要进行深入的分析研究。

（4）用回归方程作预报或控制，必须考虑问题的背景是否发生了变化。

4.6 回归变量的选择

逐步回归法。

4.7 逻辑斯谛回归模型

逻辑斯谛模型为

$$\ln\frac{p}{1-p} = \beta_0 + \sum_{i=1}^{k}\beta_i x_i$$

逻辑斯谛模型的参数估计通常有两个办法：最大似然估计法和加权最小二乘法。

最大似然估计法。逻辑斯谛模型的对数似然函数是关于参数的多元严格凸函数。在似然方程组有解时，解唯一且是对数似然函数的最大值点，因此方程组的解是参数的最大似然估计。但应注意的是，似然方程组有时无解。

加权最小二乘法。此法对数据有些特殊要求，我们以 $\beta = (\beta_0, \beta_1)$ 为例予以说明。设 $x = x_i$ 时对 Y 作了 n_i 次观测（n_i 较大），其中事件 $\{Y = 1\}$ 发生了 r_i 次（$i = 1, 2, \cdots, m$），这里 x_1，$\cdots$，x_m 两两不同。通常用

$$z_i = \ln\frac{r_i + 0.5}{n_i - r_i + 0.5}$$

作为 $\ln\frac{p(x_i)}{1-p(x_i)}$ 的估计值（$i=1,2,\cdots,m$）。令

$$v_i=\frac{(n_i+1)(n_i+2)}{n_i(r_i+1)(n_i-r_i+1)},\ \tilde{Q}(\beta_0,\beta_1)=\sum_{i=1}^{m}\frac{1}{v_i}(z_i-\beta_0-\beta_1x_i)^2$$

使 $\tilde{Q}(\beta_0,\beta_1)$ 达到最小值的 $\tilde{\beta}_0$，$\tilde{\beta}_1$ 分别称为β_0，β_1 **的加权最小二乘估计**。

加权最小二乘估计的想法，在于用 $\frac{r_i}{n_i-r_i}$ 估计 $\frac{p(x_i)}{1-p(x_i)}$ （加上 0.5 是为了避免分子和分母出现零）。基于概率论中的极限定理，可以证明 $z_i=\ln\frac{r_i+0.5}{n_i-r_i+0.5}$ 近似服从正态分布 $N\left(\ln\frac{p(x_i)}{1-p(x_i)},\frac{1}{n_ip(x_i)(1-p(x_i))}\right)$。于是 $z_i=\ln\frac{p(x_i)}{1-p(x_i)}+\varepsilon_i=\beta_0+\beta_1x_i+\varepsilon_i$，这里 ε_i 近似服从 $N\left(0,\frac{1}{n_ip(x_i)(1-p(x_i))}\right)$。再用 v_i 估计 $\frac{1}{n_ip(x_i)(1-p(x_i))}$，则有

$$\frac{1}{\sqrt{v_i}}z_i=\frac{1}{\sqrt{v_i}}(\beta_0+\beta_1x_i)+\frac{1}{\sqrt{v_i}}\varepsilon_i$$

可以使用通常的最小二乘法。

4.8 习题

5 试验设计与方差分析

6 序贯分析初步

7 统计决策与贝叶斯统计大意

7.1 统计决策问题概述

$(\Theta,\underset{\sim}{X},A,L(\theta,a))$ 叫做**决策问题的四个要素**。其中 Θ 是**参数空间**，$\underset{\sim}{X}$ 是 X 的样本，A 是**行动空间**，$L(\theta,a)$ 是定义在 $\Theta\times A$ 上的非负函数，表示参数是 θ 时采取行动 a 引起的损失，叫做**损失函数**。**统计决策问题**是：如何根据样本 $\underset{\sim}{X}$ 的值恰当地选取行动 a 使得引起的损失尽可能的小。

称样本空间到行动空间 A 的映射 $\delta=\delta(x_1,\cdots,x_n)$ 为**决策函数**，简称**决策**。

在估计问题中，通常取 $A=\Theta$，则决策就是估计量。在假设检验问题中，参数空间 $\Theta=\Theta_0\cup\Theta_1$，$\Theta_0$ 与 Θ_1 不相交，设 H_i 是 $\theta\in\Theta_i$（$i=0,1$），行动空间 $A=\{a_0,a_1\}$，其中 a_i 表示接受 H_i（$i=0,1$），则给出决策等价于给出对 H_0 的否定域。

设 δ 是一个决策，称平均损失 $R(\theta,\delta)=E[L(\theta,\delta(X_1,\cdots,X_n)]$ 为δ **的风险**。风险 $R(\theta,\delta)$ 是 θ 的函数，对一切 θ 风险最小的决策（所谓**一致最优决策**）难得存在。如果一致最优决策不存在，只好去找比较“优良”的决策了。有两个途径探讨这个问题。一是对决策 δ 作一定限制，缩小选择的范围，在较小的范围内找最优的。另一是放宽比较的要求，不要求风险函数对一切 δ 都最小。

称决策 δ^* 是 **minimax 决策**，若对一切决策 δ 成立

$$\sup_{\theta}R(\theta,\delta^*)\le\sup_{\delta}R(\theta,\delta).$$

minimax 决策是一种保守的决策，在许多情形下应尽量避免采用保守的决策。为此，应将 θ 看成一个随机变量，风险 $R(\theta,\delta)$ 关于 θ 的分布的平均值 $\rho(\delta)=E[R(\theta,\delta]$ 是评价决策 δ 的优良性指标，数值越小越好。

7.2 什么是贝叶斯统计

设总体 X 有分布密度（或概率函数）$f(x,\theta)$，其中 θ 是未知的，但知 θ 的变化范围是 Θ，经典方法把参数 θ 看作是客观常数，通过对样本 $\underset{\sim}{X}=(X_1,\cdots,X_n)$ 的研究对 θ 给出估计值或者推断 θ 属于某个给定的范围。贝叶斯学派的根本观点，是认为在关于 θ 的任何统计推断问题中，除了使用样本 $\underset{\sim}{X}$ 提供的信息外，还必须对 θ 规定一个先验分布，它是进行推断时不可缺少的要素。也即把 θ 看成随机变量，它服从某个概率分布（叫做**先验分布**），总体 X 的分布实际上是 θ 给定时 X 的条件分布。

在样本 $\underset{\sim}{X}=(X_1,\cdots,X_n)=\underset{\sim}{x}=(x_1,\cdots,x_n)$ 的条件下，θ 的条件分布密度（或条件概率函数）$\xi(\theta|\underset{\sim}{x})$ 称为**后验分布**。贝叶斯方法的关键在于所作出的任何推断都只须根据后验分布 $\xi(\theta|\underset{\sim}{x})$，而不再涉及样本 $\underset{\sim}{X}$ 的分布。但在如何使用 $\xi(\theta|\underset{\sim}{x})$ 上还有较大的灵活性，涉及到行动空间和损失函数的类型。

称 $\delta^*=\delta^*(x_1,\cdots,x_n)$ 是**贝叶斯决策**，若

$$\rho(\delta^*)=\inf_{\delta}\rho(\delta),$$

这里 $\rho(\delta)=E_\xi[R(\theta,\delta)]$ 是对风险 $R(\theta,\delta)$ 取在关于 θ 的概率分布 ξ 下的平均值。此时 $\rho(\delta^*)$ 叫做**贝叶斯风险**。

注意贝叶斯决策依赖于先验分布。先验分布变了，贝叶斯决策一般也要变。贝叶斯统计的基本任务就是针对先验分布 $\xi(\theta)$，找出贝叶斯决策。

定理 7.1 (寻找贝叶斯决策)**.** 令

$$\rho(x_1,\cdots,x_n)=\inf_{a\in A}\int_\Theta L(\theta,a)\xi(\theta|x_1,\cdots,x_n)d\theta$$

这个 $\rho(x_1,\cdots,x_n)$ 是样本值为 $x_1,\cdots,x_n$ 时的最小后验平均损失。若决策 $\delta^*=\delta^*(x_1,\cdots,x_n)$ 满足

$$\int_\Theta L(\theta,\delta^*(x_1,\cdots,x_n))\xi(\theta|x_1,\cdots,x_n)d\theta=\rho(x_1,\cdots,x_n),\ (\text{对一切}x_1,\cdots,x_n)$$

也即采用决策 δ^* 引起的后验平均损失达到最小值，则 δ^* 就是贝叶斯决策。

为了计算后验分布，可以证明

命题 7.1. 当 θ 的先验分布密度是 $\xi(\theta)$ 时，在 $\underset{\sim}{X}=(x_1,\cdots,x_n)$ 的条件下 θ 的后验分布密度是

$$\xi(\theta|x_1,\cdots,x_n)=\frac{\prod_{i=1}^n f(x_i,\theta)\xi(\theta)}{\int_\Theta\prod_{i=1}^n f(x_i,\theta)\xi(\theta)d\theta}$$

序贯统计决策由停止法则 τ 及判决法则 δ 组成。停止法则 τ 告诉我们何时停止观测，是不依赖于将来的随机变量；判决法则 δ 告诉我们，如何根据序贯样本 $X_1,\cdots,X_\tau$ 选取行动 $\delta(X_1,\cdots,X_\tau)$（$\delta$ 的值属于 A）。**序贯统计决策** (τ,δ) **的总风险**为

$$\rho(\tau,\delta)=E[L(\theta,\delta(X_1,\cdots,X_\tau))+\tau C]$$

（损失和费用之和的平均值）。这里 E 表示 θ 之先验分布为 ξ 及给定 θ 时诸 X_i 的分布密度为 $f(x,\theta)$ 时计算期望。称 (τ^*,δ^*) 是**贝叶斯序贯决策**，若

$$\rho(\tau^*,\delta^*)=\inf_{(\tau,\delta)}\rho(\tau,\delta).$$

7.3 先验分布的确定

7.4 应用实例——电视机寿命验证试验的贝叶斯方法

7.5 习题

▶ 1. 设 X 是离散或连续随机变量，EX^2 存在。试证明：为了使 $E(X-a)^2$（a 是实数）达到最小值，必须且只须 $a = EX$。

证明. 我们注意到有 $E(X-a)^2 = a^2 - 2EX \cdot a + E(X^2)$。所以 $E(X-a)^2$ 在 $a = EX$ 达到最小值。 □

▶ 2. 设 X 是离散或连续随机变量，EX 存在。试证明：为了使 $E|X-a|$（a 是实数）达到最小值，必须且只须 a 是 X 的中位数。

证明. 首先回顾中位数的定义：给定随机变量 ξ，它的中位数 $med(\xi)$ 是满足如下条件的实数 α：

$$\min\{P(\xi \le \alpha), P(\xi \ge \alpha)\} \ge \frac{1}{2}.$$

特别的，如果 ξ 的分布函数是连续函数，则 $P(\xi \le \alpha) = \frac{1}{2}$。其次，为了看出题中结论的正确性，可以形式地做如下运算：

$$\frac{d}{da}E|X-a| = E\left[\frac{d}{da}|X-a|\right] = E\left[1_{\{X\le a\}} - 1_{\{X>a\}}\right] = 2P(X \le a) - 1.$$

所以极值点是满足 $P(X \le a) = \frac{1}{2}$ 的 a。

严格的证明如下。对任意的 a，我们有

$$\begin{aligned} E|X-a| &= E[(X-a)1_{\{X>a\}} + (a-X)1_{\{X<a\}}] = E\left[\int_{-\infty}^{\infty} 1_{\{a<t\le X\}}dt + \int_{-\infty}^{\infty} 1_{\{X<t\le a\}}dt\right] \\ &= \int_a^{\infty} P(t \le X)dt + \int_{-\infty}^{a} P(X<t)dt. \end{aligned}$$

于是对任意的 $b > a$，我们有

$$E|X-b| - E|X-a| = -\int_a^b P(t\le X)dt + \int_a^b P(X<t)dt = \int_a^b [1 - 2P(t \le X)]dt$$

注意 $1-2P(t\le X)$ 是 t 的递增函数，故 $E|X-b| - E|X-a| \ge 0$ 对任意的 $b \in (a,\infty)$ 成立，当且仅当 $1-2P(a\le X)\ge 0$，也即 $P(a \le X) \le \frac{1}{2}$。所以 $P(a \ge X) \ge P(a > X) = 1 - P(a \le X) \ge \frac{1}{2}$。

类似地，对任意的 $b < a$，我们有

$$E|X-b| - E|X-a| = \int_b^a P(t\le X)dt - \int_b^a P(X<t)dt = \int_b^a [2P(t\le X) - 1]dt$$

注意 $2P(t\le X)-1$ 是 t 的递减函数，故 $E|X-b| - E|X-a| \ge 0$ 对任意的 $b \in (\infty, a)$ 成立，当且仅当 $2P(a\le X) - 1 \ge 0$，也即 $P(a \le X) \ge \frac{1}{2}$。

由中位数的定义，我们可知 $E|X-a|$ 在中位数 $med(X)$ 取得最小值。 □

注 7. 另一直观证明如下：先证明当 X 取有限个数值时命题成立，然后用一列取有限值的离散随机变量 $\{X_n\}_n$ 逼近 X，即可证明一般情形（需要用到实变函数论中的控制收敛定理）。为证明有限个取值的情形，假设 X 的取值为 $x_1, x_2, \cdots, x_N$ （$x_1 < x_2 < \cdots < x_N$）并定义 $x_{N+1} = \infty$。则当 $a < x_1$，a 每向右移动 Δ，$E|X-a|$ 就减少 $\sum_{i=1}^N p_i\Delta = \Delta$。当 a 处于 $[x_{i_0}, x_{i_0+1})$ 时（$i = 1, \cdots, N$），a 每向右移动 Δ，$E|X-a|$ 就增加 $\Delta\sum_{i=1}^{i_0} p_i$，并同时减少 $\Delta\sum_{i=i_0+1}^{N} p_i$。由此可见 a 的最佳取值为中位数。

8 抽样调查概述

参考文献

[1] 陈家鼎、孙山泽、李东风、刘力平：《数理统计学讲义（第 2 版）》。北京：高等教育出版社，2006。

[2] 陈希孺：《数理统计学简史》。长沙：湖南教育出版社，2002。

[3] 汪仁官：《概率论引论》。北京：北京大学出版社，1994。

[4] 韦博成、鲁国斌、史建清：《统计诊断引论》。南京：东南大学出版社，1991。

[5] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis*, 3rd edition. Cambridge University Press, 1920.