

第三章：参数估计与分布理论

3.1 最小二乘估计(Least Squared Estimate) 线性模型

$$Y_{n \times 1} = X_{n \times p} \beta_{p \times 1} + e_{n \times 1},$$

其中 $Ee = 0$ ，通常对误差 e 两种假定：

1. $Cov(e) = \sigma^2 I_n$ ， σ^2 未知；

2. $e \sim N(0, \sigma^2 I_n)$ ， σ^2 未知(比 1 强)。

由最小二乘法的思想， β 的估计应选择使得 $Q(\beta) = \|e\|^2 = \|Y - X\beta\|^2 = (Y - X\beta)'(Y - X\beta)$ 达到最小。

若 $\hat{\beta}$ 使得 $\|Y - X\hat{\beta}\|^2 = \min_{\beta} \|Y - X\beta\|^2$ ，则称 $\hat{\beta}$ 为 β 的一个最小二乘解。极小化 $Q(\beta)$ ，则 β 需满足方程 $\frac{\partial Q(\beta)}{\partial \beta} = 0$ ，即得到正规方程

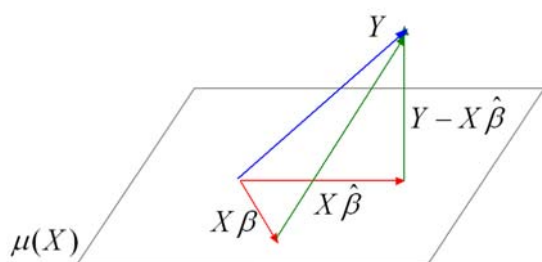
$$X'X\beta = X'Y.$$

正规方程的所有解为 $(X'X)^- X'Y$ 。

定理 3.1.1: $\hat{\beta}$ 是最小二乘解 $\Leftrightarrow \hat{\beta}$ 是正规方程的解，即 $\hat{\beta} = (X'X)^- X'Y$ 。

最小二乘法几何解释：

$$\min_{\beta} \|Y - X\beta\|^2 = \min_{\theta \in \mu(X)} \|Y - \theta\|^2$$



当 $rank(X) = p$ 时，最小二乘解有唯一解 $\hat{\beta} = (X'X)^{-1} X'Y$ ，且此时 $\hat{\beta}$ 为 β 的无偏估计，即 $E\hat{\beta} = \beta$ ，方差 $Var(\hat{\beta}) = \sigma^2 (X'X)^{-1}$ 。

当 $rank(X) < p$ 时，最小二乘解不唯一，此时最小二乘解中无 $\hat{\beta}$ 能作为 β 的无偏估计。此外可以证明此时 β 的无偏估计不存在，此时 β 称为不可估的(nonestimable)。

定义 3.1.1: $c'\beta$ 为 β 的某一线性函数 (c 已知)，若存在 Y 的线性函数 $a'Y$ 使得 $Ea'Y = c'\beta$ ， $\forall \beta$ ，则称 $c'\beta$ 是可估函数。

可估性的含义可由下面的例子形象说明。

例 1: 设两个物体重量 β_1, β_2 未知，把它们同时放在天平上称 n 次，第 i 次结果为 y_i ，模型为

$$y_i = \beta_1 + \beta_2 + e_i, i = 1, \dots, n.$$

其中 e_i 为第 i 次称量误差，且设 $e = (e_1, \dots, e_n)'$ ， $Ee = 0, Var(e) = \sigma^2 I_n$ 。

令 $c = (1, 1)'$ ， $\beta = (\beta_1, \beta_2)'$ ，则 $c'\beta = \beta_1 + \beta_2$ 是可估的，，例如 y_1 就是其一个无偏估计，但 β_1, β_2 都不可估。 β_1 (或 β_2) 不可估的理由很清楚，因为每次称量都是两物体一起称，当然无法由结果对其中单独一个物体的重量作“估计”。

定理 3.1.2: 以下三条等价

1. $c'\beta$ 可估;

2. $X\beta_1 = X\beta_2 \Rightarrow c'\beta_1 = c'\beta_2$;

3. $c \in \mu(X')$ 。

注: 1. 一切 $c'\beta$ 可估 $\Leftrightarrow \text{rank}(X) = p$ 。

2. 若 $c'_1\beta, c'_2\beta$ 可估, 则任线性组合 $\lambda_1 c'_1\beta + \lambda_2 c'_2\beta$ 也是可估。若 c_1, c_2 线性无关, 称 $c'_1\beta, c'_2\beta$ 也线性无关。对线性模型来说之多有 $\text{rank}(X)$ 个线性无关的可估函数。

3. 若 $c'\beta$ 可估, $\hat{\beta}$ 是最小二乘解, 则 $c'\hat{\beta}$ 值唯一, 与广义逆 $(X'X)^-$ 的选取无关; 此外 $c'\hat{\beta}$ 是 $c'\beta$ 的无偏估计。

定义 3.1.2: 设 $c'\beta$ 可估, 称 $c'\hat{\beta}$ 为 $c'\beta$ 的最小二乘估计。

例 2: 考虑如下模型

$$y_{ij} = \mu + \alpha_i + e_{ij}, i=1,2; j=1,2.$$

$$e_{ij} \text{ i.i.d } Ee_{ij} = 0, Ee_{ij}^2 = \sigma^2.$$

μ, α_1, α_2 不可估, 但 $\alpha_1 - \alpha_2$ 可估。

若 $c'\beta$ 可估, $a'Y$ 为一无偏估计, 对 $\forall b \in \mu(X)^\perp, (a+b)'Y$ 都是 $c'\beta$ 的无偏估计。在所有线性无偏估计中, 找出方差最小的估计, 此估计称为**最优线性无偏估计**(Best Linear Unbiased Estimate, 简写成 **BLUE**), 或称为 **Gauss-Markov 估计**(GM 估计)。

定理 3.1.3: (Gauss-Markov 定理)若 $c'\beta$ 可估, 则 $c'\hat{\beta}$ 是其唯一的 GM 估计($\hat{\beta}$ 为 β 的 LS 估计)。

注: 在一切线性无偏估计类中, $c'\hat{\beta}$ 是方差最小的, 但不排除比 $c'\hat{\beta}$ 方差更小的非线性无偏估计; 但若误差分布还是正态分布, 则这种可能性不存在。

下面考虑 σ^2 的估计。令 $\hat{e} = Y - X\hat{\beta} = (I_n - P_X)Y$, 称为**残差**(residual)向量, $E\hat{e} = 0, \text{Cov}(\hat{e}) = \sigma^2(I_n - P_X)$ 。

定理 3.1.4: 设 $r = \text{rank}(X_{n \times p})$, 则

$$\hat{\sigma}^2 = \frac{\hat{e}'\hat{e}}{n-r} = \frac{\|Y - X\hat{\beta}\|^2}{n-r} = \frac{Y'(I_n - P_X)Y}{n-r}$$

为 σ^2 的无偏估计。

3.2 分布理论

迄今为止, 关于误差的假定是满足 GM 条件, 若误差还服从正态分布即 $e \sim N_n(0, \sigma^2 I_n)$, 则可以确定一些估计量的精确分布。

定理 3.2.1: 在误差正态分布假设下, 设 $c'\beta$ 为可估函数, $\hat{\beta} = (X'X)^- X'Y$, $\text{rank}(X) = r$, 则:

1. $c'\hat{\beta}$ 为 $c'\beta$ 的极大似然估计 (MLE), 且 $c'\hat{\beta} \sim N(c'\beta, \sigma^2 c'(X'X)^- c)$;

2. $\frac{n-r}{n} \hat{\sigma}^2$ 是 σ^2 的 MLE, 且 $\frac{(n-r)\hat{\sigma}^2}{\sigma^2} \sim \chi_{n-r}^2$;

3. $c'\hat{\beta}$ 与 $\hat{\sigma}^2$ 独立。

注: 对可估函数来说, 其 LSE 与 MLE 一致; 但对误差方差的估计, LSE 是无偏的, MLE 是有偏的。

*定理 3.2.2: 在正态误差假设下

1. $T_1 = Y'Y$, $T_2 = X'Y$ 是 **完全、充分统计量**;
2. 若 $c'\beta$ 可估, 则 $c'\hat{\beta}$ 是唯一 **最小方差无偏估计** (Minimum Variance Unbiased Estimate, 简写 **MVUE**);
3. $\hat{\sigma}^2$ 为 σ^2 的 MVUE。

3.3 有线性约束时的估计

考虑线性模型, 其系数 β 满足线性约束条件 $L\beta = d$ (此约束为相容性方程)。不失一般性可假设 $d = 0$ (否则, 取 β_0 使得 $L\beta_0 = d$, 令 $\tilde{Y} = Y - X\beta_0$, $\tilde{\beta} = \beta - \beta_0$, 考虑线性模型 $\tilde{Y} = X\tilde{\beta} + e$, 此时线性约束变为 $L\tilde{\beta} = 0$)。本节考虑线性模型如下

$$Y = X\beta + e, Ee = 0, Cov(e) = \sigma^2 I_n$$

$$L_{q \times p} \beta = 0$$

由最小二乘法, 此时 β 的估计应选择在 $L\beta = 0$ 的条件下极小化 $\|Y - X\beta\|^2$, 即选择 $\hat{\beta}_L = \text{Arg} \min_{L\beta=0} \|Y - X\beta\|^2$ 。引入 Lagrange 乘子 $\lambda_{q \times 1}$, 考虑 $Q(\beta, \lambda) = \|Y - X\beta\|^2 + 2\lambda'L\beta$, 由 $\frac{\partial Q}{\partial \beta} = 0, \frac{\partial Q}{\partial \lambda} = 0$ 得到方程 $X'X\beta + L'\lambda = X'Y$, $L\beta = 0$, 即

$$\begin{pmatrix} X'X & L' \\ L & 0 \end{pmatrix} \begin{pmatrix} \beta \\ \lambda \end{pmatrix} = \begin{pmatrix} X'Y \\ 0 \end{pmatrix} \quad (*)$$

引理 3.3.1: 设 $S = \{A_{n \times m}x | B_{k \times m}x = 0, x \in R^m\}$, 则 S 为线性子空间且 $\dim S = \text{rank} \begin{pmatrix} A \\ B \end{pmatrix} - \text{rank}(B)$ 。

引理 3.3.2: 设 $V_{p \times p} \geq 0$, A 为 $p \times q$ 矩阵, 则

1. $\mu(A) \cap \mu(VA^\perp) = \{0\}$;
2. $\mu(V:A) = \mu(VA^\perp:A)$ 。

回到方程 (*), 由 $L\beta = 0$ 解得 $\beta = (I_p - L'L)z, \forall z_{p \times 1}$, 代入前一个方程得

$$X'X(I_p - L'L)z + L'\lambda = X'Y \quad (**)$$

由引理 3.3.2, 注意到 $I_p - L'L = (L')^\perp$, 有 $\mu(X') = \mu(X'X) \subset \mu(X'X:L') = \mu(X'X(I_p - L'L):L')$ 。因此方程(**)是相容得, 从而方程(*)也是相容的。

定义 3.3.1: 若存在 a 使得 $Ea'Y = c'\beta$ 对所有满足 $L\beta = 0$ 的 β 成立, 则称 $c'\beta$ 是 **条件可估函数**, $a'Y$ 为 $c'\beta$ 的 **条件无偏估计**。

定理 3.3.1: 在线性约束 $L\beta = 0$ 的条件下, 线性函数 $c'\beta$ 是条件可估函数 $\Leftrightarrow c \in \mu(X':L')$ 。

定理 3.3.2: 在线性约束 $L\beta=0$ 的条件下,

1. 约束最小二乘解为 $\hat{\beta}_L = G_{11}XY$, 其中,

$$\begin{pmatrix} XX' & L' \\ L & 0 \end{pmatrix}^{-1} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix};$$

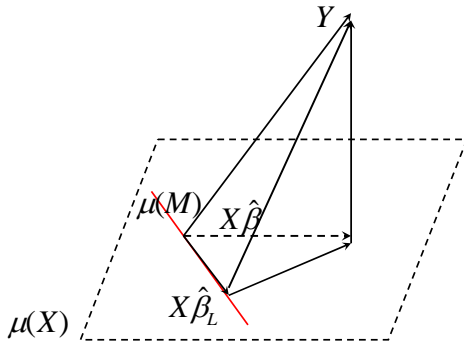
2. 对任何条件可估函数 $c'\beta$, $c'\hat{\beta}_L$ 值唯一 (不依赖广义逆的选取) 且

$$\text{Var}(c'\hat{\beta}_L) = \sigma^2 c'G_{11}c;$$

3. $c'\hat{\beta}_L$ 是条件可估函数 $c'\beta$ 的唯一 BLUE。

几何解释:

$X\hat{\beta}_L$ 是 Y 到子空间 $\mu(M) = \{X\beta | L\beta=0\}$ 的正交投影, 这里 $M = X(L')^\perp = X(I_p - LL')$ 。由于 $X\hat{\beta}_L = XG_{11}XY$, 故 $XG_{11}X'$ 为 $\mu(M)$ 上的投影算子, 即 $P_M = XG_{11}X'$ 。



推论 3.3.2: 在一些条件下 $\hat{\beta}_L$ 有简单的形式。设 $\text{rank}(L_{q \times p}) = q$,

1. 若 $\mu(L') \subset \mu(X')$, 则

$$\hat{\beta}_L = \hat{\beta} - (XX')^{-1}L'[L(XX')^{-1}L']^{-1}L\hat{\beta}$$

其中 $\hat{\beta} = (XX')^{-1}XY$;

2. 若 $\text{rank}(X_{n \times p}) = p$, 则

$$\hat{\beta}_L = \hat{\beta} - (XX')^{-1}L'[L(XX')^{-1}L']^{-1}L\hat{\beta}$$

其中 $\hat{\beta} = (XX')^{-1}XY$ 。

定理 3.3.3: 在线性约束 $L\beta=0$ 条件下,

$$\hat{\sigma}_L^2 = \frac{\|Y - X\hat{\beta}_L\|^2}{n-s} = \frac{Y'(I_n - P_M)Y}{n-s} \text{ 为 } \sigma^2 \text{ 的条件}$$

无偏估计, 这里 $M = X(L')^\perp$

$$s = \text{rank}\begin{pmatrix} X \\ L \end{pmatrix} - \text{rank}(L)。$$

定理 3.3.4: 在约束 $L\beta=0$ 的条件下, 若进一步假设误差还是正态分布, $c'\beta$ 条件可估, 则:

$$1. c'\hat{\beta}_L \sim N(c'\beta, \sigma^2 c'G_{11}c);$$

$$2. \frac{(n-s)\hat{\sigma}_L^2}{\sigma^2} \sim \chi_{n-s}^2, \quad s = \text{rank}\begin{pmatrix} X \\ L \end{pmatrix} - \text{rank}(L);$$

$$3. c'\hat{\beta}_L \text{ 与 } \hat{\sigma}_L^2 \text{ 独立。}$$

当 $\text{rank}(X_{n \times p}) < p$ 时, β 不可估, 对 $c \in \mu(X')$, $c'\beta$ 可估; 在带有约束 $L\beta = 0$ 的条件下, 对 $c \in \mu(X':L')$, $c'\beta$ (条件) 可估。增加约束后, 可估函数的选取范围“扩大”了。加上怎样的约束 $L\beta = 0$, 当然, 加上约束后要与原模型一致(等价), 可以使得 $\forall c$, $c'\beta$ (条件) 可估?

设线性模型 $Y = X\beta + e$, $r = \text{rank}(X) < p$, 由最小二乘法, 实质是找 Y 在空间 $\mu(X)$ 的投影; 加上约束 $L\beta = 0$ 后, 找 Y 在空间

$S = \{X\beta | L\beta = 0, \beta \in R^p\}$ 上的投影, 要使两模型等价, 由于 $S \subset \mu(X)$, 要求 $S = \mu(X) \Leftrightarrow \dim S = \dim \mu(X) = \text{rank}(X)$,

由于 $\dim S = \text{rank} \begin{pmatrix} X \\ L \end{pmatrix} - \text{rank}(L)$, 所以当

$$\text{rank} \begin{pmatrix} X \\ L \end{pmatrix} - \text{rank}(L) = \text{rank}(X)$$

即 $\mu(X') \cap \mu(L') = \{0\}$ 时, 加上约束与原模型本质上一致。

另一方面, 在约束 $L\beta = 0$ 下, 要对 $\forall c$, $c'\beta$ (条件) 可估, 则 $\mu(X':L') = R^p$, 即 $\text{rank} \begin{pmatrix} X \\ L \end{pmatrix} = p$ 。因此, 若 $L_{q \times p}$ 满足条件:

$$\mu(X') \cap \mu(L') = \{0\} \text{ 且 } \text{rank} \begin{pmatrix} X \\ L \end{pmatrix} = p \quad (\#)$$

则加上约束与不加约束线性模型等价, 此时对 $\forall c$, $c'\beta$ 在条件 $L\beta = 0$ 下可估。若 L 满足条件 (#), 则称约束 $L\beta = 0$ 为 **side condition (边界条件)**。

以下不失一般性, 可以假设 $L_{q \times p}$ 是行满秩的, 若 L 满足条件 (#), 则 $q = p - \text{rank}(X)$ 。

引理 3.3.1: 设 $L\beta = 0$ 为 side condition, 则

$$X'X(X'X + L'L)^{-1}X' = X',$$

$$L(X'X + L'L)^{-1}X' = 0.$$

定理 3.3.5: 设有约束 $L\beta = 0$ 的线性模型, 若 $L\beta = 0$ 为 side condition, 则

1. 约束最小二乘问题的解 $\hat{\beta}_L$ 唯一且

$$\hat{\beta}_L = (X'X + L'L)^{-1}X'Y;$$

2. $\hat{\beta}_L$ 为方程 $\begin{cases} X'X\beta = X'Y \\ L\beta = 0 \end{cases}$ 的唯一解;

3. $\hat{\beta}_L$ 为 β 在 $L\beta = 0$ 条件下的无偏估计且对任何可估函数 $c'\beta$ ($c \in \mu(X')$), $c'\hat{\beta}_L = c'\hat{\beta}$ (这里 $\hat{\beta} = (X'X)^{-}X'Y$)。

上述定理表明, 对通常线性模型, 正规方程 $X'X\beta = X'Y$ 的解 $\hat{\beta}$ 一般不唯一。若加上一个 side condition, 问题转化求解 $\begin{cases} X'X\beta = X'Y \\ L\beta = 0 \end{cases}$, 此时解 $\hat{\beta}_L$ 唯一, 且对任何可估 $c'\beta$, $c'\hat{\beta}_L = c'\hat{\beta}$ 。Side condition 让线性模型正规方程有一个特殊的解。由于满足 side condition 的 L 选择并不唯一, $\forall D_{q \times q}$ 可逆, $DL\beta = 0$ 都是一个 side condition。

可以选择一个“好”的 side condition 使得方程 $\begin{cases} X'X\beta = X'Y \\ L\beta = 0 \end{cases}$ 解很容易求出。

例 3: 考虑下面线性模型:

$$y_{ij} = \mu + \alpha_i + e_{ij}, i = 1, \dots, n; j = 1, \dots, m.$$

写成矩阵形式 $Y = X_{nm \times (n+1)}\beta + e$, $\beta = (\mu, \alpha_1, \dots, \alpha_n)'$, $\text{rank}(X) = n < n+1$, side condition 为一个方程, 选择 $L\beta = \sum_{i=1}^n \alpha_i = 0$, 易解的 $\hat{\mu} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m y_{ij}$, $\hat{\alpha}_i = \frac{1}{m} \sum_{j=1}^m y_{ij} - \hat{\mu}$ 。

3.4 Aitken 模型与广义最小二乘

到目前为止,我们都是在误差 $\text{Cov}(e) = \sigma^2 I_n$ 下讨论估计问题,此时任何可估函数的 LS 估计与其 GM 估计一致。但在许多实际情况下,误差协方差阵为 $\text{Cov}(e) = \sigma^2 \Sigma$, $\Sigma > 0$ (已知),在此条件下线性模型 $Y = X\beta + e, Ee = 0$, 称为 Aitken 模型。

由于 $\Sigma > 0$ 为已知,则作变换 $\tilde{Y} = \Sigma^{-1/2}Y, \tilde{X} = \Sigma^{-1/2}X, \tilde{e} = \Sigma^{-1/2}e$, 此时 $\tilde{Y} = \tilde{X}\beta + \tilde{e}, E\tilde{e} = 0, \text{Cov}(\tilde{e}) = \sigma^2 I_n$ 。

令 $Q(\beta) = \|\tilde{Y} - \tilde{X}\beta\|^2 = (Y - X\beta)' \Sigma^{-1} (Y - X\beta)$, 由最小二乘法原理,此时 β 的估计应选择 $\beta^* = \text{Arg min}_{\beta} Q(\beta)$ 。此时得到 Aitken 方程:

$$X' \Sigma^{-1} X \beta = X' \Sigma^{-1} Y,$$

解 $\beta^* = (X' \Sigma^{-1} X)^{-} X' \Sigma^{-1} Y$, 称为 **广义最小二乘解**。

注: 如果 Σ 为对角阵 $\text{diag}(\sigma_1, \sigma_2, \dots, \sigma_m)$, 令 $W = \Sigma^{-1} = \text{diag}(w_1, w_2, \dots, w_m)$, 这里 $w_i = \sigma_i^{-2}$, 此时 β^* 也称加权最小二乘估计。

由于可估性只涉及 Y 的均值与其方差无关,且 $\mu(\tilde{X}) = \mu(X')$, 故此时 $c'\beta$ 可估 $\Leftrightarrow c \in \mu(X')$ 。

注: 由于 $X(X' \Sigma^{-1} X)^{-} X' \Sigma^{-1}$ 不依赖于广义逆的选取, 令 $A = X(X' \Sigma^{-1} X)^{-} X' \Sigma^{-1}$, 由于 $AX = X$, 因此 $A^2 = A$ 为幂等矩阵。

定理 3.4.1: 对任何可估函数 $c'\beta$, $c'\beta^*$ 是唯一的 BLUE, 其方差 $\text{Var}(c'\beta^*) = \sigma^2 c'(X' \Sigma^{-1} X)^{-} c$ 。

定理 3.4.2: $\sigma^{*2} = \frac{(Y - X\beta^*)' \Sigma^{-1} (Y - X\beta^*)}{n - r}$,

$r = \text{rank}(X)$ 为 σ^2 的无偏估计。

定理 3.4.3: 若假设 $e \sim N(0, \sigma^2 \Sigma), \Sigma > 0$, 则

1. 对任何可估 $c'\beta$, $c'\beta^* \sim N(c'\beta, \sigma^2 c'(X' \Sigma^{-1} X)^{-} c)$;

2. $\frac{(n-r)\sigma^{*2}}{\sigma^2} \sim \chi_{n-r}^2$ 且 $c'\beta^*$ 与 σ^{*2} 独立;

3. 若 $\text{rank}(X_{n \times p}) = p$, 则

$\beta^* \sim N_p(\beta, \sigma^2 (X' \Sigma^{-1} X)^{-1})$ 且与 σ^{*2} 独立。

3.5 稳健回归与 M -估计

到目前为止,我们对线性模型都采用最小二乘法给出回归系数的参数估计。前面也指出,在误差 $e \sim N(0, \sigma^2 I_n)$ 条件下,回归系数的最小二乘估计还是极大似然估计。但也很多证据表明当误差分布偏离正态分布时,最小二乘估计可能不再是有效的估计(估计的方差会偏大)。即便误差正态分布,如果有些异常值影响,最小二乘法都会过于敏感,给出很“坏”的估计。

例 3.5.1: 下图显示异常值对线性模型最小二乘估计的影响。

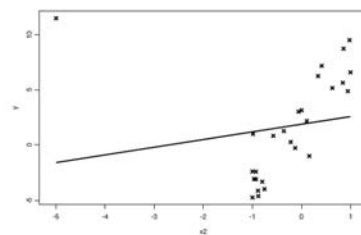
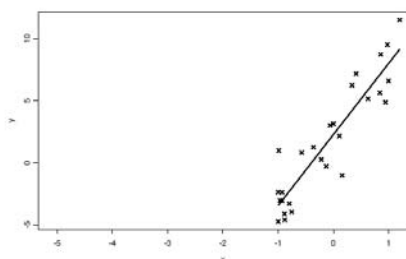


Figure Data with 27 points and the corresponding least squares regression line (top) and the sensitivity of least squares regression to an outlier in the x -direction (bottom).

Huber(1964)首先考虑位置参数的稳健(robust)估计, 提出一类现在称为 M -估计的估计方法。对线性模型来说, 给定一合适的函数 $\rho(\cdot)$, 考虑最小化问题

$$\min_{\beta} \sum_{i=1}^n \rho(y_i - x_i' \beta),$$

上式最小化问题的解 $\hat{\beta}_n$, 称为回归系数 β 的 M -估计。

注: 最小二乘法相当于 $\rho(x) = x^2$ 。

如果, 函数 ρ 的导数 $\varphi = \dot{\rho}$ 处处存在且连续, 则 M -估计 $\hat{\beta}_n$ 为下方方程的一解

$$\sum_{i=1}^n x_i \varphi(y_i - x_i' \beta) = 0.$$

注: 在一些情况下, 导函数 $\varphi = \dot{\rho}$ 不连续或者函数 ρ 在一些至多可数的点不存在, 即 $\varphi = \dot{\rho}$ 除去这些点存在, 此时上方方程可能无解。此时, M -估计的定义只能回到原始的极值问题。

M -估计的一个常用特例是取 L_1 范数, 对应函数 $\rho(x) = |x|$, 此时最小化

$$\min_{\beta} \sum_{i=1}^n |y_i - x_i' \beta|,$$

其解 $\hat{\beta}_n$, 称为 least absolute deviation (LAD) 估计。研究表明, 最小二乘法之所以不稳健是因为 $\rho(x) = x^2$, $|\varphi(x)| = |\dot{\rho}(x)| = 2|x|$, 当 $|x|$ 过大时增长太快, 会“放大”异常值的影响。而 LAD 方法是稳健的方法, 因为 $\rho(x) = |x|$, $|\varphi(x)| = |\dot{\rho}(x)| = 1$ (在 0 处导数不存在), 增长平缓。

Huber(1964)提出一类函数, 现在称为 Huber 函数, 设 $\varphi = \dot{\rho}$, 取

$$\varphi_H(x) = \begin{cases} x & |x| \leq k \\ k \operatorname{sign}(x) & |x| > k \end{cases},$$

此时得到的估计称为 Huber 估计。

一些人建议减少异常值影响, 当 $x \rightarrow \pm\infty$ 时

$$\varphi(x) \rightarrow 0, \text{ 由此可取 } \varphi(x) = \frac{2x}{1+x^2}.$$

注: 关于 M -估计的渐近理论可参见陈希孺, 赵林城(1996), 线性模型中的 M 方法, 上海科学技术出版社。

3.6*最小二乘统一理论

对线性模型：

$$Y = X\beta + e, \quad Ee = 0, \quad \text{Cov}(e) = \sigma^2 \Sigma (\Sigma \text{已知})$$

若 $\Sigma > 0$ ，按照广义最小二乘法，只需求解关于参数 β 的二次函数 $Q(\beta) = (Y - X\beta)' \Sigma^{-1} (Y - X\beta)$ 的极小值问题；若 $|\Sigma| = 0$ ，此时称为**奇异线性模型**，由于 Σ^{-1} 不存在， $Q(\beta)$ 无定义，若用广义逆 Σ^- 代替 Σ^{-1} ，把 $Q(\beta)$ 定义为 $Q(\beta) = (Y - X\beta)' \Sigma^- (Y - X\beta)$ ，

则由于 $Q(\beta)$ 与广义逆 Σ^- 的选择有关，不同的 Σ^- ， $Q(\beta)$ 不同，因此极小化 $Q(\beta)$ 无意义。因此对于奇异线性模型，一个核心问题是寻找一个新矩阵 T ，能够充当 $Q(\beta)$ 中 Σ^{-1} 所担负的作用。C.R.Rao 成功的解决了这个问题，他定义

$$T = \Sigma + XUX', \quad \text{其中 } U \geq 0,$$

且使得 $\text{rank}(T) = \text{rank}(\Sigma : X)$ ，然后令 $Q(\beta) = (Y - X\beta)' T^- (Y - X\beta)$ ， β 的估计应选择 $\beta_R^* = \text{Arg min}_{\beta} Q(\beta)$ 。

后面将证明，对于可估函数 $c'\beta$ ，其 BLUE 就是 $c'\beta_R^*$ 。而且此方法适用于设计矩阵 X 列满秩或不满秩， Σ 奇异或非奇异的所有情形，因此此方法和结果称为最小二乘的统一理论。

引理 3.6.1：对本节线性模型， $Y \in \mu(\Sigma : X)$ 以概率 1 成立。

引理 3.6.2：对上述定义的 T ，

$$1. \mu(\Sigma : X) = \mu(T);$$

2. $X'T^-X, X'T^-Y$ 和 $Q(\beta)$ 不依赖广义逆 T^- 的选取。

引理 3.6.3：对线性模型，可估函数 $c'\beta$ 的某一个无偏估计 $a'Y$ 为其 BLUE \Leftrightarrow 对任意 0 的无偏估计 $b'Y$ 总有 $\text{Cov}(a'Y, b'Y) = 0$ 。

注：满足 Rao 定义的 T 总是存在的，例如取 $U = k \cdot I_n (k > 0)$ 。特别当 $\Sigma > 0$ 或 $\mu(X) \subset \mu(\Sigma)$ 时，可以取 $U = 0$ ，此时 $T = \Sigma$ 。

极小化 $Q(\beta)$ 得相容性方程 $X'T^-X\beta = X'T^-Y$ ，解 $\beta_R^* = (X'T^-X)^- X'T^-Y$ 。

定理 3.6.1：对本节线性模型

1. 对任何可估函数 $c'\beta$ ，其 BLUE 为 $c'\beta_R^*$ ；

$$2. \text{Var}(c'\beta_R^*) = \sigma^2 c' [(X'T^-X)^- - U] c;$$

3. σ^2 的无偏估计为

$$\sigma_R^{*2} = \frac{(Y - X\beta_R^*)' T^- (Y - X\beta_R^*)}{q}$$

这里 $q = \text{rank}(T) - \text{rank}(X)$ 。

最后作为本章结束提一下两步估计方法。当 $\text{Cov}(e) = \sigma^2 \Sigma(\theta)$ ，其中 $\Sigma(\theta) > 0$ 含有未知参数 θ ，先设法对 θ 作估计，得到 $\hat{\theta}$ ，从而得到 $\Sigma(\theta)$ 的估计 $\hat{\Sigma}(\hat{\theta})$ ，最后用广义最小二乘的思想得到 β 的估计 $\hat{\beta}(\hat{\theta}) = (X' [\hat{\Sigma}(\hat{\theta})]^- X)^- X' [\hat{\Sigma}(\hat{\theta})]^- Y$ 。对于可估函数 $c'\beta$ ，在一定条件下 $c'\hat{\beta}(\hat{\theta})$ 为 $c'\beta$ 的无偏估计。