Statistical Inference

Guanxu Gleason WANG

University of Glasgow

版本：Past Paper of STAT5028 Statistical Inference, UofG

更新：December 18, 2024

目录

1 极大似然估计 (Maximum Likelihood Estimate, MLE)

1.1 One Parameter

1.1.1 用极大似然法估计参数

随机变量 X 服从关于参数 θ 的某分布, 假设它的 p.m.f. (discrete) 或 p.d.f. (continuous) 为

$$\Pr(X = x) \quad \text{or} \quad f(\theta; x),$$

那么:

- The likelihood function:

$$L(\theta) = \prod_{i=1}^{n} \Pr(X = x_i) \quad \text{or} \quad L(\theta) \propto \prod_{i=1}^{n} f(\theta; x_i);$$

- The log-likelihood function:

$$\ell(\theta) = \log(L(\theta)) = \log\left(\prod_{i=1}^{n} \Pr(X = x_i)\right) = \sum_{i=1}^{n} \log\left(\Pr(X = x_i)\right);$$

- Set $\ell'(\theta) = 0$ to find a turning point $\widehat{\theta}_{\mathrm{MLE}}$;
- Check to ensure it is a maximum turning point:

$$\ell''(\theta) < 0, \ \forall\theta.$$

1.1.2 Relative likelihood

- The relative likelihood function:
$$R(\theta) = \frac{L(\theta)}{L(\widehat{\theta}_{\mathrm{MLE}})};$$
- The relative likelihood function:
$$r(\theta) = \log(R(\theta)) = \ell(\theta) - \ell(\widehat{\theta}_{\mathrm{MLE}}).$$
- A p likelihood interval:
$$R(\theta) \geqslant p \quad \text{or} \quad r(\theta) \geqslant \log(p).$$

Specially, $1-\alpha$ confidence interval (CI) is equivalent to a likelihood interval (LI) of
$$\exp\left(-\frac{z^2}{2}\right), \text{where } z = \Phi^{-1}(1-\alpha).$$

例 1.1 (@ 2021-2022 T3-(b)) Assuming normality holds, a 95% CI is equivalent to a 14.65% likelihood interval, defined by:
$$r(\delta) \geqslant \log(0.1465) = 1.9208,$$
where
$$14.65\% = \exp\left(-\frac{\left(\Phi^{-1}(0.95)\right)^2}{2}\right) = \exp\left(-\frac{1.96^2}{2}\right) = \exp(-1.9208),$$
i.e.
$$r(\delta) \geqslant -\frac{\left(\Phi^{-1}(0.95)\right)^2}{2} = -\frac{1.96^2}{2} = -1.9208.$$

1.1.3 Wilks intervals

The pivotal (枢轴量) function:

$$\mathrm{PIV}(\theta) = -2r(\theta) = 2\left(r(\widehat{\theta}_{\mathrm{MLE}}) - r(\theta)\right) \sim \chi_1^2,$$

hence an approximate $100c\%$ confidence (Wilks) interval for θ:

$$\left\{\theta : -2r(\theta) \leqslant \chi_1^2(c)\right\}$$

1.1.4 Newton-Raphson method

The Newton-Raphson approach can be used to find the roots on an equation ($f(x) = 0$).

The following is iterated (for iteration n) until convergence (收敛) for each bound B:

$$x_B^{(n+1)} = x_B^{(n)} - \frac{f\left(x_B^{(n)}\right)}{f'\left(x_B^{(n)}\right)}, \quad \text{until } x_B^{(n+1)} = x_B^{(n)}.$$

In the case of likelihood interval, $f(\theta_B) = r(\theta_B) - s$, where s is the objective value.

1.1.5 Generalised Likelihood Ratio Test (GLRT)

The GLRT test statistic is

$$\lambda = 2(\ell(\widehat{\theta}_{H_1}) - \ell(\widehat{\theta}_{H_0})) \sim \chi_{df}^2(0.95),$$

We compare this to the χ^2_{df} distribution whose upper α point is $\chi^2_{df}(1-\alpha)$, since there is $[df]$ restriction on the parameters under H_0.

1.2 Two Parameters

1.2.1 用极大似然法估计参数

- The joint likelihood function:

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

例 1.2 (@ 2022-2023 T1-(a)) Assume that the data follow independent Binomial probability models, so that:

$$Y_1 \sim \mathrm{Bi}(30, \theta_1), 0 < \theta < 1$$
$$Y_2 \sim \mathrm{Bi}(25, \theta_2), 0 < \theta < 1$$

The observed values of the number of samples above the threshold for each reservoir are $y_1 = 16$ and

$y_2 = 13$. 由于这里只有两个独立的观测值 Y_1, Y_2, 因此联合似然函数写作

$$
\begin{aligned}
L(\theta_1, \theta_2) &= \Pr(Y_1 = y_1 \mid 30, \theta_1) \cdot \Pr(Y_2 = y_2 \mid 25, \theta_2) \\
&= \binom{30}{y_1} \theta_1^{y_1} (1-\theta_2)^{30-y_1} \cdot \binom{25}{y_2} \theta_2^{y_2} (1-\theta_2)^{25-y_2} \\
&= K \cdot \theta_1^{16} (1-\theta_2)^{14} \cdot \theta_2^{13} (1-\theta_2)^{12} \qquad K = \binom{30}{16} \cdot \binom{25}{13}
\end{aligned}
$$

- The joint log-likelihood function:

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i; \boldsymbol{\theta});$$

- Derive the first partial derivatives and set equal to zero to find a turning point:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = 0;$$

- Check to ensure it is a maximum turning point (Hessian matrix):

$$H = \begin{bmatrix} \dfrac{\partial^2 \ell}{\partial \theta_1^2} & \dfrac{\partial^2 \ell}{\partial \theta_1 \partial \theta_2} \\ \dfrac{\partial^2 \ell}{\partial \theta_2 \partial \theta_1} & \dfrac{\partial^2 \ell}{\partial \theta_2^2} \end{bmatrix}.$$

If H is a diagonal matrix and all entries on the main diagonal are negative, H will be negative definite for all values for θ_1 and θ_2 and hence the estimators θ_1 and θ_2 are maximum likelihood estimators.

1.2.2 Wald confidence interval

If we now wish to estimate $\boldsymbol{b}^\top \boldsymbol{\theta}$, we have

$$\boldsymbol{b}^\top \widehat{\boldsymbol{\theta}} \pm \Phi^{-1}\left(1 - \frac{\alpha}{2}\right) \sqrt{\boldsymbol{b}^\top K^{-1} \boldsymbol{b}},$$

where $K^{-1} = -H^{-1}$.

例 1.3 (@ 2022-2023 T1-(a)) If we now wish to estimate $\theta_1 - \theta_2$ then by setting $\boldsymbol{b} = (1, -1)^\top$, that is

$$\widehat{\theta}_1 - \widehat{\theta}_2 = \boldsymbol{b}^\top \widehat{\boldsymbol{\theta}} = \begin{bmatrix} 1 & -1 \end{bmatrix} \begin{bmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{bmatrix}.$$

Assume that

$$H = \begin{bmatrix} -h_1 & 0 \\ 0 & -h_2 \end{bmatrix}, h_1 > 0, h_2 > 0,$$

then

$$K^{-1} = -H^{-1} = \begin{bmatrix} \dfrac{1}{h_1} & 0 \\ 0 & \dfrac{1}{h_2} \end{bmatrix}.$$

so, $\boldsymbol{b}^\top K^{-1} \boldsymbol{b} = \dfrac{1}{h_1} + \dfrac{1}{h_2}$.

例 1.4 (@ 2021-2022 T2-(c)) 若所估计的参数组合并非 $\widehat{\theta}_1, \widehat{\theta}_2$ 的形式, 而是 $f(\widehat{\theta}_1), f(\widehat{\theta}_2)$, 那么

$$K^{-1} = \begin{bmatrix} \mathrm{Var}\left[f(\widehat{\theta}_1)\right] & 0 \\ 0 & \mathrm{Var}\left[f(\widehat{\theta}_2)\right] \end{bmatrix},$$

显然这种做法的前提条件是 $f(\widehat{\theta}_1), f(\widehat{\theta}_2)$ 彼此独立.

2 Bayesian Inference

For the case of one data point x and a parameter θ:

$$p(\theta \mid X) = \frac{p(\theta) \cdot L(\theta \mid X)}{p(x)} \propto p(\theta) \cdot L(\theta \mid X),$$

- $p(\theta \mid X)$: the posterior distribution (后验分布);
- $L(\theta \mid X) = p(X \mid \theta)$: the likelihood for a parameter θ given the data x (似然函数);
- $p(\theta)$: the prior distribution of the parameter θ (先验分布);
- $p(x)$: the conditional (marginal) distribution of the data.

通过似然函数找到共轭先验分布:

例 2.1 (Binomial Model) 用 Bayesian Inference 来估计成功比例, 因为若 $\theta \sim \mathrm{Beta}(\alpha, \beta)$, 则 $\mathbb{E}[\theta] = \dfrac{\alpha}{\alpha+\beta}$, 其中 α, β 分别为成功和失败次数的先验假设.

似然函数 $L(\theta \mid X)$	共轭 (conjugate) 先验分布 $p(\theta)$
$X \sim \text{Binomial}(\theta)$	$\theta \sim \text{Beta}(\alpha, \beta),\ p(\theta) = \dfrac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)}$
$X \sim \text{Poisson}(\lambda)$	$\lambda \sim \text{Gamma}(\alpha, \beta),\ p(\lambda) = \dfrac{\lambda^{\alpha-1}e^{-\frac{1}{\beta}\lambda}}{\beta^{\alpha}\Gamma(\alpha)}$
$X \sim \text{Exp}(\lambda)$	
$X \sim \text{Normal}(\mu, \sigma^2)$	$\theta \sim \text{Normal}(\mu, \sigma^2)$

- Likelihood: $L(\theta \mid X)$;
- Prior (Beta): $p(\theta)$;
- Posterior: $p(\theta \mid X) \propto p(\theta) \cdot L(\theta \mid X) \Longrightarrow \theta \mid X \sim \text{Beta}(\widetilde{\alpha}, \widetilde{\beta})$;
- The mean of the posterior distribution: $\mathbb{E}\left[\theta \mid X\right] = \dfrac{\widetilde{\alpha}}{\widetilde{\alpha} + \widetilde{\beta}}$.

3 Hypothesis Testing

- H_0: Null hypothesis;
- H_1: Alternative hypothesis (study hypothesis: 即题目中所研究的内容).

Notion:

- η: population median (总体中位数);
- μ: population mean (总体均值);
- θ: population proportion/probability (总体比例/概率).

结论要么是we reject H_0 in favour of H_1, 要么是we do not reject H_0, 绝不可能是prove H_0 from a sample.

3.1 Confidence Interval

3.1.1 One-sample

$1-\alpha$ confidence interval:

$$\overline{x} \pm t_{1-\frac{\alpha}{2}}(n-1) \cdot \frac{s}{\sqrt{n}},$$

where:

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

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right].$$

3.1.2 Two-sample

$1-\alpha$ confidence interval:

$$\overline{x}-\overline{y}\pm t_{1-\frac{\alpha}{2}}(n+m-2)\cdot s_p\sqrt{\frac{1}{n}+\frac{1}{m}},$$

where:

$$s_p^2=\frac{(n-1)s_x^2+(m-1)s_y^2}{n+m-2},$$

as $(n-1)$ and $(m-1)$ are the degree of freedom of x and y respectively, such that the total degree of freedom is $(n-1)+(m-1)=(n+m-2)$.

3.2 Wilcoxon Signed Ranks Test (non-parametric test)

例 3.1 (@ 2022-2023 T2-(b)) Test assumptions:

- Observations come from a symmetric distribution (对称分布);
- Observations are independent from one another (彼此独立).

Then H_0 :?? vs. H_1 :??.

Observed value	-1	2	5	3	2	-2	5	8
$\lvert d_i \rvert$	1	2	5	3	2	2	5	8
sign	-	+	+	+	+	-	+	+
rank	1	3	6.5	5	3	3	6.5	8

- Test statistic: $W = \min(W+, W-) = 4$, where (sum of rank)

$$W+ = 3 + 6.5 + 5 + 3 + 6.5 + 8 = 32,$$
$$W- = 1 + 3 = 4.$$

- Rejection region: $RR = \{W : W \leqslant ?\}$

3.3 Generalised Likelihood Ratio Test (GLRT)

例 3.2 (@ 2022-2023 T2-(c)) Use the Generalised Likelihood Ratio Test to investigate the hypothesis.

- $H_0 : \theta_1 = \theta, \theta_2 = 2\theta, \theta_3 = 4\theta;$
- H_1 : the multinomial probabilities will not have the specified form in H_0.

The cell probabilities must sum to 1, i.e. $\sum_{i=1}^{3} \theta_i = 1 \implies \theta + 2\theta + 4\theta = 1 \implies \theta = \frac{1}{7}$.

In this case, $n = 89 + 210 + 414 = 713$.

No.	1	2	3
Observed frequency of orders	89	210	414
Expected frequency of orders (under H_0)	101.8571	203.7143	407.4286
Reason	$713 \times \frac{1}{7} = 101.8571$	$713 \times \frac{2}{7} = 203.7143$	$713 \times \frac{4}{7} = 407.4286$

- The test statistic:

$$2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{E_i} \right).$$

- The observed value of the test statistic:

$$2 \left(89 \log \left(\frac{89}{101.8571} \right) + 210 \log \left(\frac{210}{203.7143} \right) + 414 \log \left(\frac{414}{407.4286} \right) \right) = 2 \times 0.9966 = 1.993.$$

- Rejection Region:

$$\begin{aligned} RR &= \left\{ x : 2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{E_i} \right) > \chi^2_{k-1-p}(1-\alpha) \right\} \\ &= \left\{ x : 2 \sum_{i=1}^{k} O_i \log \left(\frac{O_i}{E_i} \right) > \chi^2_{3-1-0}(1-\alpha) \right\}, \end{aligned}$$

where

  - $k =$ the number of levels (or categories) (类别个数);
  - $p =$ the number of parameters being estimated (待估参数个数).

3.4 Mann-Whitney U Test (双变量的秩检验) (non-parameter)

例 3.3 (@ 2021-2022 T3-(b)) First, order data

% (order)	62.81	63.80	64.71	66.22	67.12	85.08	91.65	94.54
rank	4	5	6	8	9	14	15	16
group	X	X	X	X	X	X	X	X
% (order)	31.15	32.45	50.89	65.05	67.20	70.89	71.01	72.65
rank	1	2	3	7	10	11	12	13
group	Y	Y	Y	Y	Y	Y	Y	Y

- Test statistic: $U = \min(U_X, U_Y) = 23$, where (sum of rank)

$$R_X = 4 + 5 + 6 + 8 + 9 + 14 + 15 + 16 = 77,$$
$$R_Y = 1 + 2 + 3 + 7 + 10 + 11 + 12 + 13 = 59.$$

理论上 R_i 的最小值为 $1 + 2 + \cdots + n_i = \dfrac{n_i(n_i+1)}{2}$, 因此通常 $U_i = R_i - \dfrac{n_i(n_i+1)}{2}$, 即

$$U_X = R_X - \frac{n_X(n_X+1)}{2} = 77 - \frac{8 \times (8+1)}{2} = 77 - 36 = 41,$$
$$U_Y = R_Y - \frac{n_Y(n_Y+1)}{2} = 59 - \frac{8 \times (8+1)}{2} = 59 - 36 = 23.$$

- Rejection region: $RR = \{U : U \leqslant ?\}$