Summary 10M

Complete Data

Complete data describes data where we know the exact value of each observation and have a full dataset.

The steps to calculate the MLE of a parameter θ are:

1. $$\begin{aligned} L(\theta) &= \prod_{\text{all } i} p(x_i) \qquad \text{(discrete)} \\ &= \prod_{\text{all } i} f(x_i) \qquad \text{(continuous)} \end{aligned}$$
2. $$l(\theta) = \ln L(\theta)$$
3. $$l'(\theta) = \frac{\mathrm{d}}{\mathrm{d}\theta} l(\theta)$$
4. $$l'(\theta) = 0$$

To estimate two parameters, take the first partial derivative of the likelihood/log-likelihood function with respect to each parameter. Then, set them both equal to zero and solve.

Incomplete Data

Left-truncation at d means the data below d was not recorded.

Right-truncation at u means the data above u was not recorded.

Left-censoring at d means the data below d is recorded as d.

Right-censoring at u means the data above u is recorded as u.

Case	Likelihood
Left-truncated at d	$\dfrac{p(x)}{\Pr(X>d)}$ (discrete) $\dfrac{f(x)}{\Pr(X>d)}$ (continuous)
Right-truncated at u	$\dfrac{p(x)}{\Pr(X<u)}$ (discrete) $\dfrac{f(x)}{\Pr(X<u)}$ (continuous)
Left-censored at d	$\Pr(X \le d)$
Right-censored at u	$\Pr(X \ge u)$
Grouped data on interval $(a, b]$	$\Pr(a < X \le b)$

Special Cases

Matching Moments

When we have complete data, the MLE of certain parameters can be calculated by matching the fitted moment to the sample moment.

Distribution	Shortcuts
Gamma, fixed α	$\hat{\theta} = \dfrac{\bar{x}}{\alpha}$
Exponential	$\hat{\theta} = \bar{x}$
Normal	$\hat{\mu} = \bar{x}$ $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2}{n}$
Lognormal	$\hat{\mu} = \dfrac{\sum_{i=1}^{n} \ln x_i}{n}$ $\hat{\sigma}^2 = \dfrac{\sum_{i=1}^{n}\left(\ln x_i - \hat{\mu}\right)^2}{n}$
Poisson	$\hat{\lambda} = \bar{x}$
Binomial, fixed m	$\hat{q} = \dfrac{\bar{x}}{m}$
Negative binomial, fixed r	$\hat{\beta} = \dfrac{\bar{x}}{r}$

Distribution	Shortcuts
Geometric	$\hat{\beta}=\bar{x}$

Uniform

$$X \sim \text{Uniform}\ (0,\ \theta)$$

$$\hat{\theta}=\max(x_1,\ x_2,\ \ldots,\ x_n)$$

Other Shortcuts

In this table,

- n is the number of uncensored data points;
- c is the number of censored data points;
- x_i is the observed value, or the censoring point for censored data;
- d_i is the truncation point.

Distribution	Shortcut
Pareto, fixed θ	$\hat{\alpha}=\dfrac{n}{\sum_{i=1}^{n+c}\left[\ln(x_i+\theta)-\ln(d_i+\theta)\right]}$
S-P Pareto, fixed θ	$\hat{\alpha}=\dfrac{n}{\sum_{i=1}^{n+c}\left\{\ln x_i-\ln\left[\max(\theta,\ d_i)\right]\right\}}$
Exponential	$\hat{\theta}=\dfrac{\sum_{i=1}^{n+c}(x_i-d_i)}{n}$
Inverse exponential	$\hat{\theta}=\dfrac{n}{\sum_{i=1}^{n}1/x_i}$
Weibull, fixed τ	$\hat{\theta}=\left(\dfrac{\sum_{i=1}^{n+c}x_i^{\tau}-\sum_{i=1}^{n+c}d_i^{\tau}}{n}\right)^{1/\tau}$
Beta, fixed $\theta, b=1$	$\hat{a}=\dfrac{n}{n\ln\theta-\sum_{i=1}^{n}\ln x_i}$
Beta, fixed θ, $a=1$	$\hat{b}=\dfrac{n}{n\ln\theta-\sum_{i=1}^{n}\ln(\theta-x_i)}$

Distribution	Shortcut
Uniform $[0,\ \theta]$, grouped data	$\hat{\theta} = c_j \left(\frac{n}{n - n_j} \right)$ n: total number of observations n_j: number of data points in the last interval $[c_j,\ \infty)$

From the table, the following shortcuts can be applied only when complete data is available:

- Inverse exponential
- Beta, fixed $\theta, b = 1$
- Beta, fixed $\theta, a = 1$

Also, the uniform shortcut cannot be used when there is a left-truncation on the