Lecture Notes II – Maximum Likelihood Estimation for Discrete Distributions

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

April, 2023

Max Likelihood Principle

ML estimation for arbitrary discrete distributions

coin flip

Other ML estimation examples

ML estimate as a random variable

Reading: Ch. 4.1, 4.2

ML estimation for arbitrary discrete distributions

ML estimate for $S = \{0, 1\}$

Data: $D: \{x^{1:n}\}$ $\quad |D| = n \Rightarrow n_0, n_1$ count

Model Family: $\mathcal{F} = \{(\theta_0, \theta_1), \theta_0 + \theta_1 = 1, \theta_{0,1} \geq 0\}$

Likelihood: $L(D | \theta_{0,1}) = \theta_0^{n_0} \theta_1^{n_1} = (1 - \theta_1)^{n_0} \theta_1^{n_1}$ $\rightarrow$ calculus

finding arg max

log-likelihood: $\underbrace{\log L(D | \theta_{0,1})}_{\ell(\theta_1)} = n_0 \log(1 - \theta_1) + n_1 \log(\theta_1)$

$(\log(x))' = \frac{1}{x}$

$(\log(1-x))' = -\frac{1}{1-x}$

$\ell'(\theta_1) = -\frac{n_0}{1-\theta_1} + \frac{n_1}{\theta_1} = 0$ to find max

$$\frac{n_1}{\theta_1} = \frac{n_0}{1-\theta_1}$$

$$n_1(1 - \theta_1) = n_0 \theta_1$$

$$n_0 \theta_1 = n_1(1 - \theta_1)$$

$$\underbrace{(n_0 + n_1)}_{n} \theta_1 = n_1$$

$$\theta_1^{ML} = \frac{n_1}{n} \Rightarrow \theta_0^{ML} = \frac{n_0}{n}$$

ML estimation for arbitrary discrete distributions

ML Estimation $S = \{1, \ldots, m\}$

Data: $\mathcal{D} = \{x^{1:n}\}$ $x^i \in S$

Model Family: $\mathcal{M} = \{(\theta_1, \ldots, \theta_m) = \theta \, ; \, \theta_j \geq 0, \, j = 1:m, \, \sum_{j=1}^{m} \theta_j = 1\}$

Likelihood: $L(D \mid \theta) = \prod_{j=1}^{m} \theta_j^{n_j} \leftarrow \#\{x^j = j\}$

Finding the (arg) max using calculus

log-likelihood: $\ell(D \mid \theta) = \sum_{j=1}^{m} n_j \log(\theta_j)$

maximize over $\theta \in \mathcal{M}$

$\max_{\theta \in \mathcal{M}} \ell(D \mid \theta)$ use lagrange multipliers

$\underbrace{\theta_1 + \theta_2 + \ldots + \theta_m = 1}_{\text{constraint}}$

$\theta_j^{ML} = \frac{n_j}{n}$ $j = 1:m$

multinomial distributed

alg: Input: $\mathcal{D} \xrightarrow{ML}$ Output: $\theta^{ML} = (\theta_1^{ML}, \ldots, \theta_m^{ML})$

ex: $S = \{1, 2, \ldots, 6\}$ $\mathcal{D} = (2, 5, 3, 1, 1, 6, 2, 4, 6, 1)$ $n = 10$ $n_2 = 2$

$\theta_1 = \frac{3}{10}$ $\theta_2 = \frac{2}{10}$ $n_1 = 3$

Other ML estimation examples

$S = \{0, 1\}$, $|S| = 2$

$S = \{1, \dots, m\}$, $|S| = m$ ← finite

ML estimation Poisson distributed dataset

$$S = \{0, 1, 2, \dots\}$$

$$\mathcal{D} = \{x^{1:n}\}$$

$$P(x \mid \underset{\uparrow}{\lambda}) = \frac{\lambda^x}{x!} e^{-\lambda}$$

expected rate of occurances

Model Family: $\mathcal{M} = \{\lambda, \lambda \geq 0\}$

$$\max_{\lambda} \mathcal{L}(\mathcal{D} \mid \lambda)$$

Likelihood: $\mathcal{L}(\mathcal{D} \mid \lambda) = \prod_{i=1}^{n} \left(\frac{\lambda^{x^i}}{x_i!} e^{-\lambda} \right)$

$$\ell(\lambda) = \sum_{i=1}^{n} \left(x^i \log(\lambda) - \underbrace{\log(x^i!)}_{\text{constant}} \right) + \overbrace{n \log(e^{-\lambda})}^{-n\lambda}$$

$$= \log \lambda \underbrace{\sum_{i=1}^{n} x^i}_{X} - n\lambda + \text{constants}$$

Other ML estimation examples

$$l'(\lambda) = \frac{1}{\lambda} S - n = 0 \Rightarrow \lambda^{ML} = \frac{S}{n}$$

In: D → Out: λ^{ML}

Discrete: $|S| = m \Rightarrow \theta_j^{ML} = \frac{n_j}{n}$ $\quad n_j = \sum_{i=1}^{n} \mathbb{I}(x^i = j)$

Poisson $\Rightarrow \lambda^{ML} = \frac{X}{n}$, $X = \sum_{i=1}^{n} x^i$

1. X, n_j's are RV

2. sufficient statistics for θ_j^{ML} & λ^{ML}

Other ML estimation examples

Censored Data:

Geometric Distribution

$S = \{0, 1, 2, \ldots\}$

Data : $\mathcal{D} = (x^{1:n})$, $x^i \in S$

Model Class : $\mathcal{M} = \{P_r : r \in (0,1)\}$: $P_r(x) = (1-r)r^x \quad x \in S$

$$r^{ML} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + n} \leftarrow X, \quad \frac{X}{X+n}$$

Censored Data for Geometric Distribution

$$z_i = \begin{cases} 0 & \text{if } x^i \text{ is even} \\ 1 & \text{if } x^i \text{ is odd} \end{cases}$$

$x^i \in S = \{0, 1, 2, \ldots\}$

$z^i \in \{0, 1\}$ $\quad \mathcal{D} = \{z^{1:n}\}$

$\mathcal{M} = \{P_r, r \in (0,1)\}$

hidden : $x^{1:6} = (2, 0, 5, 0, 0, 1)$

observed : $z^{1:6} = (0, 0, 1, 0, 0, 1)$

Want r by ML

Other ML estimation examples

Likelihood: $L(D|r) = P(z^{1:6}|r) = (P(z=0|r))^4 (P(z=1|r))^2$

$$P(z=0|r) = P_r(x \text{ even}) = P_r(0) + P_r(2) + \dots + P_r(2k) + \dots$$

$$= \sum_{k=0}^{\infty} (1-r)\, r^{2k} = (1-r) \underbrace{\sum_{k=0}^{\infty} r^{2k}}_{= \frac{1}{1-r^2}} = \frac{1}{1+r}$$

algebra

infinite sum of geometric sequence

$$P(z=1|r) = 1 - \frac{1}{1+r} = \frac{r}{1+r}$$

$$P(z^{1:6}|r) = \left(\frac{1}{1+r}\right)^4 \left(\frac{r}{1+r}\right)^2 = \frac{r^{\sum_{i=1}^{6} z^i}}{(1+r)^6}$$

in general case (n datapoints)

$$L(D|r) = P(z^{1:n}|r) = \frac{r^{\sum_{i=1}^{n} z^i}}{(1+r)^n} = \frac{r^{\tilde{s}}}{(1+r)^n}$$

Other ML estimation examples

log-likelihood

$$\ell(r) = \tilde{s}\log(r) - n\log(1+r)$$

$$\ell'(r) = \frac{\tilde{s}}{r} - \frac{n}{1+r} = 0 \Rightarrow r^{ML} = \frac{\tilde{s}}{n-\tilde{s}} \overset{?}{\in} [0,1) \quad \tilde{s} < \frac{n}{2}$$

sufficient statistic is $\tilde{s}$

ML estimate for uncensored data

$$r^{ML} = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i + n}$$

$n = 6$

hidden : $x^{1:6} = (2, 0, 5, 0, 0, 1)$

observed : $z^{1:6} = (0, 0, 1, 0, 0, 1)$

censored data $\hookrightarrow \tilde{r}^{ML} = \frac{2}{4} = \frac{1}{2}$

uncensored data $\hookrightarrow r^{ML} = \frac{4}{7}$

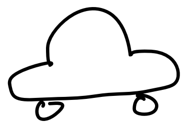
$\sum x_i = 8$

$n = 6$

$$\frac{8}{6+8} = \frac{8}{14}$$

Other ML estimation examples

ML w/ tied parameters



bad road

$$S = \{ft_1, ft_2, ft_3, ft_4, O\}$$

$$\theta_1 \quad \theta_2 \quad \theta_3 \quad \theta_4 \quad \theta_0$$

Model Class : $\{(\theta_0, \theta_1, \dots, \theta_4),\ \theta_1 = \theta_2 = \theta_3 = \theta_4 = \theta_f,\ \underbrace{\theta_0 + \theta_1 + \theta_2 + \theta_3 + \theta_4}_{\theta_0 + 4\theta_f = 1} = 1\}$

Data : $n_0 = 90 \quad n_1 = 2 \quad n_2 = 0 \quad n_3 = 5 \quad n_4 = 3$

Likelihood: $L(D \mid \theta_0, \theta_f) = \theta_0^{n_0}\, \theta_f^{n_1}\, \theta_f^{n_2}\, \theta_f^{n_3}\, \theta_f^{n_4} = \theta_0^{n_0}\, \theta_f^{\overbrace{n_1+n_2+n_3+n_4}^{n_f}}$

log-likelihood: $l(\theta_0, \theta_f) = n_0 \log(\theta_0) + n_f \log(\theta_f)$

we have $\theta_0 = 1 - 4\theta_f \Rightarrow n_0 \log(1 - 4\theta_f) + n_f \log(\theta_f)$

$$l'(\theta_f) = n_0\left(-\frac{4}{1-4\theta_f}\right) + n_f\left(\frac{1}{\theta_f}\right) = 0$$

$n = n_f + n_0$

$$\frac{n_f}{\theta_f} = \frac{4n_0}{1-4\theta_f} \Rightarrow n_f - 4n_f\theta_f = 4n_0\theta_f \Rightarrow \theta_f = \frac{n_f}{4n}$$

Other ML estimation examples

Other ML estimation examples