STAT 391

1/16/25

Lecture 4

HW 1 up

L II Max L.

Ch 4 Max L

Max Likelihood **Principle**

ML 4 discrete S

Lecture Notes II – Maximum Likelihood Estimation for Discrete Distributions

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

~~April, 20~~21

Max Likelihood Principle

ML estimation for arbitrary discrete distributions

Other ML estimation examples

ML estimate as a random variable

Reading: Ch. 4.1, 4.2

Maximum Likelihood Principle

Ex Language models

$S = \{a, b, c, \dots z\}$

$m = |S| = 26$

English

$P^E = (\theta_a^E, \theta_b^E, \dots \theta_z^E)$

$\theta_a^E = P^E(a)$

$\boxed{P^E(\{a\}) = P^E(a)}$

Spanish $P^S = (\theta_a^S, \theta_b^S, \dots)$

French $P^F = (\theta_a^F, \theta_b^F, \dots)$

Sentence statistics $= x^1 \dots x^n$ $n = 10$

Data

$x^{(1)} \in S$ $x^{(2)} \in S$

Maximum Likelihood Principle = choose "model" parameters that make data most likely (=probable)

Q: Language = ?

Answer $\in \{E, F, S, G\} = \mathcal{M}$

$P^E(\text{statistics}) = L(E)$ likelihood $\quad \ln L(E) = \ell(E)$ log-likelihood

$P^F(\quad - " - \quad) = L(F)$

. . . $\qquad\qquad$ . . .

$\log_2 \rightarrow \ell(E) = -38.6 \Longleftrightarrow P^E(\overbrace{\text{statistics}}^{\mathcal{D}}) = 2^{-38.6}$

$\ell(F) = \boxed{-38.3} \longrightarrow \max \Rightarrow$ Answer = F

$\ell(G) = -39.8$

. . .

It's a guess !!!

Maximum Likelihood Principle

Data $\mathcal{D} = \{x^1, \ldots x^n\} \sim$ iid S with unknown P on S

Model class = {all possible P's to consider} $= \mathcal{M}$

Belief about P

Inference

$P^{ML} \in \mathcal{M}$

Estimation

Max likelihood

$$P^{ML} = \underset{P \in \mathcal{M}}{\text{argmax}}\ \underset{\underset{L(P;\mathcal{D})}{\equiv}}{P(\mathcal{D})} =$$

likelihood of $\mathcal{D}$ under P

ML estimation for arbitrary discrete distributions (finite S)

Model class

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

$\mathcal{M} = \{ (\theta_{1:m}) \text{ with } \theta_j \geq 0 \text{ for all } j, \sum_{j=1}^{m} \theta_j = 1 \}$

any P on S defined by $\theta_1, \theta_2, \dots \theta_m$ parameters

$\theta_j = P(j) \geq 0$

$\sum_{j=1}^{m} \theta_j = 1$

Data $\mathcal{D} = \{x^1, \dots x^n\} \subset S$

Ex: Coin toss $S = \{0, 1\}$ $\mathcal{M} = \{ (\theta_0, \theta_1), \theta_0 + \theta_1 = 1$

$\theta_{0,1} \geq 0 \}$

$\mathcal{D} = 1, 0, 1, 1, 0 \Rightarrow L(\theta) = \theta_1 \theta_0 \theta_1 \theta_1 \theta_0 = \theta_0^2 \theta_1^3$

$n = 5$

$n_1 = 3$ $n_0 = 2$

$\overset{n_0}{2}$ $\overset{n_1}{3}$

ML estimation for arbitrary discrete distributions

0. S, Model class $\mathcal{M}$

$\theta = (\theta_1 \cdots \theta_m)$

1. Write likelihood

$$L(\theta) \equiv P(\mathcal{D}|\theta) = \prod_{i=1}^{n} \underbrace{P(x^i|\theta)}_{\theta_{x^i}} = \prod_{i=1}^{n} \theta_{x^i} = \theta_1^{n_1}\theta_2^{n_2}\cdots\theta_m^{n_m}$$

if $x_i = j \Rightarrow \theta_j$

$$= \prod_{j=1}^{m} \theta_j^{n_j}$$

- $n_j = \#\{x^i = j\}$ counts

$j = 1:m$

1.' Log likelihood

$$\ell(\theta) = \ln L(\theta) = \sum_{j=1}^{m} n_j \ln \boxed{\theta_j}$$

unknowns

data

2. Find $\boxed{\theta^{ML} = \underset{\theta\in\mathcal{M}}{\operatorname{argmax}}\, \ell(\theta) \equiv \underset{\theta\in\mathcal{M}}{\operatorname{argmax}}\, L(\theta)}$

STATISTICS

CALCULUS

ML estimation for arbitrary discrete distributions

m=2 (from Ex)

$n_0 = 2$
$n_1 = 3$

$L(\theta) = \theta_0^2 \theta_1^3$ ← 1. write L for generic θ

$\ell(\theta) = 2 \cdot \ln \theta_0 + 3 \cdot \ln(\theta_1)$ ←? $\theta_0, \theta_1 = \arg\max$

2. calculus

$\theta_0 = 1 - \theta_1$ $\quad \max_{\theta_1} 2 \ln(1-\theta_1) + 3 \ln(\theta_1) = \ell(\theta)$

$\ell'(\theta_1) = 2 \cdot \frac{-1}{1-\theta_1} + 3 \frac{1}{\theta_1} = 0 \quad \Big| \times \theta_1 (1-\theta_1)$

$-2\theta_1 + 3(1-\theta_1) = 0$

$3 = 5\theta_1 \Rightarrow$ $\boxed{\theta_1^{ML} = \frac{3}{5} = \frac{n_1}{n} \quad \theta_0^{ML} = \frac{2}{5} = \frac{n_0}{n}}$ ← in general

+ Verify $\ell(\theta^{ML})$ is max not min

e.g. $\ell''(\theta^{ML}) < 0$

$\boxed{(\ln z)' = \frac{1}{z}}$

Other ML estimation examples

$\mathcal{D} = \{x^1, \ldots x^n\} \subset S$

$S = \{1, \ldots m\}$ $\quad m > 2$

$\mathcal{M} = \{(\theta_1, \ldots \theta_m), \boxed{\theta_j \geq 0 \text{ for all } j, \ \sum \theta_j = 1}\}$ → * constraints

1. Write $L(\theta) = \theta_1^{n_1} \theta_2^{n_2} \ldots \theta_m^{n_m}$

1'. log-L $\quad l(\theta) = \underline{n_1} \ln \theta_1 + \ldots + \underline{n_m} \ln \theta_m$ ← m unknowns

from $\mathcal{D}$ (counts)

Ex $\quad S = \{1, \ldots 6\}$

die roll $\quad n = 9$

$\mathcal{D} = 5, 3, 3, 1, 2, 4, 6, 5, 1$

$n_1 = n_3 = n_5 = 2$

$n_2 = n_4 = n_6 = 1$

STAT

$l = \underbrace{2 \ln \theta_1 + 1 \ln \theta_2 + 2 \ln \theta_3 + 1 \ln \theta_4 + 2 \ln \theta_5 + \ln \theta_6}$

CALC

2. $\theta_{1:6}^{ML} = \underset{\theta_{1:6}}{\operatorname{argmax}} \ l(\theta)$ subject to * constraints

Maximization with constraints ← Lagrange multipliers

Other ML estimation examples

$\ell(\theta) = \sum_{j=1}^{m} n_j \ln \theta_j$ ← argmax s.t. *

Solution by Lagrange multiplier method

$$\max_{\lambda,\, \theta_{1:m}} \ \ell(\theta) + \lambda\left(\sum_{j=1}^{m} \theta_j - 1\right)$$

← constraint

$\lambda(\cdots)$

new variable λ

linear in λ

m+1 unknowns

m+1 variables

data

m+1 eqns

$$\frac{\partial \ell}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}\Big[\sum_{l=1}^{m} n_l \underbrace{\ln \theta_l}_{\frac{1}{\theta_j}} + \lambda \underbrace{\sum_{l=1}^{m} \theta_l - 1}_{1}\Big] = \frac{n_j}{\theta_j} + \lambda = 0$$

$$\Downarrow$$

$$\theta_j = -\frac{n_j}{\lambda}$$

$$\frac{\partial \mathcal{H}}{\partial \lambda} = \sum_{j=1}^{n} \theta_j - 1 = 0 \Rightarrow \sum \theta_j = 1 \Rightarrow \sum \theta_j = -\frac{\sum n_j}{\lambda} = -\frac{n}{\lambda}$$

$\sum n_j = n$

$$\lambda = -n \Rightarrow \boxed{\theta_j^{ML} = \frac{n_j}{n}} \geq 0$$

Other ML estimation examples

$$\theta_j^{ML} = \frac{\boxed{n_j}}{n} \quad j = 1, \ldots m$$

sufficient statistics

$n_{1:m}$ known $\longleftrightarrow$ $\theta_{1:m}$ known

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

statistic — function of $\mathcal{D}$ $\longrightarrow$ r.v. useful for estimation

function of $x \in S$

STAT

PROB $\in$ MATH

$f : S \to \mathbb{R}$ random variable

Probability

$\mathcal{D} \longrightarrow n_{1:m} \longrightarrow \theta^{ML}$

ML estimation

$|\mathcal{D}| = n$

Exponential family

$\exists$ sufficient statistics

finite S categorical

Normal

Poisson

uniform