STAT 391 4/4/'23

Lecture 3

- R.I.S → Binomial
- Maximum Likelihood

- L11 posted
- HW1 tomorrow
- OH Instructor
  Mon 2:10-3:00
  PDL C-301
  Conference room
- tell me after class

Lecture Notes I – (Discrete) Sample spaces and the Multinomial

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

March, 2023

Sample space, outcome, event, probability,. . .

Discrete sample spaces

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

OR S countable

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

Repeated independent trials, Binomial, Multinomial

Reading: Ch. 2, 3

Discrete sample spaces ← Repeated Independent Trials "Sampling"

- S sample space (outcome space)
- $x \in S$ outcome
- $E \subseteq S$ event
- $2^S = \{E \mid E \subseteq S\} \equiv \mathcal{P}(S)$
- $P : 2^S \to [0,1]$ probability distribution
- $f : S \to \mathbb{R}$ random variable
- $f : S \to \mathbb{R}$ statistic

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

P on S ← $(\theta_1, \dots \theta_m)$ parameters

$\tilde{S} = \underbrace{S \times S \times \cdots S}_{\times n}$

$\tilde{P}$ on $\tilde{S}$

$\tilde{x} = (x^1, x^2, \dots x^n)$

$$\tilde{P}(\tilde{x}) = \prod_{i=1}^{n} P(x^i)$$

independence

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

counts

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

$j \in S$

$n_j \geq 0$

$n_1 + n_2 + \dots + n_m = n$

Discrete sample spaces

S countable

for ex $\{\underset{a_1}{1}, \underset{a_2}{01}, \underset{a_3}{001}, \ldots\}$

$P[a_j]$ known

- S sample space (outcome space)
- $x \in S$ outcome
- $E \subseteq S$ event
- $2^S = \{E \mid E \subseteq S\} \equiv \mathcal{P}(S)$
- $P : 2^S \to [0,1]$ probability distribution
- $f : S \to \mathbb{R}$ random variable
- $f : S \to \mathbb{R}$ statistic

x^i ← the i-th observation

$i = 1:n$

$x_j \equiv j$ an outcome

n trials $\quad n = 4$

$\tilde{x} = \{001, \underset{a_1}{1}, 0001, \underset{a}{1}\}$

$\tilde{P}[\tilde{x}] = P[a_1]^2 P[a_3] P[a_4]$

$n_1 = 2 \quad n_4 = 1$

$n_2 = 0 \quad n_5 = n_6 = \cdots = 0$

$n_3 = 1$

Discrete sample spaces

$x \in S$, $x^i \in S$ outcomes

- S sample space (outcome space)
- $x \in S$ outcome
- $E \subseteq S$ event
- $2^S = \{E \mid E \subseteq S\} \equiv \mathcal{P}(S)$
- $P : 2^S \rightarrow [0, 1]$ probability distribution
- $f : S \rightarrow \mathbb{R}$ random variable
- $f : S \rightarrow \mathbb{R}$ statistic

Multiple sequences with same $\tilde{P}[\tilde{x}]$?

Ex:

$S = \{0, 1\}$

$n = 2$ $\quad \theta_1 > \theta_0$ unfair coin

$\tilde{x} = (x^1 \dots x^n)$ sequence (of outcomes)

$\downarrow$

counts $(n_1 \dots n_m)$

$00 \rightarrow \theta_0^2$

$\left.\begin{array}{l} 01 \rightarrow \theta_0\theta_1 \\ 10 \rightarrow \theta_0\theta_1 \end{array}\right\}$ same $\tilde{P}$

YES : $\left.\begin{array}{l} \tilde{x} \\ \tilde{x}' \end{array}\right\}$ same counts $\Rightarrow \boxed{\tilde{P}[\tilde{x}] = \tilde{P}[\tilde{x}']}$

Repeated independent trials, Binomial, Multinomial

$n=4$ $S=\{0,1\}$

$\theta_1 \equiv p$

A coin is tossed 4 times, and the probability of 1 is $p > 0.5$. The outcomes, their probability and their counts are (in order of decreasing probability):

$\theta_0 = 1-p$

outcome $\tilde{x}$	n_0	n_1	$P(\tilde{x})$	event
1111	0	4	p^4	$E_{0,4}$
1110	1	3	$p^3(1-p)^1$	$E_{1,3}$
1101	1	3	$p^3(1-p)^1$	
1011	1	3	$p^3(1-p)^1$	
0111	1	3	$p^3(1-p)^1$	
1100	2	2	$p^2(1-p)^2$	$E_{2,2}$
1010	2	2	$p^2(1-p)^2$	
1001	2	2	$p^2(1-p)^2$	
0110	2	2	$p^2(1-p)^2$	
0101	2	2	$p^2(1-p)^2$	
0011	2	2	$p^2(1-p)^2$	
0100	3	1	$p^1(1-p)^3$	$E_{3,1}$
1000	3	1	$p^1(1-p)^3$	
0010	3	1	$p^1(1-p)^3$	
0001	3	1	$p^1(1-p)^3$	
0000	4	0	$(1-p)^4$	$E_{4,0}$

$|E_{0,4}| = 1$

event $E_{1,3}$

$|E_{1,3}|$

$|E_{2,3}| = 6$

- for every (n_0, n_1) counts $\rightarrow E_{n_0 n_1}$
- $P[\tilde{X} \in E_{n_0,n_1}] = \theta_0^{n_0} \theta_1^{n_1}$

all outcomes $\in E_{n_0 n_0}$ have same probability

- $\bigcup_{(n_0,n_1)} E_{n_0,n_1} = \tilde{S}$ partition of $\tilde{S}$
- $|E_{n_0,n_1}| = \binom{n}{n_1} \equiv \binom{n}{n_0} = \dfrac{n!}{n_1!\,n_0!}$

Repeated independent trials, Binomial, Multinomial

$$\boxed{P[E_{n_0,n_1}] = P[\underset{\substack{\text{single}\\\text{sequence}\\ \in E_{n_0 n_1}}}{X}]\ |E_{n_0 n_1}| = \theta_0^{n_0}\theta_1^{n_1}\binom{n}{n_0}}$$

$$\binom{n}{n_0} = \frac{n!}{n_0!\,n_1!}$$

BINOMIAL DISTRIBUTION

for all $n_{0,1} \geq 0$, $n_0 + n_1 = n$

- distribution over counts

Repeated independent trials, Binomial, Multinomial

A coin is tossed 4 times, and the probability of 1 is $p > 0.5$. The outcomes, their probability and their counts are (in order of decreasing probability):

outcome x	n_0	n_1	$P(x)$	event
1111	0	4	p^4	$E_{0,4}$
1110	1	3	$p^3(1-p)^1$	$E_{1,3}$
1101	1	3	$p^3(1-p)^1$	
1011	1	3	$p^3(1-p)^1$	
0111	1	3	$p^3(1-p)^1$	
1100	2	2	$p^2(1-p)^2$	$E_{2,2}$
1010	2	2	$p^2(1-p)^2$	
1001	2	2	$p^2(1-p)^2$	
0110	2	2	$p^2(1-p)^2$	
0101	2	2	$p^2(1-p)^2$	
0011	2	2	$p^2(1-p)^2$	
0100	3	1	$p^1(1-p)^3$	$E_{3,1}$
1000	3	1	$p^1(1-p)^3$	
0010	3	1	$p^1(1-p)^3$	
0001	3	1	$p^1(1-p)^3$	
0000	4	0	$(1-p)^4$	$E_{4,0}$

Ex $\theta_1 = \frac{3}{4}$ $\theta_0 = \frac{1}{4}$ $n = 4$

E^*

$\tilde{P}[\tilde{x}] = \frac{3^4}{4^4} = P_1$

$P[E_{04}] = \frac{3^4}{4^4}$

$\tilde{P}[\tilde{x}] = \frac{3^2}{4^4} = P_2 < P_1$

$P[E_{2,2}] = \frac{3^2}{4^4} \cdot 6$

$P[E_{1,3}] = \frac{3^3}{4^4} \cdot 4 =$

$\frac{4}{3} P[E_{0,4}] > P[E_{04}]$

typical counts $E^* = E_{n_1, n_3}$

$\tilde{x} \in E^*$ typical sequence

Maximum Likelihood ~~Principle~~ and language models

(see also HW 1)

Models

English $(\theta_a^E, \theta_b^E, \theta_c^E, \ldots \theta_z^E)$ ← english.dat

$\theta_a^E = 0.081$

German $(\theta_a^G, \theta_b^G, \theta_c^G, \ldots \theta_z^G)$

Spanish $(\theta_a^S, \ldots)$

French $(\theta_a^F, \ldots)$

Data hello, world

ignore spaces

ignore special chars

→ $n_e = n_h = n_w = n_d = \cdots\ 1$

$n_o = 2$

$n_l = 3$

$n_a = n_b = \ldots = 0$

$$P^E(\text{hello}\ldots) = \theta_h^E \theta_e^E \theta_e^E \theta_e^E \ldots$$

$$= \theta_h^E \theta_e^E (\theta_l^E)^3 (\theta_o^E)^2 \ldots$$ ← likelihood of English

Maximum Likelihood ~~Principle~~

$P^G[\text{hello world}] = \theta^G_h \theta^G_e (\theta^G_l)^3 (\theta^G_o)^2 \cdots$ ← likelihood of German

$\ell(\text{English}) = \log_2 P^E[\text{hello world}]$ ↔ maximum ⇓ guess is English

$\ell(\text{German}) = \log_2 P^G[\text{— " —}]$

$\ell(\text{Spanish}) = \log_2 P^S[\text{— " —}]$

$\ell(\text{French}) = \log_2 P^F[\text{— " —}]$

Maximum Likelihood Estimation → random

Probability
Statistics

Data: $x^1, x^2 \dots x^n \sim$ iid P unknown given

S known, finite $S=\{1,..m\}$

Model:

→ $P = (\theta_1, \theta_2 \cdots \theta_m)$ Model family $\mathcal{F} = \{ (\theta_{1:m}), \theta_{1:m} \geq 0, \sum \theta_j = 1 \}$

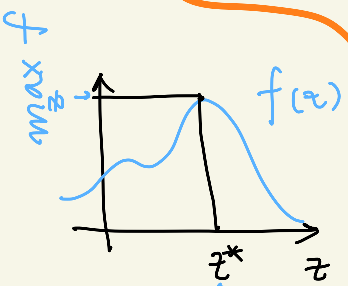
Pb: $P = ?$ ⟺ $\theta_{1:m} = ?$ estimate parameters

How solve? Max Likelihood Principle

$$\boxed{\theta^{ML}_{1:m}} = \underset{\theta_{1:m} \in \mathcal{F}}{\operatorname{argmax}} \underbrace{P[x^{1:n} \mid \theta_{1:m}]}_{L(\theta_{1:m})} = \underset{\theta_{1:m}}{\operatorname{argmax}} \overbrace{\ln P[x^{1:n} \mid \theta_{1:m}]}^{\text{log-likelihood}}$$

Calculus

$L(\theta_{1:m})$ ← Likelihood of D given $\theta_{1:m}$ (model)



$z^* = \operatorname{argmax} f(z)$

Principle → lets us derive estimation methods

a guess!!
random variables

Example: n coin tosses $\quad n = 4$

1) $\mathcal{D} = (0,1,0,0) \rightarrow n_0 = 3$, $n_1 = 1$

$S = \{0,1\}$

2) Model family $\quad \mathcal{F} = \{(\theta_0, \theta_1)$

$\theta_0 + \theta_1 = 1 \,,\; \theta_{0,1} \geq 0\}$

3) write likelihood

$$P[\mathcal{D} \mid \theta_{0,1}] = \theta_0^{n_0} \theta_1^{n_1} = \theta_0^3 \theta_1^1 = (1-\theta_1)^3 \theta_1$$

calculus

4) find (arg) max

log-likelihood

$$\ln P[\mathcal{D} \mid \theta_{0:1}] = 3 \ln (1-\theta_1) + \ln \theta_1 = \underline{\ell}(\theta_1)$$

$$\ell'(\theta_1) = 3 \frac{-1}{1-\theta_1} + \frac{1}{\theta_1} = 0$$

want max

$$\frac{1}{\theta_1} = \frac{3}{1-\theta_1}$$

$$1 - \theta_1 = 3\theta_1$$

$$1 = 4\theta_1 \Rightarrow \boxed{\theta_1^{ML} = \frac{1}{4} \quad \theta_0^{ML} = \frac{3}{4}}$$

my estimate (ML) — a guess!!

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

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

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

1) $\mathcal{D} = \{x^{1:n}\}$ $\quad |\mathcal{D}| = n \Rightarrow n_0, n_1$ counts

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

3) Likelihood $L(\theta_0, \theta_1) = \theta_0^{n_0} \theta_1^{n_1} = (1-\theta_1)^{n_0} \theta_1^{n_1}$

Calculus

$\dots \Rightarrow$ $\boxed{\theta_j^{ML} = \frac{n_j}{n} \quad j = 1:m}$