STAT 391

2/4/2025

Lecture 9

Parametric
ML estimation
density

$$\theta_0^{5} \theta_1^{7}$$
n_0 n_1

with
$n_1 = 7$
$n_0 = 5$

$0.3^5 \, 0.7^7$

Q1 Thu 12:30
10–15 min
↳ Canvas: info
Chapters 1–4 +
small probs

Sol 2
L IV
+ supplement posted

Lecture Notes IV – Continuous distributions. Parametric density estimation.

Marina Meilă
mmp@stat.washington.edu

Department of Statistics
University of Washington

January 2025

CDF and PDF. Sampling ✓

Examples of continuous distributions ✓

ML estimation for continuous distributions

ML estimation by gradient ascent

Reading: Ch.5, 6

CDF and PDF refresher

Cumulative distribution function (CDF)

$$F(x) \; = \; P[X \leq x] \qquad (1)$$

1. $F \; \geq \; 0$ positivity.
2. $\lim_{x \to -\infty} F = 0$
3. $\lim_{x \to \infty} F = 1$
4. F is an increasing function

Probability density [function] (PDF)

$$f \; = \; \frac{dF}{dx} \qquad (2)$$

$$P(a, b) \; = \; P[a, b] \; = \; F(b) - F(a) \; = \; \int_a^b f(x) dx \qquad (3)$$

normalization condition

$$\int_{-\infty}^{\infty} f(x) dx \; = \; 1 \qquad (4)$$

Examples of continuous distributions

$$\mathcal{F}_1 = \{u_{[a,b]},\ a < b\} \quad \text{uniform} \tag{5}$$

$$f(x;a,b) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

$$\mathcal{F}_2 = \{N(.;\mu,\sigma^2)\} \quad \text{normal} \tag{7}$$

$$f(x;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8}$$

$$F(x;a,b) = \frac{1}{1+e^{-ax-b}}, \quad a > 0 \quad \text{logistic} \tag{9}$$

$$f(x;a,b) = \frac{ae^{-ax-b}}{(1+e^{-ax-b})^2} \tag{10}$$

$\mathcal{F}_4$

exponential

$\mathcal{F}_3 = \{ \exp(\theta),\ \theta > 0 \}$

$f(x|\theta) = \theta e^{-\theta x}$

$x \in [0,\infty) = S$

ML estimation for continuous distributions → density estimation

$S \subseteq (-\infty, \infty)$

parametric = estimating parameters
by ML

non-parametric
not by ML

ML Principle

Given data $\mathcal{D} = \{x^1, \ldots x^n\} \subset S \subseteq (-\infty, \infty)$

Choose model class $\mathcal{F} = \{ f(x \mid \theta), \theta \in \Theta \}$

densities

parameters

ML principle

Choose $\theta^{ML} = \underset{\theta \in \Theta}{\operatorname{argmax}}\ L(\theta)$ ← Likelihood of $\mathcal{D}$

$$\boxed{L(\theta) = \prod_{i=1}^{n} f(x^i \mid \theta)}$$

ML estimation for continuous distributions exponential

Data $x^{1:n} \geq 0$

Model $\mathcal{F} = \{ f(x|\vartheta) = \vartheta e^{-\vartheta x}, \vartheta > 0 \}$ exponential family

Want $\vartheta = ?$ by ML

s.s.

1. Likelihood

$$L(\vartheta) = \prod_{i=1}^{n} f(x^i|\vartheta) = \prod_{i=1}^{n} \vartheta e^{-\vartheta x^i} = \vartheta^n e^{-\vartheta \sum_{i=1}^{n} x^i}$$

1'. log-likelihood

$$\ell(\vartheta) = \ln L(\vartheta) = n \ln \vartheta - \vartheta \boxed{\sum_{i=1}^{n} x^i}$$

sufficient stat.

STAT

CALCULUS

2. max $\ell(\vartheta)$

$$\ell'(\vartheta) = n\frac{1}{\vartheta} - \sum x^i = 0 \Rightarrow \boxed{\frac{1}{\vartheta} = \frac{\sum x^i}{n}}$$

mean of $\mathcal{D}$

$$\boxed{\vartheta = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x^i}}$$

ML estimation for continuous distributions Normal (μ, σ^2)

Data $x^{1:n} \in (-\infty, \infty)$

Model class $\mathcal{F}_2 = \{ N(\mu, \sigma^2), \ \sigma^2 > 0, \ \mu \in (-\infty, \infty) \}$

2 params

Wanted: μ, σ^2

1. Likelihood

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x^i-\mu)^2}{2\sigma^2}}$$

1'. log-l

$$l(\mu, \sigma^2) = -n \cdot \frac{1}{2} \cdot \ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(x^i - \mu)^2}{2\sigma^2}$$

↑ STAT

↓ CALC

2. maximize $l(\mu, \sigma^2)$
$\mu \ \sigma^2$

$$\frac{\partial l}{\partial \mu} = 0 \quad \cancel{-\frac{1}{2\sigma^2}} \sum_{i=1}^{n} \cancel{2}(-x^i + \mu) = 0 \Rightarrow \sum x^i = n\mu$$

$$\boxed{\mu = \frac{\sum_{i=1}^{n} x^i}{n}}$$

ML estimation for continuous distributions

Normal(μ, σ^2)

$$\boxed{\mu^{ML} = \frac{\sum_{i=1}^{n} x^i}{n}}$$

Stat

$$\ell(\mu, \sigma^2) = -n \cdot \frac{1}{2} \cdot \ln(2\pi\sigma^2) - \sum_{i=1}^{n} \frac{(x^i - \mu)^2}{2\sigma^2}$$

const

Probability: $E[X] = \mu$

Now for σ^2

$$\frac{\partial \ell}{\partial(\sigma^2)} = -\frac{n}{\cancel{2}} \cdot \frac{1}{\sigma^2} + \sum_{i=1}^{n} \frac{+1}{\sigma^4} \cdot \frac{(x^i - \mu^{ML})^2}{\cancel{2}} = 0 \Rightarrow$$

underestimates true σ^2

"BIASED"

$$\Rightarrow \quad n = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x^i - \mu^{ML})^2 \Rightarrow \boxed{(\sigma^2)^{ML} = \frac{1}{n} \sum_{i=1}^{n} (x^i - \mu^{ML})^2}$$

suff stat

sample variance

corrected estimator

$$\boxed{\tilde{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x^i - \mu)^2} = \frac{n}{\underset{>1}{n-1}} (\sigma^2)^{ML}$$

$\sigma^2 = \theta$

$$\left(\frac{1}{\sigma^2}\right)' = \left(\frac{1}{\theta}\right)' = \frac{-1}{\theta^2} = \frac{-1}{\sigma^4}$$

ML estimation for continuous distributions uniform $[\alpha, \beta]$

$\mathcal{D} = x^{1:n} \in (-\infty, \infty)$

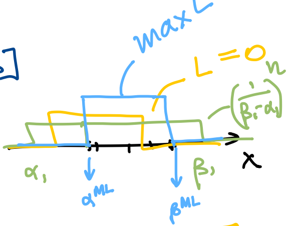
$\mathcal{F} = \{ \text{uniform}[\alpha, \beta],\ \underline{\alpha} < \underline{\beta} \}$

Wanted, by ML

1. Likelihood

$$f(x \mid \alpha, \beta) = \begin{cases} \dfrac{1}{\beta - \alpha} & x \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$

$$L(\alpha, \beta) = \begin{cases} \left(\dfrac{1}{\beta - \alpha}\right)^n = \dfrac{1}{\ell^n} & \text{if } \quad x^{1:n} \in [\alpha, \beta] \\ 0 & \text{otherwise} \end{cases}$$



$\beta - \alpha = \ell$ length of $[\alpha, \beta]$

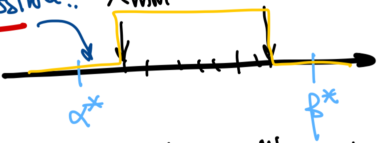
max $L(\alpha, \beta)$: min ℓ so that $x^{1:n}$ covered

$$\boxed{\begin{aligned} \alpha^{ML} &= \min x^{1:n} \\ \beta^{ML} &= \max x^{1:n} \end{aligned}}$$

ML estimation for continuous distributions uniform(α, β)

$$\boxed{\begin{aligned}\alpha^{ML} &= \min x^{1:n} \\ \beta^{ML} &= \max x^{1:n}\end{aligned}} \Longrightarrow \text{impossible!!}$$

DO NOT USE !!



$$\underline{\underline{\alpha^*}} < \alpha^{ML} < \beta^{ML} < \underline{\underline{\beta^*}}$$

Correction

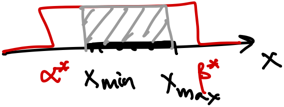
$$\tilde{\alpha} < \alpha^{ML}$$

$$\hat{\beta} > \beta^{ML}$$

I α^*, β^* known

$$\Pr[\, x_{min}, x_{max} \mid \alpha^*, \beta^* \,] =$$

out of n samples



$$= \left(\frac{x_{max} - x_{min}}{\underbrace{\beta^* - \alpha^*}_{?}} \right)^n = \underline{\underline{\gamma}} \leq 1$$

known

choose

STATISTICS confidence

e.g. $\gamma = 95\%$

CALC: $\ell^n / (\ell^*)^n = \gamma \Rightarrow \ell^* = \ell \cdot \gamma^{-\frac{1}{n}}$

ML estimation for continuous distributions

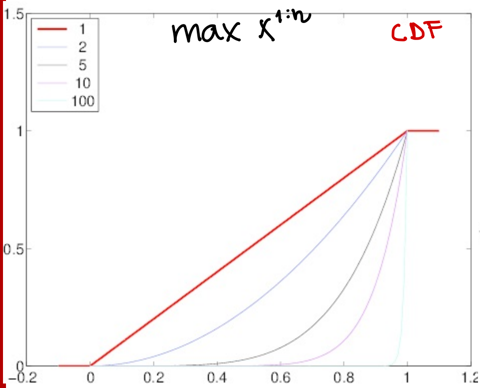
$$\ell^* = \ell \cdot \gamma^{-\frac{1}{n}} \Rightarrow \beta^* - \alpha^* = \underline{(\beta^{ML} - \alpha^{ML})} \cdot \gamma^{-\frac{1}{n}}$$

length

> 1
very close to 1

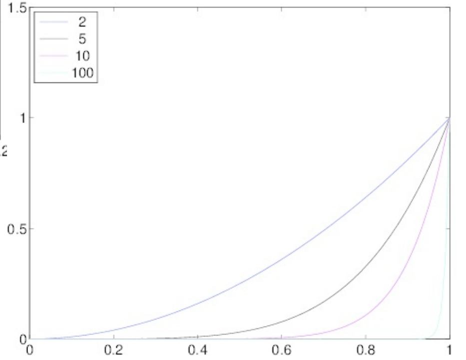
depends on n and α

ML estimation for continuous distributions



range $\underbrace{\max x^{1:n} - \min x^{1:n}}_{\ell}$

CDF



ML estimation by gradient ascent logistic (a,b)

$$f(x|a,b) = \frac{-a\,e^{-ax-b}}{(1+e^{-ax-b})^2}$$

$a>0$

max f $\quad -f\left(-\frac{b}{a}\right)$

symmetric around $-\frac{b}{a}$

prove for $b=0$

$$F(x|a,b) = \frac{1}{1+e^{-ax-b}}$$

$$l(a,b) = n\ln a - a\sum_i x_i - nb - 2\sum_{i=1}^{n}\ln(1+e^{-ax_i-b})$$

log. likelihood

$$\frac{\partial l}{\partial a} = \frac{n}{a} - \sum_{i=1}^{n} x_i \frac{1-e^{-ax_i-b}}{1+e^{-ax_i-b}} = 0$$

$$\frac{\partial l}{\partial b} = -\sum_{i=1}^{n} \frac{1-e^{-ax_i-b}}{1+e^{-ax_i-b}} = 0$$

No closed form

No sufficient statistics

x	$-\infty$	0	$-\frac{b}{a}$	∞
$ax+b$	$-\infty$	b	0	∞
e^{-ax-b}	∞		1	0
F	0			1
f	0		max	0

$a=1,\ b=0$

HW3 Pb2:

a, b, c, d are about estimating μ from $y^{1:n}$

(you don't see the original x's!)

Likelihood

$= \Pr(\text{Data} \mid \cancel{\text{model}}$ parameters$)$

expo: μ^{ML}

HW3: $\mu^{censored, ML}$

Language models: $\theta^{ML}_{a,b,c\ldots}$

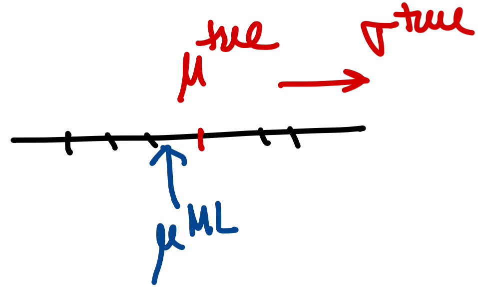
$\mathcal{D}$ = sentence "hello world"

$L(\theta) = \theta_h^1 \theta_e^1 \theta_e^3 \ldots$

$\mathcal{D}'$ = "$n_e = 3$, $n_h = 1$, $n_e = 1$, ..."

$L(\theta) = \binom{n}{n_h!\, n_e! \ldots} \theta_h^1 \theta_e^3 \theta_e^1 \ldots$

Why $(\sigma^2)^{ML}$ is underestimate of σ^2?



$$\arg\min_a \frac{1}{n}\sum_i (x^i - a)^2 = \mu^{ML}$$

$$\downarrow$$

$$(\sigma^2)^{ML}$$