7.1, 7.2 Maximum Likelihood Estimation

+ 7.6.1 continued

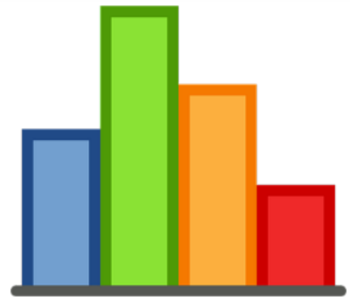
Anna Karlin
Most slides by Alex Tsun

Agenda

- Probability vs Statistics
- Likelihood
- Maximum Likelihood Estimation (MLE)
- MLE Example (Poisson)
- MLE Example (Normal)
- Unbiased estimators

Probability vs Statistics

$Ber(p = 0.5)$ → **Probability** given model, predict data → $P(THHTHH)$



$Ber(p = ???)$ ← **Statistics** given data, predict model ← $THHTHH$

Probability

1. Model

$X \sim \mathrm{Poi}(5)$

$Y \sim \mathrm{Ber}(\frac{1}{3})$

$Z \sim N(0,1)$

2. $\Pr(X=3) = e^{-5}\frac{5^3}{3!}$

$P_X(3;5)$

$\Pr(Y=1) = \frac{1}{3}$

$P_Y(1;\frac{1}{3})$

$\Pr(Z \le 3) = \int_{-\infty}^{3} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,dx$

Likelihood. (statistics)

1. Parametric model

iid. samples from $\mathrm{Poi}(\theta)$, $\mathrm{Ber}(\theta)$, $N(\theta_1, \theta_2)$

Use θ to denote parameter(s) of distribution.
Unknown to us

Goal: Given iid. samples $(x_1, x_2, \dots x_n)$
e.g. $\mathrm{Poi}(\theta)$
What is most likely value of θ?

What is value of θ that maximizes the prob that we would have seen this data?

$$L(x_1, \dots, x_n \mid \theta) = \prod_{i=1}^{n} P_X(x_i ; \theta)$$

$$= \prod_{i=1}^{n} \frac{e^{-\theta} \theta^{x_i}}{x_i!}$$

$P(x_1, \dots, x_n ; \theta)$

θ is fixed

fn of $x_1, \dots, x_n$

$L(x_1, \dots, x_n \mid \theta)$

fixed $x_1 \dots x_n$

variable θ

Maximum Likelihood Estimation (Poisson)

Let's say $x_1, x_2, \ldots, x_n$ are iid samples from $Poi(\theta)$. (might look like $x_1 = 3, x_2 = 5, x_3 = 4$, etc.) What is the MLE of θ?

$$L(\boldsymbol{x} \mid \theta) = \prod_{i=1}^{n} p_X(x_i\,;\,\theta) = \prod_{i=1}^{n} e^{-\theta}\frac{\theta^{x_i}}{x_i!}$$

find θ to maximize this

$$\boxed{LL}(x \mid \theta)$$

$$\ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n}\left[-\theta + x_i \ln\theta - \ln(x_i!)\right]$$

$$\frac{\partial}{\partial\theta}\ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n}\left[-1 + \frac{x_i}{\theta}\right]$$

$$\sum_{i=1}^{n}\left[-1 + \frac{x_i}{\hat\theta}\right] = 0 \rightarrow -n + \frac{1}{\hat\theta}\sum_{i=1}^{n} x_i = 0 \rightarrow \boxed{\hat\theta = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

use $\hat\theta$ to denote soln to $\frac{d}{d\theta} LL(x \mid \theta) = 0$

$$\frac{\partial^2}{\partial\theta^2}\ln L(\boldsymbol{x} \mid \theta) = \sum_{i=1}^{n}\left[-\frac{x_i}{\theta^2}\right] < 0 \rightarrow \text{concave down everywhere}$$

don't need to check 2nd order conditions

General recipe

Given indep samples $x_1, \dots, x_n$ from parametric model $\rightarrow \text{Ber}(\theta)$ $\rightarrow \text{Poisson}(\theta)$

Find value of θ that maximizes $L(x_1 \dots x_n | \theta)$

discrete $\quad L(x_1, \dots, x_n | \theta) \triangleq \prod_{i=1}^{n} p_X(x_i; \theta)$

cont. $\quad L(x_1, \dots, x_n | \theta) \triangleq \prod_{i=1}^{n} f_X(x_i; \theta)$

find $\hat{\theta}$ that maximizes log likelihood $LL(x_1 \dots x_n | \theta)$

discrete $\quad = \sum_{i=1}^{n} \ln[p_X(x_i; \theta)]$

cont $\quad = \sum_{i=1}^{n} \ln[f_X(x_i; \theta)]$

Find $\hat{\theta}$ to max $LL(\vec{x} \mid \theta)$

$\vec{x}$: $x_1, \ldots, x_n$

distn has 1 parameter

compute $\frac{d\, LL(\theta)}{d\theta}$

set $\frac{d LL(\theta)}{d\theta} = 0$

solve for $\hat{\theta}$

don't need to do this

[verify soln is max (2^{nd} deriv < 0)]

multiple params $\theta_1, \ldots, \theta_k$

find $\hat{\theta}_1, \ldots, \hat{\theta}_k$ that are soln to this system

$$\begin{cases} \frac{\partial LL}{\partial \theta_1} = 0 \\ \frac{\partial LL}{\partial \theta_2} = 0 \\ \vdots \\ \frac{\partial LL}{\partial \theta_k} = 0 \end{cases}$$

[Check max

check Hessian

-ve definite]

General recipe

Given indep samples $x_1, \ldots, x_n$ from parametric model

$\exp(\theta)$

$x_1 = 2.1 \quad x_2 = 3.14 \ldots$

Find value of θ that maximizes $L(x_1 \ldots x_n | \theta)$

why product of densities?

① only care about relative likelihood and density captures that

cont. $L(x_1, \ldots, x_n | \theta) \triangleq \left. \prod_{i=1}^{n} f_X(x_i; \theta) \right]$

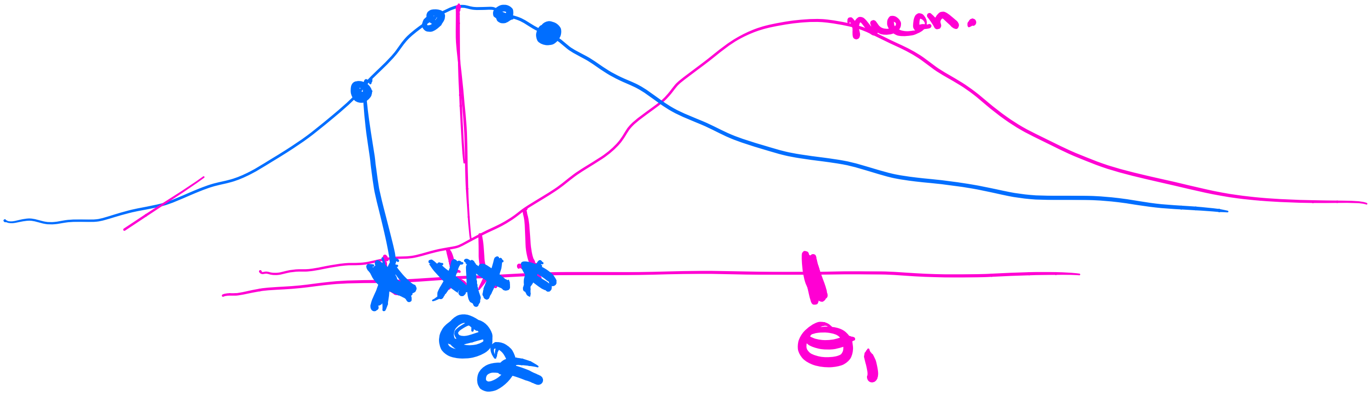
$\Pr(x \leq X \leq x+\varepsilon) \approx \underline{\varepsilon f_X(x)}$

find $\hat{\theta}$ that maximizes log likelihood $LL(x_1 \ldots x_n | \theta)$

cont $= \sum_{i=1}^{n} \ln\left[f_X(x_i; \theta)\right]$

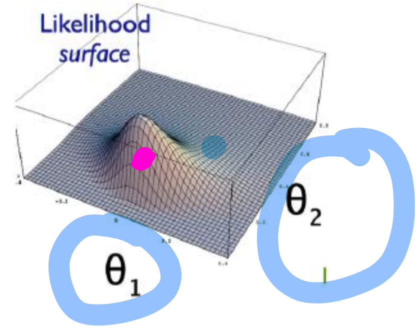
$X_1, \dots, X_n \sim N(\theta, 3)$

mean.



Random Picture

Maximum Likelihood Estimation (Normal)



$x_i \sim N(\mu, \sigma^2)$, μ, σ^2 both unknown

θ_1 θ_2

unknown. mean variance

$x_1, x_2, \ldots, x_n$

m.

$$L(x_1, \ldots, x_n \mid \theta_1, \theta_2) = \prod_{i=1}^{n} f_X(x_i; \theta_1, \theta_2)$$

$\ln(a \cdot b) = \ln(a) + \ln(b)$

$\ln(c^d) = d \ln c \ldots$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{(x_i - \theta_1)^2}{2\theta_2}}$$

Find θ_1, θ_2 to max $\rightarrow$

$$\ln L(x_1 \ldots x_n \mid \theta_1, \theta_2) = \sum_{i=1}^{n} \left[-\frac{1}{2} \ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \right]$$

$$\boxed{\frac{\partial \ln L}{\partial \theta_1} = 0} \qquad \boxed{\frac{\partial \ln L}{\partial \theta_2} = 0}$$

Solve this system $\Rightarrow$ $\hat{\theta}_1, \hat{\theta}_2$

$$\frac{\partial \ln L}{\partial \theta_1} = \sum_{i=1}^{n} \frac{2(x_i - \theta_1)}{2\theta_2} \qquad \frac{\partial \ln L}{\partial \theta_2} = \sum_{i=1}^{n}\left[-\frac{1}{2}\frac{2\pi}{2\pi\theta_2} + \frac{1}{2}\frac{(x_i-\theta_1)^2}{\theta_2^2}\right]$$

$$\sum_{i=1}^{n} \frac{(x_i - \hat{\theta}_1)}{\cancel{\hat{\theta}_2}} = 0$$

Set this $= 0$ $\Rightarrow$

$$\sum_{i=1}^{n} x_i = n\hat{\theta}_1$$

$$\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta}_1)^2$$

$$\boxed{\hat{\theta}_1 = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

sample mean.

$$\boxed{\hat{\theta}_2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \frac{1}{n}\sum_{j=1}^{n} x_j\right)^2}$$

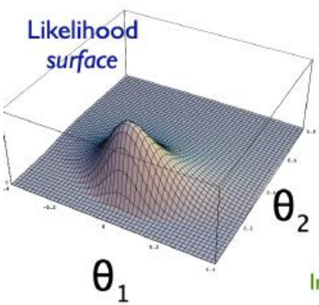
sample variance

Maximum Likelihood Estimation (Normal)

$x_i \sim N(\mu, \sigma^2),\ \mu, \sigma^2$ both unknown

$$\begin{aligned}
\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
\frac{\partial}{\partial\theta_1} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} \frac{(x_i - \theta_1)}{\theta_2} \quad = \quad 0
\end{aligned}$$

$$\boxed{\widehat{\theta}_1 \quad = \quad \left(\sum_{i=1}^{n} x_i\right)/n \quad = \quad \overline{x}}$$



Sample mean is MLE of population mean, again

In general, a problem like this results in 2 equations in 2 unknowns.
Easy in this case, since θ_2 drops out of the $\partial/\partial\theta_1 = 0$ equation

$x_i \sim N(\mu, \sigma^2),\ \mu, \sigma^2$ both unknown

$$
\begin{aligned}
\ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} -\frac{1}{2}\ln(2\pi\theta_2) - \frac{(x_i - \theta_1)^2}{2\theta_2} \\
\frac{\partial}{\partial\theta_2} \ln L(x_1, x_2, \ldots, x_n | \theta_1, \theta_2) &= \sum_{i=1}^{n} -\frac{1}{2}\frac{2\pi}{2\pi\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} = 0
\end{aligned}
$$

$$
\boxed{\widehat{\theta}_2 = \left(\sum_{i=1}^{n}(x_i - \widehat{\theta}_1)^2\right)/n = \bar{s}^2}
$$

Sample variance is MLE of *population* variance

Unbiased estimators

$N(\mu, \sigma^2)$

$\hat{\Theta} = \frac{1}{n}\sum_{i=1}^{n} x_i$ estimator for mean of normal distn

An estimator $\hat{\Theta}(X_1, \ldots, X_n)$ is unbiased if $E(\hat{\Theta}(X_1, \ldots, X_n)) =$ true param $= \mu$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$E(\bar{X}) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E(X_i)}_{\mu} = \mu$$

Our estimator for mean of normal distn unbiased

$\to X_i \sim Ber(\Theta)$

$E(X_i) = 1 \cdot \Theta + 0 \cdot (1-\Theta) = \Theta$

$X_1, \ldots X_n$ iid $Ber(\Theta)$

Is $\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ an unbiased estimator for Θ?

a) Yes

b) No

unbiased if $E(\bar{X}) = \Theta$

$X_1, \ldots, X_n$ iid $Ber(\Theta)$

Is X_1 an unbiased estimator?

a) Yes

b) No

$\rightarrow X_i \sim Ber(\theta)$

$E(X_i) = 1 \cdot \theta + 0 \cdot (1-\theta)$
$= \theta$

$X_1, \dots X_n$ iid $Ber(\theta)$

Is $\boxed{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i}$ an unbiased estimator for θ?

a) Yes

b) No

unbiased if $E(\bar{X}) = \theta$

MLE

$\text{Ber}(\theta)$

$x_1, \dots, x_n$

1 1 0 0 1

$\hat{\theta}(\underline{x_1, \dots, x_n})$ maximizes $\underline{L(x_1, \dots, x_n \mid \theta)}$

$$\hat{\theta}(x_1, \dots, x_n) = \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i}_{\bar{X}}$$

Estimator unbiased.

$$E\left[\hat{\theta}(X_1, \dots, X_n)\right] = \theta$$

true param.

$$E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \theta$$

true param

$\text{Ber}(\theta)$