

Statistics: analyzing & understanding data

Common approach: use parametric model of data

$\text{Bin}(p)$ ,  $\text{Poi}(\lambda)$ ,  $\text{Exp}(\lambda)$ ,  $N(\mu, \sigma^2)$ ,  $\text{Uni}[a, b]$

Use  $\vec{\theta}$  to denote unknown parameters

Goal: Given indep samples  $x_1, x_2, \dots, x_n$  from parametric model, determine best choice of parameters  $\vec{\theta}$

Approach: Find MLE, most likely choice of  $\vec{\theta}$

$$L(x_1, \dots, x_n | \vec{\theta}) = \prod_{i=1}^n f(x_i | \vec{\theta})$$

likelihood  
function

density cont.  
p.m.f. cont.

$$LL(\vec{x} | \vec{\theta}) = \log(L(\vec{x} | \vec{\theta})) = \sum_{i=1}^n \log[f(x_i | \vec{\theta})]$$

log-likelihood  
function

choose  $\vec{\theta}$  to maximize  $L(\vec{x}|\vec{\theta})$

$$\Leftrightarrow \text{maximize } LL(\vec{x}|\vec{\theta})$$

1 parameter

compute  $\frac{dLL}{d\theta}$

$$\text{set } \frac{dLL}{d\theta} = 0$$

Solve

verify soln is max ( $2^{\text{nd}} \text{ deriv} < 0$ )

Multiple parameters

$$\frac{\partial LL}{\partial \theta_1} = 0$$

$$\frac{\partial LL}{\partial \theta_2} = 0$$

$\vdots$

$$\frac{\partial LL}{\partial \theta_k} = 0$$

check max  
Hessian -ve definite

$$\frac{\partial^2 f}{\partial x_i \partial x_j}$$

Example:

$k_1, \dots, k_n$  samples from geometric dist'n, param  $\theta$

Find MLE of  $\theta$

$$L(\vec{k}|\theta) = \prod_{i=1}^n (1-\theta)^{k_i-1} \theta$$

$$p(k_i) = (1-\theta)^{k_i-1} \theta$$

$$LL(\vec{k}|\theta) = \sum_{i=1}^n (k_i-1) \log(1-\theta) + n \log \theta$$

$$\frac{d}{d\theta} LL(\vec{k}|\theta) = -\sum_i \frac{(k_i-1)}{(1-\theta)} + \frac{n}{\theta}$$

$$\frac{d}{d\theta} LL(\vec{k}|\theta) = 0 \Rightarrow \frac{n}{\hat{\theta}} = \sum_i \frac{(k_i-1)}{(1-\hat{\theta})}$$

$$\equiv (1-\hat{\theta})n = \hat{\theta} \sum_i (k_i-1)$$

$$\equiv \hat{\theta} = \frac{n}{\sum_i k_i}$$

$X_1, \dots, X_n$  samples from  $U[0, \theta)$

↑  
unknown

Find MLE of  $\theta$

$$f(x_i) = \begin{cases} \frac{1}{\theta} & 0 \leq x_i < \theta \\ 0 & \text{o.w.} \end{cases}$$

$$L(\vec{x}|\theta) = \frac{1}{\theta^n} \quad \text{if all } x_i\text{'s} \in [0, \theta]$$

0 otherwise

$$L(\vec{x}|\theta) \downarrow \text{ as } \theta \uparrow \Rightarrow \hat{\theta} = \max(x_i)$$

Estimator unbiased if

$$E(\hat{\theta}) = \theta$$

↑  
true value

Last time for mean of normal distn

$$E(\hat{\theta}) = E\left(\frac{\sum x_i}{n}\right) = \mu \quad \text{unbiased.}$$

for  $U[0, \theta]$   $\hat{\theta} = \max(x_1, \dots, x_n)$

$$E(\max(X_1, \dots, X_n)) = ? \quad \Pr(\hat{\theta} \leq x) = \prod_{i=1}^n \Pr(X_i \leq x)$$

$\downarrow$   
 $U[0, \theta]$

$$= \begin{cases} \frac{x^n}{\theta^n} & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$$

$$f_{\hat{\theta}}(x) = \begin{cases} \frac{n x^{n-1}}{\theta^n} & 0 \leq x \leq \theta \\ 0 & \text{o.w.} \end{cases}$$

$$E(\hat{\theta}) = \int_0^{\theta} x f_{\hat{\theta}}(x) dx = \frac{n}{\theta^n} \int_0^{\theta} x \cdot x^{n-1} dx = \frac{n}{n+1} \theta$$

$\Rightarrow \hat{\theta}$  not unbiased, but it is asymptotically unbiased

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta$$

Estimator **consistent** if  $\hat{\Theta}_n \xrightarrow[n \rightarrow \infty]{} \Theta$  no matter what  $\Theta$  is in probability

$$\Pr(\max(X_1, \dots, X_n) \leq \Theta - \varepsilon) = \left(\frac{\Theta - \varepsilon}{\Theta}\right)^n = \left(1 - \frac{\varepsilon}{\Theta}\right)^n \rightarrow 0 \text{ as } n \rightarrow \infty$$

(under mild technical conditions, MLE always consistent)

iid observations  $X_1, \dots, X_n \sim U[\Theta, \Theta+1]$

Find MLE of  $\Theta$

$$f_{X_i}(x_i) = \begin{cases} 1 & \Theta \leq x_i \leq \Theta+1 \\ 0 & \text{o.w.} \end{cases}$$

$$L(\vec{x} | \Theta) = \begin{cases} 1 & \Theta \leq \min x_i \leq \max x_i \leq \Theta+1 \\ 0 & \text{otherwise} \end{cases}$$

Any  $\hat{\Theta} \in [\max X_i - 1, \min X_i]$  equally good

consistent since  $\min X_i \rightarrow \Theta$

$\max X_i \rightarrow \Theta+1$

If choose midpoint  $\hat{\Theta} = \frac{1}{2} [\max X_i + \min X_i - 1]$   
it's unbiased

$$E(\hat{\theta}) = \frac{1}{2} \left[ \underbrace{E(\max X_i)}_{\theta + 1 - \frac{1}{n+1}} + \underbrace{E(\min X_i)}_{\theta + \frac{1}{n+1}} - 1 \right] = \theta$$

Grade dist'n for 312

$$\begin{cases} A & \frac{1}{2} \\ B & \mu \\ C & 2\mu \\ F & \frac{1}{2} - 3\mu \end{cases}$$

MLE for  $\mu$  when samples are  $x_A, x_B, x_C, x_F$   
(Assume we see precise grade each person got)

$$L(x_A, x_B, x_C, x_F | \mu) = \left(\frac{1}{2}\right)^{x_A} \mu^{x_B} (2\mu)^{x_C} \left(\frac{1}{2} - 3\mu\right)^{x_D}$$

$$\log L(\vec{x} | \mu) = x_A \log\left(\frac{1}{2}\right) + x_B \log(\mu) + x_C \log(2\mu) + x_D \log\left(\frac{1}{2} - 3\mu\right)$$

set  $\frac{dL}{d\mu} = 0$ , solve for  $\hat{\mu}$

$$\frac{x_B}{\hat{\mu}} + \frac{2x_C}{2\hat{\mu}} - \frac{3x_D}{\frac{1}{2} - 3\hat{\mu}} = 0$$

$$\frac{(x_B + x_C)}{\hat{\mu}} = \frac{3x_D}{\frac{1}{2} - 3\hat{\mu}}$$

$$\left(\frac{1}{2} - 3\hat{\mu}\right)(x_B + x_C) = 3\hat{\mu}x_D$$

$$\frac{x_B + x_C}{2} = 3\hat{\mu}(x_B + x_C + x_D)$$

$$\hat{\mu} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$$

Same problem, but suppose we don't see  $x_A$  &  $x_B$

We just see  $x_{AB} = \#$  students that got A or B

## Expectation-Maximization Algorithm EM

alg for finding MLE of parameters when there are hidden vars  
or computational issues with MLE

Idea:

If we knew  $z_A, z_B$  (where  $z_A + z_B = x_{AB}$ ) could compute  $\hat{\mu}$

If we knew  $\hat{\mu} \rightarrow$  could determine  $E(z_A), E(z_B)$

EM: Starts w/ guess for  $\mu$ :  $\mu^0$

Repeatedly  $t=0,1,\dots$

Expectation: (compute  $E(z_A | \mu^t, x_{AB})$ )

$$\bar{z}_A = E(z_A | \mu^t, x_{AB}) = \frac{x_{AB} \frac{1}{2}}{\frac{1}{2} + \mu^t}$$

$$\bar{z}_B = E(z_B | \mu^t, x_{AB}) = x_{AB} - E(z_A | \mu^t, x_{AB})$$

Maximization:

$$LL(z_A, z_B, x_C, x_D | \mu)$$

$$= z_A \log\left(\frac{1}{2}\right) + z_B \log(\mu) + x_C \log(2\mu) + x_D \log\left(\frac{1}{2} - 3\mu\right)$$

recall:  $\log L(\vec{x} | \mu) = x_A \log\left(\frac{1}{2}\right) + x_B \log(\mu) + x_C \log(2\mu) + x_D \log\left(\frac{1}{2} - 3\mu\right)$

$$E\left[LL(z_A, z_B, x_C, x_D | \mu)\right] \\ = \bar{z}_A \log\left(\frac{1}{2}\right) + \bar{z}_B \log(\mu) + x_C \log(2\mu) + x_D \log\left(\frac{1}{2} - 3\mu\right)$$

Find  $\mu$  that maximizes  $E[LL(z_A, z_B, x_C, x_D | \mu)]$

$$\text{call this } \mu^{t+1} = \frac{\bar{z}_B + x_C}{6(\bar{z}_B + x_C + x_D)}$$

until convergence



$$\text{recall } \hat{M} = \frac{x_B + x_C}{6(x_B + x_C + x_D)}$$

## Summary of structure of EM alg

- Samples  $x_1, \dots, x_n$  Missing data  $z_1, \dots, z_m$

Parameters:  $\theta_1, \dots, \theta_k$

Initialize  $\vec{\theta}^0$

repeat until convergence  $t=0, \dots$

Expectation:

$$\text{compute } E(z_i | \vec{x}, \vec{\theta}^t)$$

Maximization:

Set  $\vec{\theta}^{t+1}$  to

$$\text{maximize } E(\text{Likelihood}(\vec{x}, \vec{z} | \vec{\theta}))$$

## Mixture of Normals

Initialize  $\vec{\theta}^0 = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \tau_1)$   $\tau_2 = 1 - \tau_1$

Repeat until convergence:

Expectation:

$$\forall i \quad E(z_{i1} | x_i; \theta^+) = \frac{f_1(x_i | \vec{\theta}^+) \tau_1}{f_1(x_i | \vec{\theta}^+) \tau_1 + f_2(x_i | \vec{\theta}^+) \tau_2}$$

Maximization:

$$L(\vec{x}, \vec{z} | \theta) = \prod_{i=1}^n \left( \tau_1 f_1(x_i | \vec{\theta}) \right)^{z_{i1}} \left( \tau_2 f_2(x_i | \vec{\theta}) \right)^{z_{i2}}$$

$$= \prod_{i=1}^n \left( \frac{\tau_1}{\sqrt{2\pi\sigma_1^2}} \right)^{z_{i1}} \left( \frac{\tau_2}{\sqrt{2\pi\sigma_2^2}} \right)^{z_{i2}} e^{\frac{-z_{i1}(x_i - \mu_1)^2}{2\sigma_1^2}} e^{\frac{-z_{i2}(x_i - \mu_2)^2}{2\sigma_2^2}}$$

$$E_{\vec{z}}(\log L(\vec{x}, \vec{z} | \vec{\theta})) =$$

$$E\left(\sum_{i=1}^n z_{i1} \left[ \ln\left(\frac{\tau_1}{\sqrt{2\pi\sigma_1^2}}\right) - \frac{(x_i - \mu_1)^2}{2\sigma_1^2} \right] + \sum_{i=1}^n z_{i2} \left[ \ln\left(\frac{\tau_2}{\sqrt{2\pi\sigma_2^2}}\right) - \frac{(x_i - \mu_2)^2}{2\sigma_2^2} \right] \right)$$

$$= \sum_{i=1}^n E(z_{i1}) \left[ \ln\left(\frac{\tau_1}{\sqrt{2\pi\sigma_1^2}}\right) - \frac{(x_i - \mu_1)^2}{2\sigma_1^2} \right] + \sum_{i=1}^n E(z_{i2}) \left[ \ln\left(\frac{\tau_2}{\sqrt{2\pi\sigma_2^2}}\right) - \frac{(x_i - \mu_2)^2}{2\sigma_2^2} \right]$$

find  $\vec{\theta}^{t+1}$  to maximize this

using  $E(z_{i1}) E(z_{i2})$   
computed before

Turns out  $\mu_j = \frac{\sum_i E(z_{ij}) x_i}{\sum_i E(z_{ij})}$