Statistics: analyzing & understanding data

Common approach: use parametric model of data

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

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

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

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

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

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 $L(\vec{x}|\vec{\theta})$

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

1 parameter

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

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

Solve

verify soln is max (2^{nd} 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 θ

Find MLE of θ

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

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

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

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

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

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

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

$X_1, \ldots, X_n$ samples from $U[0, \underset{\substack{\uparrow \\ \text{unknown}}}{\theta})$

Find MLE of θ

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

$$L(\vec{x} \mid \theta) = \begin{cases} \frac{1}{\theta^n} & \text{if all } x_i\text{'s} \in [0, \theta] \\ 0 & \text{otherwise} \end{cases}$$

$L(\vec{x} \mid \theta) \downarrow$ as $\theta \uparrow$ $\implies \hat{\theta} = \max(x_i)$

Estimator $\boxed{\text{unbiased}}$ if $E(\hat{\theta}) = \underset{\substack{\uparrow \\ \text{true value}}}{\theta}$

Last time for mean of normal distn

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

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

$$E(\max(X_1,\dots,X_n)) = ?$$

$$Pr(\hat{\theta} \le x) = \prod_{i=1}^{n} \underbrace{Pr(X_i \le x)}_{U[0,\theta]}$$

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

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

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

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

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

Estimator $\boxed{\text{consistent}}$ if $\hat{\theta}_n \xrightarrow[n\to\infty]{} \theta$ no matter what θ is
in probability

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

(under mild technical conditions, MLE always consistent

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

Find MLE of θ

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

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

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

consistent since $\min X_i \longrightarrow \theta$

$\max X_i \to \theta+1$

If choose midpoint $\hat{\theta} = \frac{1}{2}[\max X_i + \min X_i - 1]$

it's unbiased

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

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 μ when samples are x_A, x_B, x_C, x_F

(Assume we see precise grade each person got)

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

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

set $\frac{dLL}{d\mu} = 0$, 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

$\boxed{\text{Expectation-Maximization Algorithm EM}}$

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 $\hat{\mu}$ $\longrightarrow$ could determine $E(z_A), E(z_B)$

EM: Starts w/ guess for μ: μ^0

Repeatedly $t = 0, 1, \ldots$

$\boxed{\text{Expectation:}}$ Compute $E(z_A \mid \mu^t, x_{AB})$

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

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

$\boxed{\text{Maximization:}}$

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

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

recall: $\log L(\vec{x} \mid \mu) = x_A \log(\frac{1}{2}) + x_1 \quad (\mu) + x_C \log(2\mu) + x_D \log(\frac{1}{2} - 3\mu)$

$$E\left[LL(z_A, z_B, x_C, x_D \mid \mu)\right]$$

$$= \bar{z}_A \log(\tfrac{1}{2}) + \bar{z}_B \log(\mu) + x_C \log(2\mu) + x_D(\log(\tfrac{1}{2} - 3\mu))$$

Find μ th- aximizes $E[LL(z_A, z_B, x_C, x_D \mid \mu)]$

call this μ^{t+1} $= \dfrac{\bar{z}_B + x_C}{6(\bar{z}_B + x_C + x_D)}$

until convergence

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

Summary of structure of EM alg

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

Parameters: $\theta_1, \ldots, \theta_k$

Initialize $\vec{\theta}^0$

repeat until convergence $t = 0, \ldots$

Expectation:

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

Maximization:

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

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

Mixture of Normals

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

Repeat until convergence:

Expectation:

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

Maximization:

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

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

$$E_{\vec{z}}\left(\log L(\vec{x}, \vec{z} \mid \vec{\theta})\right) =$$

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

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

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

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

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