DATE

Parametric Model:

PMF or PDF

$P(x) = f(x)$

$\mathcal{F} := \{ p(x;\theta) : \theta \in \Theta \}$

such that $\dim[\Theta] < \infty$

$\mathcal{F}_{Bern} = \{ \theta^x (1-\theta)^{1-x} : \theta \in (0,1) \}$

Joint Mass Function / Joint Density Function:

$$p(x_1, x_2, x_3, \ldots, x_n ; \theta) = \prod_{i=1}^{n} p_i(x_i;\theta) = \prod_{i=1}^{n} p(x_i;\theta)$$

If $x_1, x_2, x_3, \ldots, x_n$ are iid | independent, iid

In the real world, you observe data ex: <0, 0, 1, 0, 1, 0>

We say, $n = 6$ (# of observations)

Assumption 1: pick a parametric model $\mathcal{F}$

(pretend $\mathcal{F}$ = Bernouilli)

Jan 31, 2019

3 goals in Statistical Inference: $x_1, x_2, x_3 \ldots x_6$ iid Bern(θ)

① point estimation - best guess

② confidence set - interval of possible values of θ

③ test theories about θ (Hypothesis Testing)

$$P(x_1, \ldots, x_6; \theta) = \prod_{i=1}^{6} P(x_i;\theta)$$

Ex: $p(\langle 0\,0\,1\,0\,1\,0 \rangle; \theta) = (\theta^0(1-\theta)^{1-0})(\theta^0(1-\theta)^{1-0})(\theta^1(1-\theta)^{1-1}) \ldots$

$= (1-\theta)^4 \theta^2$

what if $\theta = \frac{1}{2}$ $= (\frac{1}{2})^6 = .0156$

what if $\theta = \frac{1}{4}$ $= (\frac{1}{4})^2 (\frac{3}{4})^4 = .0198$

smaller # is more likely

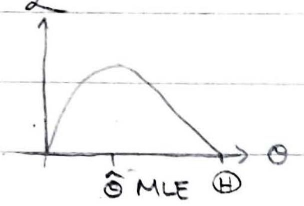
$P(\vec{x};\theta) = \mathcal{L}(\theta;\vec{x})$ understanding θ based on the data

JMF, Likelyhood

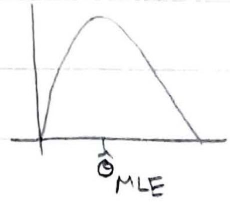
$\text{argmax}\{\mathcal{L}(\theta;\vec{x})\} = \hat{\theta}_{MLE}$ $\theta \in \Theta$

(maximum likelyhood estimate)

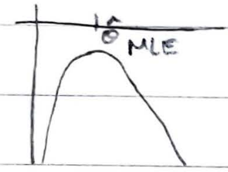
$\text{argmax}\{g(\mathcal{L}(\theta;x))\}$ $\theta \in \Theta$



$1.73\mathcal{L}$ ← doesn't change the MLE



$\ln(\mathcal{L})$



$\ell(\theta;x) := \ln(\mathcal{L}(\theta;x))$

$$\ell(\theta;x) = \ln(\theta^2(1-\theta)^4)$$

$$= 2\ln\theta + 4\ln(1-\theta)$$

$$\ell'(\theta;x) = \frac{2}{\theta} - \frac{4}{1-\theta} = 0$$

$$\frac{2}{\theta} = \frac{4}{1-\theta}$$

$$2-2\theta = 4\theta$$

$$2 = 6\theta$$

$$\boxed{\theta = \frac{2}{6}}$$

$x_1, \dots x_n$ is iid Bern(θ). Find $\hat{\theta}_{MLE}$

$$\ell(\theta;x) = \ln(\mathcal{L}(\theta;x)) = \ln\left(\prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$= \ln\left(\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}\right)$$

$$= (\Sigma x_i)(\ln\theta) + (n-\Sigma x_i)\ln(1-\theta)$$

$$\ell'(\theta;x) = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{n-\Sigma x_i}{1-\theta} = 0$$

$$\frac{n\bar{x}}{\theta} - \frac{n-n\bar{x}}{1-\theta} = 0$$

$$\frac{\bar{x}}{\theta} = \frac{1-\bar{x}}{1-\theta}$$

$$(\bar{x})(1-\theta) = \theta(1-\bar{x})$$

$$\boxed{\hat{\theta}_{MLE} = \bar{x}}$$

DATE

- $\hat{\theta}_{MLE}$ is consistant ($\hat{\theta}_{MLE}$ converges to θ)
- $\hat{\theta}_{MLE}$ converges to $N(\theta, SE[\hat{\theta}_{MLE}]^2)$ (asymptotic normality)
- $\hat{\theta}_{MLE}$ is efficient (lowest variance among all consistant variables)

Ex: $X_1, \ldots, X_n$ is iid $Geom(\theta) := (1-\theta)^x \theta$

Supposed you are running iid Bernoullis until the 1st "1". Then you stop + count the "0"s. You observe 49 "0"s

$x = 49, \quad \theta = 1\%$

$$P(x=49) = (1-1\%)^{49}(1\%) = .0061$$

$Supp(x) = \{0, 1, 2, \ldots\}$

Find the MLE:

$$\ell(\theta; x) = \ln\left(\prod_{i=1}^{n} (1-\theta)^{x_i}\theta\right)$$

$$= \ln\left((1-\theta)^{\sum x_i}\theta^n\right)$$

$$= \sum x_i \ln(1-\theta) + n\ln\theta$$

$$= n\bar{x}\ln(1-\theta) + n\ln\theta$$

$$\ell'(\theta; x) = -\frac{n\bar{x}}{1-\theta} + \frac{n}{\theta} = 0$$

$$-\frac{\bar{x}}{1-\theta} + \frac{1}{\theta} = 0$$

$$\frac{\bar{x}}{1-\theta} = \frac{1}{\theta}$$

$$\bar{x}\theta = 1-\theta$$

$$\theta(\bar{x}+1) = 1$$

$$\boxed{\hat{\theta}_{MLE} = \frac{1}{\bar{x}+1}}$$

What if $\bar{x} = 99$? $\hat{\theta}_{MLE} = 1\%$

What if $\bar{x} = 0$? $\hat{\theta}_{MLE} = 100\%$

In iid Bern, $\hat{\theta}_{MLE} \approx N(\theta, SE[\bar{x}]^2) = N\left(\theta, \sqrt{\frac{\theta(1-\theta)}{n}}^2\right)$

In iid Geom, $\hat{\theta}_{MLE} \approx N\left(\theta, SE\left[\frac{1}{1+\bar{x}}\right]^2\right)$

DATE

Confidence Interval:

$$CI_{\theta,1-\alpha} = \left[\hat{\theta}_{MLE} \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\right]$$

↑ Standard Norm Distr (1.96)

Hypothesis Test:

$H_0: \theta = \theta_0$ Null

$H_a: \theta \neq \theta_0$ Alternative

2 outcomes < retain H_0 / reject H_0

Retainment Region = $\left[\theta_0 \pm Z_{\frac{\alpha}{2}} SE[\hat{\theta}_{MLE}]\Big|_{\theta=\theta_0}\right]$

$\hat{\theta}_{MLE} \in$ Retainment Region < retain H_0 / reject H_0

MLEs are popular in frequentist inference

Frequentism: Believing

① θ is fixed and cannot be a RV

② Reliable or repeated sampling

PROBLEM: It's only normal when n is large

Ex: iid Bern case: $X = \langle 0,0,0 \rangle$

$\hat{\theta}_{MLE} = 0$ (but 0 isn't in the parameter space)

$CI_{\theta,1-\alpha} = [0,0] = \{0\}$

Ex: you know $\theta \notin [.1, .2]$

$\hat{\theta}_{MLE} = .14$ (not in Θ!)