Math 650
02/03/2021
Lec 02

In the real world you see $X = \langle 0,0,1,0,1,0 \rangle$, the "data". Then you pick a $\mathcal{F}$ (i.e. a Parametric model). But you don't know θ! So you have to guess θ. This guessing is "inference". There are typically three goals of "Satistical inference":

① Point estimate: Give me your best guess of θ (one value)

② Confidence sets: Give me a range of likely θ's.

③ Theory testing. Evaluate a theory about the value of θ

Assume, $\mathcal{F}$ = Bernoulli. Once you make an assumtion of the Parametric model, you can compute the JMF or JDF:

$$P(X;\theta) = \prod_{i=1} P(x_i,\theta)$$

$$P(\langle 0,0,1,0,1,0 \rangle; \theta) = (\theta^0(1-\theta)^{1-0})(\theta^0(1-\theta)^{1-0})(\theta^1(1-\theta)^{1-1})\ldots$$
$$= \theta^2(1-\theta)^4$$

If $\theta = 0.5 \Rightarrow P(X;\theta) = 0.5^2(1-0.5)^4 = 0.0156$
If $\theta = 0.25 \Rightarrow P(X;\theta) = 0.25^2(1-0.25)^4 = 0.0198$

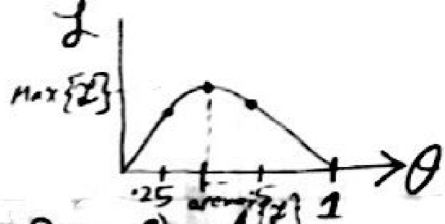
$\Rightarrow \theta = 0.25$ seems "more likely" than $\theta = 0.5$

$\mathcal{L}(\theta;X) = P(X;\theta)$, likelyhood function

Probability of the data with known θ.

likelyhood function. Probability of θ given x known or the likelyhood of "seeing" the parameter at a certain value.

MLE = Maximum Likelihood Estimator
Estimates



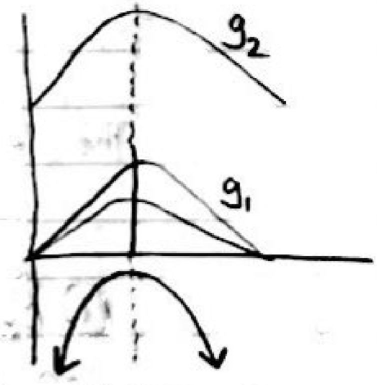
$$\int_{\Theta} \mathcal{L}(\theta;x)\,d\theta = \text{no rule} \qquad \sum_{Supp[X]} P(x;\theta) = 1$$

"^" means "estimator"

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

Let g be a strictly increasing function.

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



let $g = \ln$

$$= \text{argmax}\{\ln(\mathcal{L}(\theta;x))\}$$ → log-likelyhood

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

$$= \text{argmax}\{l(\theta;x)\}$$

$$l(\theta;x) = \ln(P(x;\theta)) \overset{\text{indp.}}{=} \ln\left(\prod_{i=1}^{n} P(x_i;\theta)\right)$$

$$= \sum_{i=1}^{n} \ln(P(x;\theta))$$

In our example of $x = \overbrace{\langle 0,0,1,0,1,0\rangle}^{n=6}$....

$$l(\theta;x) = \sum_{i=1}^{6} \ln\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)$$

$$= \sum_{i=1}^{6=n} \left(x_i \ln(\theta) + (1-x_i)\ln(1-\theta)\right)$$

$$= \left(\sum x_i\right)\ln(\theta) + \left(6-\sum x_i\right)\ln(1-\theta)$$

Note: $\bar{x} := \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x}$

$$= 6\bar{x}\ln(\theta) + (6-6\bar{x})\ln(1-\theta)$$

$$= 6\left(\bar{x}\ln(\theta) + (1-\bar{x})\ln(1-\theta)\right)$$

We need to find the argmax of this function...

$\hat{\hat{\theta}}$ ← this represents estimate (a realization from the estimator)

$\hat{\theta}$ = take derivative of the log likelyhood wrt θ and set = 0 and solve.

Scanned with CamScanner

$$\frac{d}{d\theta}[l(\theta;x)] = \cancel{6}\left(\frac{\bar{x}}{\theta} - \frac{1-\bar{x}}{1-\theta}\right) = 0 \Rightarrow \frac{\bar{x}}{\theta} = \frac{1-\bar{x}}{1-\theta}$$

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

$$\Rightarrow \hat{\theta}_{MLE} = \bar{x} = \frac{2}{6} = 0.33$$

The estimator, $\hat{\theta}_{MLE} = \bar{X}$ is a rv whose realizations are the estimates. This rv has nice properties:

① $\hat{\theta}_{MLE}$ is "consistent". This means that this estimator can provide arbitrary precision on θ given enough n.

② $\hat{\theta}_{MLE} \sim N(\theta, SE[\hat{\theta}_{MLE}]^2)$ asymptotic normality.

③ "Efficiency" means that among all consisten estimators, it has minimum variance.

Consider $X \sim Geom(\theta) := (1-\theta)^x\theta \Rightarrow Supp[X] = \{0,1,2,\ldots\}$ $\Theta = (0,1)$

Consider a sequence of iid Bernoulli thetas. This rv tells you the number of failures (realizations of zero) before the first success. (realizations of one).

If $\theta = 1\%$

$\underset{1^{st}}{0}, 0, 0, \ldots\ldots, \underset{49th}{0}, \underset{50th}{1} \Rightarrow x = 49$ | $P(X=49; \theta=0.01) = (0.99)^{49}(0.01)$

$\mathcal{F}$ = iid Geometric. n realization

$$\mathcal{L}(\theta;x) = \prod_{i=1}^{n}(1-\theta)^{x_i}\theta = (1-\theta)^{\Sigma x_i}\theta^n$$

$$l(\theta;x) = \ln(1-\theta)^{\Sigma x_i}\theta^n = (\Sigma x_i)\ln(1-\theta) + n\ln(\theta)$$

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

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

Let's find the MLE. we take the derivative of the log-likelyhood wrt θ and set it equal to zero and solve.

$$\frac{d}{d\theta}[\ell] = n\left(-\frac{\bar{X}}{1-\theta} + \frac{1}{\theta}\right) \overset{set}{=} 0 \Rightarrow \frac{1}{\theta} = \frac{\bar{X}}{1-\theta} \Rightarrow 1-\theta = \theta\bar{X}$$

$$\Rightarrow \frac{1-\theta}{\theta} = \bar{X} \Rightarrow \frac{1}{\theta} - 1 = \bar{X} \Rightarrow \frac{1}{\theta} = \bar{X}+1$$

$$\Rightarrow \hat{\hat{\theta}}_{MLE} = \frac{1}{\bar{X}+1}$$

Consider $\bar{X} = 49 \Rightarrow \hat{\hat{\theta}}_{MLE} = \frac{1}{49+1} = 2\%$

Let's examine MLE Property #2:

$$\hat{\theta}_{MLE} \sim N(\theta, SE[\hat{\theta}_{MLE}]^2) = N\left(\theta, \sqrt{\frac{\theta(1-\theta)}{n}}^{\,2}\right)$$

In the curly-F: iid Binomial case

$$\hat{\theta}_{MLE} = \bar{X}, \; SE[\hat{\theta}_{MLE}] = SE[\bar{X}] = \sqrt{Var[\bar{X}]} = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\theta(1-\theta)}{n}}$$

In the $\mathcal{F}$: iid Geometric case,

$$\hat{\theta}_{MLE} = \frac{1}{\bar{X}+1}, \; SE\left[\frac{1}{\bar{X}+1}\right] = \text{?}$$ difficult without more mathematics.

We now use property 2 to attack the other goals of inference:

Confidence **Sets**: we use a method called the "confidence interval":

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

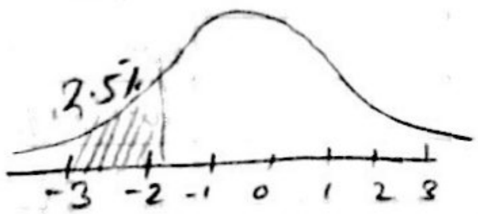
Parameter level of confidence std normal quantile at $\frac{\alpha}{2}$

For the iid Bernuli case:

$$CI_{\theta, 1-\alpha} = \left[\bar{X} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{x}(1-\bar{X})}{n}}\right]$$

$1-\alpha = 95\% \Rightarrow \alpha = 5\%$

$$CI_{\theta, 95\%} = \left[\bar{x} \pm 1.96 \sqrt{\frac{\bar{x}(1-\bar{x})}{n}} \right]$$



$$Z_{2.5\%} = -1.96$$