In the real world you see x = <0,0,1,0,1,0>, the "data".

Then, you pick a curly-F (i.e. a parametric model).
But you don't know theta! So you have to guess theta. This guessing is "inference". There are typically three goals of "statistical inference":
(1) Point estimation. Give me your best guess of theta (one value).
(2) Confidence sets. Give me a range of likely theta's.
(3) Theory testing. Evaluate a theory about the value of theta.

Assume curly-F = Bernoulli iid. Once you make an assumption of the parametric model, you can compute the JMF or JDF:

$$p(x;\theta) = \prod_{i=1}^{6} p(x_i;\theta)$$

$$p(<0,0,1,0,1,0>;\theta) = \left(\theta^0(1-\theta)^{1-0}\right)\left(\theta^0(1-\theta)^{1-0}\right)\left(\theta^1(1-\theta)^{1-1}\right)\cdots = \theta^2(1-\theta)^4$$

$$\text{if } \theta = 0.5 \Rightarrow p(x;\theta) = 0.5^2(1-0.5)^4 = 0.0156$$
$$\text{if } \theta = 0.25 \Rightarrow p(x;\theta) = 0.25^2(1-0.25)^4 = 0.0198$$

=> theta = 0.25 seems "more likely" than theta = 0.5.

$$\mathcal{L}(\theta;x) = p(x;\theta)$$

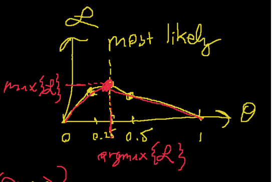
probability of the data with theta known

likelihood function, probability of theta given x known or the likelihood of "seeing" the parameter at a certain value.

$$\sum_{\Theta} \mathcal{L}(\theta;x) = \infty$$

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

Maximum Likelihood Estimator / Estimate

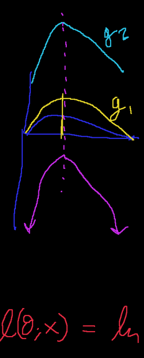


"hat" means "estimator"

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

Let g be a strictly increasing function.

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



let $g = \ln$

log-likelihood

$$= \text{argmax} \{\ln(\mathcal{L}(\theta;x))\}$$

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

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

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

$n = 6$

In our example of x =<0,0,1,0,1,0>...

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

$$= \left(\sum x_i\right)\ln(\theta) + \left(6 - \sum x_i\right)\ln(1-\theta) \qquad \text{Note: } \bar{x} := \frac{1}{n}\sum x_i \Rightarrow \sum x_i = n\bar{x}$$

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

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

$\hat{\hat{\theta}}_{MLE} =$ take derivative of the log likelihood wrt theta and set = 0 and solve.

this represents estimate (a realization from the estimator)

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

$$\Rightarrow \bar{x}(1-\theta) = (1-\bar{x})\theta \Rightarrow \bar{x} - \cancel{\bar{x}\theta} = \theta - \cancel{\bar{x}\theta} \Rightarrow \hat{\hat{\theta}}_{MLE} = \bar{x} = \frac{2}{6} = 0.33\ldots$$

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 theta given enough n.

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

standard error

(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 $\theta = 1\%$

$$\underset{1^{st}}{0}, \underset{2^{nd}}{0}, \underset{3^{rd}}{0}, \ldots, \underset{49^{th}}{0}, \underset{50^{th}}{1} \Rightarrow x = 49 \qquad P(X = 49; \theta = 0.01) = 0.99^{49}\,0.01$$

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

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

$$\ell(\theta;x) = \ln(\quad) = \left(\sum x_i\right)\ln(1-\theta) + n\ln(\theta) = n\bar{x}\ln(1-\theta) + n\ln(\theta)$$
$$= n\left(\bar{x}\ln(1-\theta) + \ln(\theta)\right)$$

Let's find the MLE. We take the derivative of the log-likelihood wrt theta and set it equal to zero and solve.

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

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

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

Let's examine MLE property #2:

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

In the curly-F: iid Bernoulli case,

$$\hat{\theta}_{MLE} = \bar{X}, \quad SE[\hat{\theta}_{MLE}] = SE[\bar{X}] = \sqrt{Var[\bar{X}]} \overset{241}{=} \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{\theta(1-\theta)}{n}}$$

In the curly-F: iid Geometric case,

$$\hat{\theta}_{MLE} = \frac{1}{\bar{X}+1}, \quad SE\left[\frac{1}{\bar{X}+1}\right] = ?$$

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":

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

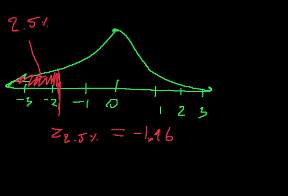
parameter

level of confidence

standard normal quantile at alpha/2

For the iid Bernoulli case,

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



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

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