$X_1, \dots, X_n \overset{iid}{\sim} \text{Bin}(\theta_1, \theta_2)$

Ecologists love this. They call it the capture-recapture estimation problem. Put down fishing rods. A fish comes up to the rod and get caught with prob θ_2. We also don't know how many fish come up which is θ_1. We only see x, the # of fish caught. You are really after θ_1, the population of fish.

Let's use MM estimation in this DGP for these two parameters.

$$\mu_1 = E[X] = \theta_1 \theta_2 \overset{\alpha_1}{=} \Rightarrow \theta_1 = \frac{\mu_1}{\theta_2}$$

$$\mu_2 = E[X^2] = Var[X] + \mu_1^2 = \theta_1 \theta_2 (1-\theta_2) + \theta_1^2 \theta_2^2 = \alpha_2$$

Let's invert the two alphas to get the beta functions:

$$\mu_2 = \theta_1\theta_2 - \theta_1\theta_2^2 + \theta_1^2\theta_2^2 = \frac{\mu_1}{\theta_2}\theta_2 - \frac{\mu_1}{\theta_2}\theta_2^2 + \frac{\mu_1^2}{\theta_2^2}\theta_2^2$$

$$= \mu_1 - \mu_1\theta_2 + \mu_1^2 \Rightarrow \mu_1\theta_2 = \mu_1 + \mu_1^2 - \mu_2 \Rightarrow \theta_2 = \frac{\mu_1 + \mu_1^2 - \mu_2}{\mu_1} = \frac{\mu_1 - (\mu_2 - \mu_1^2)}{\mu_1} = \beta_2$$

$$\Rightarrow \mu_1 = \theta_1 \left(\frac{\mu_1 - (\mu_2 - \mu_1^2)}{\mu_1}\right) \Rightarrow \theta_1 = \frac{\mu_1^2}{\mu_1 - (\mu_2 - \mu_1^2)} = \beta_1$$

Now we sub in the moment estimators for the mu's to get the MM estimators for both θ's:

$$\hat{\theta}_1^{MM} = \frac{\hat{\mu}_1^2}{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)} = \frac{\bar{x}^2}{\bar{x} - \hat{\sigma}^2}, \quad \hat{\theta}_2^{MM} = \frac{\hat{\mu}_1 - (\hat{\mu}_2 - \hat{\mu}_1^2)}{\hat{\mu}_1} = \frac{\bar{x} - \hat{\sigma}^2}{\bar{x}}$$

Imagine the data for n=5 is <3, 7, 5, 5, 6> $\Rightarrow \bar{x} = 5.2, \hat{\sigma}^2 = 2.64$

$$\hat{\hat{\theta}}_1^{M} = \frac{5.2^2}{5.2 - 2.64} = 10.56, \quad \hat{\hat{\theta}}_2^{m} = \frac{5.2 - 2.64}{5.2} = 0.49$$

You would probably round thetahathat1MM up to 11.
These estimates make sense. This data reasonably is realized from the DGP iid Bin(11, 0.49).

Imagine the data for n=5 is <3,7,5,11,6> $\Rightarrow \bar{x} = 6.4, \hat{\sigma}^2 = 10.56$

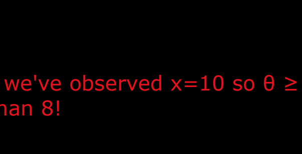
$$\hat{\hat{\theta}}_1^{MM} = \frac{6.4^2}{6.4 - 10.56} = -9.8, \quad \hat{\hat{\theta}}_2^{MM} = \frac{6.4 - 10.56}{6.4} = -0.65$$

These estimates do not make sense! There are not in the parameter space for θ_1 or θ_2.

We did not make a mistake. MM estimators just aren't very good sometimes. They have high error. And, they don't need to respect the parameter space so you can get "illegal" estimates.

Let's see another example of MM estimators failing.

$X_1, \dots, X_n \overset{iid}{\sim} U(0, \theta)$



$$\mu_1 = E[X] = \frac{0 + \theta}{2} = \frac{\theta}{2} = \alpha_1 \Rightarrow \beta_1 = \theta = 2\mu_1 \Rightarrow \hat{\theta}^{MM} = 2\hat{\mu}_1 = 2\bar{X}$$

Imagine the data for n=4 is <1,2,3,10> $\Rightarrow \bar{x} = 4$

$$\Rightarrow \hat{\hat{\theta}}^{MM} = 2 \cdot 4 = 8$$

This is clearly nonsensical since we've observed x=10 so θ ≥ 10 but we said it can't be greater than 8!

We've seen many problems. Let's explore another technique that can create estimators for parameters, the "maximum likelihood method" (Fisher popularized this between 1912-1922).

$$X_1, \dots, X_n \overset{iid}{\sim} DGP(\theta_1, \dots, \theta_K) \overset{\text{discrete}}{=} p(x; \theta_1, \dots, \theta_K) \quad PMF$$

$$\overset{\text{continuous}}{=} f(x; \theta_1, \dots, \theta_K) \quad PDF$$

I will just use the f notation. For discrete rv's just sub in p for f.

$$f(x_1, \dots, x_n; \theta_1, \dots, \theta_K) \overset{iid}{=} \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_K)$$

joint density function (JDF).

The θ's are values you need to know to calculate the value of f. But in statistics, θ's are not only unknown, their values are what we are trying to figure out! So we do the following conceptual inversion:

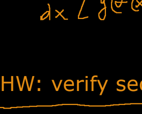
$$\prod_{i=1}^n \mathcal{L}(\theta_1, \dots, \theta_K; x_i) = \mathcal{L}(\theta_1, \dots, \theta_K; x_1, \dots, x_n) = f(x_1, \dots, x_n; \theta_1, \dots, \theta_K) = \prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_K)$$

We vary the θ's and ask "which value of θ gives the highest density (probability for discrete DGPs) and that θ is called the "maximum likelihood estimate" (MLE), $\hat{\hat{\theta}}^{MLE}$.

$$\hat{\theta}_1^{MLE}, \dots, \hat{\theta}_K^{MLE} := \underset{\vec{\theta} \in \Theta}{\text{argmax}}(\mathcal{L}) \overset{\text{if iid}}{=} \underset{\vec{\theta} \in \Theta}{\text{argmax}}\left(\prod_{i=1}^n \mathcal{L}(\theta_1, \dots, \theta_K; x_i)\right)$$

searching only the parameter space forces the max lik estimates to be legal!

$$f(x) = -x^2 + 4x + 1 = -(x-2)^2 + 5, \quad \max_{x \in \mathbb{R}}(f(x)) = 5$$



$$x_* = \text{argmax}(f(x)) := \{x : f(x) = \max(f(x))\}$$

$= 2$ $\quad f'(x) \overset{set}{=} 0$ and solve for x_*, check $f''(x_*) < 0$

Thm: the argmax is unaffected by f(x) being strung through a monotonically increasing function g. $\Rightarrow g'(y) > 0 \; \forall y$

WTS $\text{argmax}(g(f(x))) = \text{argmax}(f(x))$

$$\frac{d}{dx}[g(f(x))] = \underbrace{g'(f(x))}_{\text{constant} > 0} f'(x) \overset{set}{=} 0 \Rightarrow f'(x) = 0 \overset{\text{solving}}{\Rightarrow} x_*$$

HW: verify second derivative is negative?

Let g be the natural log function which is monotonically increasing.

$$\ell(\theta_1, \dots, \theta_K; x_1, \dots, x_n) := \ln(\mathcal{L}(\theta_1, \dots, \theta_K; x_1, \dots, x_n))$$ log-likelihood function

$$\hat{\hat{\theta}}_1^{MLE}, \dots, \hat{\hat{\theta}}_K^{MLE} = \text{argmax}(\ell)$$

$$\overset{\text{if iid}}{=} \text{argmax}\left(\ln\left(\prod_{i=1}^n f(x_i; \theta_1, \dots, \theta_K)\right)\right)$$

$$= \text{argmax}\left(\sum_{i=1}^n \ln(f(x_i; \theta_1, \dots, \theta_K))\right)$$

To solve for the argmax (the MLE's), we take each of the partial derivatives and set it = 0 and solve. The natural log makes everything super easy. Because taking the derivative of a sum of n functions is way better than the nightmare of taking the derivateve of a product of n functions. NOTE: this works if the max is not a boundary point (calculus 101).

$$\sum_{i=1}^n \frac{\partial}{\partial \theta_1}[\ln(f(x_i; \theta_1, \dots, \theta_K))] \overset{set}{=} 0 \quad \text{solve for } \hat{\hat{\theta}}_1^{MLE}$$

$$\sum_{i=1}^n \frac{\partial}{\partial \theta_2}[\ln(f(x_i; \theta_1, \dots, \theta_K))] \overset{set}{=} 0 \quad \text{solve for } \hat{\hat{\theta}}_2^{MLE}$$

$$\vdots$$

$$\sum_{i=1}^n \frac{\partial}{\partial \theta_K}[\ln(f(x_i; \theta_1, \dots, \theta_K))] \overset{set}{=} 0 \quad \text{solve } \hat{\hat{\theta}}_K^{MLE}$$

system of K equations

$X_1, \dots, X_n \overset{iid}{\sim} \text{Bern}(\theta), \quad K = 1$ Let's derive the maximum likelihood estimator for θ.

$$\sum_{i=1}^n \frac{\partial}{\partial \theta}[\ln(p(x_i; \theta))] = \sum_{i=1}^n \frac{\partial}{\partial \theta}\left[\ln\left(\theta^{x_i}(1-\theta)^{1-x_i}\right)\right]$$

$$= \sum_{i=1}^n \frac{\partial}{\partial \theta}[x_i \ln(\theta) + (1-x_i)\ln(1-\theta)] = \sum_{i=1}^n \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta}$$

$$= \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} \overset{set}{=} 0 \text{ find MLE}$$

$$\Rightarrow \frac{\sum x_i}{\theta} = \frac{n - \sum x_i}{1-\theta} \Rightarrow (1-\theta)\sum x_i = \theta(n - \sum x_i) \Rightarrow \sum x_i - \cancel{\theta \sum x_i} = \theta n - \cancel{\theta \sum x_i}$$

$$\Rightarrow \hat{\theta}^{MLE} = \frac{\sum x_i}{n} = \bar{x}$$

$X_1, \dots, X_n \overset{iid}{\sim} N(\theta_1, \theta_2)$

$$\frac{1}{\sqrt{2\pi\theta_2}} = (2\pi\theta_2)^{-\frac{1}{2}} = (2\pi)^{-\frac{1}{2}}\theta_2^{-\frac{1}{2}}$$

$$\sum_{i=1}^n \frac{\partial}{\partial \theta_1}\left[\ln\left(\frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x_i - \theta_1)^2}\right)\right] =$$

$$(x_i - \theta_1)^2 = x_i^2 - 2x_i\theta_1 + \theta_1^2$$

$$= \sum_{i=1}^n \frac{\partial}{\partial \theta_1}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta_2) - \frac{1}{2\theta_2}(x_i - \theta_1)^2\right] =$$

$$= \sum_{i=1}^n -\frac{1}{2\theta_2}\frac{\partial}{\partial \theta_1}[\cancel{x_i^2} - 2x_i\theta_1 + \theta_1^2] = \sum_{i=1}^n \frac{\cancel{2}x_i}{\cancel{2}\theta_2} - \frac{\cancel{2}\theta_1}{\cancel{2}\theta_2} = \frac{\sum x_i}{\cancel{\theta_2}} - \frac{n\theta_1}{\cancel{\theta_2}} \overset{set}{=} 0$$

$$\Rightarrow \hat{\theta}_1^{MLE} = \frac{\sum x_i}{n} = \bar{X}$$

For θ_2...

$$\dots = \sum \frac{\partial}{\partial \theta_2}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta_2) - \frac{1}{2\theta_2}(x_i - \theta_1)^2\right]$$

$$= \sum -\frac{1}{2\theta_2} + \frac{(x_i - \theta_1)^2}{2\theta_2^2} = -\frac{n}{\cancel{2\theta_2}} + \frac{\sum(x_i - \theta_1)^2}{\cancel{2}\theta_2^{\cancel{2}}} \overset{set}{=} 0$$

System of eq's

$$\Rightarrow n = \frac{\sum(x_i - \theta_1)^2}{\theta_2} \Rightarrow \hat{\theta}_2^{MLE} = \frac{1}{n}\sum(x_i - \theta_1)^2 = \frac{1}{n}\sum(x_i - \bar{x})^2 = \hat{\sigma}^2$$

once again is the sample variance without Bessel's correction