Method of Moments

$X_n \overset{iid}{\sim} f_\theta$ where $\theta = (\theta_1, \ldots, \theta_K)$

Let $\mu_1, \ldots, \mu_K$ be the first K moments

and $m_1, \ldots, m_K$ are the first K sample moments

We form a system of eqns

$$\begin{aligned} \mu_1 &= m_1 \\ \mu_2 &= m_2 \\ &\vdots \\ \underbrace{\mu_K}_{\text{depend on } \theta} &= m_K \end{aligned}$$

We then solve this system for $\theta_1, \ldots, \theta_K$ in terms of $\underset{\sim}{X}$.

Ex. Let $X_n \overset{iid}{\sim} Bin(k, p)$

Let's find MoM ests for k and p.

1. Get pop. moments

$$\mu_1 = \mathbb{E}X_n = kp$$

$$\mu_2 = \mathbb{E}X_n^2 = Var(X_n) + \mathbb{E}[X_n]^2$$

$$= kp(1-p) + k^2p^2$$

② Form sys. of eqns

$$\begin{cases} \mu_1 = kp = \overline{X} = m_1 \\ \mu_2 = kp(1-p) + k^2p^2 = \overline{X^2} = m_2 \end{cases}$$

③ Solve for k and p in terms of $\underset{\sim}{X}$.

$$\left\{ kp = \overline{X} \quad \text{and} \quad kp(1-p) + k^2p^2 = \overline{X^2} \right.$$

$$kp(1-p) + k^2p^2 = \overline{X^2}$$

$$\Downarrow \quad kp = \overline{X}$$

$$\overline{X}(1-p) + \overline{X}^2 = \overline{X^2}$$

$$\Downarrow \quad \overline{X}(1-p) = \overline{X^2} - \overline{X}^2$$

$$\Downarrow \quad 1-p = \frac{\overline{X^2} - \overline{X}^2}{\overline{X}}$$

$$\Downarrow \quad \boxed{\hat{p} = 1 - \frac{\overline{X^2} - \overline{X}^2}{\overline{X}}}$$

$$kp = \overline{X}$$

$$\Downarrow$$

$$\boxed{\hat{k} = \frac{\overline{X}}{\hat{p}}}$$

Lecture 5: Maximum Likelihood Estimation (MLE)

MoM examples

Ex. $X_n \overset{iid}{\sim} U(0,\theta)$

Let's get the MoM est. for θ.

$$\mu_1 = \mathbb{E}[X_n] = m_1 = \frac{1}{N}\sum_{n=1}^{N} x_n = \bar{X}$$

$$\mu_1 = \mathbb{E}[X_1] = \frac{0+\theta}{2} = \theta/2 = \bar{X}$$

Solve for θ then $\boxed{\hat{\theta}_{MOM} = 2\bar{X}}$

Ex. $X_n \overset{iid}{\sim} Beta(\alpha, 1)$

What is the MoM est. for α?

$$\mu_1 = \mathbb{E}X_n = m_1 = \bar{X}$$

$$\mu_1 = \mathbb{E}X_n = \frac{\alpha}{\alpha+1} = \bar{X}$$

Solve for α: $\alpha = (\alpha+1)\bar{X}$

$$\Rightarrow \alpha = \alpha\bar{X} + \bar{X}$$

$$\Rightarrow \alpha - \alpha\bar{X} = \bar{X}$$

$$\Rightarrow \alpha(1-\bar{X}) = \bar{X}$$

$$\Rightarrow \boxed{\hat{\alpha}_{MOM} = \frac{\bar{X}}{1-\bar{X}}}$$

Maximum Likelihood Estimator (MLE)

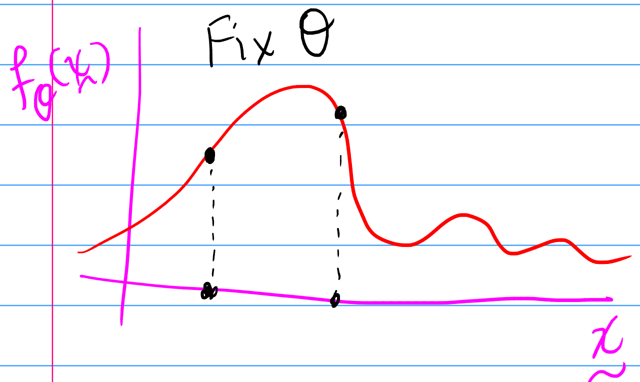
Assume $X_n \overset{iid}{\sim} f_\theta$ where $\theta \in \Theta$

recall: joint dist of my data

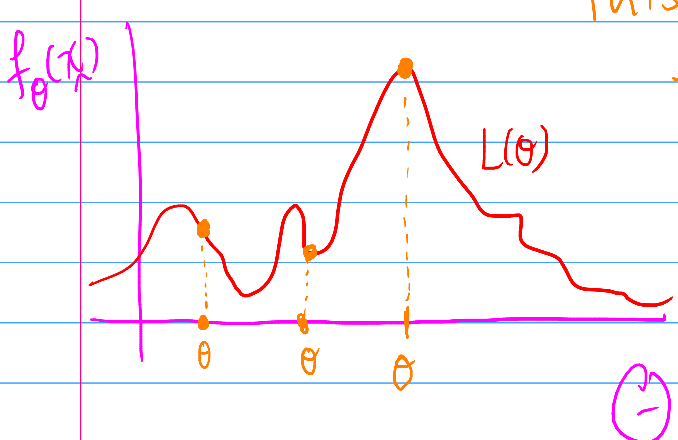
$$f_\theta(\underset{\sim}{x}) = \prod_{n=1}^{N} f_\theta(x_n)$$

typically we look at this as a fn of $\underset{\sim}{x}$

way 1: $f_\theta : \mathbb{R}^N \to \mathbb{R}$



We can also fix $\underset{\sim}{x}$ and think of this as a fn of θ



This view is called the likelihood function

$L : \Theta \to \mathbb{R}$

$L(\theta) = f_\theta(\underset{\sim}{x})$

MLE says choose $\hat{\theta}_{MLE}$ as the val. of θ that maximizes L

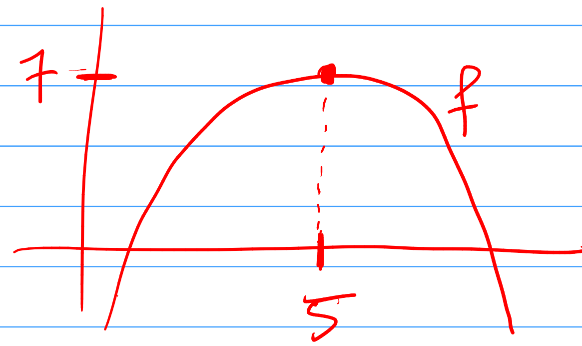
Defn: Maximum Likelihood estimator

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, L(\theta)$$

Ex,

$$\max_x f(x) = 7$$

$$\underset{x}{\operatorname{argmax}} \, f(x) = 5$$



Often we work with the log-likelihood fn

$$\ell(\theta) = \log L(\theta)$$

log = natural log

Alt. defn of $\hat{\theta}_{MLE}$

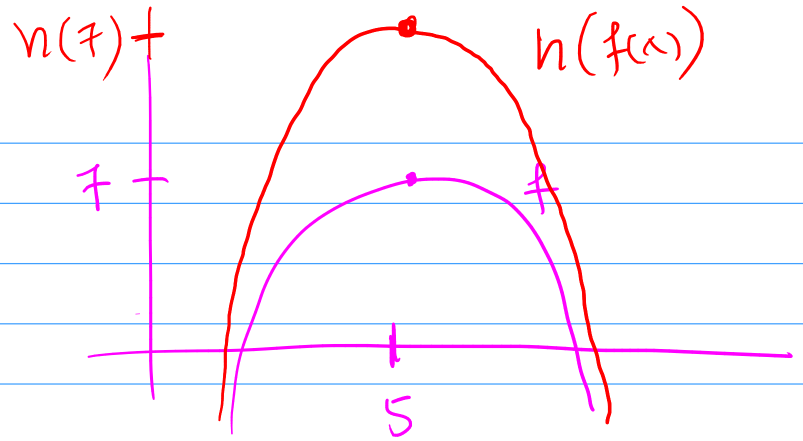
$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}} \, \ell(\theta)$$

equivalent defn.

b/c log is an increasing function

$$\underset{x}{\operatorname{argmax}} \, f(x) = \underset{x}{\operatorname{argmax}} \, h(f(x))$$

inc. fn.



by defn an inc. fn says $x_1 < x_2$ then $h(x_1) < h(x_2)$

Ex. $X_n \overset{iid}{\sim} N(\theta, 1)$ where $\theta \in \mathbb{R}$

what's the MLE?

(1) let's get $\ell(\theta)$

$$L(\theta) = f_\theta(\underset{\sim}{x}) = \prod_{n=1}^{N} f_\theta(x_n) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_n-\theta)^2\right)$$

$$= (2\pi)^{-N/2} \exp\left(\sum_{n=1}^{N} -\frac{1}{2}(x_n-\theta)^2\right)$$

$$L(\theta) = (2\pi)^{-N/2} \exp\left(-\frac{1}{2}\sum_{n=1}^{N}(x_n-\theta)^2\right) \propto e^{-\theta^2}$$

$$\ell(\theta) = \log L(\theta) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{n=1}^{N}(x_n-\theta)^2$$

(2) let's take a derivative $\frac{\partial \ell}{\partial \theta}$

$$\frac{\partial \ell}{\partial \theta} = 0 - \frac{1}{2}\sum_n \underbrace{\frac{\partial}{\partial \theta}(x_n-\theta)^2}_{2(x_n-\theta)(-1) = -2(x_n-\theta)}$$

$$= -\frac{1}{2}\sum_n -2(x_n-\theta)$$

$$= \sum_n (x_n-\theta)$$

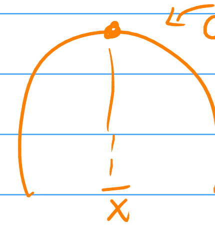
$$= \sum_n x_n - N\theta = 0$$

Calc says set $\frac{\partial \ell}{\partial \theta} = 0$ and solve for θ

$$\Rightarrow \sum_n x_n = N\theta$$

$$\Rightarrow \boxed{\hat{\theta}_{MLE} = \frac{1}{N}\sum_n x_n = \bar{x}}$$

concave down



Technically need to look at $\frac{\partial^2 \ell}{\partial \theta^2} < 0$

and also need to consider $\lim_{\theta \to \pm\infty} L(\theta) = 0$

$$\frac{\partial^2 \ell}{\partial \theta^2} = \frac{\partial}{\partial \theta}\left(\sum_n x_n - N\theta\right) = -N < 0$$

<u>Theorem: MLE are based on sufficient stats</u>

$$\hat{\theta}_{MLE} = \text{function}(T)$$

sufficient for θ

Pf. By factorization theorem if T is sufficient then

x only shows up through T

$$L(\theta) = f_\theta(\underset{\sim}{x}) = \underbrace{g(T,\theta)\, h(\underset{\sim}{x})}$$

and

$$\hat{\theta}_{MLE} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta) = \underset{\theta \in \Theta}{\operatorname{argmax}}\, g(T,\theta)\, \cancel{h(\underset{\sim}{x})}$$

$$= \underset{\theta \in \Theta}{\operatorname{argmax}}\, g(T,\theta)$$

$$= \text{function}(T)$$

Ex. Let $X_n \overset{iid}{\sim} \text{Bern}(p)$, $p \in [0,1]$

what is $\hat{p}_{MLE}$?

Bern:

$$f_p(x) = p^x (1-p)^{1-x}\, \mathbb{1}(x=0 \text{ or } 1)$$

$$= \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

1. lets get $L(p)$ and $\ell(p)$

$$L(p) = f_p(\underset{\sim}{x}) = \prod_{n=1}^{N} f_p(x_n)$$

$$= \prod_{n=1}^{N} p^{x_n}(1-p)^{1-x_n}\, \mathbb{1}(x_n = 0 \text{ or } 1)$$

$$\boxed{\sum_{n=1}^{N} x_n = N\bar{x}}$$

$$= p^{\sum_n x_n} (1-p)^{\sum_n (1-x_n)} \prod_n \mathbb{1}(x_n = 0 \text{ or } 1)$$

$$= p^{N\bar{x}} (1-p)^{N-N\bar{x}} \prod_n \mathbb{1}(x_n = 0 \text{ or } 1)$$

$$\ell(p) = \log L(p) = N\bar{x}\log p + (N - N\bar{x})\log(1-p)$$
$$+ \underbrace{\log\left(\prod_n \mathbb{1}(x_n = 0 \text{ or } 1)\right)}$$

always disappear if parameter doesn't change support

(2) take a derivative wrt. p

$$\frac{\partial \ell}{\partial p} = \frac{N\bar{x}}{p} - \frac{N - N\bar{x}}{1-p}$$

(3) Set $\frac{\partial \ell}{\partial p} = 0$ and solve for p

$$\frac{N\bar{x}}{p} - \frac{N - N\bar{x}}{1-p} = 0 \Rightarrow \frac{N\bar{x}}{p} = \frac{N - N\bar{x}}{1-p}$$
$$\Rightarrow N\bar{x}(1-p) = (N - N\bar{x})p$$
$$\Rightarrow N\bar{x} - \cancel{pN\bar{x}} = Np - \cancel{pN\bar{x}}$$
$$\Rightarrow N\bar{x} = Np$$
$$\Rightarrow \boxed{\hat{p}_{MLE} = \bar{x}} = \frac{\sum x_n}{N} = \frac{\#\text{ of 1s}}{\text{total }\#}$$

= pct. of 1s in my data

Consider

$$\eta = \frac{p}{1-p} = \text{odds}$$

I might want $\hat{\eta}_{MLE}$.