Likelihood

One of the most fundamental concepts in <u>statistical Inference</u>.

- → <u>Point Estimation</u> (single best guess of <u>some quantity of interest</u> θ^*)
- → <u>Set Estimation</u> (guess a set that traps some quantity of interest with a high probability <u>eg</u>:- confidence interval)
- → <u>hypothesis testing</u> (set up to falsify null hypothesis of interest)

<u>other Point/set Estimation problems</u>

- → Regression
- → Classification.
- → Risk Minimization
- → ...

our current focus.

<u>Two types of Estimation</u>: $\boxed{\text{Parametric}}$ and $\boxed{\text{Non-parametric}}$ <u>Experiments</u>.

<u>Dfn</u> Likelihood Function.

Suppose $X_1, X_2, \cdots, X_n$ have joint density $f(x_1, x_2, \cdots, x_n; \theta)$ specified by parameter $\theta \in \Theta$. Let the observed data be $x_1, x_2, \cdots, x_n$. The likelihood function is:

$$L_n(\theta) : \Theta \longrightarrow \mathbb{R} \quad , \quad L_n(\theta) = L_n(x_1, \cdots, x_n; \theta)$$
$$\propto f(x_1, \cdots, x_n; \theta)$$

proportional to

The log-likelihood function is:

$$\ell_n(\theta) := \log(L_n(\theta))$$

Example: Likelihood of IID Bernoulli(θ^*) experiment

$X_1, X_2, \ldots, X_n \overset{\text{IID}}{\sim}$ Bernoulli (θ^*) with joint density:

$$f(x_1, \ldots, x_n ; \theta) = \prod_{i=1}^{n} f(x_i ; \theta) = \prod_{i=1}^{n} \mathbb{1}_{\{0,1\}} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n - \sum_{i=1}^{n} x_i}$$

Let us define the following statistic of our data:

$$T_n(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i : \mathbb{X}_n \longrightarrow \mathbb{T}_n .$$

with realization $T_n(x_1, \ldots, x_n) = t_n = \sum_{i=1}^{n} x_i$

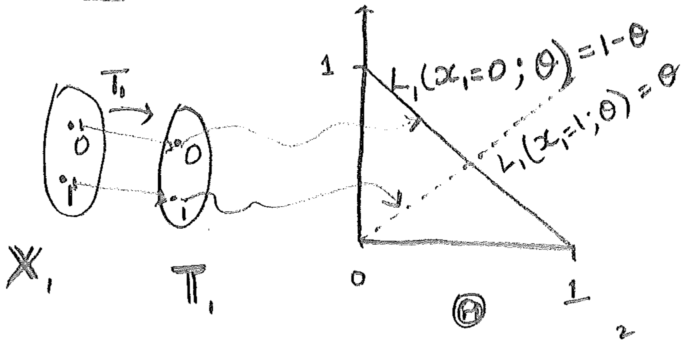
Therefore,

$$L_n(\theta) = L_n(x_1, \ldots, x_n ; \theta) = f(x_1, \ldots, x_n ; \theta)$$

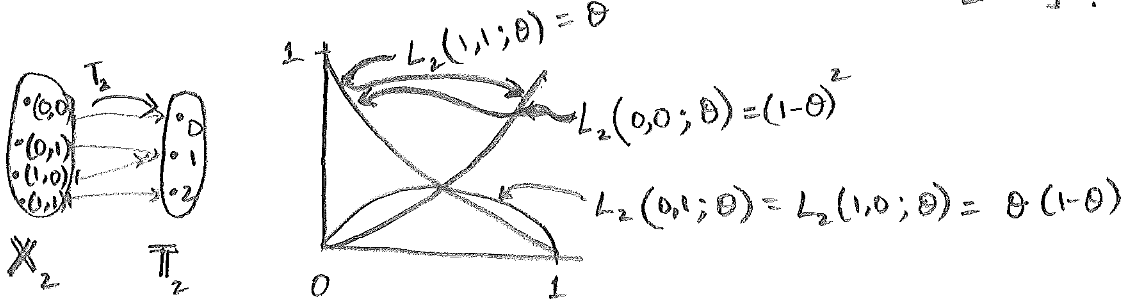
$$= \theta^{t_n} (1-\theta)^{n-t_n}$$

and $\ell_n(\theta) = t_n \log(\theta) + (n-t_n) \log(1-\theta)$.

Pictures:



Exercise: Draw the picture for $\mathbb{X}_3$, $\mathbb{T}_3$, and 4 likelihood functions over $\Theta = [0,1]$.



<u>Example</u> Recall the Lotto Data of "1^{st} ball being odd/even".
Let us look at the Likelihood function and log-likelihood function:

$L_n(\theta)$ and $l_n(\theta)$ for $n = 1, 2, \dots, 1114$.

<u>SAGE</u>. <u>interact</u>. Recall Model: $X_1, X_2, \dots, X_{1114} \overset{IID}{\sim} \text{Bernoulli}(\theta^*)$

fixed & unknown
we suspect it to be $1/2$

What are our observations from this example?

1. as n increases what is happening to the $L_n(\theta)$ and $l_n(\theta)$?
2. Behaviour of t_n/n? as n increases.
3. How can we use t_n to estimate θ^*?

— x —

<u>Maximum</u> <u>Likelihood</u> <u>Estimator</u> (MLE).

Let $X_1, X_2, \dots, X_n \sim f(x_1, \dots, x_n; \theta^*)$

The MLE $\widehat{\Theta}_n$ of the fixed and possibly unknown parameter $\theta^* \in \Theta$ is the value of θ that maximises the likelihood function:

$$\widehat{\Theta}_n := \widehat{\Theta}_n(X_1, X_2, \dots, X_n) := \underset{\theta \in \Theta}{\operatorname{argmax}}\, L_n(\theta).$$

since log is monotone

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

The realization is max lkl estimate (MLe) $\hat{\theta}_n = \widehat{\Theta}_n(x_1, \dots, x_n)$.

Example Bernoulli experiment. $X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(\theta^*)$

What is MLE, $\hat{\Theta}_n \overset{?}{=} \underset{\theta \in [0,1]}{\operatorname{argmax}} \, l_n(\theta)$.

$$l_n(\theta) = \log\left(\theta^{t_n}(1-\theta)^{n-t_n}\right) = t_n \log(\theta) + (n-t_n)\log(1-\theta).$$

Let us take derivative w.r.t. θ.

$$\frac{\partial}{\partial\theta} l_n(\theta) = \frac{\partial}{\partial\theta}\left(t_n \log(\theta) + (n-t_n)\log(1-\theta)\right)$$

$$= \frac{t_n}{\theta} + \frac{n-t_n}{(1-\theta)}(-1) = \frac{t_n}{\theta} - \frac{n-t_n}{1-\theta}$$

Now, set $\frac{\partial}{\partial\theta} l_n(\theta) = 0$ and solve for θ to obtain MLE

$$\frac{\partial}{\partial\theta} l_n(\theta) = 0 \iff \frac{t_n}{\theta} = \frac{n-t_n}{1-\theta} \iff \frac{1-\theta}{\theta} = \frac{n-t_n}{t_n}$$

$$\iff \frac{1}{\theta} - 1 = \frac{n}{t_n} - 1 \iff \theta = \frac{t_n}{n}$$

$\therefore$ MLE is:

$$\hat{\Theta}_n(X_1, \ldots, X_n) = \frac{1}{n} T_n(X_1, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}_n.$$

Application

For the 1114 Bernoulli Odd/Even Lotto 1^{st} Balls,

$t_{1114} = 546$ and the maximum likelihood estimate

(MLE) $\hat{\theta}_{1114} = \frac{546}{1114} = 0.49012567\ldots$ (see SAGE interact).

Example. Consider the de Moivre Experiment:

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{de Moivre}\left(\tfrac{1}{40}, \tfrac{1}{40}, \ldots, \tfrac{1}{40}\right)$$

what is the MLE of the parameter here?

Tricky Qn...

$$\hat{\Theta}_n = \underset{\theta \in \Theta := \{(\frac{1}{40}, \frac{1}{40}, \ldots, \frac{1}{40})\}}{\arg\max} \; l_n(\theta) = \left(\tfrac{1}{40}, \tfrac{1}{40}, \ldots, \tfrac{1}{40}\right)$$

since the parameter space Θ is a singleton or only has a single element.

However, if we asked the following question then you have two parameters to choose from:

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{de Moivre}(\theta^*),$$

where, $\theta^* \in \Theta = \left\{ \left(\tfrac{1}{40}, \tfrac{1}{40}, \ldots, \tfrac{1}{40}\right), \left(\tfrac{1}{40}, \tfrac{1}{40}, \ldots, \tfrac{1}{40}, \tfrac{1}{39}, 1 - \left(\tfrac{39}{40} + \tfrac{1}{39}\right)\right) \right\}$

So, you need to find the loglikelihood of each parameter in Θ and choose the one with the highest likelihood as the MLE $\hat{\Theta}_n$.

Let us look at this case in detail for a simpler problem.

Example The most likely of the three coins

Suppose there are three coins in a bag, where,

coin 1 has $\theta = 1/4$

coin 2 has $\theta = 3/4$

coin 3 has $\theta = 1/2$

Now, suppose you are given one particular coin from the bag and asked to guess (your single best guess) the coin (which one of the three coins is it?). You are allowed to toss the given coin three times.

<u>Experiment</u>. $X_1, X_2, X_3 \overset{\text{iid}}{\sim} \text{Bernoulli}(\theta^*)$,

$$\theta^* \in \left\{\tfrac{1}{4}, \tfrac{3}{4}, \tfrac{1}{2}\right\}$$

Suppose your 3 tosses are:

$$x_1 = 1, \quad x_2 = 0, \quad x_3 = 0$$

and MLE $\hat{\theta}_3 = \dfrac{1+0+0}{3} = \dfrac{1}{3} \notin \left\{\tfrac{1}{4}, \tfrac{2}{3}, \tfrac{1}{2}\right\}$.

So, the MLE $\hat{\theta}_n = \dfrac{\sum_{i=1}^{n} x_i}{n}$ is not a good idea if we have a finite set of θ's. Recall that the MLE for the product Bernoulli Experiment was derived on the continuous parameter space $[0,1]$.

So, we will compute the likelihood for each θ:

$$L\left(\tfrac{1}{4}\right) = f\left(x_1, x_2, x_3; \theta = \tfrac{1}{4}\right) = f\left(1,0,0; \theta = \tfrac{1}{4}\right) = \tfrac{1}{4}\cdot\tfrac{3}{4}\cdot\tfrac{3}{4} = \tfrac{9}{64}$$

$$L\left(\tfrac{2}{3}\right) = f\left(x_1, x_2, x_3; \theta = \tfrac{3}{4}\right) = f\left(1,0,0; \theta = \tfrac{3}{4}\right) = \tfrac{3}{4}\,\tfrac{1}{4}\,\tfrac{1}{4} = \tfrac{3}{64}$$

$$L\left(\tfrac{1}{2}\right) = f\left(x_1, x_2, x_3; \theta = \tfrac{1}{2}\right) = f\left(1,0,0; \theta = \tfrac{1}{2}\right) = \tfrac{1}{2}\cdot\tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{8} = \tfrac{8}{64}$$

Since $\frac{9}{64}$ is larger than $\frac{3}{64}$ and $\frac{8}{64}$

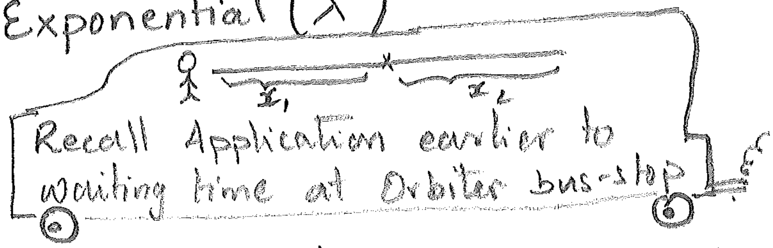
The most likely coin on the basis of the maximum likelihood <u>principle</u> is $\hat{\theta}_3 = \frac{1}{4}$.

Recall from last week that

MLE $\hat{\Lambda}_n = \frac{n}{\sum_{i=1}^{n} X_i}$ can be used to find a

point estimate $\hat{\lambda}_n = \frac{n}{\sum_{i=1}^{n} x_i}$ of $\lambda^* \in \mathbb{\Lambda} = (0, \infty)$ when,

$$X_1, X_2, \ldots, X_n \overset{IID}{\sim} \text{Exponential}(\lambda^*)$$



Application

Suppose we look at the times between Earth-quakes in NZ between 18-Jan-2008 02:23:44 to 18-Aug-2008 19:29:29 under a simplifying modelling assumption that

$$X_1, X_2, \ldots, X_{6128} \overset{IID}{\sim} \text{Exponential}(\lambda^*)$$

By measuring the times between earth-quakes in units of days, the sample mean $\overline{x}_{6128} = 0.0349$.

observed

- What is the MLE (maximum likelihood estimate) of the unknown and fixed rate parameter λ^* in our simple model of inter-earth-quake times?

$$\hat{\lambda}_{6128} = \frac{n}{\sum_{i=1}^{6128} x_i} = \frac{1}{\overline{x}_{6128}} = \frac{1}{0.0349} = 28.6694$$

- What is the mean # minutes between Earth Quakes?

$.0349 * 24 * 60 = 50.26$ minutes.