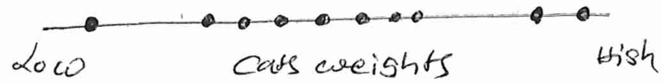
Motivation for Maximum Likelihood Estimation

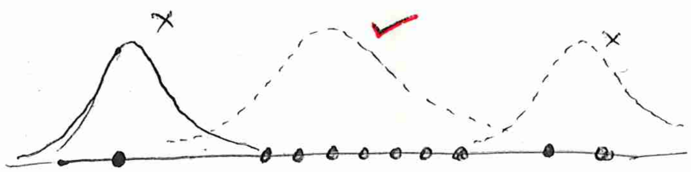
Let's say we weighted a bunch of cats

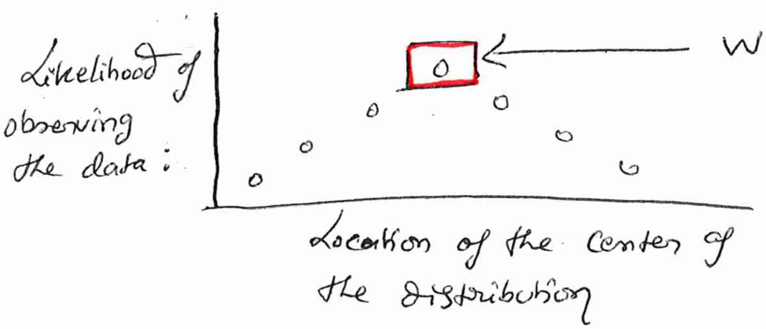


There are lots of different types of distributions for different types of data

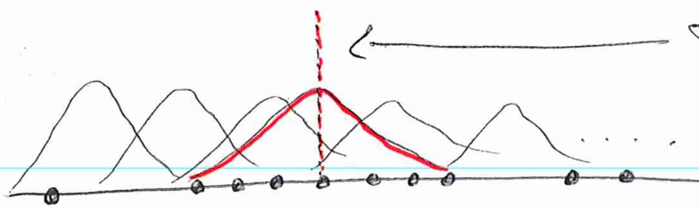
The reason you want to fit a distribution to your data is it can be easier to work with and it also more general - it applies to every experiment of same type.

↳ In this case, you can think that the weights ~~are~~ might be Normally Distributed. Now, "Normally distributed" means we might expect most of the measurements to be close to mean. and the measurements ~~Lxxxe~~ to be relatively symmetrical around the mean.





We want the location that "maximizes the likelihood" of observing the weights we measured.



This location for the mean "maximizes the likelihood" of observing the weights we measured.

Thus, it is the "maximum likelihood estimate for the mean". ✯

Similar thing happens for the "MLE for the standard deviation".

In statistics, "likelihood" refers to the situation, where we are trying to find the optimal value for the parameters of a distⁿ given a bunch of measured observations

Introduction to Maximum Likelihood Estimation.

↳ Maximum Likelihood Principle :

Given a dataset (assumed to be from a distribution), choose the parameter of interest θ, in a way that the data are most likely.

parameter (assumed known)

Consider a pdf / pmf : $f(x;\theta) = C \cdot P_X(x;\theta)$

Random variable — Coefficient $\perp \theta$ — ↳ Everything else

Likelihood function : $L(\theta) = P_X(x;\theta) = \frac{1}{C} f(x;\theta)$

$$0 \le L(\theta; x) < \infty$$

For Example :- Let, $X \sim Bin(n, p)$

Then, pmf, $f(x) = \binom{n}{x} p^x (1-p)^{n-x}$; $x \in \{0, 1, \cdots, n\}$
$0 \le p \le 1$.

$L(p) = \cancel{\binom{n}{x}} p^x (1-p)^{n-x}$; $0 \le p \le 1$
(Likelihood function)

$\Rightarrow L(p) = p^x (1-p)^{n-x}$; $0 \le p \le 1$

Maximum Likelihood Estimate (MLE):

↳ The MLE of a parameter θ is the value $\hat{\theta}$ that maximizes the likelihood, $L(\theta)$, given that x.

i.e., $\hat{\theta} = \arg\max_{\theta} L(\theta) \Rightarrow \frac{d}{d\theta} L(\theta) \Big|_{\hat{\theta}} = 0$

$$\Rightarrow L(\hat{\theta}) \ge L(\theta)$$

Log-likelihood : $l(\theta) = \log[L(\theta)]$ or $\ln[L(\theta)]$

It is valid because, $\log(x)$ is monotonically increasing

$\star$ Exercise

↳ Back to $X \sim Bin(n, p)$

$L(p) = p^x (1-p)^{n-x}$; Find the MLE $\hat{p}$.

Taking log on both sides,

$$\log[L(p)] = \log[p^x (1-p)^{n-x}]$$

or, $$l(p) = \log(p^x) + \log((1-p)^{n-x})$$

$$= x \log(p) + (n-x)\log(1-p) \; ; \; 0<p<1$$

Now, differentiating both sides,

$$l'(p) = \frac{x}{p} - \frac{n-x}{1-p} \; ; \; 0<p<1$$

$$l'(p)\Big|_{\hat{p}} = \left[\frac{x}{p} - \frac{n-x}{1-p}\right]\Bigg|_{\hat{p}} = 0$$

$$\Rightarrow \frac{x}{\hat{p}} - \frac{n-x}{1-\hat{p}} = 0 \Rightarrow \boxed{\hat{p} = x/n}$$ Sample proportion

2nd derivative test : We want to show $l''(\hat{p}) < 0$

$$l''(p) = \frac{-x}{p^2} - \frac{n-x}{(1-p)^2} < 0 \quad \forall p \in (0,1).$$

For boundaries; $L(0) = 0$, $L(1) = 0$; $L(\hat{p}) \geq 0$

Some more defns:

The score function: $S(\theta) = \frac{d}{d\theta} l(\theta)$ or $l'(\theta)$

Information function: $I(\theta) = -\frac{d^2}{d\theta} l(\theta) = -\frac{d}{d\theta} S(\theta)$

or, $S'(\theta)$.

Comments: On MLE $\hat{\theta}$, $S(\hat{\theta}) = 0 \leftarrow$ defn of MLE

$I(\hat{\theta}) > 0 \leftarrow$ 2nd derivative test.

Example:- Let X be a random variable with pmf.

$$f(x;\mu) = \frac{e^{-\mu x}(\mu x)^{x-1}}{x!} \; ; \; x \in \{1, 2, \cdots\}$$

or

$x \in \mathbb{N}^+$; $\mu \in [0,1]$

Find MLE $\hat{\mu}$.

↳ Step 1 :- The likelihood function, $\left[L(\theta) = \frac{1}{c} \cdot f(\cdots)\right]$

$\mu \in [0,1]$

$$L(\mu) = e^{-\mu x} \mu^{x-1}$$

Now, the log-likelihood function,

$$l(\mu) = \ln\left[e^{-\mu x} \mu^{x-1}\right]$$

$$\Rightarrow l(\mu) = -\mu x + (x-1)\ln(\mu) \quad , \quad 0 < \mu < 1 \quad \ldots (i)$$

Step-2 :- Now, differentiate both sides of Eq (i),

$$l'(\mu) = -x + \frac{x-1}{\mu}$$

Step-3 :- $l'(\hat{\mu}) = 0$

$$\Rightarrow -x + \frac{x-1}{\mu} = 0 \Rightarrow \boxed{\hat{\mu} = \frac{x-1}{x}}$$

Step-4 :- 2nd derivative test !!