Section 6.1: Maximum Likelihood Estimation (MLE)

1. A strategic gambler believes they have identified a faulty slot machine which pays out significantly more money than the other slot machines. She and her friends watch the machine 24 hours a day for 7 days and observed the slot machine paid out the $1,000,000 jackpot prize 10 times during the week. How can she figure out whether the machine is faulty or whether the number of jackpot prizes are within reason?

Random Variable

Parameter = Fixed

(a) Collect data: The decide to compare the performance of the suspect slot machine to other slot machines. They pick a random sample of 4 other slot machines and record how many jackpot prizes each machine pays over a one week time frame:

pdf → $\boxed{f(x;\lambda)=\frac{\lambda^x e^{-\lambda}}{x!}}$

$$\boxed{x_1 = 1\ ,\ x_2 = 3\ ,\ x_3 = 4\ ,\ x_4 = 8}$$

Most frequent

Get less and less frequent as $x \to \infty$

(b) What model best fits the data?

Poisson Distribution

Discrete

λ = mean # jackpots in one week X

(c) Determine the value of the parameter(s) of the model: Given the observed data, what are the most likely values of the parameters?

$$P(X_1=1, X_2=3, X_3=4, X_4=8) = P(X_1=1)\cdot P(X_2=3)\cdot P(X_3=4)\cdot P(X_4=8)$$

$$= \left(\frac{\lambda^1 e^{-\lambda}}{1!}\right)\cdot\left(\frac{\lambda^3 e^{-\lambda}}{3!}\right)\left(\frac{\lambda^4 e^{-\lambda}}{4!}\right)\left(\frac{\lambda^8 e^{-\lambda}}{8!}\right) = P(\text{of our sample})$$

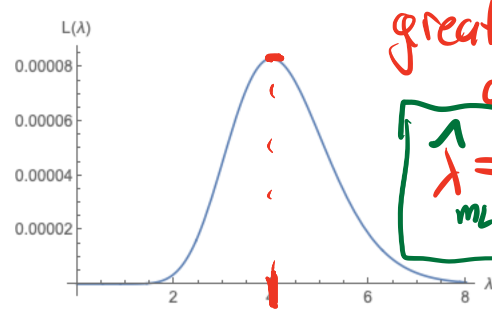
$f(1;\lambda)$

> The **likelihood function** $L(\theta) = L(\theta \mid x_1, x_2, \ldots x_n)$ gives the likelihood of the parameter θ given the observed data. A **maximum likelihood estimate,** $\hat{\theta}_{\text{MLE}}$, is a value of θ that maximizes the likelihood function.
> **MLE is a process for finding the best parameter(s) for a model based on a given dataset.**

2. Find the value of θ that maximizes the likelihood function from question 1.

$$L(\lambda \mid x_1=1, x_2=3, x_3=4, x_4=8)$$

$$L(\lambda \mid x_1=1, x_2=2, x_3=4, x_4=8) = \frac{\lambda^{16} e^{-4\lambda}}{1!\,3!\,4!\,8!}$$



$\lambda = 4$

$L(\lambda)$ is greatest at

$\boxed{\hat{\lambda}_{MLE} = 4}$

$$\boxed{L(\lambda) = \frac{\lambda^{16} e^{-4\lambda}}{1!\,3!\,4!\,8!}}$$ — Input = λ

Output = $P(\text{our sample if } \lambda = \lambda_0)$

Find value of λ that makes our sample most likely

Deriving the Likelihood Function

Let $f(x;\theta)$ denote the pdf of a random variable X with associated parameter θ. Suppose $X_1, X_2, \ldots, X_n$ are random samples from this distribution, and $x_1, x_2, \ldots, x_n$ are the corresponding observed values.

$$L(\theta \mid x_1, x_2, \ldots, x_n) = f(x_1;\theta)f(x_2;\theta)\ldots f(x_n;\theta) = \prod_{i=1}^{n} f(x_i;\theta).$$

3. For the following random samples, find the likelihood function:

Discrete

(a) $(x_1, x_2, x_3, x_4) = (1, 3, 3, 2)$ comes from $X \sim \text{Binom}(3, p)$.

$f(x;p) = \binom{3}{x} p^x (1-p)^{3-x}$

$n=3$

$$L(p \mid x_1=1, x_2=3, x_3=3, x_4=2) = \boxed{(3\cdot1\cdot1\cdot3)p^9(1-p)^3 = L(p)}$$

$$\underbrace{\binom{3}{1}p^1(1-p)^2}_{x_1=1}\ \underbrace{\binom{3}{3}p^3(1-p)^0}_{x_2=3}\ \underbrace{\binom{3}{3}p^3(1-p)^0}_{x_3=3}\ \underbrace{\binom{3}{2}p^2(1-p)^1}_{x_4=2}$$

$$\boxed{L(p) = 9p^9(1-p)^3}$$

(b) $x_1, x_2, x_3, \ldots, x_n$ come from $X \sim \text{Exp}(\lambda)$.

$f(x;\lambda) = \lambda e^{-\lambda x}$ $\boxed{\lambda = 1/\mu}$

μ = mean time between occurrences.

$$L(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \left(\lambda e^{-\lambda x_1}\right)\left(\lambda e^{-\lambda x_2}\right)\cdots\left(\lambda e^{-\lambda x_n}\right)$$

$$\boxed{L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}}$$

Maximizing the Likelihood Function

4 samples of 3 trials = 12 trials
9 successes
$\hat{p} = \frac{9}{12}$

4. Find the MLE for p when $(x_1, x_2, x_3, x_4) = (1, 3, 3, 2)$ comes from $X \sim \text{Binom}(3, p)$.

1. $L(p) = 9p^9(1-p)^3$ Find Likelihood Function

2. $dL/dp = 81p^8(1-p)^3 - 27p^9(1-p)^2$ (by prod rule)

$dL/dp = 27p^8(1-p)^2[3(1-p) - p]$

$dL/dp = 27p^8(1-p^2)(3-4p) = 0$ when

Critical points

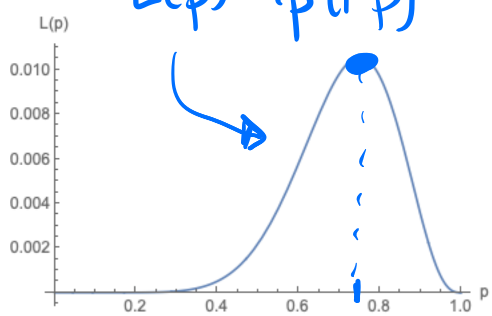
$p=0$ $L(0)=0$

$p=1$ $L(1)=0$

$p=-1$ $L(-1)=\text{und}$

$\boxed{p = 3/4 \quad L(3/4) \approx 0.01}$

$L(p) = 9p^9(1-p)^3$



3/4

$\boxed{\hat{p}_{MLE} = \frac{3}{4}}$

Steps for finding MLE, $\hat{\theta}_{\text{MLE}}$:

1. Find a formula the likelihood function.

$$L(\theta \mid x_1, x_2, \ldots, x_n) = f(x_1; \theta) f(x_2; \theta) \ldots f(x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta)$$ ✓

1A. Find log-likelihood function.

2. Maximize the likelihood function.
   (a) Take the derivative of L with respect to θ
   (b) Find critical points of L where $\frac{dL}{d\theta} = 0$ (or is undefined).
   (c) Evaluate L at each critical point and identify the MLE.

constant

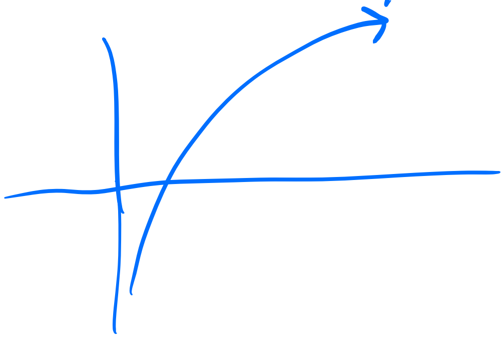
$L(0) = 0$

$L\left(\frac{\sum x_i}{n}\right) > 0$

5. Find the MLE for λ when $x_1, x_2, x_3, \ldots, x_n$ comes from $X \sim \text{Exp}(\lambda)$.

1. $L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$ $\Leftarrow$ Likelihood function $= \prod_{i=1}^{n} f(x_i ; \lambda)$

$\frac{dL}{d\lambda} = ?$



The greater $\ln(x)$ implies the greater x is. If we find the greatest value of $\ln(L(\lambda))$ that will correspond to value of λ that makes $L(\lambda)$ greatest.

value of λ that maximizes $L(\lambda)$ must also maximize $y = \ln(L(\lambda))$.

$$y = \ln\left(\lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}\right) = \ln(\lambda^n) + \ln\left(e^{-\lambda \sum_{i=1}^{n} x_i}\right)$$

$\ln(A \cdot B) = \ln(A) + \ln(B)$

log-likelihood function $\boxed{y = n \ln(\lambda) - \lambda \sum_{i=1}^{n} x_i}$

$$\frac{dy}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

$\lambda = 0$ is a critical point

$$\boxed{\hat{\lambda}_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i}} = \frac{1}{\bar{X}}$$

The value of θ that maximizes the **log-likelihood function** $y = \ln\Big(L(\theta \mid x_1, x_2, \ldots, x_n)\Big)$ will also the value that maximizes $L(\theta \mid x_1, x_2, \ldots, x_n)$.

Practice

6. Suppose a random variable with $X_1 = 5$, $X_2 = 9$, $X_3 = 9$, and $X_4 = 10$ is drawn from a distribution with pdf

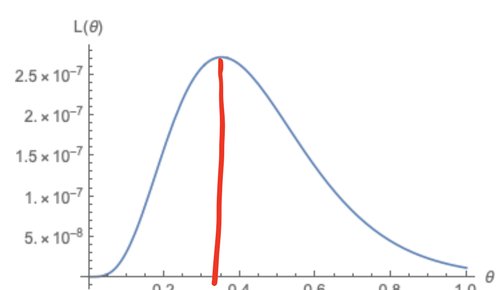
$$f(x;\theta) = \frac{\theta}{2\sqrt{x}} e^{-\theta\sqrt{x}}, \qquad \text{where x} > 0.$$

Find an MLE for θ.

1. Find Likelihood Function

1B. Find log-likelihood function

2 Maximize log-likelihood function.



0.351

Practice

6. Suppose a random variable with $X_1 = 5$, $X_2 = 9$, $X_3 = 9$, and $X_4 = 10$ is drawn from a distribution with pdf
$$f(x;\theta) = \frac{\theta}{2\sqrt{x}} e^{-\theta\sqrt{x}}, \qquad \text{where x} > 0.$$
Find an MLE for θ.

$L(\theta \mid x_1=5, x_2=9, x_3=9, x_4=10) = f(5;\theta)\,f(9;\theta)\,f(9;\theta)\,f(10;\theta)$

$$= \left(\frac{\theta}{2\sqrt{5}} e^{-\theta\sqrt{5}}\right)\left(\frac{\theta}{2\sqrt{9}} e^{-\theta\sqrt{9}}\right)\left(\frac{\theta}{2\sqrt{9}} e^{-\sqrt{9}}\right)\left(\frac{\theta}{2\sqrt{10}} e^{-\theta\sqrt{10}}\right)$$

$$L(\theta) = \frac{\theta^4 e^{-\theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10})}}{2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10}}$$

$$y = \ln(L(\theta)) = \ln\left(\frac{\theta^4 e^{-\theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10})}}{2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10}}\right)$$

$\ln(\theta^4) + \ln\left(e^{-\theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10})}\right) - \ln\left(2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10}\right)$

$$\boxed{y = 4\ln(\theta) - \theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}) - \ln(2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10})}$$

$$dy/d\theta = 4/\theta - (\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}) - 0 = 0$$

when $\theta = \dfrac{4}{\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}} \approx 0.351$ or $\theta = 0$

$$\boxed{\hat{\theta}_{MLE} \approx 0.351}$$

Practice

6. Suppose a random variable with $X_1 = 5$, $X_2 = 9$, $X_3 = 9$, and $X_4 = 10$ is drawn from a distribution with pdf

$$f(x;\theta) = \frac{\theta}{2\sqrt{x}} e^{-\theta\sqrt{x}}, \qquad \text{where x} > 0.$$

Find an MLE for θ.

$$L(\theta \mid x_1=5, x_2=9, x_3=9, x_4=10) = f(5;\theta)\, f(9;\theta)\, f(9;\theta)\, f(10;\theta)$$

$$= \left(\frac{\theta}{2\sqrt{5}} e^{-\theta\sqrt{5}}\right)\left(\frac{\theta}{2\sqrt{9}} e^{-\theta\sqrt{9}}\right)\left(\frac{\theta}{2\sqrt{9}} e^{-\sqrt{9}}\right)\left(\frac{\theta}{2\sqrt{10}} e^{-\theta\sqrt{10}}\right)$$

$$\boxed{L(\theta) = \frac{\theta^4 e^{-\theta(\sqrt{5}+\sqrt{9}+\sqrt{\ }+\sqrt{10})}}{2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10}}}$$

$$y = \ln(L(\theta)) = \ln\left(\frac{\theta^4 e^{-\theta(\sqrt{5}+\sqrt{9}+9+\sqrt{10})}}{2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10}}\right)$$

$$\ln(\theta^4) + \ln\left(e^{-\theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10})}\right) - \ln(2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10})$$

$$\boxed{y = 4\ln(\theta) - \theta(\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}) - \ln(2^4\sqrt{5}\sqrt{9}\sqrt{9}\sqrt{10})}$$

$$dy/d\theta = 4/\theta - (\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}) - 0 = 0$$

when $\theta = \dfrac{4}{\sqrt{5}+\sqrt{9}+\sqrt{9}+\sqrt{10}} \approx 0.351$ or $\theta = 0$

$\theta = 0$

$$\boxed{\hat{\theta}_{MLE} \approx 0.351}$$

Summary of Results

So far we have observed:

Distribution	$\hat{\theta}_{\text{MLE}}$
Binomial	$\hat{p}_{\text{MLE}} = \hat{p}$
Exponential	$\hat{\lambda}_{\text{MLE}} = \dfrac{1}{\bar{x}}$

Theorem 3. *Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from $N(\mu, \sigma)$. The maximum likelihood estimates of μ and θ are*

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x} \quad \textit{and} \quad \hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}.$$

□

- **MLE's give reasonable estimates that make sense!**
- MLE's are often good estimators since they satisfy several nice properties
  - *Consistency:* As we get more data (sample size goes to infinity), the estimator becomes more and more accurate and converges to the actual value of θ.
  - *Normality:* As we get more data, the MLE's converge to a normal distribution.
  - *Efficiency:* They have the smallest possible variance for a consistent estimator.
- **The downside is finding MLE's are not always easy (or possible).**