MH1820 Week 9

population

θ

e.g. $\theta = \mu, \sigma^2$ etc

$\to X_1, X_2, \dots, X_n$

i.i.d

$$\hat{\theta} = T(X_1, \dots, X_n)$$

1. Population, random samples, statistics and sampling distribution

$X_1, \dots, X_n$ μ, σ^2

2. Law of large numbers and CLT

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

n large

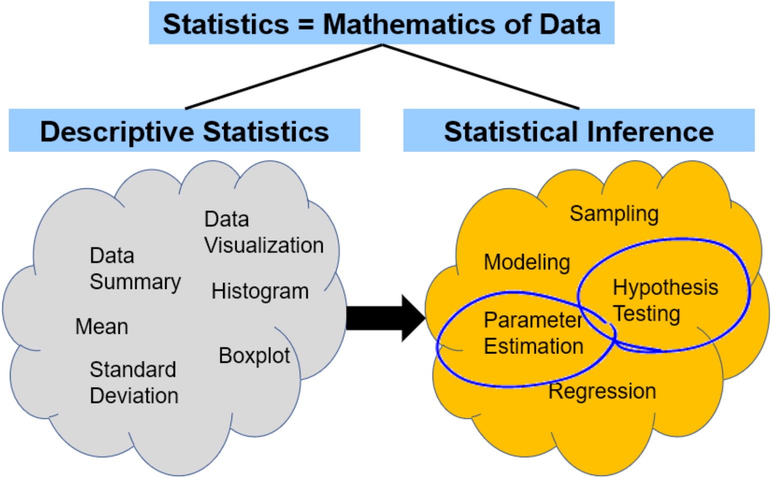
3. Parameter Estimation: Point Estimation

$$\bar{x} = \frac{X_1 + \dots + X_n}{n}$$

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{x})^2$$

Parameter Estimation: Point Estimation

- ad-hoc $\rightarrow$ try to connect $\hat{\theta}$ to $\bar{X}, S^2$ if possible.
- maximum likelihood


Statistics = Mathematics of Data
Descriptive Statistics
Statistical Inference
Data Summary
Data Visualization
Histogram
Mean
Boxplot
Standard Deviation
Sampling
Modeling
Hypothesis Testing
Parameter Estimation
Regression

Random sample $X_1, \ldots, X_n$ i.i.d.

Often the type of distribution ($N(\mu, \sigma^2)$, $Exp(\theta)$ etc.) of X_i is known, but its parameters μ, σ, θ etc. are unknown.

Parameter estimation: Extract information from $X_1, \ldots, X_n$ on these parameters.

$\hat{\theta}$

- A **point estimator** is a random variable that provides a "best guess" for a parameter.
- An **interval estimate** produces an (confidence) interval with random endpoints such that the true parameter (hopefully) with high probability is contained in the interval

$[\hat{\theta}_L, \hat{\theta}_U]$ contains θ with high probability.

Process of Point Estimation.

Given data/observations $x_1, \ldots, x_n$ as realizations of $X_1, \ldots, X_n$.

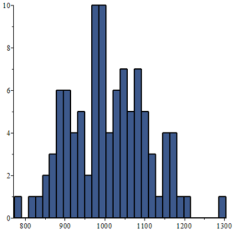
- Modeling: Identify a suitable type of distribution for X_i, which depends on parameter θ.
- Point estimation: Find functions $\widehat{\theta}(X_1, \ldots, X_n)$ which approximates θ.
- Substitute data $X_1 = x_1, \ldots, X_n = x_n$ into these functions to get estimates for θ.

Example: Data $x_1, \ldots, x_{100}$ (measurements in a physics experiment)

[1067., 773.2, 1119., 938.2, 1166., 1006., 881.4, 995.9, 1102., 1056.,
1045., 1091., 1170., 1085., 893.9, 1097., 1054., 959.3, 975.3, 969.4,
971.6, 1024., 984.2, 929.4, 1061., 998.4, 1209., 901.8, 864.2, 978.0,
1025., 1143., 858.0, 890.2, 1110., 1195., 944.0, 846.7, 872.7, 925.9,
1028., 980.5, 870.3, 1071., 1057., 1044., 987.0, 999.8, 981.4, 911.6,
1014., 1012., 825.4, 991.1, 1034., 944.8, 1001., 1097., 1149., 929.0,
1081., 994.1, 1174., 1050., 1162., 1081., 976.1, 1109., 1127., 1053.,
899.9, 1080., 941.4, 947.5, 1033., 912.1, 912.5, 1077., 1072., 1082.,
1005., 914.0, 1054., 883.9, 1164., 925.0, 1305., 1036., 998.7, 885.4,
998.2, 955.3, 883.7, 1155., 1095., 827.5, 993.0, 1152., 968.4, 976.6]

Step 1: Modeling: Normal distribution $N(\mu, \sigma^2)$ seems appropriate for this data. Want to estimate two parameters: μ, σ.

Histogram:



Step 2: Find functions to estimate

- Use sample mean to estimate μ (in view of Law of Large Number):

$$\mu \approx \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

- Use sample variance to estimate σ (not clear at this point why this is a good estimate):

$$\sigma \approx S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

Step 3: Sub in the data

- $\mu \approx \overline{x} = \frac{1}{100} \sum_{i=1}^{100} x_i = 1010.45.$
- $\sigma \approx S = \sqrt{\frac{1}{99} \sum_{i=1}^{100} (x_i - 1010.45)^2} = 98.54.$

$\square$

Write: $\hat{\theta} = \begin{cases} T(\bar{x}, s^2) \\ T(\bar{x}) \\ T(s^2) \end{cases}$

Idea:

When estimating a parameter θ which can be expressed as a **function of mean or variance**, we expect

$$\text{(sample mean)} \quad \overline{X} \approx \mathbb{E}[X_i] \quad \text{(population mean);}$$

$$\text{(sample variance)} \quad S^2 \approx \mathrm{Var}[X_i] \quad \text{(population variance).}$$

From the above, we then deduce an estimate $\widehat{\theta}$ of the parameter.

Example 12

Let $X_1, \ldots, X_m$ be i.i.d $\sim Binomial(n, p)$, where n and p are both unknown.

- Given: Observations $x_1, \ldots, x_m$
- Goal: Estimate n and p from $x_1, \ldots, x_m$ $\quad \hat{n}, \hat{p}$

$$X_i \sim \text{Binomial}(n, p)$$

$$\bar{X} \approx \mathbb{E}[X_i] = np \quad (1)$$

$$S^2 \approx Var[X_i] = np(1-p) \quad (2)$$

$$\frac{(2)}{(1)} \quad \frac{S^2}{\bar{X}} \approx \frac{\cancel{np}(1-p)}{\cancel{np}} \Rightarrow p \approx \boxed{1 - \frac{S^2}{\bar{X}}}$$

Choose $\hat{p} = \boxed{1 - \dfrac{s^2}{\overline{X}}}$ ✓

From ①: $n \approx \dfrac{\overline{X}}{p} \approx \dfrac{\overline{X}}{\hat{p}}$

$\approx \dfrac{\overline{X}}{1 - \dfrac{s^2}{\overline{X}}}$

So can choose $\hat{n} = \boxed{\dfrac{\overline{X}}{1 - s^2/\overline{X}}}$ ✓

Sample data: $x_1, \ldots, x_{1000}$ drawn from $Binomial(n, p)$ with unknown n and p.

12 13 7 10 6 11 12 6 7 10 10 9 10 15 11 14 9 8 9 12 10 16 16 135 11 7 6 9 19 7 16 9 11 14 9 8 13 16 8 7 11 10
10 7 12 11 1312 18 8 8 12 16 11 14 7 7 10 13 4 8 7 6 8 9 12 12 11 14 8 3 8 9 12 9 7 10 8 10 8 5 10 9 13 10
10 8 12 6 8 9 11 10 14 7 7 9 8 12 12 7 9 5 9 9 10 6 8 10 12 13 13 11 10 15 9 9 12 12 10 12 9 6 8 12 3 7 9
11 8 7 10 10 8 15 9 11 12 11 6 7 9 7 8 11 13 3 12 10 11 9 6 11 13 14 10 10 10 8 18 7 15 11 7 6 10 710 7 7 13
7 9 14 12 7 14 10 15 13 12 7 5 14 13 8 8 9 8 9 8 9 8 8 7 12 14 12 6 8 12 6 4 9 10 11 14 7 6 13 9 10 10 8 8
7 9 9 10 8 10 9 13 10 11 13 11 7 9 6 11 8 13 13 9 16 15 9 11 6 11 13 12 12 16 8 10 10 17 7 9 11 9 9 10 12 8 12
3 12 6 10 10 12 14 5 10 3 9 9 14 10 12 14 7 13 11 15 8 4 7 8 8 14 8 11 9 11 8 10 11 8 9 10 11 14 9 13 7 6 8
9 11 7 9 7 9 10 5 9 11 11 7 8 11 3 13 9 8 9 5 11 15 11 4 8 11 14 13 8 14 10 5 10 13 12 5 17 10 8 14 10 10 11
13 8 16 9 10 11 9 6 6 12 10 12 12 9 8 11 11 16 12 8 7 11 12 14 10 16 10 9 8 11 15 11 7 10 8 12 12 9 12 11 9 8
11 9 10 12 7 11 12 12 10 12 6 8 9 13 5 4 15 11 10 10 11 10 12 15 12 11 11 7 8 14 7 14 9 9 7 12 6 10 10 6 12 11
10 9 9 11 11 11 4 12 13 11 11 10 8 7 8 14 9 12 12 13 13 4 10 8 8 10 6 10 16 12 13 10 8 12 9 13 11 9 8 7 8 6
18 6 9 6 10 12 10 9 12 10 7 11 6 8 4 11 9 9 16 10 8 10 8 12 11 13 11 14 14 5 11 7 11 7 9 10 10 9 10 12 12 8 8
9 11 12 9 9 14 8 11 10 12 11 11 12 15 10 11 16 7 11 14 15 12 10 9 12 11 8 17 11 13 10 7 10 12 12 9 11 7 9 11 13
10 10 13 9 6 12 9 12 8 9 14 11 5 8 13 13 15 8 11 5 9 5 14 14 11 14 10 11 11 9 13 10 8 12 7 9 11 7 7 7 12 10
10 9 6 11 13 11 10 10 12 12 9 9 11 12 13 9 7 7 8 8 8 9 13 10 11 10 8 7 10 10 9 11 6 12 8 10 14 8 11 8 17 18
11 9 9 11 7 12 12 16 15 7 7 6 14 15 8 5 9 12 10 7 9 13 8 7 11 11 15 9 8 11 8 9 6 5 11 6 8 5 11 15 7 8 6 7
9 10 10 5 4 4 10 9 7 7 5 6 8 9 4 11 12 13 9 9 8 10 9 9 10 11 11 7 13 13 12 5 10 6 11 10 9 11 12 10 8 11 9 8
14 8 14 7 8 6 16 10 8 7 11 13 12 12 12 9 6 10 7 10 9 12 14 7 7 8 8 6 8 13 8 12 11 8 12 10 7 7 12 13 10 10 13
5 13 13 9 15 10 10 11 13 13 11 9 12 10 12 7 9 10 6 9 16 9 16 8 11 7 13 4 7 10 11 18 10 12 4 10 9 15 11 9 8 15
10 9 16 11 7 6 10 6 9 12 14 11 11 9 10 15 8 13 11 8 9 10 8 12 8 11 9 7 10 6 10 7 9 10 9 7 10 9 6 8 12 11 13
8 14 8 8 7 12 10 15 7 7 10 9 8 8 12 14 4 12 2 10 10 10 13 4 12 12 9 9 12 11 9 13 613 8 8 10 9 9 10 13 9 14
12 7 2 9 12 7 10 15 14 7 12 10 12 10 12 9 13 12 7 14 10 13 13 6 12 9 7 6 7 13 6 6 10 7 11 15 9 15 11 12 9 11
16 9 13 10 10 11 12 11 6 9 10 10 9 10 10 5 8 10 12 10 14 7 5 7 14 9 9 10 7 14 10 9 12 12 13 8 7 10 12 12 13 7
10 12 17 9 7 8 12 12 13 11 10 9 6 14 13 13 11 11

- $\overline{x} = \frac{1}{1000}\sum_{i=1}^{1000} x_i = 9.959$
- $s^2 = \frac{1}{999}\sum_{i=1}^{1000}(x_i - \overline{x})^2 = 7.749068$
- $p \approx 1 - s^2/\overline{x} \approx 0.22$
- $n \approx \frac{\overline{x}}{1-s^2/\overline{x}} \approx 44.88.$

Observations where actually drawn from $Binomial(50, 0.2)$. □

Conclusion:

- Sample mean and variance can be useful to estimate unknown parameters of the populations distribution.
- However, the arguments used so far are "ad-hoc" and we do not yet have a way to measure the accuracy of the estimation.
- More systematics methods are needed for parameter estimation in general.

MH1820 Week 10

1. Bias and Standard Error of an Estimator

2. Maximum Likelihood Estimator

3. Interval Estimator

Bias and Standard Error of an Estimator

Bias

how far is the estimate (on average) from the true value.

Let $\widehat{\theta}$ be an estimator of θ. The **bias** of $\widehat{\theta}$ is defined by

$$\text{Bias}(\widehat{\theta}) = \underline{\mathbb{E}[\widehat{\theta}] - \theta}.$$

Here, the expectation is computed under the population distribution parametrized by θ.

$\text{Bias}(\hat{\theta}) > 0 \Rightarrow$ overestimating.

$\text{Bias}(\hat{\theta}) < 0 \Rightarrow$ underestimating.

Intepretation: $\text{Bias}(\widehat{\theta})$ is the expected distance of $\widehat{\theta}$ from the true parameter θ.

A good estimator must have bias zero or at least its bias should tend to zero for increasing sample size. n.

$\widehat{\theta}$ is **unbiased** if $\text{Bias}(\widehat{\theta}) = 0$.

Example 1

- Population distribution: *Bernoulli*(p)
- Estimator: $\widehat{p} = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ (for p)

Find the bias of $\widehat{p}$.

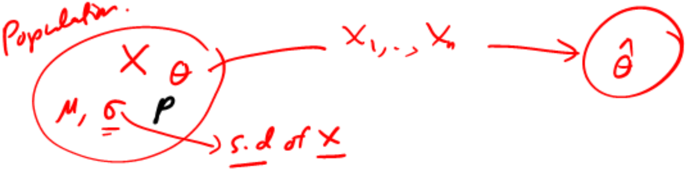
Solution.

Bernoulli (p)

X	0	1
$p(x)$	$1-p$	p

$$\mathbb{E}[X] = p$$

$$
\begin{aligned}
\mathrm{Bias}(\hat{p}) &= \mathbb{E}[\hat{p}] - p \\
&= \mathbb{E}\left[\frac{\sum_{i=1}^{n} X_i}{n}\right] - p \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] - p \\
&= \frac{1}{n}\sum_{i=1}^{n} p - p \\
&= \frac{1}{\cancel{n}}\cancel{n}p - p = 0.
\end{aligned}
$$



Standard error of $\widehat{\theta}$:

$$SE(\widehat{\theta}) = \sqrt{\mathrm{Var}[\widehat{\theta}]}.$$

Here the variance is computed under the population distribution parametrized by θ.

e.g $SE(\bar{x})$

$SE(s) = SE(\hat{\sigma})$ $\quad \hat{\sigma} = s = \sqrt{\frac{1}{n-1}\sum(x_i-\bar{x})^2}$

$SE(\hat{p})$ etc.

- $SE(\widehat{\theta})$ measures variability of our estimate, i.e. standard deviation of sampling distribution.
- Rule of thumb: For large samples, the true θ will be in the interval
  - $[\widehat{\theta} - SE(\widehat{\theta}), \widehat{\theta} + SE(\widehat{\theta})]$ in around 70% of the cases
  - $[\widehat{\theta} - 2SE(\widehat{\theta}), \widehat{\theta} + 2SE(\widehat{\theta})]$ in around 95% of the cases,

  if $\widehat{\theta}$ is used repeatedly to estimate

Compare this to $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

Example 2

Let $X_1, \ldots, X_n$ be i.i.d with population distribution $N(0, \sigma^2)$.

μ

Estimator for σ^2: $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

Find $SE(\hat{\sigma}^2)$. $= \sigma^2\sqrt{\frac{2}{n}}$. $\longrightarrow 0$ as $n \to \infty$.

$$SE(\hat{\sigma}^2) = \sqrt{Var(\hat{\sigma}^2)}$$

Solution. Note that

$$\frac{X_i}{\sigma} \sim N(0,1) \Longrightarrow \left(\frac{X_i}{\sigma}\right)^2 \sim \chi^2(1) = Gamma(\frac{1}{2}, 2).$$

Recall that if $X \sim Gamma(\alpha, \theta)$, then $\mathrm{Var}[X] = \alpha\theta^2$.

$$\mathrm{Var}\left(\frac{X_i}{\sigma}\right)^2 = \frac{1}{2} \cdot (2)^2 = 2.$$

$$\mathrm{Var}(\hat{\sigma}^2) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right)$$

$$= \mathrm{Var}\left(\boxed{\frac{\sigma^2}{n}}\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2\right)$$

$$= \left(\frac{\sigma^2}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n}\left(\frac{X_i}{\sigma}\right)^2\right)$$

Here : $\mathrm{Var}(aX) = a^2\,\mathrm{Var}(X)$

$$\mathrm{Var}(\hat{\sigma}^2) = \left(\frac{\sigma^2}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}\left[\left(\frac{X_i}{\sigma}\right)^2\right]$$

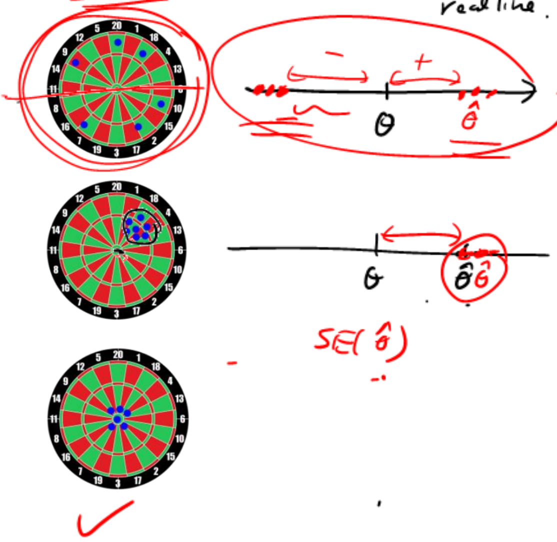
$$= \left(\frac{\sigma^2}{n}\right)^2 \sum_{i=1}^{n} 2 \qquad (\text{by } \star).$$

$$= \frac{\sigma^4}{n^2} \cdot 2n$$

$$= \frac{2\sigma^4}{n}. \implies SE(\hat{\sigma}^2) = \sqrt{\frac{2\sigma^4}{n}} = \sigma^2\sqrt{\frac{2}{n}}.$$

$Bias(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta$

- Point estimator on average should be close to the true parameter $\Longrightarrow$ bias must be small.
- The values of the estimator should not be spread out too far $\Longrightarrow$ standard error must be small.
- Ideal situation: both *Bias* and *SE* as small as possible.



Maximum Likelihood Estimator

Suppose we observed the following data (sample size 10) drawn from *Bernoulli*(p):

$$x_1 = x_2 = x_3 = 1, \; x_4 = \cdots = x_{10} = 0.$$

- It seems that $p = 0.3$ is quite likely. But we could not decide since the data could have come from $p = 0.5$ or $p = 0.9$, though $p = 0.3$ seems much more plausible.
- Is there a way to pick the "most probable" p?
- Problem: There is no "most probable" value of p (since p is not a random variable!)

Main idea of maximum likelihood:
Reverse the approach: Find a value of p that makes the observations most likely, i.e. maximizing the probability of observing the data!

Find $\hat{p}$ that maximizes $\mathbb{P}(x_1, \ldots x_n \mid p)$.

In general:

Find $\hat{\theta}$ " " $\mathbb{P}(x_1, \ldots, x_n \mid \theta)$.

Set-up for maximum likelihood:

- Let $X_1, \ldots, X_n$ be i.d. with PMF or PDF $f(x|\theta)$, depending on an unknown parameter θ.
- Observations $x_1, \ldots, x_n$ are given.
- The idea of the maximum likelihood method is to choose the value for θ as estimator which maximizes the following **maximum likelihood function**:

$$L(X_1 = x_1, \ldots, X_n = x_n|\theta) \quad = \quad \underline{\underline{\prod_{i=1}^{n} f(x_i|\theta)}}$$

Interpretation. The maximum likelihood function $L(x_1,\ldots,x_n|\theta)$ is the probability of observing the data assuming θ is the real value. E.g. in the discrete case, we have

$$\begin{aligned}\mathbb{P}(X_1=x_1,\ldots,X_n=x_n|\theta) &= \prod_{i=1}^{n}\mathbb{P}(X_i=x_i|\theta) \quad \text{(since } X_i \text{ are independent)}\\ &= \prod_{i=1}^{n} f(x_i|\theta) \quad \text{(since } X_i\text{'s are identical)}\\ &= L(x_1,\ldots,x_n|\theta).\end{aligned}$$

$$\ln\left(\prod_{i=1}^{n} f(x_i|\theta)\right) = \sum_{i=1}^{n} \ln f(x_i|\theta)$$

Finding Maximum Likelihood Estimator (MLE):

The maximum likelihood estimator i.e. the value of θ that maximizes $L(x_1, \ldots, x_n|\theta)$ can be found by solving

- $\dfrac{d}{dp}(L) = 0$ or
- $\dfrac{d}{dp}(\ln L) = 0$

L is $\underline{\underline{\text{max}}} \Leftrightarrow \ln L$ is $\underline{\underline{\text{max}}}$

Both solution methods are valid, but sometimes the second method often is faster. There are likelihood functions for which the maximizer cannot be found in this way, but such cases will not occur in this course.

Example 3

- $X_1, \ldots, X_{10}$ i.i.d. $\sim$ *Bernoulli*(p), $0 < p < 1$.
- Observations: $x_1 = x_2 = x_3 = 1$, $x_4 = \cdots = x_{10} = 0$

Find p that maximizes the likelihood function.

$$f(x_i \mid p) = \begin{cases} p & \text{if } x_i = 1 \\ 1-p & \text{if } x_i = 0 \end{cases}$$

x	0	1
$f(x\mid p)$	$1-p$	p.

$$L(x_1, \dots, x_{10} | p) = \prod_{i=1}^{n} f(x_i | p)$$

$$= \overbrace{p \cdot p \cdot p}^{3} \overbrace{(1-p) \dots (1-p)}^{7}$$

$$= p^3 (1-p)^7$$

$$f(\underline{\underline{x_1}}|p) f(x_2|p) f(x_3|p) f(x_4|p) \dots f(x_{10}|p)$$

$$= f(1|p) f(1|p) f(1|p) f(0|p) \dots f(0|p)$$

$$= \underline{p \cdot p \cdot p} \cdot \underline{(1-p) \dots (1-p)}$$

Solution. Recall that the PMF for *Bernoulli*(p) is $f(x|p) = p^x(1-p)^{1-x}$, $x = 0, 1$. The likelihood function is

$$L = L(x_1, \ldots, x_{10}|p) = \prod_{i=1}^{10} f(x_i|p) = f(1|p)^3 f(0|p)^7 = p^3(1-p)^7.$$

To find the maximum of L, we set the derivative (with respect to p) to zero:

$$\begin{aligned} \frac{dL}{dp} &= 0 \\ 3p^2(1-p)^7 - 7p^3(1-p)^6 &= 0 \\ p^2(1-p)^6(3(1-p)-7p) &= 0 \end{aligned}$$

This implies that $p = 0$, or $p = 1$, or $p = \frac{3}{10}$.

Thus, $p = \frac{3}{10}$ is the maximizer, i.e. the maximum likelihood estimator for p is $\frac{3}{10}$. $\square$

trials until the first success

Example 4

- $X_1, \ldots, X_n$ i.i.d $\sim Geom(p)$, $0 < p < 1$.
- PMF: $f(x) = (1-p)^{x-1} p$, $x = 1, 2, \ldots$.
- Given the observation $x_1 = 2$ $(n = 1)$, what is the *MLE* for p?

$$f(x_i | p) = (1-p)^{x_i - 1} p.$$

$$L(x_1, .., x_n | p) = \quad L(\underset{2}{\underset{\shortparallel}{x_1}} | p) = (1-p)^{x_1 - 1} p = (1-p)^{2-1} p = (1-p)p.$$

$$L = (1-p)p$$
$$= p - p^2$$

$$\frac{dL}{dp} = 0$$
$$1 - 2p = 0$$
$$p = \frac{1}{2}.$$

waiting time
until first arrival (Poisson(λ))
$\theta = \frac{1}{\lambda}$

Example 5

- $X_1, \ldots, X_n$ i.i.d $\sim Exp(\theta)$, $\theta > 0$.
- PDF: $f(x|\theta) = \frac{1}{\theta}e^{-x/\theta}$ for $x > 0$ and $f(x|\theta) = 0$ otherwise.
- Find MLE for θ based on the observations $\underline{1, 2, 5, 1, 1}$ ($n = 5$).

① Compute L or $\ln L$.

$$L = \prod_{i=1}^{n} f(x_i|\theta)$$

$$= \prod_{i=1}^{5}\left(\frac{1}{\theta}e^{-x_i/\theta}\right) = \frac{1}{\theta^5}e^{-\sum_{i=1}^{5}\frac{x_i}{\theta}}$$

$$= \frac{1}{\theta^5}e^{-\frac{1}{\theta}\sum_{i=1}^{5}x_i}$$

$$= \frac{1}{\theta^5}e^{-\frac{1}{\theta}(1+2+5+1+1)}.$$

$$= \frac{1}{\theta^5}e^{-10/\theta}$$

Switch to $\ln L$

$$\ln L = \ln\left(\frac{1}{\theta^5} e^{-10/\theta}\right)$$

$$= \ln\left(\frac{1}{\theta^5}\right) + \ln e^{-10/\theta}$$

$$= \ln \theta^{-5} + \left(-\frac{10}{\theta}\right)$$

$$= -5 \ln\theta - \frac{10}{\theta}$$

(2) Differentiate :

$$\frac{d \ln L}{d\theta} = -\frac{5}{\theta} + \frac{10}{\theta^2} = 0$$

$$-5\theta + 10 = 0$$

$$\theta = 2 .$$

Solution. The maximum likelihood function is

$$L(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} f(x_i|\theta) = \prod_{i=1}^{n} \underbrace{\left(\frac{1}{\theta} e^{-x_i/\theta}\right)} = \frac{1}{\theta^n} e^{-\frac{\sum_{i=1}^{n} x_i}{\theta}}.$$

$$\frac{1}{\theta} e^{-x_1/\theta} \cdot \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}$$

From the obeservations, we have $n = 5$ and $\sum_{i=1}^{5} x_i = 1 + 2 + 5 + 1 + 1 = 10$, and so

$$L = \theta^{-5} e^{-10/\theta} \implies \ln L = -5 \ln \theta - 10/\theta.$$

$$\boxed{e^a e^b = e^{a+b}}$$

Differentiating $\ln L$ with respect to θ and set it to 0, we get

$$\begin{aligned}
\frac{d}{d\theta}(\ln L) &= 0 \\
-\frac{5}{\theta} + \frac{10}{\theta^2} &= 0 \\
\theta &= 2.
\end{aligned}$$

The MLE for θ is 2. $\square$