INTRODUCTION TO STATISTICS

LECTURE 2

LAST TIME

- DESCRIPTIVE STATISTICS
  - Data about **the whole population** is available.
  - Summarize it with
    - *summary statistics;*
    - *tables;*
    - *plots.*

LAST TIME

- DESCRIPTIVE STATISTICS
  - Data about **the whole population** is available.
  - Summarize it with
    - *summary statistics;*
    - *tables;*
    - *plots.*

- INFERENTIAL STATISTICS
  - Data about ***a sample* of the whole population** is available.
  - Reason about the whole population.

LAST TIME

- DESCRIPTIVE STATISTICS
  - Data about **the whole population** is available.
  - Summarize it with
    - *summary statistics;*
    - *tables;*
    - *plots.*

- INFERENTIAL STATISTICS
  - Data about ***a sample* of the whole population** is available.
  - Reason about the whole population.

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p**	**1-p**

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p**	**1-p**

- **How many men and women are there in the world?**

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p**	**1-p**

- **How many men and women are there in the world?**

- There is precise data! ☺

https://knoema.com/gndibag/sex-ratio-around-the-world-men-per-100-women-1950-2100

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p = 0.5**	**1-p = 0.5**

- **How many men and women are there in the world?**

- There is precise data! ☺

https://knoema.com/gndibag/sex-ratio-around-the-world-men-per-100-women-1950-2100

p = 0.5

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p = 0.5**	**1-p = 0.5**

- *If we take 100 random people $X_1 \ldots X_{100}$, how many men do we expect to see?*

- **How many men and women are there in the world?**
- There is precise data! ☺

World Male Population
3,720,696
thousands

World Female Population
3,659,101
thousands

https://knoema.com/gndibag/sex-ratio-around-the-world-men-per-100-women-1950-2100

p = 0.5

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p** = **0.5**	**1-p** = **0.5**

- *If we take 100 random people $X_1 \ldots X_{100}$, how many men do we expect to see?*

$$N_{men} = X_1 + \ldots + X_{100} \sim \mathrm{Bi}(100, \mathbf{p})$$

- **How many men and women are there in the world?**
- There is precise data! ☺

https://knoema.com/gndibag/sex-ratio-around-the-world-men-per-100-women-1950-2100

$$\mathbf{p} = \mathbf{0.5}$$

MEN AND WOMEN

- *If we take a random person X, what's the probability that it's a man?*

X	1	0
P(X)	**p = 0.5**	**1-p = 0.5**

- *If we take 100 random people $X_1 \ldots X_{100}$, how many men do we expect to see?*

$$N_{men} = X_1 + \ldots + X_{100} \sim Bi(100, \mathbf{p})$$
$$E(N_{men}) = 100 * \mathbf{p} = 50$$

- **How many men and women are there in the world?**

- There is precise data! ☺

World Male Population	World Female Population
3,720,696 thousands	3,659,101 thousands

https://knoema.com/gndibag/sex-ratio-around-the-world-men-per-100-women-1950-2100

p = 0.5

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

$$X \sim \text{Bernoulli}(\mathbf{p})$$
$$P(X = 1) = \mathbf{p} = ?$$

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

  $X \sim \text{Bernoulli}(\mathbf{p})$
  $P(X = 1) = \mathbf{p} = ?$

- **How many vegetarians are there in the world?**

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

$X \sim \text{Bernoulli}(\mathbf{p})$
$P(X = 1) = \mathbf{p} = ?$

- **How many vegetarians are there in the world?**
- No precise data! ☹

What would you do?

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

$$X \sim \text{Bernoulli}(\mathbf{p})$$
$$P(X = 1) = \mathbf{p} = ?$$

- **How many vegetarians are there in the world?**
- No precise data! ☹

What would you do?

- Survey:
  - 100 people: $X_1 \ldots X_{100}$
  - 13 are vegetarians

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

  X ~ Bernoulli(**p**)
  P(X = 1) = **p** = ?

- *What is a good guess for the value of* **p**?

- **How many vegetarians are there in the world?**
- No precise data! ☹

  What would you do?

- Survey:
  - 100 people: $X_1 \ldots X_{100}$
  - 13 are vegetarians

(NON-)VEGETARIANS

- *If we take a random person X, what's the probability that it's a vegetarian?*

$$X \sim \text{Bernoulli}(\mathbf{p})$$
$$P(X = 1) = \mathbf{p} = ?$$

- *What is a good guess for the value of* **p***?*

$$\hat{p} = 13 / 100 = 0.13$$

- **How many vegetarians are there in the world?**
- No precise data! ☹

What would you do?

- Survey:
  - 100 people: $X_1 \ldots X_{100}$
  - 13 are vegetarians

(NON-)VEGETARIANS

INFERENTIAL STATISTICS

- *If we take a random person X, what's the probability that it's a vegetarian?*

$X \sim \text{Bernoulli}(\mathbf{p})$
$P(X = 1) = \mathbf{p} = ?$

- *What is a good guess for the value of* **p***?*

$\hat{\mathbf{p}} = 13 / 100 = 0.13$

...get the parameter estimate

- **How many vegetarians are there in the world?**
- No precise data! ☹

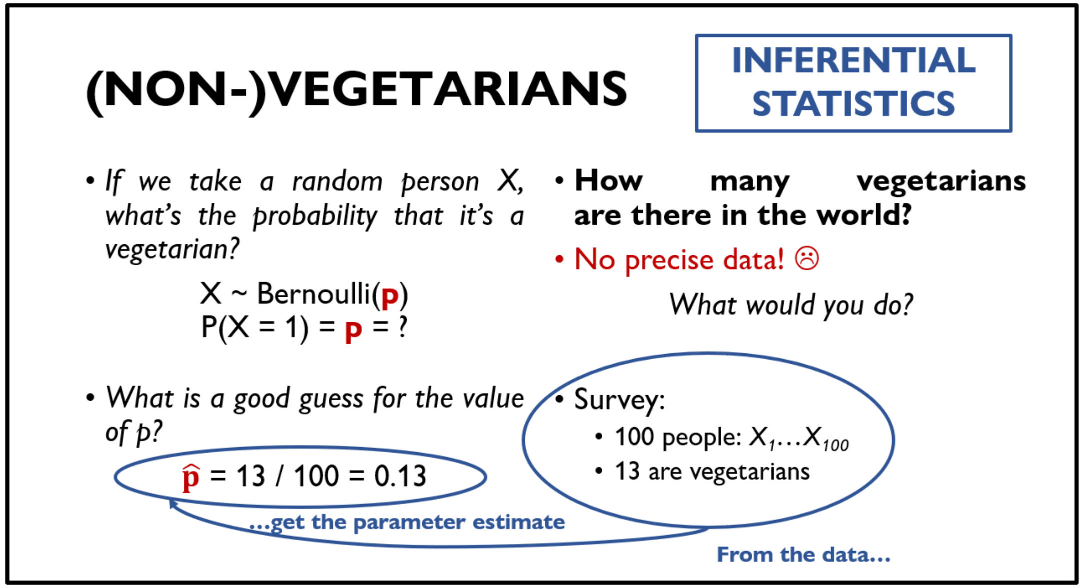
What would you do?

- Survey:
  - 100 people: $X_1 \ldots X_{100}$
  - 13 are vegetarians

From the data...

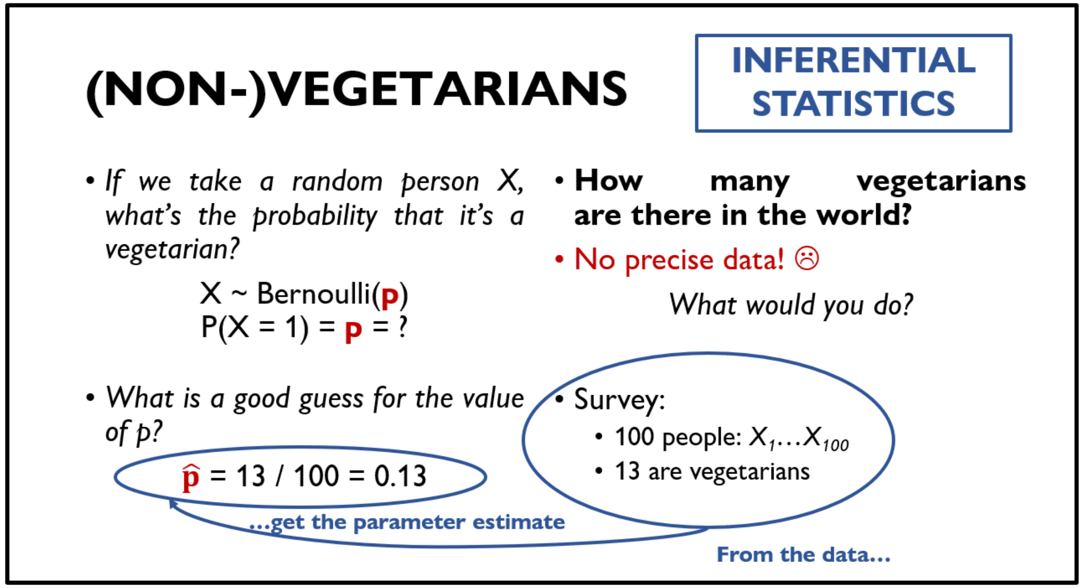
SOME QUESTIONS TO ASK OURSELVES

- Was it a good way to obtain an estimate?



SOME QUESTIONS TO ASK OURSELVES

- Was it a good way to obtain an estimate?
- How precise is our estimate?

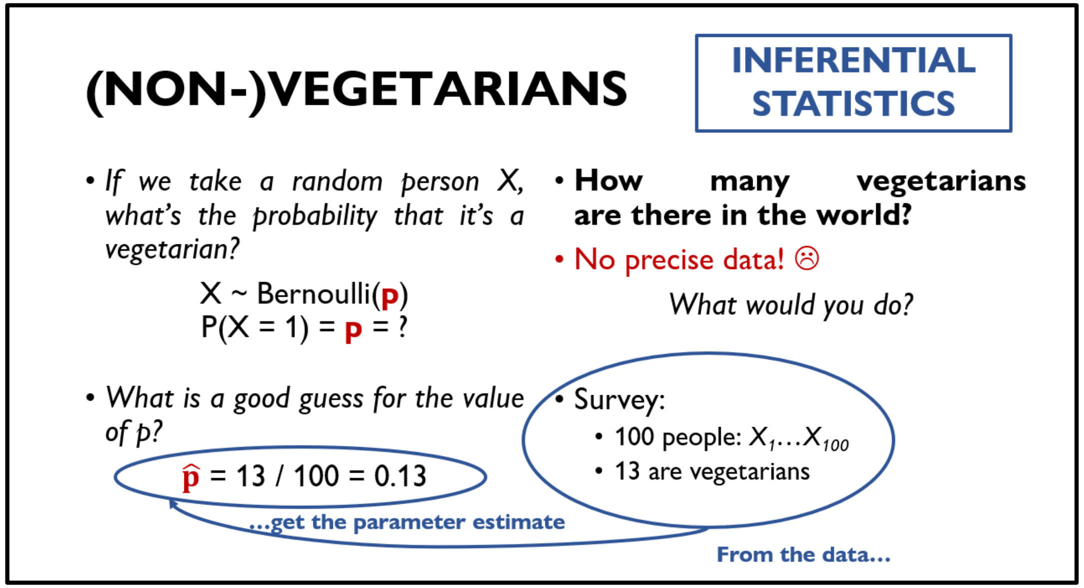


SOME QUESTIONS TO ASK OURSELVES

- Was it a good way to obtain an estimate?

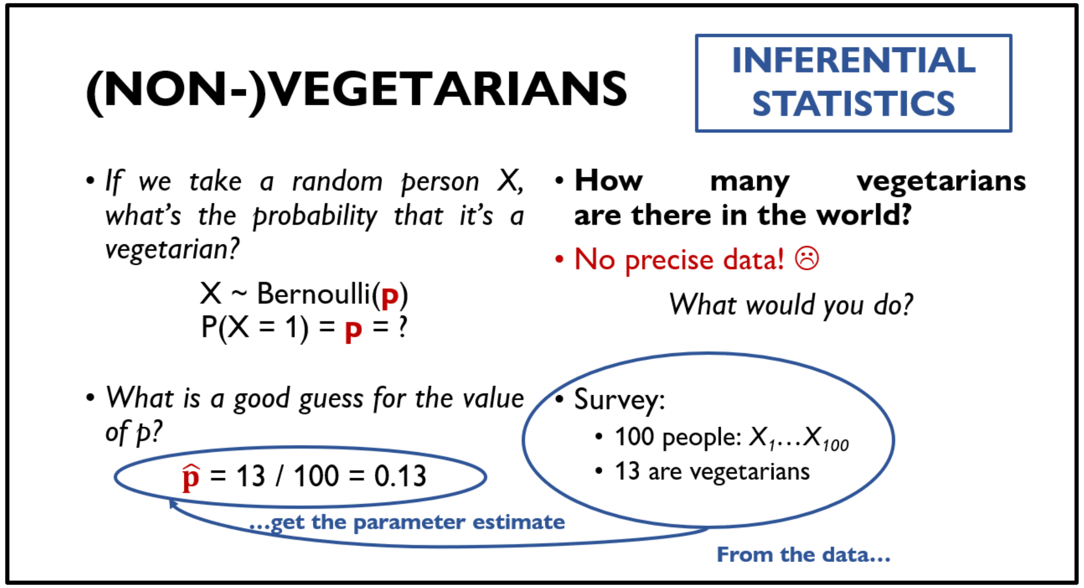
- How precise is our estimate?

- Would it be better if there were more people in the survey?



SOME QUESTIONS TO ASK OURSELVES

- **Was it a good way to obtain an estimate?**
- How precise is our estimate?
- Would it be better if there were more people in the survey?



THE METHOD OF MAXIMUM LIKELIHOOD

A general framework for parameter estimation

THE METHOD OF MAXIMUM LIKELIHOOD

A general framework for parameter estimation

LIKELIHOOD FUNCTION

- X ~ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p*** **-** the true proportion of the vegetarians.

LIKELIHOOD FUNCTION

- $X \sim \text{Bernoulli}(p)$ – whether a person is a vegetarian (1) or not (0).
- ***p* -** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

LIKELIHOOD FUNCTION

- X ~ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p*** **-** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given ***p***?

LIKELIHOOD FUNCTION

- X ~ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p* -** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given ***p***?

$$P(X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0 \mid p) =$$

LIKELIHOOD FUNCTION

- X ~ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p*** **-** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given ***p***?

$$P(X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0 \mid p) =$$

{$X_1, \ldots, X_{100}$ are independent}

LIKELIHOOD FUNCTION

- X ~ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p*** **-** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given ***p***?

$$P(X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0 \mid p) =$$

$\{X_1, \ldots, X_{100}$ are independent$\}$

$$= P(X_1 = 1 \mid p) * \ldots * P(X_{13} = 1 \mid p) * P(X_{14} = 0 \mid p) * \ldots * P(X_{100} = 0 \mid p) =$$

LIKELIHOOD FUNCTION

- $X \sim$ Bernoulli(***p***) – whether a person is a vegetarian (1) or not (0).
- ***p*** **-** the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given ***p***?

$$P(X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0 \mid p) =$$

$\{X_1, \ldots, X_{100}$ are independent$\}$

$$= P(X_1 = 1 \mid p) * \ldots * P(X_{13} = 1 \mid p) * P(X_{14} = 0 \mid p) * \ldots * P(X_{100} = 0 \mid p) =$$

$$= p^{13}(1-p)^{87}$$

LIKELIHOOD FUNCTION

- $X \sim \text{Bernoulli}(p)$ – whether a person is a vegetarian (1) or not (0).
- p - the true proportion of the vegetarians.
- Survey results: 13 out of 100 participants are vegetarians:

$$X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0$$

- What is the probability to obtain such results given p? **Likelihood:**

$$L(p) = P(X_1 = 1, \ldots, X_{13} = 1, X_{14} = 0, \ldots, X_{100} = 0 \mid p) =$$

$\{X_1, \ldots, X_{100}$ are independent$\}$

$$= P(X_1 = 1 \mid p) * \ldots * P(X_{13} = 1 \mid p) * P(X_{14} = 0 \mid p) * \ldots * P(X_{100} = 0 \mid p) =$$

$$= p^{13}(1-p)^{87}$$

Likelihood function is the joint probability of realized sample given the parameters.

MAXIMUM LIKELIHOOD

- The probability to observe such data given p is:

$$L(p) = p^{13}(1-p)^{87}$$

MAXIMUM LIKELIHOOD

- The probability to observe such data given *p* is:

$$L\,(p) = p^{13}(1-p)^{87}$$

- How to chose the best p?

MAXIMUM LIKELIHOOD

- The probability to observe such data given *p* is:

$$L\,(p) = p^{13}(1\text{-}p)^{87}$$

- How to chose the best p?
  - We want the data that we obtained to be the most likely one.

MAXIMUM LIKELIHOOD

- The probability to observe such data given *p* is:

$$L\ (p) = p^{13}(1-p)^{87}$$

- How to chose the best p?
  - We want the data that we obtained to be the most likely one.
  - **Maximize L(p) w.r.t. p!**

CALCULUS 101

- **How to optimize a function?**

CALCULUS 101

- **How to optimize a function?**

1. Compute its derivative.
2. Set it to zero.
3. Get the critical point(s).
4. Chose the point of maximum.

CALCULUS 101

- **How to optimize a function?**

$$f(x) = \log x + \log(1 - x)$$

1. Compute its derivative.
2. Set it to zero.
3. Get the critical point(s).
4. Chose the point of maximum.

CALCULUS 101

- **How to optimize a function?**

$$f(x) = \log x + \log(1 - x)$$

1. Compute its derivative.
2. Set it to zero.

$$\frac{d}{dx} f(x) = \frac{1}{x} - \frac{1}{1 - x} = 0$$

3. Get the critical point(s).

4. Chose the point of maximum.

CALCULUS 101

- **How to optimize a function?**

1. Compute its derivative.
2. Set it to zero.
3. Get the critical point(s).
4. Chose the point of maximum.

$$f(x) = \log x + \log(1 - x)$$

$$\frac{d}{dx} f(x) = \frac{1}{x} - \frac{1}{1 - x} = 0$$

Critical points:

$$x = 0, \qquad \mathrm{x} = 1, \qquad \mathrm{x} = \frac{1}{2}$$

CALCULUS 101

- **How to optimize a function?**

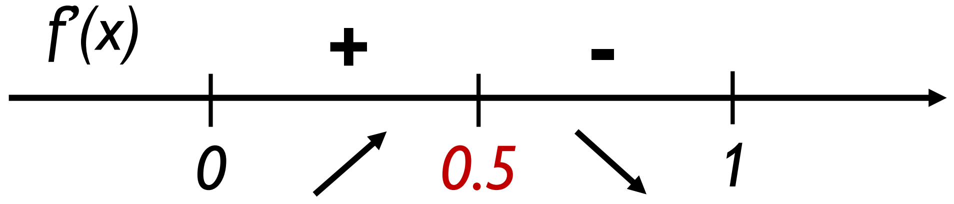
1. Compute its derivative.
2. Set it to zero.
3. Get the critical point(s).
4. Chose the point of maximum.

$$f(x) = \log x + \log(1 - x)$$

$$\frac{d}{dx} f(x) = \frac{1}{x} - \frac{1}{1 - x} = 0$$

Critical points:

$$x = 0, \qquad \mathrm{x} = 1, \qquad \mathrm{x} = \frac{1}{2}$$



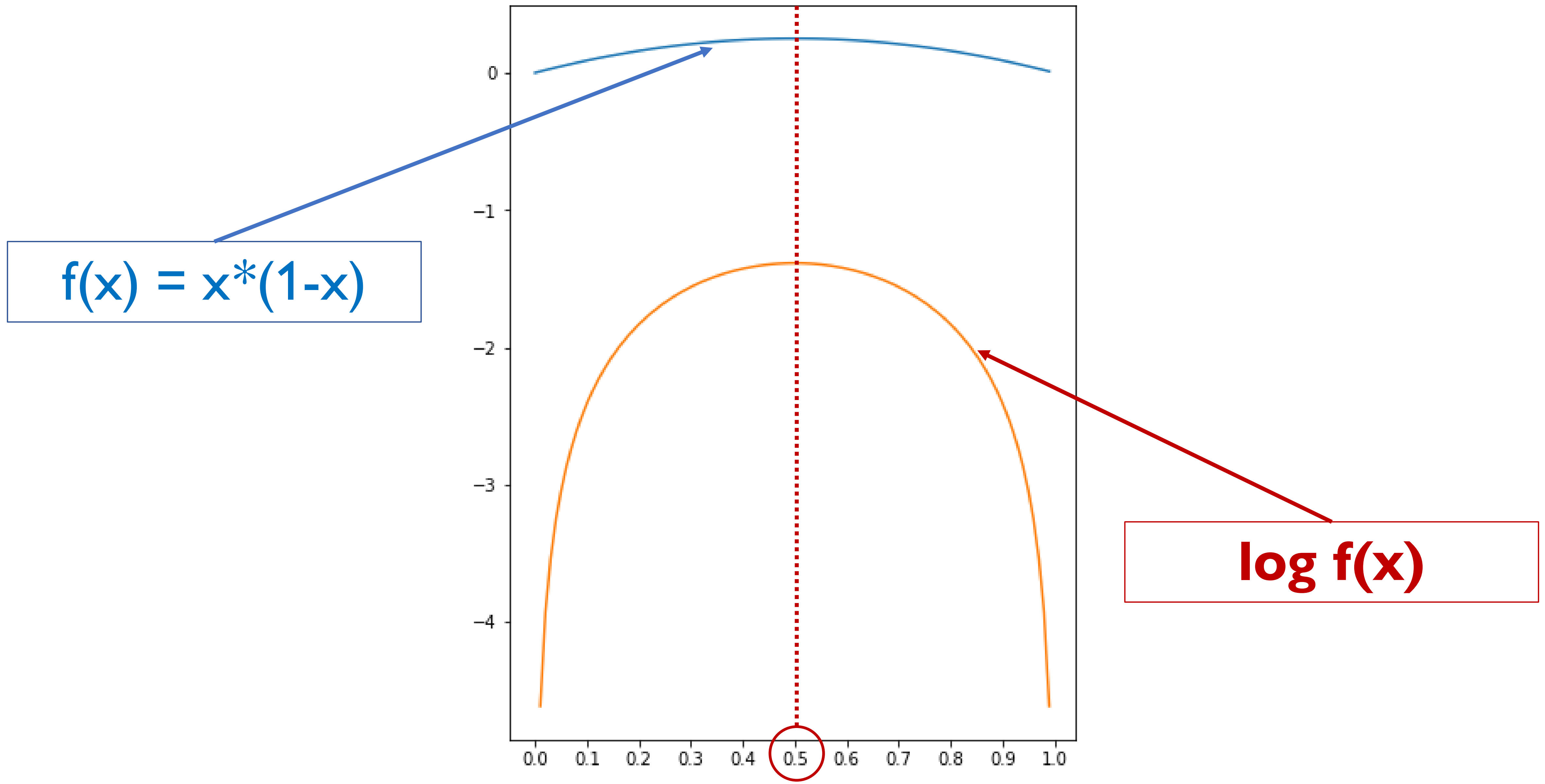
MAXIMUM LIKELIHOOD

$$\textit{maximize } L(p) = p^{13}(1-p)^{87} \qquad \text{w.r.t. } p$$

MAXIMUM LIKELIHOOD

$$\textit{maximize } L(p) = p^{13}(1-p)^{87} \qquad \text{w.r.t. } p$$

USEFUL TRICK: MAXIMIZE LOG-LIKELIHOOD INSTEAD



f(x) = x*(1-x)
log f(x)
0
−1
−2
−3
−4
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

MAXIMUM LIKELIHOOD

$$\textit{maximize } L(p) = p^{13}(1-p)^{87} \qquad \text{w.r.t. } p$$

USEFUL TRICK: MAXIMIZE LOG-LIKELIHOOD INSTEAD

$$\textit{maximize } \log L(p) = 13 \log(p) + 87 \log(1-p) \quad \text{w.r.t. } p$$

MAXIMUM LIKELIHOOD

maximize $L(p) = p^{13}(1-p)^{87}$ w.r.t. p

USEFUL TRICK: MAXIMIZE LOG-LIKELIHOOD INSTEAD

maximize $\log L(p) = 13 \log(p) + 87 \log(1-p)$ w.r.t. p

$$\frac{d}{dp} \log L(p) =$$

MAXIMUM LIKELIHOOD

maximize $L(p) = p^{13}(1-p)^{87}$ w.r.t. p

USEFUL TRICK: MAXIMIZE LOG-LIKELIHOOD INSTEAD

maximize $\log L(p) = 13 \log(p) + 87 \log(1-p)$ w.r.t. p

$$\frac{d}{dp}\log L(p) = \frac{13}{p} - \frac{87}{1-p} = 0$$

MAXIMUM LIKELIHOOD

maximize $L(p) = p^{13}(1-p)^{87}$ w.r.t. p

USEFUL TRICK: MAXIMIZE LOG-LIKELIHOOD INSTEAD

maximize $\log L(p) = 13 \log(p) + 87 \log(1-p)$ w.r.t. p

$$\frac{\mathrm{d}}{\mathrm{dp}} \log \mathrm{L(p)} = \frac{13}{\mathrm{p}} - \frac{87}{1-\mathrm{p}} = 0$$

$$\hat{p} = \frac{13}{100}$$

MAXIMUM LIKELIHOOD ESTIMATE

1. Write down the likelihood function:

$$L(\theta) = P(X_1, \ldots, X_n \mid \theta) = \prod_{i=1}^{n} P(Xi \mid \theta)$$

2. Find its maximum w.r.t. the unknown parameter θ:

$$\widehat{\Theta} = \text{argmax } L(\theta) \quad \text{w.r.t. } \theta$$

*(!) In many cases, it's easier to maximize **log-likelihood**:*

$$\log L(\theta) = \log \prod_{i=1}^{n} P(Xi \mid \theta) = \sum_{i=1}^{n} \log P(Xi \mid \theta)$$

$$\widehat{\Theta} = \text{argmax } \log L(\theta)$$

Maximum Likelihood Estimate (MLE) is the value which …

Maximum Likelihood Estimate (MLE) is the value which maximizes the probability of observing the realized sample.

BREAK

MLE FOR SOME DISCRETE DISTRIBUTIONS

BERNOULLI DISTRIBUTION

X ~ Bernoulli(p)

E(X) = p

Var(X) = p(p-1)

- Models the probability of success in an experiment with two outcomes.

BERNOULLI DISTRIBUTION

X ~ Bernoulli(p)

E(X) = p

Var(X) = p(p-1)

- Models the probability of success in an experiment with two outcomes.

- *(General case of the previous example)*
- You want to estimate the proportion of the vegetarians.

BERNOULLI DISTRIBUTION

$X \sim \text{Bernoulli}(p)$

$E(X) = p$

$Var(X) = p(p-1)$

- Models the probability of success in an experiment with two outcomes.

- *(General case of the previous example)*
- You want to estimate the proportion of the vegetarians.
- Survey: you ask N people whether they are vegetarians and get responses

$$X_1, X_2, \ldots X_N$$

BERNOULLI DISTRIBUTION

$X \sim \text{Bernoulli}(p)$

$E(X) = p$

$Var(X) = p(p-1)$

- Models the probability of success in an experiment with two outcomes.

- *(General case of the previous example)*
- You want to estimate the proportion of the vegetarians.
- Survey: you ask N people whether they are vegetarians and get responses

$$X_1, X_2, \dots X_N$$

- What's the MLE of p?

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) =$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1 - X_i) \log(1-p)$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1-X_i)\log(1-p)$$

$$\frac{d}{dp}\log L(p) =$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1 - X_i) \log(1-p)$$

$$\frac{d}{dp} \log L(p) = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{N - \sum_{i=1}^{N} X_i}{1-p} = 0$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1 - X_i)\log(1-p)$$

$$\frac{d}{dp}\log L(p) = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{N - \sum_{i=1}^{N} X_i}{1-p} = 0$$

$$\hat{p} =$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1 - X_i)\log(1-p)$$

$$\frac{d}{dp}\log L(p) = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{N - \sum_{i=1}^{N} X_i}{1-p} = 0$$

$$\hat{p} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

MLE FOR BERNOULLI DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} X_i \log p + (1 - X_i) \log(1-p)$$

$$\frac{d}{dp} \log L(p) = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{N - \sum_{i=1}^{N} X_i}{1-p} = 0$$

$$\boxed{\hat{p} = \frac{1}{N} \sum_{i=1}^{N} X_i}$$

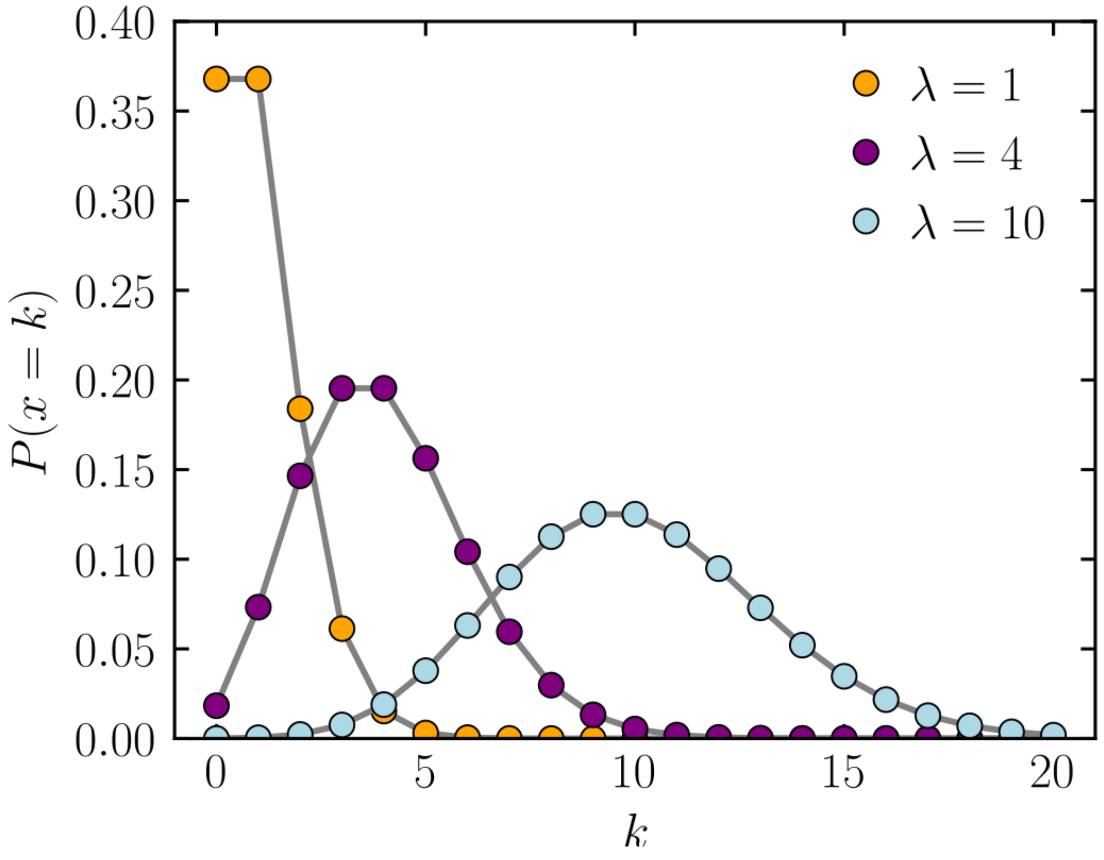
POISSON DISTRIBUTION

$X \sim Po(\lambda), \quad \lambda > 0$

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$E(X) = \lambda, \quad Var(X) = \lambda$

- Models number of events occurring in a fixed interval of time.
- Assumptions: events occur
  - with a known constant mean rate;
  - independently of the time since the last event.



POISSON DISTRIBUTION

$X \sim Po(\lambda), \quad \lambda > 0$

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$E(X) = \lambda, \quad Var(X) = \lambda$

- Models number of events occurring in a fixed interval of time.
- Assumptions: events occur
  - with a known constant mean rate;
  - independently of the time since the last event.

- Suppose you want to model the number of phone calls a call center receives in an hour.

POISSON DISTRIBUTION

$X \sim Po(\lambda), \quad \lambda > 0$

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$E(X) = \lambda, \quad Var(X) = \lambda$

- Models number of events occurring in a fixed interval of time.
- Assumptions: events occur
  - with a known constant mean rate;
  - independently of the time since the last event.

- Suppose you want to model the number of phone calls a call center receives in an hour.
- You've been recording it for the past *n* hours:

X1, X2, …, Xn

POISSON DISTRIBUTION

$X \sim Po(\lambda), \quad \lambda > 0$

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$E(X) = \lambda, \quad Var(X) = \lambda$

- Models number of events occurring in a fixed interval of time.
- Assumptions: events occur
  - with a known constant mean rate;
  - independently of the time since the last event.

- Suppose you want to model the number of phone calls a call center receives in an hour.
- You've been recording it for the past *n* hours:

X1, X2, …, Xn

- What's the MLE of λ?

MLE FOR POISSON DISTRIBUTION

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$$L(\lambda) = P(X_1, \dots, X_n | \lambda) =$$

MLE FOR POISSON DISTRIBUTION

$$P(X = k) = \begin{cases} \frac{e^{-\lambda} \cdot \lambda^k}{k!}, k \geq 0 \\ 0, \text{otherwise} \end{cases}$$

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n | \lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda} \lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) =$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i=1}}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i=1}}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X_i} \log \lambda - \log \mathrm{X_i}!]$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i}=1}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X_i} \log \lambda - \log \mathrm{X_i}!]$$

$$= -\mathrm{n}\lambda + \log \lambda \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \log \mathrm{X_i}!$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i}=1}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X_i} \log \lambda - \log \mathrm{X_i}!]$$

$$= -\mathrm{n}\lambda + \log \lambda \textstyle\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \log \mathrm{X_i}!$$

$$\frac{d}{d\lambda} L(\lambda) =$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i}=1}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X_i} \log \lambda - \log \mathrm{X_i}!]$$

$$= -\mathrm{n}\lambda + \log \lambda \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \log \mathrm{X_i}!$$

$$\frac{d}{d\lambda} L(\lambda) = -n + \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}}{\lambda} = 0,$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \dots, X_n|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i}=1}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X_i} \log \lambda - \log \mathrm{X_i}!]$$

$$= -\mathrm{n}\lambda + \log \lambda \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \log \mathrm{X_i}!$$

$$\frac{d}{d\lambda} L(\lambda) = -n + \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}}{\lambda} = 0, \qquad \hat{\lambda} = \frac{1}{n} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}$$

MLE FOR POISSON DISTRIBUTION

$$\mathrm{L}(\lambda) = P(X_1, \ldots, X_2|\lambda) = \prod_{\mathrm{i}=1}^{\mathrm{n}} \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{X_i}}}{\mathrm{X_i}!}$$

$$\log \mathrm{L}(\lambda) = \sum_{\mathrm{i}=1}^{\mathrm{n}} [-\lambda \log \mathrm{e} + \mathrm{X}_1 \log \lambda - \log \mathrm{X_i}!]$$

$$= -\mathrm{n}\lambda + \log \lambda \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i} - \sum_{\mathrm{i}=1}^{\mathrm{n}} \log \mathrm{X_i}!$$

$$\frac{d}{d\lambda} L(\lambda) = -n + \frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}}{\lambda} = 0,$$

$$\boxed{\hat{\lambda} = \frac{1}{n} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{X_i}}$$

MLE: YET ANOTHER EXAMPLE

- Random variable X takes three values with unknown probabilities:

X	1	2	3
P(X)	?	?	?

MLE: YET ANOTHER EXAMPLE

- Random variable X takes three values with unknown probabilities:

X	1	2	3
P(X)	?	?	?

- How many parameters does such a distribution have?

MLE: YET ANOTHER EXAMPLE

- Random variable X takes three values with unknown probabilities:

X	1	2	3
P(X)	p	q	1-p-q

- How many parameters does such a distribution have?

MLE: YET ANOTHER EXAMPLE

- Random variable X takes three values with unknown probabilities:

X	1	2	3
P(X)	p	q	1-p-q

- How many parameters does such a distribution have?

- We observed it N times:

X	1	2	3
#	n_1	n_2	n_3

MLE: YET ANOTHER EXAMPLE

- Random variable X takes three values with unknown probabilities:

X	1	2	3
P(X)	p	q	1-p-q

- How many parameters does such a distribution have?

- We observed it N times:

X	1	2	3
#	n_1	n_2	n_3

- What is the MLE of the parameters?

CALCULUS 101

- **How to optimize a function of several variables?**

CALCULUS 101

- **How to optimize a function of several variables?**

1. Compute partial derivatives.
2. Set them to zero.

3. Solve the equations and get the critical points.

CALCULUS 101

- **How to optimize a function of several variables?**

$$f(x,y) = x^2 + 2xy - 2x - 4y$$

1. Compute partial derivatives.
2. Set them to zero.

$$\frac{d}{dx} f(x,y) = 2x + 2y - 2 = 0$$

$$\frac{d}{dy} f(x,y) = 2x - 4 = 0$$

3. Solve the equations and get the critical points.

$$x^* = 2, \qquad y^* = 1$$

MLE: YET ANOTHER EXAMPLE

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

- What's the probability of observing such data? Likelihood:

$$L(p, q) =$$

$$\Leftrightarrow$$

MLE: YET ANOTHER EXAMPLE

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

- What's the probability of observing such data? Likelihood:

$$L(p, q) = p^{n_1} \cdot q^{n_2} \cdot (1 - p - q)^{n_3}$$

maximize L(p, q) w.r.t. p, q

$\Leftrightarrow$

MLE: YET ANOTHER EXAMPLE

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

- What's the probability of observing such data? Likelihood:

$$L(p, q) = p^{n_1} \cdot q^{n_2} \cdot (1 - p - q)^{n_3}$$

maximize L(p, q) w.r.t. p, q

$$\Leftrightarrow$$

log L(p, q) =

MLE: YET ANOTHER EXAMPLE

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

- What's the probability of observing such data? Likelihood:

$$L(p, q) = p^{n_1} \cdot q^{n_2} \cdot (1 - p - q)^{n_3}$$

maximize L(p, q) w.r.t. p, q

⇔

$$\log L(p, q) = n_1 \log p + n_2 \log q + n_3 \log(1 - p - q)$$

maximize log L(p, q) w.r.t. p, q

MLE: EXAMPLE 2

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

maximize $\mathrm{n}_1 \log \mathrm{p} + \mathrm{n}_2 \log \mathrm{q} + \mathrm{n}_3 \log(1 - \mathrm{p} - \mathrm{q})$ w.r.t. p, q

$$\frac{d}{dp} \log L(p,q) = \frac{n_1}{p} - \frac{n_3}{1-p-q} = 0$$

$$\frac{d}{dq} \log L(p,q) = \frac{n_2}{q} - \frac{n_3}{1-p-q} = 0$$

MLE: EXAMPLE 2

X	1	2	3
P(X)	p	q	1-p-q

Value	1	2	3
#	n_1	n_2	n_3

maximize $\mathrm{n}_1 \log \mathrm{p} + \mathrm{n}_2 \log \mathrm{q} + \mathrm{n}_3 \log(1 - \mathrm{p} - \mathrm{q})$ w.r.t. p, q

$$\frac{d}{dp} \log L(p,q) = \frac{n_1}{p} - \frac{n_3}{1-p-q} = 0$$

$$\frac{d}{dq} \log L(p,q) = \frac{n_2}{q} - \frac{n_3}{1-p-q} = 0$$

$$\Rightarrow$$

$$\hat{p} = \frac{n_1}{N}$$

$$\hat{q} = \frac{n_2}{N}$$

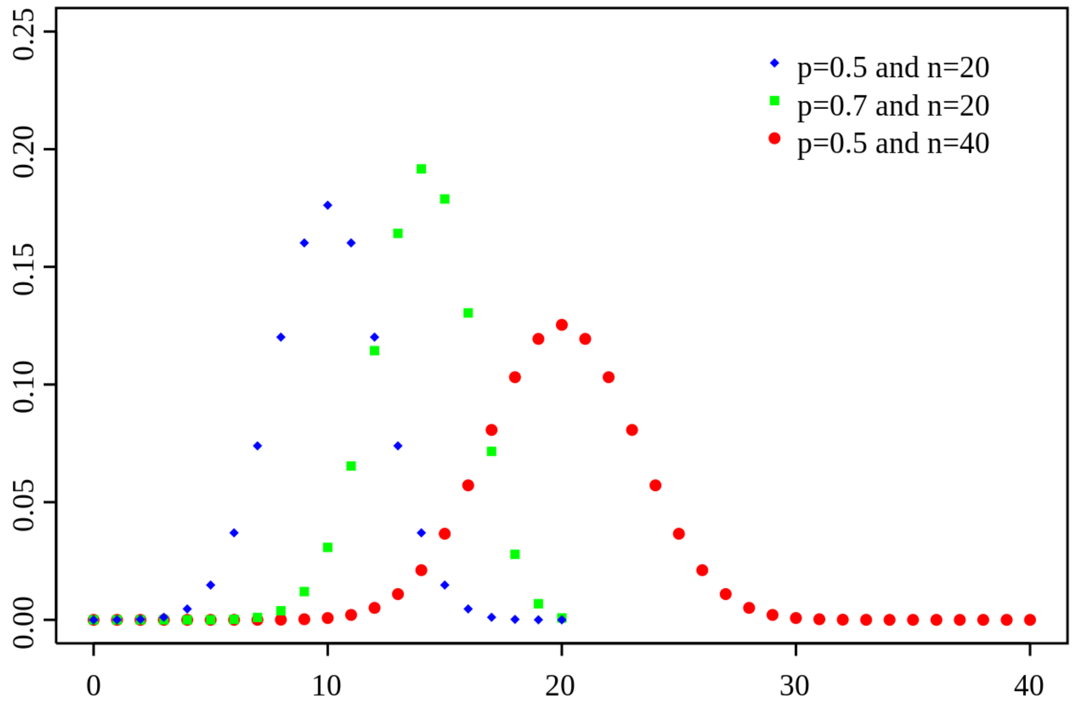
BINOMIAL DISTRIBUTION

$X \sim Bi(n, p), \;\; n = 1, 2, \ldots, \;\; 0 < p < 1$

$P(X = k) = C_n^k \, p^k (1 - p)^{n-k}$

$E(X) = np, \quad Var(X) = np(1-p)$

- Models the number of successes in a series of n independent Bernoulli trials, each of which has a success probability p.



BINOMIAL DISTRIBUTION

$X \sim Bi(n, p), \;\; n = 1, 2, \ldots, \;\; 0 < p < 1$

$P(X = k) = C_n^k \, p^k (1 - p)^{n-k}$

$E(X) = np, \quad Var(X) = np(1-p)$

- Models the number of successes in a series of *n* independent Bernoulli trials, each of which has a success probability *p*.

- You sell two types of sandwiches: chicken and vegetarian. Which one people like more?

- For the past N days, you were selling n=100 sandwiches every day and recorded the number of the chicken ones:

$$X_1, X_2, \ldots, X_N$$

- What is the MLE of the p parameter?

MLE FOR BINOMIAL DISTRIBUTION

$$\mathrm{L}(\mathrm{p}) =$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p)^{100 - X_i}$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p)^{100 - X_i}$$

$$\log L(p) =$$

MLE FOR BINOMIAL DISTRIBUTION

$$\mathrm{L(p)} = \prod_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{C}_{\mathrm{n}}^{\mathrm{X_i}} \mathrm{p}^{\mathrm{X_i}} (1 - \mathrm{p}\ )^{100 - \mathrm{X_i}}$$

$$\log \mathrm{L(p)} = \sum_{\mathrm{i=1}}^{\mathrm{N}} \left[\log \mathrm{C}_{\mathrm{n}}^{\mathrm{X_i}} + \mathrm{X_i} \log \mathrm{p} + (100 - \mathrm{X_i}) \log(1 - \mathrm{p})\right]$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p\ )^{100 - X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1 - p)\right]$$

$$\frac{d}{dp} L(p) =$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p\ )^{100 - X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1 - p) \right]$$

$$\frac{d}{dp} L(p) = \sum_{i=1}^{N} \frac{X_i}{p} - \frac{100 - X_i}{1 - p} = \qquad\qquad = 0$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p)^{100 - X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1 - p)\right]$$

$$\frac{d}{dp} L(p) = \sum_{i=1}^{N} \frac{X_i}{p} - \frac{100 - X_i}{1 - p} = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{100N - \sum_{i=1}^{N} X_i}{1 - p} = 0$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p\ )^{100 - X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1 - p)\right]$$

$$\frac{d}{dp} L(p) = \sum_{i=1}^{N} \frac{X_i}{p} - \frac{100 - X_i}{1 - p} = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{100N\ - \sum_{i=1}^{N} X_i}{1 - p} = 0$$

$$\hat{p} =$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1 - p\ )^{100 - X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1 - p)\right]$$

$$\frac{d}{dp} L(p) = \sum_{i=1}^{N} \frac{X_i}{p} - \frac{100 - X_i}{1 - p} = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{100N\ - \sum_{i=1}^{N} X_i}{1 - p} = 0$$

$$\hat{p} = \frac{\sum_{i=1}^{N} X_i}{100N}$$

MLE FOR BINOMIAL DISTRIBUTION

$$L(p) = \prod_{i=1}^{N} C_n^{X_i} p^{X_i} (1-p)^{100-X_i}$$

$$\log L(p) = \sum_{i=1}^{N} \left[\log C_n^{X_i} + X_i \log p + (100 - X_i) \log(1-p)\right]$$

$$\frac{d}{dp} L(p) = \sum_{i=1}^{N} \frac{X_i}{p} - \frac{100 - X_i}{1-p} = \frac{\sum_{i=1}^{N} X_i}{p} - \frac{100N - \sum_{i=1}^{N} X_i}{1-p} = 0$$

$$\boxed{\hat{p} = \frac{\sum_{i=1}^{N} X_i}{100N}}$$

BREAK

Task: find a coin ☺

RANDOMIZED RESPONSE

Asking embarrassing questions

MOTIVATION

- BEFORE:
    - How many vegetarians are in the world?
    - Survey: ask "*are you a vegetarian?*", estimate the true proportion.

MOTIVATION

- BEFORE:
  - How many vegetarians are in the world?
  - Survey: ask "*are you a vegetarian?*", estimate the true proportion.
- BUT WHAT IF the question is very sensitive?
  - People won't tell the truth.

MOTIVATION

- BEFORE:
  - How many vegetarians are in the world?
  - Survey: ask *"are you a vegetarian?"*, estimate the true proportion.
- BUT WHAT IF the question is very sensitive?
  - People won't tell the truth.
- EXAMPLE: *Do you find this course boring?*
  - How do I find out what my students *actually* think?

STRATEGY: RANDOMIZED RESPONSE

- Watch the video about the randomized response strategy.
- In groups, discuss the strategy and complete the assignment.
- See Google Classroom.

LET’S TRY THIS OUT!

- Toss a coin…

LET'S TRY THIS OUT!

- Toss a coin…
- Now, answer one of the following questions:

LET’S TRY THIS OUT!

- Toss a coin…
- Now, answer one of the following questions:

If you got HEADS:	DO YOU FIND THIS CLASS BORING?
If you got TAILS:	ARE YOU AT THE STATISTICS CLASS RIGHT NOW?

TO SUM UP

- **Likelihood function** is the joint probability of realized sample given the parameters.

- **Maximum Likelihood Estimate (MLE)** is the value which maximizes the probability of observing the realized sample.