Bayesian estimation

Chrysafis Vogiatzis

Department of Industrial and Enterprise Systems Engineering
University of Illinois at Urbana-Champaign

Lecture 19

ILLINOIS
ISE | Industrial & Enterprise Systems Engineering
GRAINGER COLLEGE OF ENGINEERING

Previously..

We have discussed two methods to identify “good” estimators $\hat{\Theta}$ for unknown parameters in the distribution of a population:

- **the method of moments.**
  1. Compute the moments of the population: $E\left[X^k\right]$.
  2. Compute the moments of the sample: $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.
  3. Equate them and solve for the unknown parameters.
- **maximum likelihood estimation.**
  1. Calculate the likelihood (or log-likelihood) function as

     $$L(\theta) = f(X_1, \theta) \cdot f(X_2, \theta) \cdot \ldots \cdot f(X_n, \theta) \cdot$$

  2. Find the maximizer (usually by setting the derivative equal to 0).

Today we will discuss **Bayesian estimation.**

Previously..

We have discussed two methods to identify "good" estimators $\hat{\Theta}$ for unknown parameters in the distribution of a population:

- **the method of moments.**
  1. Compute the moments of the population: $E\left[X^k\right]$.
  2. Compute the moments of the sample: $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.
  3. Equate them and solve for the unknown parameters.
- **maximum likelihood estimation.**
  1. Calculate the likelihood (or log-likelihood) function as

$$L(\theta) = f(X_1, \theta) \cdot f(X_2, \theta) \cdot \ldots \cdot f(X_n, \theta) \cdot$$

  2. Find the maximizer (usually by setting the derivative equal to 0).

Today we will discuss **Bayesian estimation**.

ILLINOIS

Previously..

We have discussed two methods to identify “good” estimators $\hat{\Theta}$ for unknown parameters in the distribution of a population:

- **the method of moments.**
  1. Compute the moments of the population: $E\left[X^k\right]$.
  2. Compute the moments of the sample: $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.
  3. Equate them and solve for the unknown parameters.

- **maximum likelihood estimation.**
  1. Calculate the likelihood (or log-likelihood) function as
     $$L(\theta) = f(X_1, \theta) \cdot f(X_2, \theta) \cdot \ldots \cdot f(X_n, \theta) \cdot$$
  2. Find the maximizer (usually by setting the derivative equal to 0).

Today we will discuss **Bayesian estimation.**

ILLINOIS

Previously..

We have discussed two methods to identify "good" estimators $\hat{\Theta}$ for unknown parameters in the distribution of a population:

- **the method of moments.**
  1. Compute the moments of the population: $E\left[X^k\right]$.
  2. Compute the moments of the sample: $\frac{1}{n}\sum_{i=1}^{n} X_i^k$.
  3. Equate them and solve for the unknown parameters.

- **maximum likelihood estimation.**
  1. Calculate the likelihood (or log-likelihood) function as

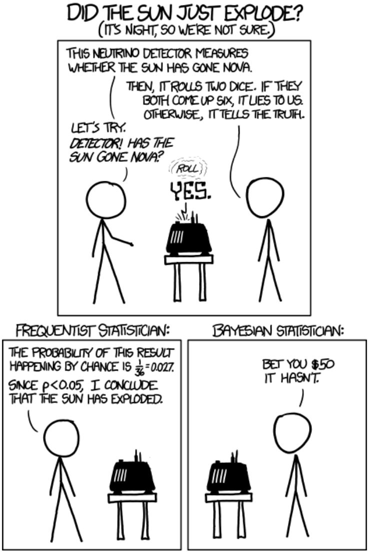
$$L(\theta) = f(X_1, \theta) \cdot f(X_2, \theta) \cdot \ldots \cdot f(X_n, \theta) \cdot$$

  2. Find the maximizer (usually by setting the derivative equal to 0).

Today we will discuss **Bayesian estimation**.

ILLINOIS

Bayesian estimation (from xkcd)



Taken from `https://xkcd.com/1132/`.

Bayesian estimation: motivation

- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{array}{l} E[X] = p \\ \overline{X} = \frac{10}{10} = 1 \end{array}\right\} p = 1.$$

Maximum likelihood estimation:

- $L(p) = p^{10}$.
- Maximized at $p = 1$.

This might be unrealistic, though.

Bayesian estimation: motivation

- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{aligned} E[X] &= p \\ \overline{X} &= \frac{10}{10} = 1 \end{aligned}\right\} p = 1.$$

Maximum likelihood estimation:

- $L(p) = p^{10}$.
- Maximized at $p = 1$.

This might be unrealistic, though.

Bayesian estimation: motivation

- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{array}{l} E[X] = p \\ \overline{X} = \frac{10}{10} = 1 \end{array}\right\} p = 1.$$

Maximum likelihood estimation:

- $L(p) = p^{10}$.
- Maximized at $p = 1$.

This might be unrealistic, though.

Bayesian estimation: motivation

- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{array}{l} E[X] = p \\ \overline{X} = \frac{10}{10} = 1 \end{array}\right\} p = 1.$$

Maximum likelihood estimation:

- $L(p) = p^{10}$.
- Maximized at $p = 1$.

This might be unrealistic, though.

ILLINOIS

Bayesian estimation: motivation

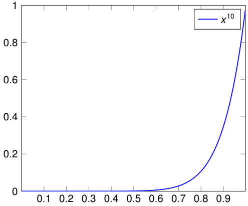
- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{array}{l} E[X] = p \\ \overline{X} = \frac{10}{10} = 1 \end{array}\right\} p = 1.$$

Maximum likelihood estimation:

- $L(p) = p^{10}$.
- Maximized at $p = 1$.



This might be unrealistic, though.

ILLINOIS

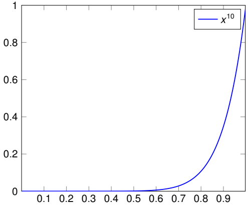
Bayesian estimation: motivation

- Assume we throw some coin with probability of Heads equal to p.
- What if we get Heads 10 times in a row?
- We *should* expect the coin always comes up Heads!

Method of moments:

$$\left.\begin{array}{l} E[X] = p \\ \overline{X} = \frac{10}{10} = 1 \end{array}\right\} p = 1.$$

Maximum likelihood estimation:



- $L(p) = p^{10}$.
- Maximized at $p = 1$.

This might be unrealistic, though.

Bayesian estimation: discrete case

Suppose you know that I carry three coins with me every time. It is equally likely I pick any one of them from my pocket.

- Coin 1: a fair coin (50%-50%).
- Coin 2: an unfair coin that favors Tails (75%).
- Coin 3: an unfair coin that favors Heads (75%).

Which coin did we “see” earlier?

Our intuition tells us that it is most probably Coin 3.

Bayesian estimation: discrete case

Suppose you know that I carry three coins with me every time. It is equally likely I pick any one of them from my pocket.

- Coin 1: a fair coin (50%-50%).
- Coin 2: an unfair coin that favors Tails (75%).
- Coin 3: an unfair coin that favors Heads (75%).

Which coin did we "see" earlier?

Our intuition tells us that it is most probably Coin 3.

ILLINOIS

Bayesian estimation: discrete case

Suppose you know that I carry three coins with me every time. It is equally likely I pick any one of them from my pocket.

- Coin 1: a fair coin (50%-50%).
- Coin 2: an unfair coin that favors Tails (75%).
- Coin 3: an unfair coin that favors Heads (75%).

Which coin did we "see" earlier?

Our intuition tells us that it is most probably Coin 3.

ILLINOIS

Back to Bayes theorem

This type of prior information is invaluable; and it comes as *extra information* on top of our observations.

We define three types of probabilities:

- *priors*: i.e., the probability we see a certain parameter. $P(\theta)$
- *likelihoods*: i.e., the probability we see an observation given a certain parameter. $P(X = x|\theta)$
- *posteriors*: i.e., the multiplication of the two. $P(\theta) \cdot P(X = x|\theta)$

The higher the posterior probability, the higher the probability that we have that specific parameter!

ILLINOIS

Back to Bayes theorem

This type of prior information is invaluable; and it comes as *extra information* on top of our observations.

We define three types of probabilities:

- *priors*: i.e., the probability we see a certain parameter. $P(\theta)$
- *likelihoods*: i.e., the probability we see an observation given a certain parameter. $P(X = x|\theta)$
- *posteriors*: i.e., the multiplication of the two. $P(\theta) \cdot P(X = x|\theta)$

The higher the posterior probability, the higher the probability that we have that specific parameter!

Back to Bayes theorem

This type of prior information is invaluable; and it comes as *extra information* on top of our observations.

We define three types of probabilities:

- *priors*: i.e., the probability we see a certain parameter. $P(\theta)$
- *likelihoods*: i.e., the probability we see an observation given a certain parameter. $P(X = x|\theta)$
- *posteriors*: i.e., the multiplication of the two. $P(\theta) \cdot P(X = x|\theta)$

The higher the posterior probability, the higher the probability that we have that specific parameter!

Back to Bayes theorem

This type of prior information is invaluable; and it comes as *extra information* on top of our observations.

We define three types of probabilities:

- *priors*: i.e., the probability we see a certain parameter. $P(\theta)$
- *likelihoods*: i.e., the probability we see an observation given a certain parameter. $P(X = x|\theta)$
- *posteriors*: i.e., the multiplication of the two. $P(\theta) \cdot P(X = x|\theta)$

The higher the posterior probability, the higher the probability that we have that specific parameter!

ILLINOIS

Back to Bayes theorem

This type of prior information is invaluable; and it comes as *extra information* on top of our observations.

We define three types of probabilities:

- *priors*: i.e., the probability we see a certain parameter. $P(\theta)$
- *likelihoods*: i.e., the probability we see an observation given a certain parameter. $P(X = x|\theta)$
- *posteriors*: i.e., the multiplication of the two. $P(\theta) \cdot P(X = x|\theta)$

The higher the posterior probability, the higher the probability that we have that specific parameter!

ILLINOIS

Bayesian estimation

parameter θ	prior $P(\theta)$	likelihood $P(X=10\|\theta)$	posterior $P(\theta) \cdot P(X=10\|\theta)$
0.25	1/3	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	1/3	$0.5^{10} = 0.00098$	0.000327
0.75	1/3	$0.75^{10} = 0.0563$	0.01877

Looking at the highest posterior, we can estimate that the coin used seems to be the one with θ equal to 75% Heads.

ILLINOIS

Bayesian estimation

parameter θ	prior $P(\theta)$	likelihood $P(X=10\|\theta)$	posterior $P(\theta) \cdot P(X=10\|\theta)$
0.25	$1/3$	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	$1/3$	$0.5^{10} = 0.00098$	0.000327
0.75	$1/3$	$0.75^{10} = 0.0563$	0.01877

Looking at the highest posterior, we can estimate that the coin used seems to be the one with θ equal to 75% Heads.

ILLINOIS

Bayesian estimation

We may prefer to normalize the posterior probabilities:

parameter θ	prior $P(\theta)$	likelihood $P(X = 10\|\theta)$	posterior $P(\theta) \cdot P(X = 10\|\theta)$
0.25	$1/3$	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	$1/3$	$0.5^{10} = 0.00098$	0.000327
0.75	$1/3$	$0.75^{10} = 0.0563$	0.01877
			$= 0.019097$

This leads to:

- $0.3179 \cdot 10^{-7} / 0.019097 = 0.000002$.
- $0.000327/0.0191 = 0.017123$.
- $0.01877/0.0191 = 0.982877$.

There is a 98.29% chance that the coin used is indeed the 75% Heads unfair coin.

If we had k coins, we would produce the same table but with k different parameter θ and still report the one with maximum posterior.

Bayesian estimation

We may prefer to normalize the posterior probabilities:

parameter θ	prior $P(\theta)$	likelihood $P(X = 10\|\theta)$	posterior $P(\theta) \cdot P(X = 10\|\theta)$
0.25	1/3	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	1/3	$0.5^{10} = 0.00098$	0.000327
0.75	1/3	$0.75^{10} = 0.0563$	0.01877
			$= 0.019097$

This leads to:

- $0.3179 \cdot 10^{-7}/0.019097 = 0.000002$.
- $0.000327/0.0191 = 0.017123$.
- $0.01877/0.0191 = 0.982877$.

There is a 98.29% chance that the coin used is indeed the 75% Heads unfair coin.

If we had k coins, we would produce the same table but with k different parameter θ and still report the one with maximum posterior.

Bayesian estimation

We may prefer to normalize the posterior probabilities:

parameter θ	prior $P(\theta)$	likelihood $P(X = 10\|\theta)$	posterior $P(\theta) \cdot P(X = 10\|\theta)$
0.25	$1/3$	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	$1/3$	$0.5^{10} = 0.00098$	0.000327
0.75	$1/3$	$0.75^{10} = 0.0563$	0.01877
			$= 0.019097$

This leads to:

- $0.3179 \cdot 10^{-7}/0.019097 = 0.000002$.
- $0.000327/0.0191 = 0.017123$.
- $0.01877/0.0191 = 0.982877$.

There is a 98.29% chance that the coin used is indeed the 75% Heads unfair coin.

If we had k coins, we would produce the same table but with k different parameter θ and still report the one with maximum posterior.

Bayesian estimation

We may prefer to normalize the posterior probabilities:

parameter θ	prior $P(\theta)$	likelihood $P(X = 10\|\theta)$	posterior $P(\theta) \cdot P(X = 10\|\theta)$
0.25	$1/3$	$0.25^{10} = 0.00000095$	$0.3179 \cdot 10^{-7}$
0.50	$1/3$	$0.5^{10} = 0.00098$	0.000327
0.75	$1/3$	$0.75^{10} = 0.0563$	0.01877
			$= 0.019097$

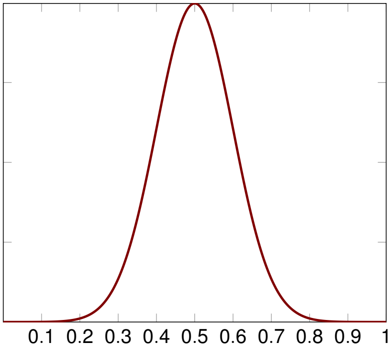
This leads to:

- $0.3179 \cdot 10^{-7}/0.019097 = 0.000002.$
- $0.000327/0.0191 = 0.017123.$
- $0.01877/0.0191 = 0.982877.$

There is a 98.29% chance that the coin used is indeed the 75% Heads unfair coin.

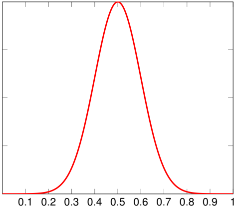
If we had k coins, we would produce the same table but with k different parameter θ and still report the one with maximum posterior.

What if we had a probability distribution $f(\theta)$ to represent the pdf of parameter θ? For example, assume that coins are produced to have a probability of Heads with pdf:

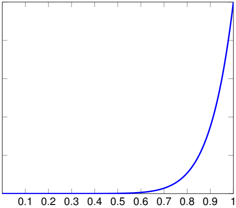


Visually

Our prior beliefs for θ ($f(\theta)$):



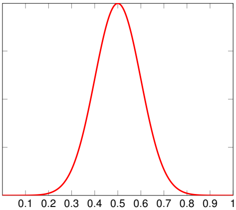
Our likelihood function ($L(\theta)$):



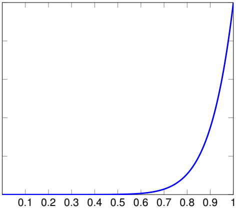
The combination of the two ($f(\theta) \cdot L(\theta)$):

Visually

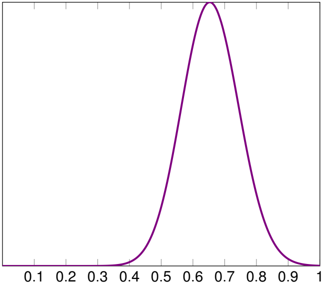
Our prior beliefs for θ ($f(\theta)$):



Our likelihood function ($L(\theta)$):



The combination of the two ($f(\theta) \cdot L(\theta)$):



ILLINOIS

Bayesian estimation: quick review

- When provided discrete cases for θ:
  1. Obtain the prior belief distribution.
     $P(\theta)$ for every possible θ.
  2. Compute the likelihood function based on the observations.
     $L(\theta)$
  3. Multiply them.
     $P(\theta) \cdot L(\theta)$.
  4. Find the maximizer $\hat{\theta}$.
- When provided a pdf for θ:
  1. Obtain the prior belief distribution.
     $f(\theta)$.
  2. Compute the likelihood function based on the observations.
     $L(\theta)$
  3. Multiply them.
     $f(\theta) \cdot L(\theta)$.
  4. Find the maximizer $\hat{\theta}$.

ILLINOIS