Statistics I - Parameter Estimation

Stochastics

Illés Horváth

2021/11/23

1. Definition, examples
2. Properties of statistics
3. Parameter estimation I - moment estimator
4. Likelihood function
5. Parameter estimation II - ML estimator

Motivation

"On a low traffic road, on average 1.8 cars pass per minute."

Motivation

"On a low traffic road, on average 1.8 cars pass per minute."

How is information like that obtained?

Motivation

"On a low traffic road, on average 1.8 cars pass per minute."

How is information like that obtained?

Usually, they make some observations, and based on those observations, an estimate is made for the rate of cars.

Motivation

"On a low traffic road, on average 1.8 cars pass per minute."

How is information like that obtained?

Usually, they make some observations, and based on those observations, an estimate is made for the rate of cars.

Example. We count the cars in 5 different 1-minute intervals, and we get the sample 1, 4, 0, 3, 1. Try to make a naive estimate for the rate of cars based on this sample.

Sample, background distribution

The general setup is as follows. A *sample* $X_1, X_2, \ldots, X_n$ is a collection of iid random variables from an unknown background distribution. n is the sample size.

We usually write $X_1, X_2, \ldots, X_n$ when emphasizing that the sample is random, and $x_1, x_2, \ldots, x_n$ for a specific realization.

Sample, background distribution

The general setup is as follows. A *sample* $X_1, X_2, \dots, X_n$ is a collection of iid random variables from an unknown background distribution. n is the sample size.

We usually write $X_1, X_2, \dots, X_n$ when emphasizing that the sample is random, and $x_1, x_2, \dots, x_n$ for a specific realization.

Usually, the background distribution is from a parametric family of distributions, denoted by $\mathbb{P}_\theta(.)$ (or the pdf $f_\theta(.)$ for continuous distributions). θ is the parameter, with possible values from a domain (the most typical domains are $\mathbb{R}, \mathbb{R}^+$ or $\mathbb{Z}^+$). Multiple parameters are also possible.

Sample, background distribution

The general setup is as follows. A *sample* $X_1, X_2, \ldots, X_n$ is a collection of iid random variables from an unknown background distribution. n is the sample size.

We usually write $X_1, X_2, \ldots, X_n$ when emphasizing that the sample is random, and $x_1, x_2, \ldots, x_n$ for a specific realization.

Usually, the background distribution is from a parametric family of distributions, denoted by $\mathbb{P}_\theta(.)$ (or the pdf $f_\theta(.)$ for continuous distributions). θ is the parameter, with possible values from a domain (the most typical domains are $\mathbb{R}, \mathbb{R}^+$ or $\mathbb{Z}^+$). Multiple parameters are also possible.

For the previous example with the cars, the background distribution is POI(λ), where λ is unknown.

A *statistic* is a $T = T(x_1, \ldots, x_n)$ function of the sample.

Statistics

A *statistic* is a $T = T(x_1, \ldots, x_n)$ function of the sample.

Notable statistics:

- sample mean:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

Statistics

A *statistic* is a $T = T(x_1, \ldots, x_n)$ function of the sample.

Notable statistics:

- sample mean:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

- median: the middle element of the sample when ordered from smallest to largest. Half of the sample elements are larger than or equal, and half of the sample elements are smaller than or equal to the median.

The mean and the median both aim to describe the typical behaviour of the sample, but they can still be quite different.

- sample minimum and maximum:

$$x_{\min} = \min(x_1, \dots, x_n), \qquad x_{\max} = \max(x_1, \dots, x_n).$$

- sample range:

$$x_{\max} - x_{\min}.$$

- sample variance (or empirical variance):

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

- sample deviation (or empirical deviation):

$$s_n = \sqrt{s_n^2}.$$

The range and deviation both aim to describe the dispersion of the sample, but once again, they can be quite different.

Properties of statistics

A point estimator or simply *estimator* is a statistic T used to estimate the unknown parameter θ or a function of the parameter $f(\theta)$.

Properties of statistics

A point estimator or simply *estimator* is a statistic T used to estimate the unknown parameter θ or a function of the parameter $f(\theta)$.

Properties of estimators:

- an estimator T is an *unbiased* estimator for θ (or $f(\theta)$) if

$$\mathbb{E}_\theta(T(X_1, \ldots, X_n)) = \theta \quad (\text{or } f(\theta)).$$

Example. If the background distribution is POI(λ), then the sample mean $\bar{X}$ is an unbiased estimator for λ:

$$\mathbb{E}_\lambda(\bar{X}) = \mathbb{E}_\lambda\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\lambda + \cdots + \lambda}{n} = \lambda.$$

Properties of statistics

A point estimator or simply *estimator* is a statistic T used to estimate the unknown parameter θ or a function of the parameter $f(\theta)$.

Properties of estimators:

- an estimator T is an *unbiased* estimator for θ (or $f(\theta)$) if

$$\mathbb{E}_\theta(T(X_1,\dots,X_n)) = \theta \quad (\text{or } f(\theta)).$$

Example. If the background distribution is POI(λ), then the sample mean $\bar{X}$ is an unbiased estimator for λ:

$$\mathbb{E}_\lambda(\bar{X}) = \mathbb{E}_\lambda\left(\frac{X_1+\dots+X_n}{n}\right) = \frac{\lambda+\dots+\lambda}{n} = \lambda.$$

So is just a single sample element:

$$\mathbb{E}_\lambda(X_1) = \lambda.$$

Properties of statistics

Lemma

For any background distribution $\mathbb{P}_\theta(.)$, $\bar{X}$ is an unbiased estimator for

$$f(\theta) = \mathbb{E}_\theta(X_1).$$

For any background distribution $\mathbb{P}_\theta(.)$,

$$s_n^{*2} = \frac{n}{n-1} s_n^2$$

is an unbiased estimator for

$$g(\theta) = \mathbb{D}_\theta^2(X_1).$$

No proof.

Properties of statistics

Lemma

For any background distribution $\mathbb{P}_\theta(.)$, $\bar{X}$ *is an unbiased estimator for*

$$f(\theta) = \mathbb{E}_\theta(X_1).$$

For any background distribution $\mathbb{P}_\theta(.)$,

$$s_n^{*2} = \frac{n}{n-1} s_n^2$$

is an unbiased estimator for

$$g(\theta) = \mathbb{D}_\theta^2(X_1).$$

No proof.

s_n^{*2} is known as the *corrected empirical variance*. See also Bessel's correction.

There exist other unbiased estimators; for example, X_1 is also an unbiased estimator for λ. In some sense, $\bar{X}$ is a better choice. We formalize this next.

- if T_1 and T_2 are both unbiased estimators for θ, then T_1 is said to be more efficient than T_2 if

$$\mathbb{D}_\theta^2(T_1) \leq \mathbb{D}_\theta^2(T_2).$$

- An unbiased statistic T is called *efficient* if no other unbiased statistics are more efficient.

There exist other unbiased estimators; for example, X_1 is also an unbiased estimator for λ. In some sense, $\bar{X}$ is a better choice. We formalize this next.

- if T_1 and T_2 are both unbiased estimators for θ, then T_1 is said to be more efficient than T_2 if
$$\mathbb{D}_\theta^2(T_1) \le \mathbb{D}_\theta^2(T_2).$$
- An unbiased statistic T is called *efficient* if no other unbiased statistics are more efficient.

Example. For POI(λ) background distribution,

$$\mathbb{D}_\lambda(\bar{X}) = \frac{\sqrt{\lambda}}{\sqrt{n}}, \qquad \mathbb{D}_\lambda(X_1) = \sqrt{\lambda},$$

so $\bar{X}$ is a more efficient estimator for λ than X_1.

- if $(T_n)_{n=1,2,\dots}$ is a sequence of statistics for all sample sizes n, then this sequence is a consistent estimator for θ if

$$\lim_{n\to\infty} \mathbb{P}(|T_n - \theta| > \varepsilon) = 0 \qquad \forall \varepsilon > 0,$$

and it is a consistent estimator for $f(\theta)$ if

$$\lim_{n\to\infty} \mathbb{P}(|T_n - f(\theta)| > \varepsilon) = 0 \qquad \forall \varepsilon > 0.$$

The LLN guarantees that $\bar{X}$ is a consistent estimator for $f(\theta) = \mathbb{E}_\theta(X_1)$ in general.

- if $(T_n)_{n=1,2,\dots}$ is a sequence of statistics for all sample sizes n, then this sequence is a consistent estimator for θ if

$$\lim_{n\to\infty} \mathbb{P}(|T_n - \theta| > \varepsilon) = 0 \qquad \forall \varepsilon > 0,$$

  and it is a consistent estimator for $f(\theta)$ if

$$\lim_{n\to\infty} \mathbb{P}(|T_n - f(\theta)| > \varepsilon) = 0 \qquad \forall \varepsilon > 0.$$

The LLN guarantees that $\bar{X}$ is a consistent estimator for $f(\theta) = \mathbb{E}_\theta(X_1)$ in general.

Specifically for POI(λ) background distribution, $\bar{X}$ is a consistent estimator for λ.

A general method to estimate the parameter of the background distribution is the moment estimator.
The idea is to estimate the unknown parameter θ so that

$$\mathbb{E}_\theta(X_1)) = \bar{x}.$$

Parameter estimation I - moment estimator

A general method to estimate the parameter of the background distribution is the moment estimator.
The idea is to estimate the unknown parameter θ so that

$$\mathbb{E}_\theta(X_1)) = \bar{x}.$$

Formally: if the function

$$g(\theta) = \mathbb{E}_\theta(X_1)$$

is invertible, then the moment estimator for θ is

$$\hat{\theta} = g^{-1}(\bar{X}).$$

Example. We count the cars in 5 different 1-minute intervals, and we get the sample 1, 4, 0, 3, 1.

Example. We count the cars in 5 different 1-minute intervals, and we get the sample 1, 4, 0, 3, 1.

The background distribution is $X_1 \sim \mathrm{POI}(\lambda)$, and

$$g(\lambda) = \mathbb{E}_\lambda(X_1) = \lambda,$$

so g is the identity function, and the moment estimator in general is

$$\hat{\lambda} = g^{-1}(\bar{X}) = \bar{X},$$

Example. We count the cars in 5 different 1-minute intervals, and we get the sample 1, 4, 0, 3, 1.

The background distribution is $X_1 \sim \mathrm{POI}(\lambda)$, and

$$g(\lambda) = \mathbb{E}_\lambda(X_1) = \lambda,$$

so g is the identity function, and the moment estimator in general is

$$\hat{\lambda} = g^{-1}(\bar{X}) = \bar{X},$$

and for the specific example, the moment estimator for λ is

$$\hat{\lambda} = \bar{x} = \frac{1+4+0+3+1}{5} = 1.8.$$

Moment estimator for 2 parameters

If the background distribution has 2 parameters, that is, $\mathbb{P}_{\theta_1,\theta_2}(.)$, then we consider the $\mathbb{R}^2 \to \mathbb{R}^2$ function

$$g(\theta_1, \theta_2) = (\mathbb{E}_{\theta_1,\theta_2}(X_1), \mathbb{D}^2_{\theta_1,\theta_2}(X_1)).$$

Moment estimator for 2 parameters

If the background distribution has 2 parameters, that is, $\mathbb{P}_{\theta_1,\theta_2}(.)$, then we consider the $\mathbb{R}^2 \to \mathbb{R}^2$ function

$$g(\theta_1, \theta_2) = (\mathbb{E}_{\theta_1,\theta_2}(X_1), \mathbb{D}^2_{\theta_1,\theta_2}(X_1)).$$

If g is invertible as an $\mathbb{R}^2 \to \mathbb{R}^2$ function, then the moment estimator for the parameters is

$$(\hat{\theta}_1, \hat{\theta}_2) = g^{-1}(\bar{x}, s_n^2)$$

where $\bar{x}$ is the sample mean and s_n^2 is the empirical variance.

Likelihood function

The *likelihood function* for a given sample $x_1, \ldots, x_n$ is

$$L_x(\theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i)$$

when the background distribution $\mathbb{P}_\theta(.)$ is discrete, and

$$L_x(\theta) = \prod_{i=1}^{n} f_\theta(X_i = x_i)$$

when the background distribution is continuous with pdf $f_\theta(.)$.

Likelihood function

The *likelihood function* for a given sample $x_1, \ldots, x_n$ is

$$L_x(\theta) = \prod_{i=1}^{n} \mathbb{P}_\theta(X_i = x_i)$$

when the background distribution $\mathbb{P}_\theta(.)$ is discrete, and

$$L_x(\theta) = \prod_{i=1}^{n} f_\theta(X_i = x_i)$$

when the background distribution is continuous with pdf $f_\theta(.)$.

Essentially, the likelihood function is equal to the probability (or density) of the sample, but viewed as the function of the parameter θ.

Maximum likelihood (ML) estimator

The likelihood function gives the idea for another parameter estimation method: for a given sample

$$\hat{\theta} = \arg\max_\theta \{L_x(\theta)\};$$

in plain words, the estimate for the parameter θ is the parameter value for which $L_x(\theta)$ is maximal.

Maximum likelihood (ML) estimator

The likelihood function gives the idea for another parameter estimation method: for a given sample

$$\hat{\theta} = \arg\max_\theta \{L_x(\theta)\};$$

in plain words, the estimate for the parameter θ is the parameter value for which $L_x(\theta)$ is maximal.

The idea is that the actual sample that occurred has a probability depending on θ; we select the θ according to which it has the highest probability.

Maximum likelihood (ML) estimator

The likelihood function gives the idea for another parameter estimation method: for a given sample

$$\hat{\theta} = \arg\max_{\theta}\{L_x(\theta)\};$$

in plain words, the estimate for the parameter θ is the parameter value for which $L_x(\theta)$ is maximal.

The idea is that the actual sample that occurred has a probability depending on θ; we select the θ according to which it has the highest probability.

The ML estimator is a concept different from Bayes! The actual value of θ is not random, we just don't know it (so we need to make an estimate).

Maximum likelihood (ML) estimator

The ML estimator is the maximum point of the function $L_x(\theta)$. If the parameter θ is from a continuous domain, then this can be computed by solving

$$\frac{\mathrm{d}}{\mathrm{d}\theta} L_x(\theta) = 0,$$

then selecting the maximum from among the solutions.

Maximum likelihood (ML) estimator

The ML estimator is the maximum point of the function $L_x(\theta)$. If the parameter θ is from a continuous domain, then this can be computed by solving

$$\frac{\mathrm{d}}{\mathrm{d}\theta} L_x(\theta) = 0,$$

then selecting the maximum from among the solutions.

There is a practical trick that generally makes the above computation easier. The *log-likelihood function* for a given sample $x_1, \ldots, x_n$ is

$$\ell_x(\theta) = \log(L_x(\theta)) = \begin{cases} \sum_{i=1}^{n} \log(\mathbb{P}_\theta(X_i = x_i)) \\ \sum_{i=1}^{n} \log(f_\theta(X_i = x_i)) \end{cases}$$

Maximum likelihood (ML) estimator

Since log is strictly increasing, $\ell_x(\theta)$ has its maximum at the same point as $L_x(\theta)$, and

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0$$

is typically easier to solve.

Maximum likelihood (ML) estimator

Since log is strictly increasing, $\ell_x(\theta)$ has its maximum at the same point as $L_x(\theta)$, and

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0$$

is typically easier to solve.

Also, points where $L_x(\theta)$ would have a local minimum at 0 do not appear as a solution of $\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0$.

Maximum likelihood (ML) estimator

Since log is strictly increasing, $\ell_x(\theta)$ has its maximum at the same point as $L_x(\theta)$, and

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0$$

is typically easier to solve.

Also, points where $L_x(\theta)$ would have a local minimum at 0 do not appear as a solution of $\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0$.

Rule of thumb: if

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0,$$

has a single solution and it is not on the border of the domain of θ, then that solution is the maximum and it is the ML estimator for θ.

Maximum likelihood (ML) estimator

If θ may only take integer values, then instead of

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0,$$

we solve

$$\frac{L_x(\theta+1)}{L_x(\theta)} = 1$$

instead.

Maximum likelihood (ML) estimator

If θ may only take integer values, then instead of

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ell_x(\theta) = 0,$$

we solve

$$\frac{L_x(\theta+1)}{L_x(\theta)} = 1$$

instead.

The idea is that up to a certain value of θ, $\frac{L_x(\theta+1)}{L_x(\theta)} > 1$ will hold, and the maximum will be at the point when the inequality changes to $\frac{L_x(\theta+1)}{L_x(\theta)} < 1$.

Problem 1

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. We roll 10 times and get the numbers 3, 6, 5, 6, 1, 4, 2, 6, 6, 4. Give a maximum likelihood estimate for p.

Problem 1

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. We roll 10 times and get the numbers 3, 6, 5, 6, 1, 4, 2, 6, 6, 4. Give a maximum likelihood estimate for p.

Solution. The sample size is $n = 10$. For each sample, the only relevant information is whether it is a 6 or not, so $\mathbb{P}(X_i = 6) = p$ and $\mathbb{P}(X_i \neq 6) = 1 - p$.

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. We roll 10 times and get the numbers 3, 6, 5, 6, 1, 4, 2, 6, 6, 4. Give a maximum likelihood estimate for p.

Solution. The sample size is $n = 10$. For each sample, the only relevant information is whether it is a 6 or not, so $\mathbb{P}(X_i = 6) = p$ and $\mathbb{P}(X_i \neq 6) = 1 - p$.

Accordingly, the likelihood function of the sample is

$$L_x(p) = (1-p) \cdot p \cdot (1-p) \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p) \cdot p \cdot p \cdot (1-p) = p^4 (1-p)^6$$

and the log-likelihood function is

$$\ell_x(p) = \log(L_x(p)) = 4 \log p + 6 \log(1 - p).$$

Problem 1

In order to compute the ML estimator, we solve

$$\frac{\mathrm{d}}{\mathrm{d}p}\ell_x(p) = 0,$$

Problem 1

In order to compute the ML estimator, we solve

$$\frac{\mathrm{d}}{\mathrm{d}p}\ell_x(p) = 0,$$

which gives

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}(4\log p + 6\log(1-p)) &= 0\\ \frac{4}{p} - \frac{6}{1-p} &= 0\\ 4(1-p) - 6p &= 0\\ 4 - 10p &= 0\\ p &= \frac{4}{10},\end{aligned}$$

so the ML estimator is

$$\hat{p} = \frac{4}{10}.$$

Problem 1

Let's compare this with what we get if we solve

$$\frac{\mathrm{d}}{\mathrm{d}p} L_x(p) = 0.$$

Problem 1

Let's compare this with what we get if we solve

$$\frac{\mathrm{d}}{\mathrm{d}p} L_x(p) = 0.$$

This gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p}\left(p^4(1-p)^6\right) &= 0 \\ 4p^3(1-p)^6 + p^4 \cdot 6(1-p)^5 \cdot (-1) &= 0 \\ p^3(1-p)^5\left(4(1-p) - 6p\right) &= 0. \end{aligned}$$

Let's compare this with what we get if we solve

$$\frac{\mathrm{d}}{\mathrm{d}p} L_x(p) = 0.$$

This gives

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}p}\left(p^4(1-p)^6\right) &= 0 \\ 4p^3(1-p)^6 + p^4 \cdot 6(1-p)^5 \cdot (-1) &= 0 \\ p^3(1-p)^5\left(4(1-p) - 6p\right) &= 0. \end{aligned}$$

Actually, this has 3 solutions: $p = 0$, $p = 1$ and $p = \frac{4}{10}$. Here, we need to check which is the maximum out of the three. $L_x(p) = 0$ for $p = 0$ and $p = 1$, so those are actually minimum points, and

$$\hat{p} = \frac{4}{10}.$$

Problem 2

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. Out of 10 rolls, we get 4 sixes. Give a maximum likelihood estimate for p. Give a moment estimate for p.

Problem 2

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. Out of 10 rolls, we get 4 sixes. Give a maximum likelihood estimate for p. Give a moment estimate for p.

Solution. Here, the sample size is $n = 1$ with $X_i \sim \mathrm{BIN}(10, p)$ and $x_i = 4$.

Problem 2

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. Out of 10 rolls, we get 4 sixes. Give a maximum likelihood estimate for p. Give a moment estimate for p.

Solution. Here, the sample size is $n = 1$ with $X_i \sim \mathrm{BIN}(10, p)$ and $x_i = 4$.

The likelihood function is

$$L_x(p) = \mathbb{P}_p(X_1 = 4) = \binom{10}{4} p^4 (1-p)^6.$$

Problem 2

We have a (possibly loaded) six-sided die where the probability p of rolling a 6 is unknown. Out of 10 rolls, we get 4 sixes. Give a maximum likelihood estimate for p. Give a moment estimate for p.

Solution. Here, the sample size is $n = 1$ with $X_i \sim \mathrm{BIN}(10, p)$ and $x_i = 4$.

The likelihood function is

$$L_x(p) = \mathbb{P}_p(X_1 = 4) = \binom{10}{4} p^4 (1-p)^6.$$

Notice that this differs from the likelihood function in Problem 1 only in a positive constant factor, so the two functions must have a maximum at the same point. This is also reflected in further calculations.

Problem 2

The log-likelihood function is

$$\ell_x(p) = \log(L_x(p)) = \log \binom{10}{4} + 4 \log p + 6 \log(1-p)$$

and we need to solve

$$\frac{\mathrm{d}}{\mathrm{d}p} \ell_x(p) = 0,$$

Problem 2

The log-likelihood function is

$$\ell_x(p) = \log(L_x(p)) = \log \binom{10}{4} + 4 \log p + 6 \log(1-p)$$

and we need to solve

$$\frac{\mathrm{d}}{\mathrm{d}p} \ell_x(p) = 0,$$

which gives

$$0 + \frac{4}{p} - \frac{6}{1-p} = 0.$$

The equation to solve at this point is identical to the equation of Problem 1, so once again the ML estimator is

$$\hat{p} = \frac{4}{10}.$$

Problem 2

For the moment estimator, we need

$$g(p) = \mathbb{E}_p(X_1) = 10p$$

for $X_1 \sim \mathrm{BIN}(10, p)$.

Problem 2

For the moment estimator, we need

$$g(p) = \mathbb{E}_p(X_1) = 10p$$

for $X_1 \sim \mathrm{BIN}(10, p)$. Then

$$g^{-1}(x) = \frac{x}{10},$$

and the moment estimator is

$$g^{-1}(\bar{x}) = g^{-1}(4) = \frac{4}{10}.$$

Problem 2

For the moment estimator, we need

$$g(p) = \mathbb{E}_p(X_1) = 10p$$

for $X_1 \sim \mathrm{BIN}(10, p)$. Then

$$g^{-1}(x) = \frac{x}{10},$$

and the moment estimator is

$$g^{-1}(\bar{x}) = g^{-1}(4) = \frac{4}{10}.$$

So in this case, the moment estimator and the ML estimator are equal.

Problem 5

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is
$0.212, 0.255, 0.300, 0.165, 0.068.$

(a) Calculate the moment estimate for a.

(b) Calculate the ML estimate for a. (Take into account that the likelihood function is not continuous.)

Problem 5

Solution.

(a) For the moment estimator, we need the expectation of $X_1 \sim U([0, a])$:

$$g(a) = \mathbb{E}_a(X_1) = \frac{a}{2},$$

so

$$g^{-1}(x) = 2x,$$

and the moment estimator for a is

$$\hat{a} = 2 \cdot \bar{x} = 2 \cdot \frac{0.212 + 0.255 + 0.300 + 0.165 + 0.068}{5} = 0.4.$$

Problem 5

Solution.

(b) The pdf of U$(0, a)$ is

$$f_a(x) = \begin{cases} \frac{1}{a} & \text{if } x \in [0, a] \\ 0 & \text{if } x \notin [0, a] \end{cases}$$

and

$$L_x(a) = \prod_{i=1}^{5} f_a(x_i).$$

Problem 5

Solution.

(b) The pdf of $\mathrm{U}(0, a)$ is

$$f_a(x) = \begin{cases} \frac{1}{a} & \text{if } x \in [0, a] \\ 0 & \text{if } x \notin [0, a] \end{cases}$$

and

$$L_x(a) = \prod_{i=1}^{5} f_a(x_i).$$

How does the function $L_x(a)$ look like? Let's evaluate it at $a = 0.25$.

Problem 5

Solution.

(b) The pdf of U$(0, a)$ is

$$f_a(x) = \begin{cases} \frac{1}{a} & \text{if } x \in [0, a] \\ 0 & \text{if } x \notin [0, a] \end{cases}$$

and

$$L_x(a) = \prod_{i=1}^{5} f_a(x_i).$$

How does the function $L_x(a)$ look like? Let's evaluate it at $a = 0.25$.

$$f_a(x_1) = f_{0.25}(0.212) = \frac{1}{0.25}$$

because $0.212 \in [0, 0.25]$.

Problem 5

(b)

$$f_a(x_2) = f_{0.25}(0.255) = 0$$

because $0.255 \notin [0, 0.25]$.

Problem 5

(b)

$$f_a(x_2) = f_{0.25}(0.255) = 0$$

because $0.255 \notin [0, 0.25]$.

This means that

$$L_x(0.25) = 0.$$

(b)

$$f_a(x_2) = f_{0.25}(0.255) = 0$$

because $0.255 \notin [0, 0.25]$.

This means that

$$L_x(0.25) = 0.$$

Actually, $L_x(a) = 0$ will hold for any $a < 0.300$. On the other hand, for $a \geq 0.300$,

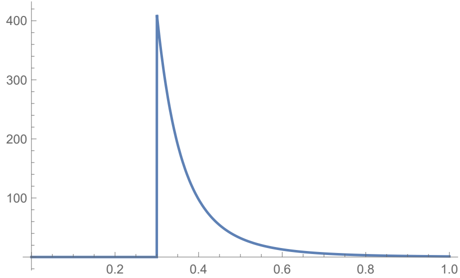
$$f_a(x_i) = \frac{1}{a} \qquad i = 1, \ldots, 5,$$

and

$$L_x(a) = \begin{cases} \frac{1}{a^5} & \text{if } a \geq 0.300 \\ 0 & \text{if } x < 0.300 \end{cases}$$

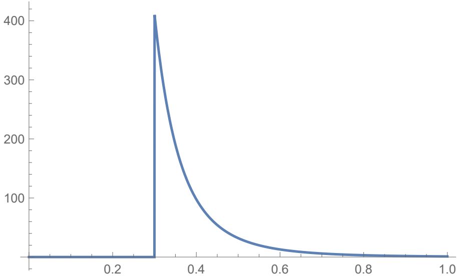
Problem 5

(b) The plot of $L_x(a)$ looks like



Problem 5

(b) The plot of $L_x(a)$ looks like

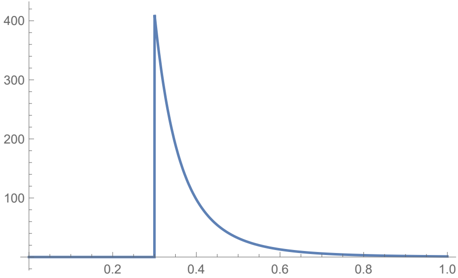


Since it is not even continuous, we cannot take the derivative. However, $\arg\max_a(L_x(a)) = 0.300$ can be determined from the plot, so the ML estimator is

$$\hat{a} = 0.300.$$

Problem 5

(b) The plot of $L_x(a)$ looks like



Since it is not even continuous, we cannot take the derivative. However, $\arg\max_a(L_x(a)) = 0.300$ can be determined from the plot, so the ML estimator is

$$\hat{a} = 0.300.$$

In general, for background distribution $\mathrm{U}(0, a)$, the ML estimator is

$$\hat{a} = \max(x_i).$$

Problem 6

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is $0.12, 0.08, 0.40, 0.05, 0.10$. Compute the moment estimator for a. Explain the result.

Problem 6

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is $0.12, 0.08, 0.40, 0.05, 0.10$. Compute the moment estimator for a. Explain the result.

Solution. Similar to the previous problem, the moment estimator for $X_1 \sim \mathrm{U}([0, a])$ is

$$\hat{a} = 2 \cdot \bar{x} = 2 \cdot \frac{0.12 + 0.08 + 0.40 + 0.05 + 0.10}{5} = 0.3.$$

Problem 6

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is $0.12, 0.08, 0.40, 0.05, 0.10$. Compute the moment estimator for a. Explain the result.

Solution. Similar to the previous problem, the moment estimator for $X_1 \sim \mathrm{U}([0, a])$ is

$$\hat{a} = 2 \cdot \bar{x} = 2 \cdot \frac{0.12 + 0.08 + 0.40 + 0.05 + 0.10}{5} = 0.3.$$

The issue with this is that it is actually not possible because the sample includes a value 0.40.

Problem 6

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is $0.12, 0.08, 0.40, 0.05, 0.10$. Compute the moment estimator for a. Explain the result.

Solution. Similar to the previous problem, the moment estimator for $X_1 \sim \mathrm{U}([0, a])$ is

$$\hat{a} = 2 \cdot \bar{x} = 2 \cdot \frac{0.12 + 0.08 + 0.40 + 0.05 + 0.10}{5} = 0.3.$$

The issue with this is that it is actually not possible because the sample includes a value 0.40.

In general, the moment estimator is not guaranteed to deliver an estimate that's actually possible.

Problem 6

A sample of 5 values were taken from a uniform distribution on the interval $[0, a]$, where a is unknown. The sample is $0.12, 0.08, 0.40, 0.05, 0.10$. Compute the moment estimator for a. Explain the result.

Solution. Similar to the previous problem, the moment estimator for $X_1 \sim \mathrm{U}([0, a])$ is

$$\hat{a} = 2 \cdot \bar{x} = 2 \cdot \frac{0.12 + 0.08 + 0.40 + 0.05 + 0.10}{5} = 0.3.$$

The issue with this is that it is actually not possible because the sample includes a value 0.40.

In general, the moment estimator is not guaranteed to deliver an estimate that's actually possible.

On the other hand, the ML estimator always gives an estimate that's possible.

Problem 7

Historically, 60% of students pass the exam of a certain course. Last semester, 14 students passed the exam, but the N number of students who took the exam is unknown. Give a ML estimate for N. Can we give a moment estimate for N?

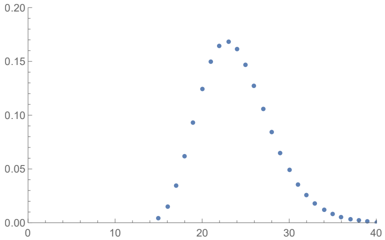
Problem 7

Historically, 60% of students pass the exam of a certain course. Last semester, 14 students passed the exam, but the N number of students who took the exam is unknown. Give a ML estimate for N. Can we give a moment estimate for N?

Solution. The sample size is $n = 1$ with $X_1 \sim \mathrm{BIN}(N, 0.6)$, and

$$L_x(N) = \binom{N}{14} 0.6^{14} \cdot 0.4^{N-14}.$$

The plot of $L_x(N)$ looks like



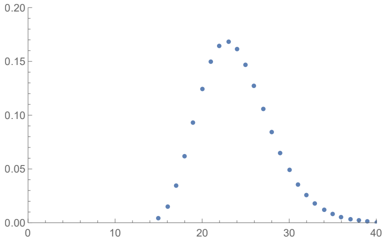
Problem 7

Historically, 60% of students pass the exam of a certain course. Last semester, 14 students passed the exam, but the N number of students who took the exam is unknown. Give a ML estimate for N. Can we give a moment estimate for N?

Solution. The sample size is $n = 1$ with $X_1 \sim \mathrm{BIN}(N, 0.6)$, and

$$L_x(N) = \binom{N}{14} 0.6^{14} \cdot 0.4^{N-14}.$$

The plot of $L_x(N)$ looks like



Problem 7

To compute the maximum point, we look to solve

$$\frac{L_x(N+1)}{L_x(N)} = 1.$$

since N is a discrete parameter now.

$$\frac{L_x(N+1)}{L_x(N)} = \frac{\binom{N+1}{14} 0.6^{14} \cdot 0.4^{N+1-14}}{\binom{N}{14} 0.6^{14} \cdot 0.4^{N-14}} =$$

$$\frac{\frac{(N+1)!}{14!(N+1-14)!} 0.6^{14} \cdot 0.4^{N+1-14}}{\frac{N!}{14!(N-14)!} 0.6^{14} \cdot 0.4^{N-14}} =$$

$$\frac{(N+1) \cdot 0.4}{(N-13)} = 1.$$

The solution of this is $N \approx 22.33$.

Problem 7

This means that for $14 \leq N < 22$,

$$\frac{L_x(N+1)}{L_x(N)} > 1,$$

so $L_x(N)$ is increasing, and for $23 \leq N$,

$$\frac{L_x(N+1)}{L_x(N)} < 1,$$

so $L_x(N)$ is decreasing.

This means that for $14 \leq N < 22$,

$$\frac{L_x(N+1)}{L_x(N)} > 1,$$

so $L_x(N)$ is increasing, and for $23 \leq N$,

$$\frac{L_x(N+1)}{L_x(N)} < 1,$$

so $L_x(N)$ is decreasing.

Altogether, $L_x(N)$ is maximal at $N = 23$, and the ML estimator is

$$\hat{N} = 23.$$

Problem 7

The moment estimator for N can be computed using

$$g(N) = \mathbb{E}_N(X_1) = 0.6N,$$

so

$$g^{-1}(\bar{x}) = \frac{1}{0.6}14 \approx 23.33.$$

Problem 7

The moment estimator for N can be computed using

$$g(N) = \mathbb{E}_N(X_1) = 0.6N,$$

so

$$g^{-1}(\bar{x}) = \frac{1}{0.6}14 \approx 23.33.$$

Any estimate for N should give an integer number, but the moment method might give a non-integer estimate. Rounding it to the nearest integer is the best we can do, so the moment estimator is

$$\hat{N} = 23.$$

Problem 9

In an M/M/1 queue, the number of jobs in the buffer has distribution PGEO$(1-\rho)$, where ρ is the load of the queue $(0 < \rho < 1)$. We check the number of jobs in the queue at 5 different points in time, and obtain the sample 2, 0, 4, 1, 1. Give a moment estimate for the load of the queue. Give a maximum likelihood estimate for the load of the queue.

Problem 9

In an M/M/1 queue, the number of jobs in the buffer has distribution PGEO($1-\rho$), where ρ is the load of the queue ($0 < \rho < 1$). We check the number of jobs in the queue at 5 different points in time, and obtain the sample 2, 0, 4, 1, 1. Give a moment estimate for the load of the queue. Give a maximum likelihood estimate for the load of the queue.

Solution. For $X_i \sim \mathrm{PGEO}(1-\rho)$,

$$g(\rho) = \mathbb{E}_\rho(X_1) = \frac{\rho}{1-\rho} = \frac{1}{1-\rho} - 1,$$

Problem 9

In an M/M/1 queue, the number of jobs in the buffer has distribution PGEO$(1-\rho)$, where ρ is the load of the queue $(0 < \rho < 1)$. We check the number of jobs in the queue at 5 different points in time, and obtain the sample 2, 0, 4, 1, 1. Give a moment estimate for the load of the queue. Give a maximum likelihood estimate for the load of the queue.

Solution. For $X_i \sim \mathrm{PGEO}(1-\rho)$,

$$g(\rho) = \mathbb{E}_\rho(X_1) = \frac{\rho}{1-\rho} = \frac{1}{1-\rho} - 1,$$

which has inverse

$$g^{-1}(x) = 1 - \frac{1}{1+x},$$

Problem 9

In an M/M/1 queue, the number of jobs in the buffer has distribution PGEO$(1-\rho)$, where ρ is the load of the queue $(0<\rho<1)$. We check the number of jobs in the queue at 5 different points in time, and obtain the sample 2, 0, 4, 1, 1. Give a moment estimate for the load of the queue. Give a maximum likelihood estimate for the load of the queue.

Solution. For $X_i \sim \mathrm{PGEO}(1-\rho)$,

$$g(\rho) = \mathbb{E}_\rho(X_1) = \frac{\rho}{1-\rho} = \frac{1}{1-\rho} - 1,$$

which has inverse

$$g^{-1}(x) = 1 - \frac{1}{1+x},$$

so the moment estimator is

$$\hat{\rho} = g^{-1}(\bar{x}) = 1 - \frac{1}{1+\bar{x}} = 1 - \frac{1}{1+8/5} = \frac{8}{13}.$$

Problem 9

The likelihood function and log-likelihood function are

$$L_x(\rho) = (1-\rho)\rho^2 \cdot (1-\rho)\rho^0 \cdot (1-\rho)\rho^4 \cdot (1-\rho)\rho^1 \cdot (1-\rho)\rho^1 = \rho^8(1-\rho)^5,$$

$$\ell_x(\rho) = 8\log(\rho) + 5\log(1-\rho).$$

Problem 9

The likelihood function and log-likelihood function are

$$\begin{aligned} L_x(\rho) &= (1-\rho)\rho^2 \cdot (1-\rho)\rho^0 \cdot (1-\rho)\rho^4 \cdot (1-\rho)\rho^1 \cdot (1-\rho)\rho^1 = \\ &\quad \rho^8(1-\rho)^5, \\ \ell_x(\rho) &= 8\log(\rho) + 5\log(1-\rho). \end{aligned}$$

Solving $\frac{\mathrm{d}}{\mathrm{d}\rho}\ell_x(\rho) = 0$ gives the ML estimator

$$\hat{\rho} = \frac{8}{13},$$

which is thus equal to the moment estimator in this case.