UPPSALA UNIVERSITET

Föreläsningsanteckningar

Inferensteori

Rami Abou Zahra

Inlämningsdatum
November 21, 2022

Contents

1. TODO

- Experiment in r (QQ-plot of exp vs n(0,1) data)
- Understand .dat files
- Add proof from book of theorem 4.9
- Problems 7.2.2 in the book
- Stora talens lag
- MLE better than methods of moments
- Derivatan av binomial
- Statistical significant change
- Pivotal storhet

2. Data Analysis (K6)

Vi kommer undersöka statistisk säkerställd skillnad (Opinion polls example), hypotestestning (räknar sannolikheten att hypotesen är sann).

Anmärkning:
Vanligtvis antar vi att datan är normalfördelad, men inte i alla fall (såsom stickprov av lön)

2.1. **Location Measures.**

A data set is given by $x_1, \cdots, x_n$

Definition/Sats 2.1: Sample mean

Sample mean is given by:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition/Sats 2.2: Median

The "middle value" of the sorted data. Different from the mean.
If n is even, the median is defined as the mean of the two middle values

Definition/Sats 2.3: Mode

This doesnt work if its continous data but it can be made discrete (such as age/time)
Mode is the most common data value

Example:
Let our data points be:

$$32\ 34\ 41\ 44\ 45\ 50\ 50\ 54\ 55\ 57\ 58\ 60\ 63$$

Find mean, median mode:

Mean: 13 data sets $\Rightarrow n = 13$:

$$\frac{1}{13}(32 + 34 + 41 + 44 + 45 + 50 + 50 + 54 + 55 + 57 + 58 + 60 + 63) \approx 49.46$$

Median: The middle value is 50

Mode: 50 is the only datavalue appearing more than once.

Anmärkning:
In this example, the median = mode. This is not always the case!

2.2. **Dispersion measures.**

Describes the "spread" of the data, such as the variance. We have the following:

Definition/Sats 2.4: Sample variance

The sample variance is given by:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

A different (yet equivalent way) of writing the sample variance is by:

$$(n-1)s^2 \Rightarrow \sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$\Leftrightarrow \sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + \overbrace{\sum_{i=1}^{n} \overline{x}^2}^{n\overline{x}^2}$$

$$= \sum_{i=1}^{n} x_i^2 - 2\overline{x}n\overline{x} + n\overline{x}^2$$

$$= \sum_{i=1}^{n} x_i^2 - 2n\overline{x}^2 + n\overline{x}^2 = \sum_{i=1}^{n} x_i^2 - n \underbrace{\overline{x}^2}_{\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^2}$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i^2 - \frac{n}{n^2}\left(\sum_{i=1}^{n} x_i\right)$$

$$\Leftrightarrow \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

$$\Rightarrow s^2 = \frac{\sum_{i=1}^{n} x_i^2}{n-1} - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n(n-1)}$$

Definition/Sats 2.5: Sample standard variance

Is given by:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Also called *sample standard deviation*

Definition/Sats 2.6: Range

Variationsbredden. The differnece between the largest and the smallest values of the data

Definition/Sats 2.7: Inter quartile range

Kvartilavståndet is the difference between the upper and lower quartiles.
If we have an odd amount of data it is including the median!

Definition/Sats 2.8: Mid-range

The mean between the biggest and smallest value in the sample

Definition/Sats 2.9: Lower/Upper quartile

The *lower quartile* is the median of the lower half of the data material including the median if n is odd

The *upper quartile* is the median of the upper half of the data material including the median if n is odd

Example:

$$0\ 0\ 1\ 1\ 2\ 2$$

Here, the mean is given by $\dfrac{(1+1+2+2)}{6} = 1.$

Therefore, the sample variance is given by $\dfrac{4}{5}$ and the sample standard deviation $\sqrt{\dfrac{4}{5}}$

We can find the inter quartile range by looking at the half, like this:

$$[0 \ \underbrace{0}_{\Delta} \ 1]\ [1 \ \underbrace{2}_{\Delta} \ 2]$$

Therefore, the inter quartile range here is $2 - 0 = 2$

2.3. **Graphical illustration.**

Stem av leafplots:

```
u = c(32,34,...)
stem(u)
```

Boxplots:
Uses quartiles, max min, and median. Useful if you want a quick look at the dispersion of data.

Bar chart:
Good for illustrating the frequency of each data point, but for large data points the data is hard to read

Histogram:
Attemps to fix the readability issues with the bar chart and is easier to compare with probability density functions.
Easier to manipulate data for readability (use bigger/smaller intervals) (one can ask what the optimal width for a histogram would be)
Very often you can ask if the data follow a normal distribution, which can be hard by just looking at the histogram (because the width varies)

Thoughts:
Dynamically widths on histograms, the more sparse data the greater the width and the more dense, the less the width

QQ-plot:
Is the data normally distributed? You order your data and construct a table with your data and compare it with if it was normally distributed:

$$\Phi(z) = \frac{i - 0.5}{n}$$

If data was perfectly normal on both axis, x_i would be a linear funcfion of z, ie. normally distributed $N(0,1)$
We plot z on the x-axis and x_i on the y-axis

The name comes from quantile-quantile-plot (QQ-plot). It is a graphical way of comparing two probability distributions (sannolikhetsfördelning)

2.4. **Data materials in several dimensions.**

We can calculate correlation through sample covariance:

Definition/Sats 2.10: Sample covariance

$$c_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

Anmärkning:
Not scale invariant (if you measure x in meters and go to cm then it is not the same). Therefore we need to norm it with something, which is where the correlation comes in:

Definition/Sats 2.11: Sample correlation coefficient

$$r_{xy} = \frac{c_{xy}}{s_x s_y}$$

Where s_x and s_y are the sample standard deviations for x and y

Definition/Sats 2.12: Sample correlation satisfies

The sample correlation coefficient satisfies

$$-1 \leq r_{xy} \leq 1$$

If it is 1, then there is a strong positive correlation (the linear regression has a line with positive derivative), similarly for negative.
When it is 0 there is no *linear* relation. There might be other, for example quadratic relation.

Bevis 2.1: Sample correlation satisfaction

$$0 \leq \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}-\frac{y_i-\overline{y}}{s_y}\right)^2$$

$$= \frac{1}{s_x^2}\underbrace{\frac{1}{n-1}\sum_i(x_i-\overline{x})^2}_{s_x^2} + \frac{1}{s_y^2}\underbrace{\frac{1}{n-1}\sum_i(y_i-\overline{y})^2}_{s_y^2} - 2\frac{1}{s_x s_y}\underbrace{\frac{1}{n-1}\sum_i(x_i-\overline{x})(y_i)(\overline{y})}_{c_{xy}}$$

$$= 2-2r_{xy} \Rightarrow r_{xy} \leq 1$$

$$0 \leq \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i-\overline{x}}{s_x}+\frac{y_i-\overline{y}}{s_y}\right)^2 = 2+2r_{xy}$$

$$\Rightarrow -1 \leq r_{xy}$$

□

Bevis 2.2: Sample correlation satsifaction

There is yet another proof that may be more intuitive for the boundedness of the sample correlation coefficient satisfaction that goes as follows:

Recall that the covariance is an inner-product, and that the variance $Var(x) = Cov(x,x)$

The "worst" that can happen is that x, y are independent, whereby the covariance $Cov(x,y) = E(XY) - E(X)E(Y) = 0$ since $E(XY) = E(X)E(Y)$

This gives the following:

$$\left|\frac{c_{x,y}}{s_x s_y}\right| = \left|\frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}\right| \geq 0$$

Next, we will use the fact that the covariance is an inner product, and therefore the Cauchy-Schwartz inequality holds
Notice that $Cov(x,x) = Var(x)$

$$|Cov(x,y)|^2 \leq |Var(x)Var(y)| \Leftrightarrow |Cov(x,y)| \leq \left|\sqrt{Var(x)}\sqrt{Var(y)}\right|$$

$$= \left|\frac{Cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}\right| \leq 1$$

From the properties of absolute values, we therefore get $-1 \leq r_{x,y} \leq 1$ □

3. Important notes from the book

3.1. Definitions/Theorems.

Definition/Sats 3.1: Mean/Medelvärde

Given n samples $x_1, \cdots, x_n$, the mean:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition/Sats 3.2: Median

Given n samples $x_1, \cdots, x_n$, the median is *the middle value* of the sorted sample

If the middle value contains 2 values (if n is even), the median is the mean of the two middle values

Definition/Sats 3.3: Mode/Typvärde

The most common number in the data set

Definition/Sats 3.4: Sample variance

Denoted by $s^2 = \sigma^2$

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

Definition/Sats 3.5: Sample standard deviation/Standardavvikelse

Given by $\sqrt{s^2}$:

$$\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Definition/Sats 3.6: Range/Variationsbredd

The difference between the largest number and the smallest number in the data set

Definition/Sats 3.7: Quartile/Kvartil

The median in the upper resp. lower half of the sorted data

Definition/Sats 3.8: Inter quartile range/Kvartilavstånd

The difference between the upper and lower quartile

Definition/Sats 3.9: Sample covariance/Kovarians

Let the data set be 2-dimensional tuples $(x_1, y_1), \cdots, (x_n, y_n)$:

$$c_{xy} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

Anmärkning:

$$S_{xy} = \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})$$

Definition/Sats 3.10: Sample correlation coefficient

Let the data set be 2-dimensional tuples $(x_1, y_1), \cdots, (x_n, y_n)$

$$r_{xy} = \frac{c_{xy}}{s_x \cdot s_y}$$

Anmärkning:

$$-1 \leq c_{xy} \leq 1$$

3.2. **Problems and Solutions.**

3.2.1. *601.*
Given $x_1, \cdots, x_5$ and $y_1, \cdots, y_9$ we have:

$$\overline{x} = 12.2 \qquad s_x = 2.1$$
$$\overline{y} = 15.8 \qquad s_y = 2.9$$

We want to combine this data into one variable $z = x_1, \cdots, x_5, y_1, \cdots, y_9$

Calculate the mean and standard deviation for z

Solution:
The mean $\frac{1}{5+9}\sum z_i = \frac{1}{5+9}\left(\sum x_i + \sum y_i\right)$
Notice we are given the mean for each variable, through algebraic manipulation we get:

$$\overline{x} = \frac{1}{5}\sum_{i=1}^{5} x_i \Leftrightarrow 5\overline{x} = \sum_{i=1}^{5} x_i = 61$$
$$\overline{y} = \frac{1}{9}\sum_{i=1}^{9} y_i \Leftrightarrow 9\overline{y} = \sum_{i=1}^{9} y_i = 142.2$$

Therefore:

$$\overline{z} = \frac{1}{14}(61 + 142.2) = 14.514$$

The standard deviation is a little trickier, but still follows from algebraic manipulation:

$$s_x = 2.1 \Rightarrow s_x^2 = 4.41 = \frac{1}{5-1}\sum_{i=1}^{5}(x_i - 12.2)^2$$
$$4.41 \cdot 4 = \sum_{i=1}^{5}(x_i - 12.2)^2 = \sum_{I=1}^{5}(x_i^2 - 2 \cdot 12.2 \cdot x_i + 12.2^2)$$
$$\Rightarrow \sum_{i=1}^{5} x_i^2 - 2 \cdot 12.2 \sum_{i=1}^{5} x_i + \sum_{i=1}^{5} 12.2^2$$
$$= \sum_{i=1}^{5} x_i^2 - 2 \cdot 12.2 \underbrace{\sum_{i=1}^{5} x_i}_{5\overline{x}} + 5 \cdot 12.2^2$$
$$= \sum_{i=1}^{5} x_i^2 - 12.2^2 = 4 \cdot 4.41 = 17.64$$
$$\Leftrightarrow \sum_{i=1}^{5} x_i^2 = 17.64 + 5 \cdot 12.2^2 = 761.84$$

Same done for y gives:

$$\Leftrightarrow \sum_{i=1}^{9} y_i^2 = 67.28 + 9 \cdot 15.6^2 = 2314.04$$

The variance for z:

$$s_z = \frac{1}{14-1}\sum_{i=1}^{14}(z_i - \overline{z})^2 = \frac{1}{13}\sum_{i=1}^{14} z_i^2 - 2\cdot 14\overline{z}^2 + 14\overline{z^2}$$
$$= \frac{1}{13}\sum_{i=1}^{14} z_i^2 - 14\overline{z^2} = 3.14176$$

3.2.2. *602.*

We essentially proceed the same way as the did for the previous problem, but take into account that we need to *remove* 19 and *add* 91.

We are given $n = 100$:

$$\overline{x} = 91.28 = \frac{1}{100}\sum_{i=1}^{100} x_i$$
$$s = 7.5$$

Find the correct mean and standard deviation

To find the mean, we proceed as follows:

$$\frac{1}{100}\sum x_i = 91.28 \Leftrightarrow 91.28\cdot 100 = \sum x_i$$
$$\sum x_i - 19 + 91 = 9200 \Leftrightarrow \overline{x} = \frac{1}{100}9200 = 92$$

Using the same trick for the standard deviation:

$$s^2 = 56.25 = \frac{1}{100-1}\sum (x_i\overline{x})^2$$
$$\Rightarrow 5568.75 = \sum (x_i^2 - 2\overline{x}x_i + \overline{x}^2)$$
$$\sum x_i^2 - 100\overline{x}^2 = 5568.75 \Leftrightarrow \sum x_i^2 = 5568.75 + 100\overline{x}^2$$
$$= 5568.75 + 100\cdot (91.28)^2 = 838772.59 = \sum x_i^2$$

Here, we correct with the squares of 19 and 91 respectively, since the summands are squared

$$8838772.59 - 19^2 + 91^2 = 846692.59 = \sum x_i^2$$

Now we can start using the real values:

$$\sum x_i^2 - 100\cdot 92^2 = 292.59$$
$$s_x = \sqrt{\frac{1}{100}292.59} = 1.71791455$$

3.2.3. *605.*

Here things get a little trickier. It is greatly encouraged to look at example 6.10 in the book.

We begin by splitting the data into intervala 0-4, 4-8,$\cdots$ and finding the middle point of those intervals (class middle):

2	6	10	14	18	22	26	30

Looking at our data, we convert it into *how many* components are breaking in an interval, and not how many we have left (frequency):

3	7	6	4	2	1	1	1

Now we can use the estimate that $\sum x_i \approx$ the sum of the frequency (f_i)·class middle (k_i):

$$\sum_{i=1}^{8} f_i k_i = 278 \approx \sum_{i=1}^{25} x_i$$

$$\overline{x} \approx \frac{278}{25} = 11.12$$

In order to calculate the standard deviation, we need to find the variance and in order to find the variance, we need to find $\sum x_i^2$, so let us do that

That sum is the same as squaring the class middle, we therefore have:

$$\sum_{i=1}^{25} x_i^2 \approx \sum_{i=1}^{8} k_i^2 f_i = 2^6 \cdot 3 + \cdots 30^2 \cdot 1 = 4356$$

$$s_x = \sqrt{\frac{1}{25-1} \sum_{i=1}^{25} x_i^2 - 25\overline{x}^2} \approx 7.259$$

4. Statistical Inference

Anmärkning:
If $X \sim Hyp$, then for a large population $X \sim Bin$ (because the chance of picking the same one in a large population is so small)

Anmärkning:
If $X_1, \cdots, X_n$ are independent and equally distributed $N(\mu, \sigma^2)$ variables, then

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2)$$

In the same way, the mean of these random variables is normally distributed with

$$\overline{X}_n \sim N(\mu, \sigma^2/n)$$

Example (*):
In an opinion poll, 1000 randomly selected voters are asked about their politcal sympathies. Let X be the number of these voters who sympathise with the party P. $X \sim Hyp$, but because the number of voters is so large, we may assume that $X \sim Bin(1000, p)$
Suppose we know that 100 of the selcted voters sympathise with P, what can p be?
Well, it makes sense that $p = \dfrac{100}{1000}$, the probability of choosing 100 from our 1000 people.

Definition/Sats 4.1: Sample (stickprov)

$x_1, x_2, \cdots, x_n$ is a *sample* from the random variable X with distribution F_X

If $X = (X_1, \cdots, X_n)$ are independent, we have a *random sample* from X

We can purposely choose our random variable in such way that makes it easier for us to analyse. It also allows us to compare these observations.

Example:
Using the same environment as example (*), we can either have a random variable $X \sim Bin(1000, p)$, *or* we can have let the 1000 people all have an associated Bernoulli distributed random variable.
In the first case, our observed sample has size $n = 1$, so our $x_1 = 100$ is an observation of the random variable X_1
In the second case, our observed sample has size $n = 1000$, with each $X_1, \cdots, X_{1000} \sim Be(p)$ and

$$\sum_{i=1}^{1000} x_i = 100$$

The pros in splitting up our situation into smaller random variables, is as touched upon previously, easier to analyse since it is easier to find independent random variables.

5. Estimation

Suppose we already know what our distribution function F is. We know this F is associated to our random variable, that we have our measured observations on.
We also know that our distribution function normally takes on certain *parameters*, for example, the Poisson distribution

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Is actually a function of both x and λ.
We call λ in our case, an *unknown parameter.* Given enough data, and knowing the distribution function, we should be able to estimate what the value of this parameter is.

We can write the following for our data:

$$x = (x_1, x_2, \cdots, x_n)$$
$$X = (X_1, X_2, \cdots, X_n)$$

Definition/Sats 5.1: Estimate (skattning)

An *estimate* $\theta^* = \theta^*(x)$ is a funciton of the sample x

The estimate is an observation of the estimator $\theta^*(X)$

This attempts to put the previous paragraphs into "functions". Given a sample from our random variable (given in $x = x_1, \cdots$), we want to find the unknown parameter θ. We can then construct a general formula for *any* data given that it comes from the random variable we have agreed upon beforehand.
This is the *estimate* function, as defined above, of a given sample x

Example:
Let $x = (x_1, x_2, x_3, x_4, x_5) = (200, 185, 210, 190, 190)$ be a random sample from $X \sim N(\mu, 100)$ (every $X_i \sim N(\mu, 100)$)

In order to estimate μ by the sample mean, we induce the function μ^* on our sample set and calculate the mean:

$$\mu^*(x) = \overline{x} = 195$$

Since the estimate is an observation of the *estimator*, lets look at the estimator:

$$\mu^*(X) = \overline{X} = \frac{1}{5}\sum_{i=1}^{5} X_i \sim N(\mu, 100) \Rightarrow \overline{X} \sim N(\mu, 100/5) = N(\mu, 20)$$

Example:
Using the poll example from above, what we really did when we said that p reasonably has to be $\frac{100}{1000}$ is determine an estimate p^*

The greater the stickprov the better the estimate (because less and less variance)

Anmärkning:
The estimator is not distributed with the same distribution, since the estimator is not always an integer. We saw this in the above example, where the estimator was $N(\mu, 20)$ distributed but our estimate was $N(\mu, 100)$ distributed.

If we have different estimates, we need to make a reasonable choice such that our error is as little as possible (this is why we introduce estimators)

5.1. Properties of estimates.

The purpose of our estimates is to estimate θ. When we calculate using our function θ^* on our sample data we get a value that may or may not deviate from the actual value θ

We can take $\theta^* - \theta = E(\theta^*(X)) - \theta + (\theta^* - E(\theta^*(X)))$
It turns out, this is equal to the systematic error + random error

Definition/Sats 5.2: Unbiased (väntevärdesriktigt)

An estimate θ^* is said to be *unbiased* if it satisfies $E(\theta^*(X)) - \theta = 0$

This is the same as saying it has no systematic error (therefore, we only have the random error left)

Example:
We show this by:
$$E(\mu^*(X)) = E(\overline{X}) = \mu$$

Let
$$p^*(X) = \frac{X}{1000} \qquad X \sim Bin(1000, p)$$
Is $p^*(x)$ an unbiased estimate of p?
Take the expected value function on both sides:
$$E\left(\frac{X}{1000}\right) = E(p^*(X)) \qquad X \sim Bin(1000, p)$$
$$\frac{1}{1000}E(X) = \frac{1}{1000} \cdot 1000 \cdot p = p$$
$$\Rightarrow E(p^*(X)) = p \Rightarrow p - p = 0$$

If we have more than one unbiased estimate, which is the best one? Well, in that case we need to start looking at the random error. We can study this by looking at the variance

Definition/Sats 5.3: Efficiency comparison of estimates

If θ_1^* and θ_2^* are unbiased estimates of θ and
$$V(\theta_1^*(X)) \leq V(\theta_2^*(X))$$

For all θ with strict inequality for some. We say that θ_1^* is more *efficient* than θ_2^* (less random error)

Example:
See slide 6 & 7

Example: (stratification)
We are interested in the proportion p of Swedish citizens that last year have traveled by plane in connection with work. Take a sample consisting of $n = 1000$ people
It is safe to assume that the number of men that take the plane will be greater than the number of woman, we can denote this using $p \pm a$. Since we are looking at the combinations of ways we can choose men and women from our set, a reasonable distribution would be a binomial distribution for men $Bin(500, p + a)$ while for women $Bin(500, p - a)$. One can also look at Bernoulli distributions and get the same results.
$$p_1^*(X) = \frac{1}{1000}X \Rightarrow E(p_1^*) = p$$
$$p_2^*(y, z) = \frac{1}{1000}y + \frac{1}{1000}z$$
$$\Rightarrow \frac{1}{1000}E(y) + \frac{1}{1000}E(z)$$
$$\frac{1}{1000}500(p + a) + \frac{1}{1000}500(p - a) = p$$
$$V(p_1^*) = \frac{p(1-p)}{1000} \geq V(p_2^*) = \frac{p(1-p)}{1000} - \frac{a^2}{1000}$$

Definition/Sats 5.4: Standard error (medelfelet)

We want to assign a numerical value to the dispersion of an estimate.
Therefore, we define the *standard error* of the estimate θ^* is an estimate of the standard deviation $D(\theta^*)$

Denoted by $d(\theta^*(x)) = d(\theta^*)$

Recall that the standard deviation is given by $\sqrt{Var}$

Example:
Given $x = 100$ (one observation) of the random variable $X \sim Bin(1000, p)$, estimate $p^* = \dfrac{x}{1000} = 0.1$ and calculate the standard error of this estimate.

5.2. **Asymptotic properties.**

The accuracy of an estimate should improve as the sample size increases, seems reasonable

Definition/Sats 5.5: Bias

The *bias* (väntevärdesfelet) for the estimate θ^* is defined as

$$B(\theta^*) = E(\theta^*) - \theta$$

Anmärkning:
An unbiased estimate has bias 0

Definition/Sats 5.6: Asymptotically unbiased

If the bias $B(\theta^*_n)$ tends to zero as $n \to \infty$ for all θ, the estimate θ^*_n is said to be *asymptotically unbiased*

Think of it like, the more data you gather, the less the bias.

Example:
Let $x_1, \cdots, x_n$ be a random sample from $N(\mu, \sigma^2)$ where μ is unknown. We want to estimate σ^2

The estimate $\sigma_n^{2*} = \dfrac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$ is biased, but it is asymptotically unbiased

The estimate s_n^2 is unbiased for σ^2, why?

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2\overline{x}x_i + \overline{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\overline{x}\underbrace{\sum_{i=1}^{n} x_i}_{nx} - n\overline{x}^2$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\overline{x}^2\right)$$

$$E(s^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n} E(x_i^2) - nE(\overline{X}^2)\right)$$

$$E(x_i^2) = V(x_i) + (E(x_i))^2 = \sigma^2 + \mu^2$$

$$E(\overline{x}^2) = V(\overline{x}) + (E(\overline{x}))^2 = \frac{\sigma^2}{n} + \mu^2$$

$$E(s^2) = \frac{1}{n-1}\left(\sum_{i=1}^{n}(\sigma^2+\mu^2) - n\left(\frac{\sigma^2}{n}+\mu^2\right)\right)$$

$$= \frac{1}{n-1}\left(n(\sigma^2+\mu^2) - \sigma^2 - n\mu^2\right) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

Definition/Sats 5.7: Consistent estimate

The estimate θ_n^* is said to be *consistent* for θ if the corresponding estimator converges to θ in probability for all θ

Definition/Sats 5.8: Convergence in probability

The estimator θ_n^* converges to θ in probability if $\forall\varepsilon > 0$:

$$\lim_{n\to\infty} P(|\theta_n^* - \theta| > \varepsilon) = 0$$

Definition/Sats 5.9

If the estimate θ_n^* is asymptotically unbiased and

$$\lim_{n\to\infty} V(\theta_n^*) = 0$$

Then it is consistent

Example:
Let $x_1, \cdots, x_n$ be a random sample from $N(\mu, \sigma^2)$ where σ^2 is known.
Estimate μ by

$$\mu_n^* = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

Show that the estimate is unbiased: $\mu - \mu = 0$

Calculate the variance for the corresponding estimator: $\dfrac{\sigma^2}{n}$

Show that the estimate is consistent: $\lim_{n\to\infty} \dfrac{\sigma^2}{n} = 0$

If an estimate is not consistent, then it does not matter if the sample size is increased, it wont yield better results.

6. Important notes from the book

6.1. Definitions/Theorems.

Definition/Sats 6.1: Sample/Stickprov

A sample $x_1, \cdots, x_n$ is of size n and is an observation from the random variable $X = X_,1 \cdots, X_n$ with distribution F

Definition/Sats 6.2: Random Sample

If the random variables $X_1, \cdots, X_n$ are independent, then the sample is a *random sample*

Definition/Sats 6.3: Estimate/Skattning

Given a sample from random variables with known distribution function but unknown "distribution function input", an *estimate* $\theta^*(x)$ is a function of the sample attempting to decote the unknown input (parameter)

Anmärkning:
The correct value one attempts to find is denoted by θ

Definition/Sats 6.4: Estimator

The estimation observed in the previous theorem, is an observation from the *estimator*; an observation of observed values of a random variable. The estimator is what the estimate observates, denoted by $\theta^*(X)$

Definition/Sats 6.5: Bias

$$E(\theta^*(X)) - \theta$$

Definition/Sats 6.6: Random error

$$\theta^* - E(\theta^*(X))$$

Definition/Sats 6.7: Total error

$$\theta^* - \theta = E(\theta^*(X)) - \theta + \theta^* - E(\theta^*(X))$$

Definition/Sats 6.8: Unbiased/Väntevärdesriktig

$$E(\theta^*(X)) - \theta = 0$$

Definition/Sats 6.9: Efficiency

Suppose θ_1^* and θ_2^* are unbiased estimates of θ and

$$V(\theta_1^*(X)) \leq V(\theta_2^*(X))$$

Then θ_1^* is *more efficient* than θ_2^*

Definition/Sats 6.10: Standard error/Medelfel

Estimate of the standard deviation, which is $\sqrt{Var}$:

$$D(\theta^*(X)) = d(\theta^*)$$

Definition/Sats 6.11: Mean squared error

$$M(\theta^*) = E((\theta^*(X) - \theta)^2)$$

Anmärkning:
Recall that $V(X) = E(X^2) - (E(X))^2 \Leftrightarrow E(X^2) = V(X) + (E(X))^2$
Therefore, MSE can be written as $E((\theta^*(X) - \theta)^2) = V(\theta^*(X) - \theta) + \underbrace{(E(\theta^*(X) - \theta))^2}_{\text{bias}^2}$

Definition/Sats 6.12: Asymptotically unbiased/Asymptotiskt Väntevärdesriktig

If the bias $B(\theta_n^*)$ goes to 0 as $n \to \infty$, then it is *asymptotically unbiased*

(for all θ in the parameter-space)

Definition/Sats 6.13: Convergence/Konvergens

The estimator $\theta_n^*(X)$ *converges* to θ:

- **In probability**:
  - If for every $\varepsilon > 0$ $P(|\theta_n^*(X) - \theta| > 0) = 0$ as $n \to \infty$
  - Notice the comparison sign, we are saying "the probability that our estimate is off from the true value by a lot goes to zero"
- **In square means**:
  - If the mean squared error $M(\theta_n^*) \to 0$ as $n \to \infty$

Anmärkning:
If the estimator converges in square means, then it converges in probability

Definition/Sats 6.14: Consistent

The estimate θ_n^* is said to be *consistent* if the estimator $\theta_n^*(X)$ converges in probability for all θ

Definition/Sats 6.15

If the estimate θ_n^* is asymptotically unbiased and $V(\theta_n^*(X)) \to 0$ as $n \to \infty$ for all θ, then our estimate is consistent

Bevis 6.1

The mean square error goes to zero, by an earlier remark it therefore converges in probability and can be written in the following way:

$$M(\theta_n^*) = V(\theta_n^*) + B^2(\theta_n^*)$$

Since the estimate is unbiased, the bias = 0, and per the theorem, the variance goes to 0 as $n \to \infty$. Then, by theorem 6.13 it converges in square means and by the remark, it also converges in probability. $\square$

7. Methods of estimation

7.1. Methods of moments.

Often, this method works quite well but it also tends to fail (**CHECK**)

Definition/Sats 7.1: Method of moments

Let $x_1, \cdots, x_n$ be a random sample from the random variable X with $E(X) = m(\theta)$ where θ is the parameter in the distribution for X

If θ is one dimensional, the moment estimate $\theta = \theta^*$ solves the equation $m(\theta) = \overline{x}$

Example:
Slide 1

$m(\beta) = E(X) = \frac{1}{\beta}$
Solve $\overline{x} = m(\beta) = \frac{1}{\beta} \Rightarrow \beta = \frac{1}{\overline{x}}$

Therefore, moment estimate is $\beta^* = \frac{1}{\overline{x}}$

Example:
Slide 2

$m(p) = E(X) = np$
Solve $x = \overline{x} = m(p) = np \Rightarrow p = \frac{x}{n}$ (recall political party example)

Moment estimation is $p^* = \frac{x}{n}$

Example:
Slide 3 (a third of those 15 birds have rings, so we can get 30 from there)

The number of birds captured with ring $= X \sim Hyp(N, n, m)$, where N = number of birds, n = how many were captured the second day (15) and m = how many with a ring in total the second day = 10:

$$p_X(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

We want to estimate N using the method of moments:

$$E(X) = n\frac{m}{N} = 15\frac{10}{N} = \frac{150}{N}$$

Solve for $f = x = \overline{x} = \frac{150}{N} \Rightarrow N = \frac{150}{5} = 30$. Moment estimation is $N^* = 30$

Definition/Sats 7.2: Method of moments with multiple parameters

If the parameter $\theta = (\theta_1, \theta_2)$, then the moment estimates solves the system:

$$E(X) = m_1(\theta_1, \theta_2) = \overline{x}$$

$$E(X^2) = m_2(\theta_1, \theta_2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Example:
Slide 4

$$m_1(\mu, \sigma^2) = E(X) = \mu$$
$$m_2(\mu, \sigma^2) = E(X^2) = V(X) + (E(X))^2 = \sigma^2 + \mu^2$$

Solve

$$\begin{cases} \mu = \overline{x} \\ \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 \end{cases}$$
$$\Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$$

Almost looks like s^2, but here in the denominator we have n instead of $n-1$

Moment estimates are:

$$\begin{cases} \mu^* = \overline{x} \\ \sigma^{2^*} = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 \end{cases}$$

Definition/Sats 7.3

Let $x_1, \cdots, x_n$ be a random sample from the random variable X where $V(X) = \sigma^2$

Then the sample variance is given by:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2$$

is an *unbiased estimate of* σ^2

7.2. **Maximum likelihood.**

Example:
Slide 5

Definition/Sats 7.4: Maximum likelihood

Let $x_1, \cdots, x_n$ be a random sample from X which has distribution $F(X;\theta)$

The likelihood function $L(\theta)$ is defined by:

$$L(\theta) = \begin{cases} \prod_{i=1}^n p(x_i;\theta) & X \text{ discrete} \\ \prod_{i=1} nf(x_i;\theta) & X \text{ continous} \end{cases}$$

The *maximum likelihood* estimate (MLE, ML-skattning) of θ is the θ that maximizes the likelihood function

Example:
Slide 6

Example:
Slide 7

Example:
Slide 8

Example:
Let $x_1, \cdots, x_n$ be a random sample from $X \sim N(\mu, \sigma^2)$ where μ and σ^2 are both unknown.

Estimate μ and σ^2 by MLE:

$$f_X(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{1}{2\sigma^2}(x-\mu)^2}$$

$$L(\mu,\sigma^2) = \prod_{i=1}^{n} f_X(x_i;\mu,\sigma^2)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\dfrac{1}{2\sigma^2}(x-\mu)^2}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$l(\mu,\sigma^2) = ln\left((2\pi\sigma^2)^{-n/2} e^{-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}\right)$$

$$= -\frac{n}{2} ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{v}\left(\sum_{i=1}^{n} x_i - n\mu\right) = \frac{n}{v}(\overline{x}-\mu)$$

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{\partial l}{\partial \mu} = 0 \Rightarrow \mu = \overline{x}$$

$$\frac{\partial l}{\partial \sigma^2} = 0 \Rightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2 \qquad (\mu = \overline{x})$$

MLE is therefore:

$$\mu^* = \overline{x}$$

$$\sigma^{2^*} = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2$$

Example:

Slide 9 (not supposed to follow, research level) (ibland för att få fram skattning måste man ta till med numeriska metoder)

Example:

Slide 1, täthetsfunk given by $f_X(x) = \dfrac{1}{\theta}$ if $0 \le x \le \theta$ and 0 otherwise, this can be shown using the indicator function $I\left\{0 \le x \le \theta\right\}$

Likelihood is given by

$$\prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} \frac{1}{\theta} I\left\{0 \le x_i \le \theta\right\}$$

This product is only $\dfrac{1}{\theta^n} \neq 0$ if all $x_i \in [0,\theta]$, otherwise it is 0

This can be expressed using an indicationr function as follows:

$$\frac{1}{\theta^n} I\left\{0 \le \min x_i \le \max x_i \le \theta\right\}$$

Anmärkning:

The max-point is when $\theta = \max x_i$, which is therfore $\theta^* = 3.0$ (from slide 1)

The method of moment is given by $\overline{x} = E(X) = m(\theta) = \dfrac{\theta}{2}$

Then $\theta = 2\overline{x}$ and $\overline{x} = 1.25 \Rightarrow \theta^* = 2.5$

But 2.5 is an unreasonable number since $3 > 2.5$

Anmärkning:
MLE is always an underestimate, so we can scale it up a little

7.3. **Least Squares.**

Definition/Sats 7.5: Least Squares

Let $x_1, \cdots, x_n$ be a random sample frmo the random variable X with $E(X) = m(\theta)$
Moreoever, let
$$Q(\theta) = \sum_{i=1}^{n} \{x_i - m(\theta)\}^2$$
The value of θ that minimizes $Q(\theta)$ is called the *least squares* estimates of θ

Example:
Slide 2.
$$m(\beta) = E(X) = \frac{1}{\beta}$$
$$Q(\beta) = \sum_{i=1}^{n} (x_i \frac{1}{\beta})^2$$
$$Q'(\beta)\frac{2}{\beta^2}\sum_{i=1}^{n}(x_i - \frac{1}{\beta}) = \frac{2}{\beta^2}\sum_{i=1}^{n} x_i - \frac{n}{\beta}$$
$$= \frac{2n}{\beta^2}(\overline{x} - \frac{1}{\beta}) = \frac{2n\overline{x}}{\beta^2} - \frac{2n}{\beta^3}$$
$$Q''(\beta) = \frac{-4n\overline{x}}{\beta^3} + \frac{6n}{\beta^4}$$
$$0 = Q'(\beta) \Rightarrow \overline{x} = \frac{1}{\beta} \Rightarrow \beta = \frac{1}{\overline{x}}$$
$$Q''\left(\frac{1}{\overline{x}}\right) = -\frac{4n\overline{x}}{(1/\overline{x})^3} + \frac{6n}{(1/\overline{x})^4} = 2n\overline{x}^4 > 0 \text{ is a minimum-point}$$
Therefore, the LSE (least square estimate) is $\beta^* = \frac{1}{\overline{x}}$, same as moment estimator.

The fact that it was equal to the moment estimator was no coincidence, this is actually a special case.

Example:
Slide 3
$$E(X) = m(\theta) \qquad m'(\theta) \neq 0$$
$$Q(\theta) = \sum_{i=1}^{n} \{x_i - m(\theta)\}^2$$
$$Q'(\theta) = -2m'(\theta)\sum_{i=1}^{n} \{x_i - m(\theta)\}$$
$$= -2n\underbrace{m'(\theta)}_{\neq 0}\{\overline{x} - m(\theta)\}$$

If the sign of the second derivative is positive, then we will see that this method of estimation yields the same answer as the moments of estimation, so let us verify the second derivative:

$$
\begin{aligned}
Q'(\theta) = 0 &\Leftrightarrow \overline{x} = m(\theta) \\
Q'(\theta) &= -2n\overline{x}m'(\theta) + 2nm'(\theta)m(\theta) \\
Q''(\theta) &= -2nm''(\theta) + 2nm''(\theta)m(\theta) + 2nm'(\theta)m'(\theta) \\
&= -2nm''(\theta)\underbrace{\{\overline{x} - m(\theta)\}}_{=0} + 2n\underbrace{\{m'(\theta)\}^2}_{>0} > 0
\end{aligned}
$$

Ok! LSE estimation solves the equation $\overline{x} = m(\theta)$ $\square$

Therefore, if the question is to calculate with LSE, you can do it using the moment of estimation which is easier

Definition/Sats 7.6

Let $x_1, \cdots, x_n$ be a random sample where the corresponding random variables $X_1, \cdots, X_n$ have $E(X_i) = m_i(\theta)$

Moreoever, let

$$Q(\theta) = \sum_{i=1}^{n} \{x_i - m_i(\theta)\}^2$$

Anmärkning:
This situation may not be handled by the method of moments nor by maximum likelihood, unless the distribution of X_i is specified (**is not the distribution of X_i given by definition of estimation?**)

Used in regression

Example:
slide 4

$$
\begin{aligned}
m_i(v) &= E(Y_i) = \text{time} = \frac{x_i}{v} \\
Q(v) &= \sum_{i=1}^{n} \left(y_i - \frac{x_i}{v}\right)^2 \\
Q'(v) &= \frac{2}{v^2}\sum_{i=1}^{n} x_i\left(y_i - \frac{x_i}{v}\right) = \frac{2}{v^2}\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{v}\sum_{i=1}^{n} x_i^2\right) \\
0 = Q'(v) &\Rightarrow v = \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i y_i} = \frac{S_{xx}}{S_{xy}} \\
Q'(v) &= \frac{2}{v^2}S_{xy} - \frac{2}{v^3}S_{xx} \\
Q''(v) &= -\frac{4}{v^3}S_{xy} + \frac{6}{v^4}S_{xx} \\
Q''\left(\frac{S_{xx}}{S_{xy}}\right) &= -\frac{4}{(S_{xx}/S_{xy})^3}S_{xy} + \frac{6}{(S_{xx}/S_{xy})^4}S_{xx} = \frac{2S_{xy}^4}{S_{xx}^3} > 0 \\
v^* &= \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i y_i} = \frac{60^2 + \cdots + 400^2}{60 \cdot 6.39 + \cdots + 400 \cdot 43.03} \approx 9.54
\end{aligned}
$$

7.4. **Parameter estimation in standard distributions.**

Example:
Two normally distributed random samples with $v = \sigma^2$ for both of them

$$L(\mu_1, \mu_2, \sigma^2) = \prod_{i=1}^{n_1} f_X(x_i) \prod_{j=1}^{n_2} f_Y(y_i)$$

$$\prod_{i=1}^{n_1} \frac{1}{\sqrt{2\pi v}} e^{-\dfrac{1}{2v}(x_i-\mu_1)^2} \cdot \prod_{j=1}^{n_2} \frac{1}{\sqrt{2\pi v}} e^{-\dfrac{1}{2v}(y_j-\mu_2)^2}$$

$$= C v^{-(n_1+n_2)/2} e^{-\dfrac{1}{2v}A}$$

$$A = \sum_{i=1}^{n_1} (x_i - \mu_1)^2 - \sum_{j=1}^{n_2} (y_j - \mu_2)^2$$

$$l = \log(L(\mu_1, \mu_2, v)) = -\log(C) - \frac{n_1+n_2}{2}\log(v) - \frac{1}{2v}A$$

$$\frac{\partial l}{\partial \mu_1} = -\frac{1}{2v}\frac{\partial}{\partial \mu_1}A = \frac{1}{v}\sum_{i=1}^{n_1}(x_i - \mu_1) = \frac{n_1}{v}(\overline{x} - \mu_1) = 0 \Rightarrow \mu_1 = \overline{x}$$

$$\frac{\partial l}{\partial \mu_2} = 0 \Rightarrow \mu_2 = \overline{y}$$

$$\frac{\partial l}{\partial v} = -\frac{n_1+n_2}{2v} + \frac{1}{2v^2}A = 0$$

$$\Leftrightarrow v = \frac{A}{n_1+n_2}$$

$$\mu_1^* = \overline{x} \quad \mu_2^* = \overline{y} \quad v^* = \frac{1}{n_1+n_2}\left(\underbrace{\sum_{i=1}^{n}(x_i - \overline{x})^2}_{(n-1)S_x^2} + \underbrace{\sum_{j=1}^{n}(y_j - \overline{y})^2}_{(n-1)S_y^2} \right)$$

$$= \frac{(n_1-1)S_x^2 + (n_2-1)S_y^2}{n_1+n_2}$$

$$E(S_x^2) = \sigma^2 = E(S_y^2)$$

$$E(V^*) = \frac{(n_1-1)E(S_x^2) + (n_2-1)E(S_y^2)}{n_1+n_2} = \frac{n_1 - 1 + n_2 - 1}{n_1+n_2}\sigma^2 \neq \sigma^2$$

$$\text{Men } S_p^2 = \frac{(n_1-1)S_x^2 + (n_2-1)S_x^2}{n_1+n_2-2} \text{ är väntevärdesriktigt}$$

7.5. **Estimation of distributions.**

Example:
slide 6
We have a random sample from a random variable X with an unknown distribution
More generally, estimate the distribution function $F_X(x) = P(X \leq x) \quad \forall x$

Definition/Sats 7.7: Empirical distribution function

Let $x_1, \cdots, x_n$ be a random sample from a random variable X. The *empirical distribution function* for X is defined as

$$F_n(x) = \frac{A_n(x)}{n}$$

where $A_n(x)$ is the number of observations in the sample that are smaller than or equal to x

Anmärkning:
For every given x, $F_n(x)$ is unbiased and consistent estimate of $F_X(x)$

8. Important notes from the book

8.1. Definitions/Theorems.

Definition/Sats 8.1: Method of moments/Momentmetoden

Let $x_1, \cdots, x_n$ random sample from X with $E(X) = m(\theta)$
If θ is one dimensional, the moment estimate $\theta = \theta^*$ solves equation $m(\theta) = \overline{x}$

Definition/Sats 8.2

Let $x_1, \cdots, x_n$ random sample from X with $E(X) = \theta$
The estimate $\theta^* = \overline{x}$ is unbiased and if $\sigma^2 = V(X) < \infty$ then it is consistent as well

Bevis 8.1

$$E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{n}{n}E(X_i) = \theta$$
$$V(\overline{X}) = V\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i) = n\frac{\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

By theorem 6.15, the estimate is unbiased (per def. in this case) and the variance goes to 0 as n increases, therefore it is consistent. □

Definition/Sats 8.3: Multivariate method of moments

$\theta = (\theta_1, \theta_2)$, moment estimates solve the system:

$$E(X) = m_1(\theta_1, \theta_2) = \overline{x}$$
$$E(X^2) = V(X) + (E(X))^2 = m_2(\theta_1, \theta_2) = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

Definition/Sats 8.4: Sample variance is unbiased

Let $x_1, \cdots, x_n$ random sample from a random variable X with variance σ^2

The sample variance s^2 is an unbiased estimation of σ^2

Anmärkning:
If $E(X^4) < \infty$ then s^2 is a consistent estimate of σ^2

8.2. Problems and Solutions.

8.2.1. *7.2.5.*

8.2.2. *7.2.6.*

8.2.3. *7.2.7.*

8.2.4. *7.2.8.*

8.2.5. *7.2.10.*

8.2.6. *7.2.12.*

9. Lesson 1

9.1. **727.**

Kalle lägger patiens, en gång per kväll, tills den går ut för första gången.
Under en vecka får han observationerna

$$3\ 7\ 10\ 5\ 12\ 8\ 4$$

Bestäm ML-skattningen av $p = P(\text{patiensen går ut})$

Lösning:
Här är slumpvariabeln ffg fördelad.
Låt X vara antalet gånger tills patiensen går ut, då är fördelningsfunktionen:

$$p_X(k) = (1-p)^{k-1}$$

Vi räknar med ML-skattning, vilket är:

$$\begin{aligned} L(p) &= \prod_{i=1}^{n} p_X(x_i) = \prod_{i=1}^{n} (1-p)^{x_i - 1} p \\ &= (1-p)^{\sum_{i=1}^{n} x_i - n} p^n \end{aligned}$$

Vi logaritmerar:

$$\begin{aligned} l(p) = \ln\{L(p)\} &= (\sum_{i=1}^{n} x_i - n)\ln(1-p) + n\ln(p) \\ l'(p) &= -\left(\sum_{i=1}^{n} x_i - n\right)\frac{1}{1-p} + \frac{n}{p} \\ l''(p) &= -\left(\sum_{i=1}^{n} x_i - n\right)\left(\frac{1}{1-p}\right)^2 - \frac{n}{p^2} < 0 \Rightarrow \ \max \\ 0 = l'(p) \Rightarrow p &= \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}} = \frac{1}{7} \end{aligned}$$

ML-skattningen är $p^* = \dfrac{1}{7}$ vilket är samma som momentskattningen (så blir det ofta men inte alltid)

9.2. **7.210.**

En enarmad bandit (spel) med vinstchans p
Albert has spelat 10 ggr och fått 2 vinster
Beata spelade tills första vinst, gång 7
ML-skatta p

Lösning:
Låt X_1 = antal vinster under 10 spel. Denna slumpvariabel är Binomialfördelad med 10,p
Låt X_2 = antalet spel till första vinst. Denna slumpvariabel är ffg fördelad med parameter p

$$
\begin{aligned}
L(p) &= p_{X_1}(x_1;p)p_{X_2}(x_2;p) \qquad x_1 = 2 \quad x_2 = 7\\
&= \binom{10}{x_1} p^{x_i}(1-p)^{10-x_1} \cdot (1-p)^{x_2-1}p\\
&= \binom{10}{x_1} p^{x_1+1}(1-p)^{9-x_1+x_2}
\end{aligned}
$$

$$l(p) = \ln\{L(p)\} = \ln\binom{10}{x_1} + (x_1+1)\ln(p) + (9-x_1+x_2)\ln(1-p)$$

$$l'(p) = (x_1+1)\frac{1}{p} - (9-x_1+x_2)\frac{1}{1-p}$$

$$l''(p) = -(x_1+1)\frac{1}{p^2} - (9-x_1+x_2)\frac{1}{(1-p)^2} < 0 \quad \text{om } 9-x_1+x_2>0 \quad \text{vilket vi har eftersom} 9-2+7>0$$

$$0 = l'(p) \Rightarrow p = \frac{x_1+1}{x_2+10} = \frac{2+1}{7+10} = \frac{3}{17} \approx 0.176$$

ML-skattningen är då $p^* = \dfrac{3}{17}$

9.3. **7.2.12.**

Taxi problemet. 7 taxibilar observeras. De är numrerade $1, \cdots, N$
Obs, numren $07o, 234, 166, 7, 65, 17, 4$

ML-skatta N

$X =$ numret på en taxibil, diskret likformigt fördelad på $(1, 2, \cdots, N)$

Sannolikhetsfunktionen är då:

$$p_X(k) = \begin{cases} \dfrac{1}{N}, 1 \le k \le N \\ 0, \text{ annars} \end{cases}$$

N kan vara hur stort som helst, $N \in \mathbb{N}^+ =$ rummet av alla positiva heltal

ML-skatta N (observationer $x_1, \cdots, x_n$):

$$L(N) = \prod_{i=1}^{n} p_X(x_i) = \begin{cases} \left(\dfrac{1}{N}\right)^N & \text{om } \forall x_i \le N \\ 0 & \text{annars} \end{cases}$$

(Logaritmera/derivera funkar ej här, man måste tyvärr tänka)

ML-skattningen N^* inträffar i $\max x_i = 234$

Momentskattning: $m(n) = E(X) = \dfrac{N+1}{2}$
Lös $\overline{x} = \dfrac{N+1}{2} \Rightarrow N = 2\overline{x} - 1$
Vi har $\overline{x} = 84.3$, momentskattningen blir 167.6, vilket blir en orimlig skattning eftersom vi har en observation som är större.

9.4. **7.2.14.**

x_1, x_2 mätningar av en storhet med värdet μ
x_3 mätning av en storhet med värdet 2μ
Mätningar saknar systematiska fel, men har en slumpfel standardavvikelse σ

Bestäm MK-skattningen av μ och visa att den är väntevärdesriktigt

Minimera $Q(\mu) = (x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - 2\mu)^2$, detta kan vi lösa med derivering:

$$\begin{aligned}
Q'(\mu) &= -2(x_1 - \mu) - 2(x_2 - \mu) - 4(x_3 - 2\mu) \\
&= -2x_1 - 2x_2 - 4x_3 + 12\mu \\
Q''(\mu) &= 12 > 0 \quad \text{ger min} \\
0 = Q'(\mu) \Leftrightarrow \mu &= \frac{1}{12}(2x_1 + 2x_2 + 4x_3) = \frac{1}{6}x_1 + \frac{1}{6}x_2 + \frac{1}{3}x_3
\end{aligned}$$

MK-skattningen μ^*
Estimatorn $\mu^*(X_1 + X_2 + X_3) = \frac{1}{6}X_1 + \frac{1}{6}X_2 + \frac{1}{3}X_3$
Då blir:

$$\begin{aligned}
E\left\{\mu^*(X_1; X_2; X_3)\right\} &= \frac{1}{6}E(X_1) + \frac{1}{6}E(X_2) + \frac{1}{3}E(X_3) \\
&= \frac{1}{6}\mu + \frac{1}{6}\mu + \frac{1}{3}2\mu = \mu
\end{aligned}$$

Då är μ^* väntevärdesriktigt

En annan skattning är:

$$\mu' = \frac{2x_1 + 2x_2 + x_3}{6}$$

Är den väntevärdesriktigt? Vi tittar på motsvarande estimator:

$$\begin{aligned}
\mu'(X_1, X_2, X_3) &= \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{6}X_3 \\
E\left\{\mu'(X_1, X_2, X_3)\right\} &= \frac{1}{3}E(X_1) + \frac{1}{3}E(X_2) + \frac{1}{6}E(X_3) \\
&= \frac{1}{3}\mu + \frac{1}{3}\mu + \frac{1}{6}2\mu = \mu
\end{aligned}$$

Ok!

Vilken skattning är effektivast?
Då jämför vi variansena:

$$\begin{aligned}
V(\mu) &= \frac{6}{36}\sigma^2 \\
V(\mu') &= \frac{9}{36}\sigma^2
\end{aligned}$$

9.5. **702.**

Observationer: 4.0, 1.1, 0.2, 1.2, 2.5, 2.0, 0.7, 1.0 är ett stickprov från en Raylerghfördelning med täthetsfunktion:

$$F_X(x) = axe^{-\frac{ax^2}{2}} \qquad x \leq 0$$

ML-skatta a:

$$L(a) = \prod_{i=1}^{n} f_X(x_i) = \prod_{i=1}^{n} a x_i e^{-\dfrac{a x_i^2}{2}}$$

$$= a^n \left(\prod_{i=1}^{n} x_i \right) e^{-\dfrac{a}{2} \sum_{i=1}^{n} x_i^2}$$

$$l(a) = \ln \{L(a)\} = n \ln(a) + \sum_{i=1}^{n} \ln x_i - \frac{a}{2} \sum_{i=1}^{n} x_i^2$$

$$l'(a) = \frac{n}{a} - \frac{1}{2} \sum_{i=1}^{n} x_i^2$$

$$l''(a) = -\frac{n}{a^2} < 0 \Rightarrow \ \max$$

$$0 = l'(a) \Rightarrow a = \frac{2n}{\sum_{i=1}^{n} x_i^2}$$

ML skattning $a^* = \dfrac{2n}{\sum_{i=1}^{n} x_i^2} = \dfrac{2 \cdot 8}{30.43} \approx 0.526$

10. Konfidensintervall/Intervallskattning

Instead of estimating the parameter, we estimate an interval in which the parameter is most likely to "be" in.

Attempts to answer the question of "was it random?"

Recall:
Let $Z \sim N(0,1)$. The α quantile λ_α is defined through $\alpha = P(Z > \lambda_\alpha)$

By symmetry, this gives $P(-\lambda_{\alpha/2} < Z < \lambda_{\alpha/2}) = 1 - \alpha$

If $1 - \alpha = 0.95$ then we have $\lambda_{\alpha/2} = \lambda_{0.025} = 1.96$ and:

$$P(-1.96 < Z < 1.96) = 0.95$$

Example:
Slide 7 example 2
We always start with an estimator:

$$\mu^*(X) = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5} \sim N(\mu, 20)$$

Then standardize it