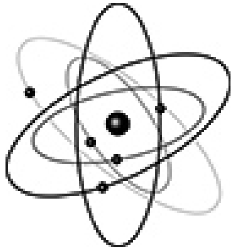
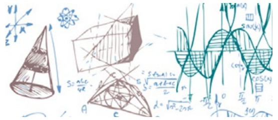
Introduction to probability, statistics and data handling

Agnieszka Obłąkowska-Mucha
Tomasz Szumlak

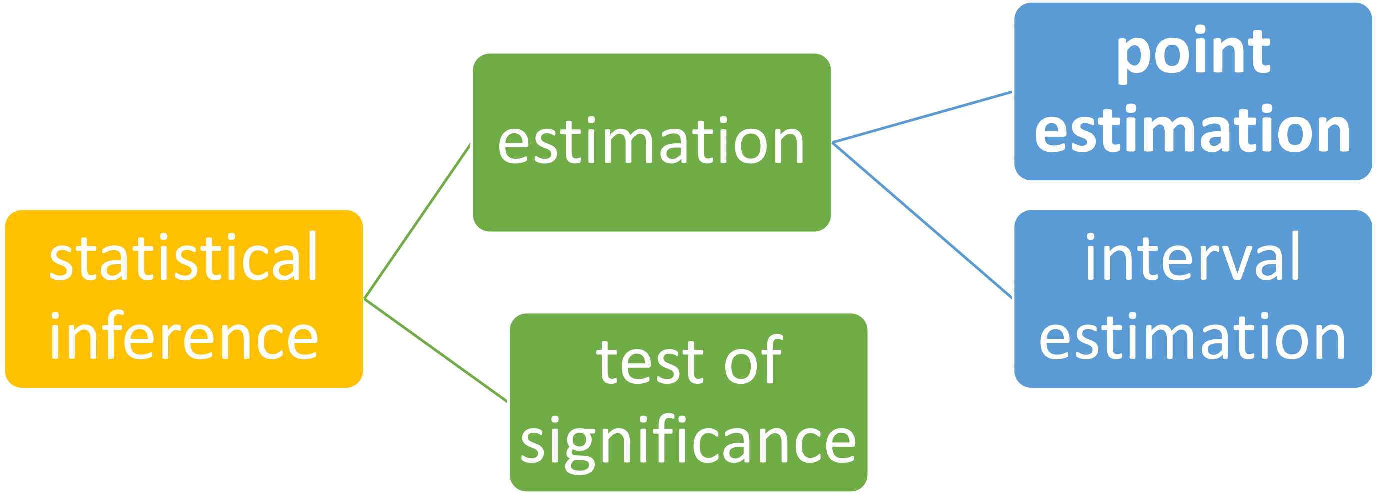
Faculty of Physics and Applied Computer Science
AGH University of Krakow





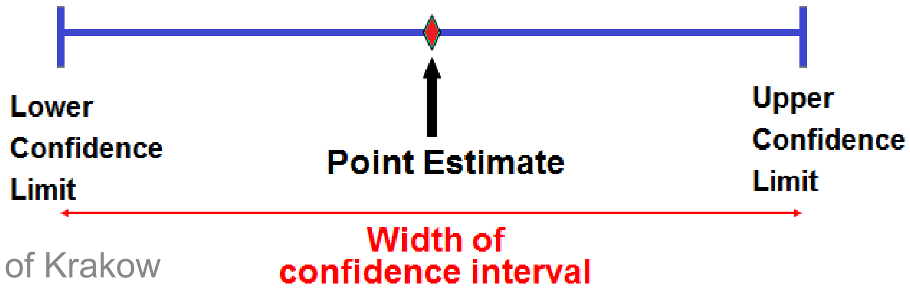
Statistical Inference

- The statistical inference consists in arriving at (quantitative) conclusions concerning a **population** where it is impossible or impractical to examine the entire set of observations that make up the population. Instead, we depend on a **subset** of observations - a **sample**.



Example:
I checked 5 restaurants in Milano and claim that mean cost of pizza is 15 €.

In Milano pizza costs betweeen 10 and 20 €.

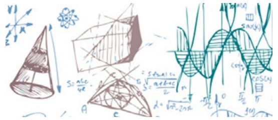


Point estimation

- **Point estimation** in statistics involves using sample data to calculate a single value that serves as the **best approximation** of an unknown population parameter.

Suppose we want to estimate the **mean weight** of a certain species of turtles in Florida. Collecting data on every individual turtle in Florida would be impractical due to the large population size. Instead, we take a **random sample** of 50 turtles and use the **sample mean weight** to estimate the true population mean.

- **Population Parameter**: The true average weight of all turtles in Florida (population mean, denoted as µ).
- **Sample Data**: We weigh 50 turtles and find the **sample mean weight** to be **150.4 pounds**.
- Our **point estimate** for the true population mean weight is **150.4 pounds**. This value represents our best guess based on the sample data.



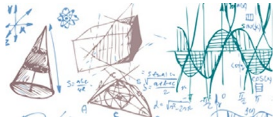
Statistical Sample and Population

- We start with two estimators:

  - estimator of a **mean value**
  - estimator of a **variance**

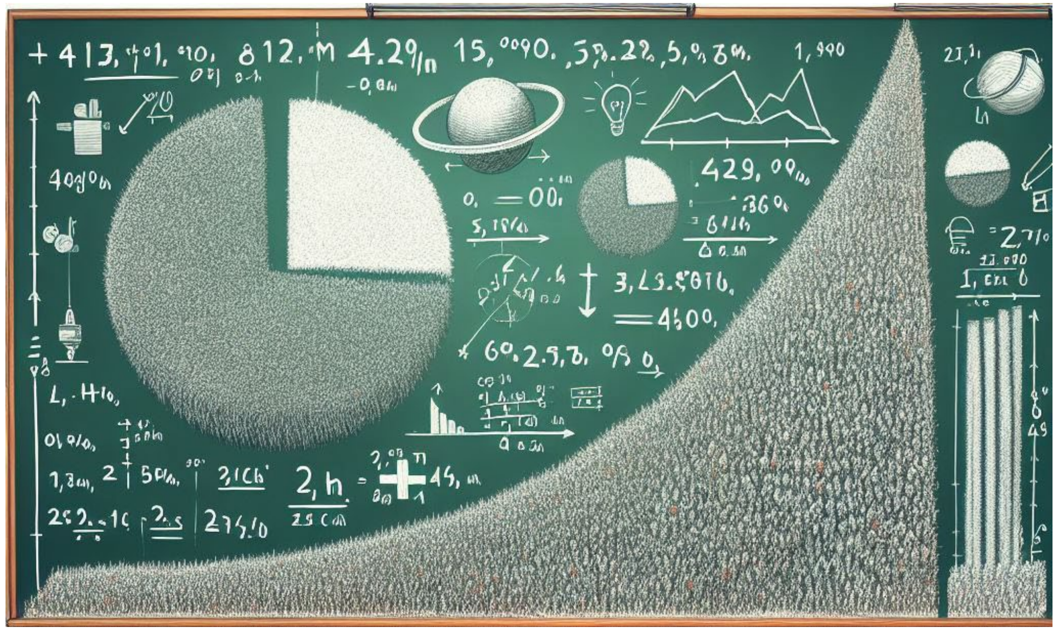
  we want to estimate μ and σ^2 of a **population** with a use of **sample**

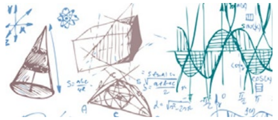
- Later we will develop methods for the estimation of unknown parameter of a model (linear, or any other) based on samples (method of moments, method of least squares, maximum likelihood estimation)



Point estimation

- Let’s think about the following: we are looking at some phenomena (took a data sample), now what we like to do is to try describe the data using a model (have we already discussed any models?)
- Using the statistics lingo we would say: we want to estimate the parameters for the hypothesised population model
- **As usual there are a lot of methods, we are going to have a look at a few of them**
- Estimators should have specific features (we will discuss it today)
- **BUT**
- Let’s start with some **examples** first!





Statistical Sample and Population

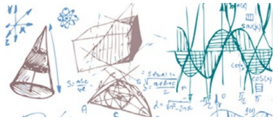
- Sample posses a property X (our RV); $X \rightarrow f(x, \boldsymbol{\lambda})$ (probability density function), $\boldsymbol{\lambda}$ – set of parameters of the population to be determined from the sample (e.g. μ, σ, etc.).
- Any function of the random variables constituting a random sample that is used for **estimation** of unknown distribution parameters $\boldsymbol{\lambda}$ is called a **statistic $\boldsymbol{S}$**:

$$S = S(X_1, X_2, \dots, X_n)$$

$$\lambda_i = E[S(X_1, X_2, \dots, X_n)] \equiv \hat{S}$$

- We say: the estimated value of a statistic $\hat{S}$ is said to be estimator of the parameter λ ; the estimation is carried out on the basis of an n-element **sample**.

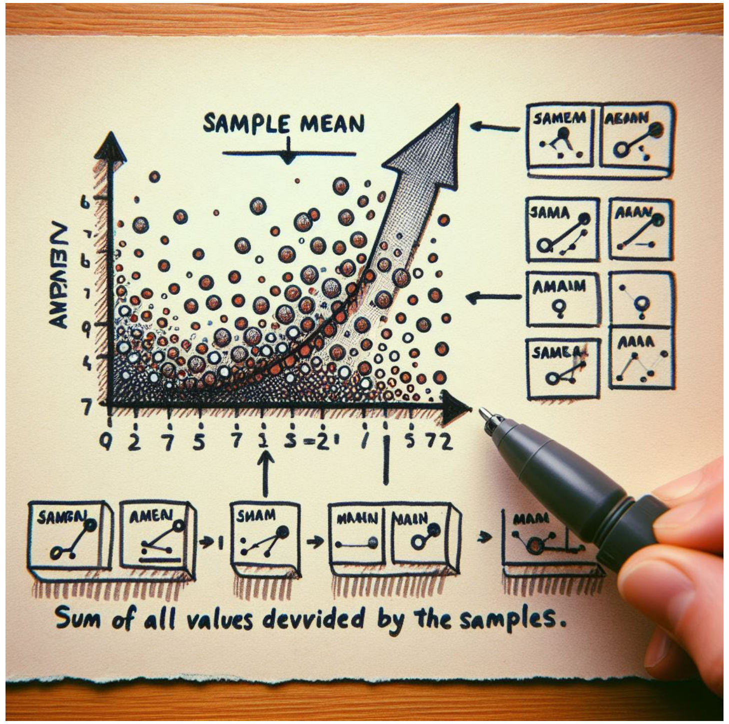
do we know any statistic?



Estimators

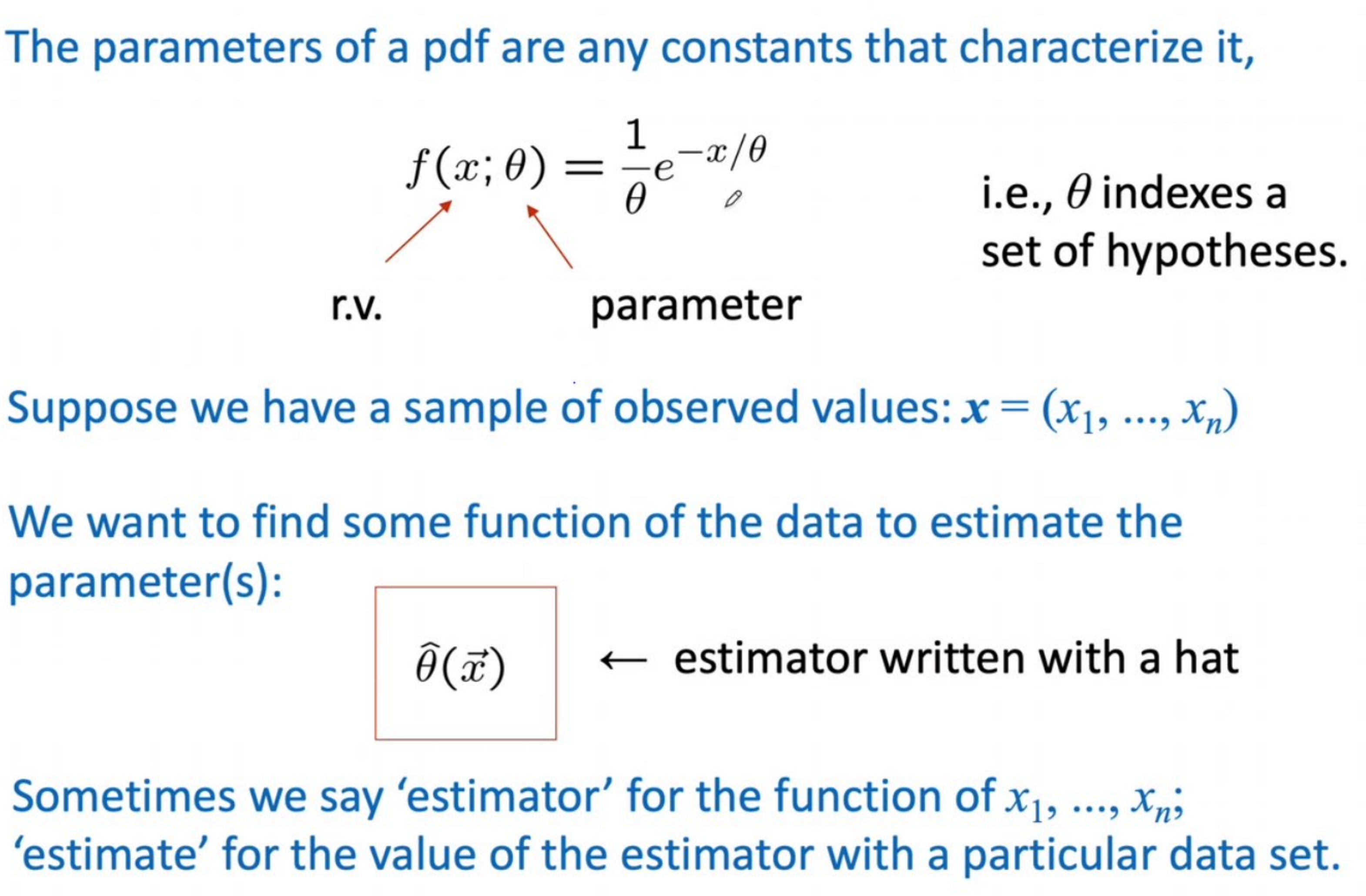
- Let's set a **generic procedure** using this simple example
- First, we **pick the parameter** to be estimated
- Next, we need to **collect data** and **compute a sampling statistics** using a formula corresponding to the parameter we are interested in
- In our example that is a **sample mean**

$$\boxed{\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}}$$



- This, in turn, we call an **estimator** of true parameter, in our case this would be: $\mu \to \bar{X} = \hat{\mu}$ (we use the caret symbor "^")
- Remember – the estimator is a random variable, for different sample we are going to get different value
- The estimator will follow its own distribution – **sampling distribution of the estimator**

Parameter estimation



The parameters of a pdf are any constants that characterize it,

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$$

r.v.

parameter

i.e., θ indexes a set of hypotheses.

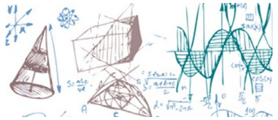
Suppose we have a sample of observed values: $\boldsymbol{x} = (x_1, \ldots, x_n)$

We want to find some function of the data to estimate the parameter(s):

$$\boxed{\hat{\theta}(\vec{x})}$$

← estimator written with a hat

Sometimes we say 'estimator' for the function of $x_1, \ldots, x_n$;
'estimate' for the value of the estimator with a particular data set.



Estimators

- Consider the following: to check the water for contamination by a micro-organism a number of samples were taken, the results are summarised as follow

Counts	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**>9**
Frequency	53	25	13	2	2	1	1	0	1	0

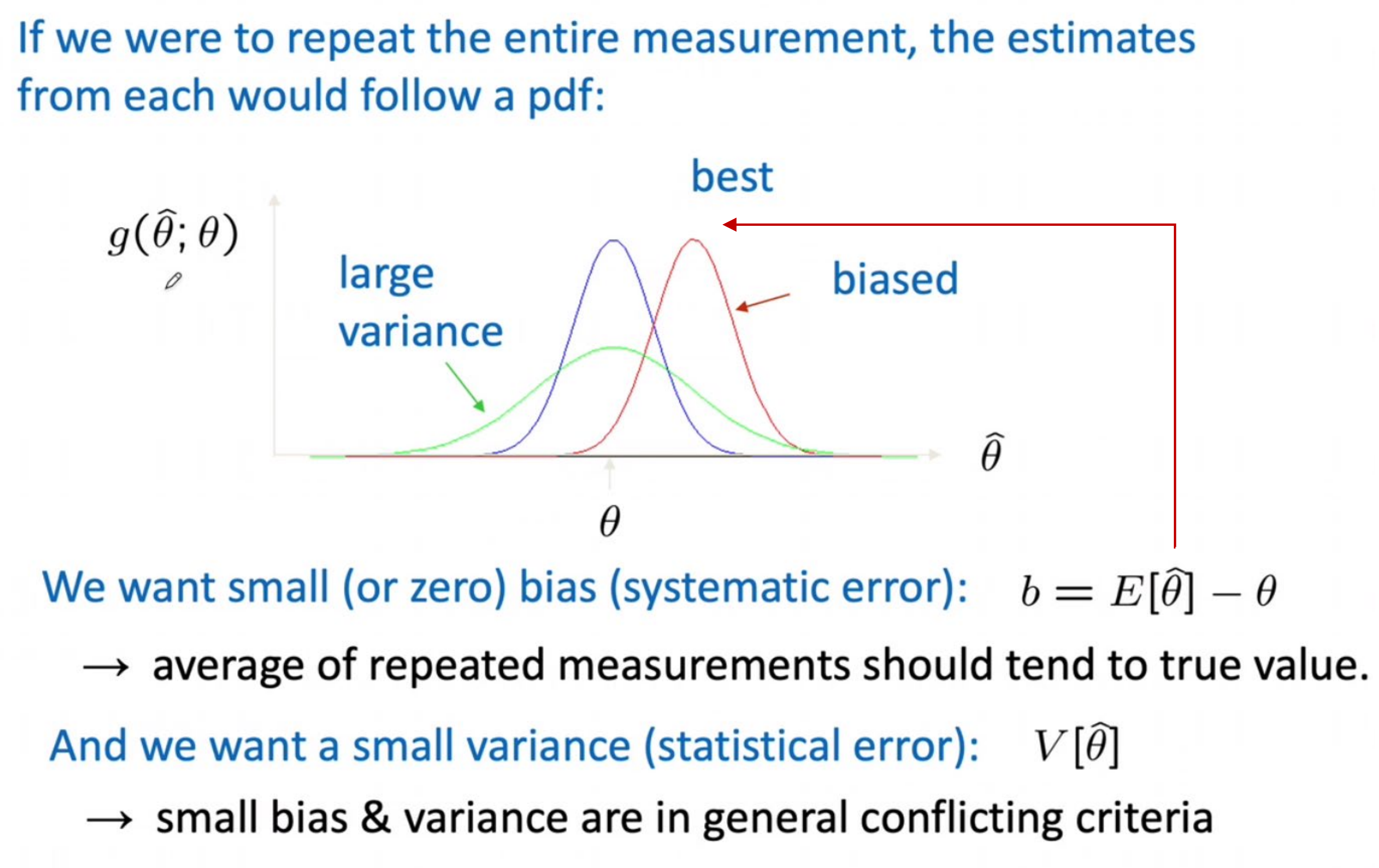
- One can assume that the data follow the Poisson distribution with an unknown parameter μ (each water sample is an independent observation on the same random variable!)
- For these particular data, we can estimate the μ as:

$$\bar{x} = \frac{0 \cdot 53 + 1 \cdot 25 + \cdots + 8 \cdot 1}{58 + 25 + \cdots + 1} = \frac{84}{103} = 0.816$$

$$\{X_1, X_2, \dots, X_{103}\} \rightarrow X \equiv Poisson(\mu)$$

$$\bar{X}_{(1)} = \frac{X_1 + X_2 + \cdots + X_{103}}{103} \rightarrow \qquad \bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Properties of estimators



If we were to repeat the entire measurement, the estimates from each would follow a pdf:

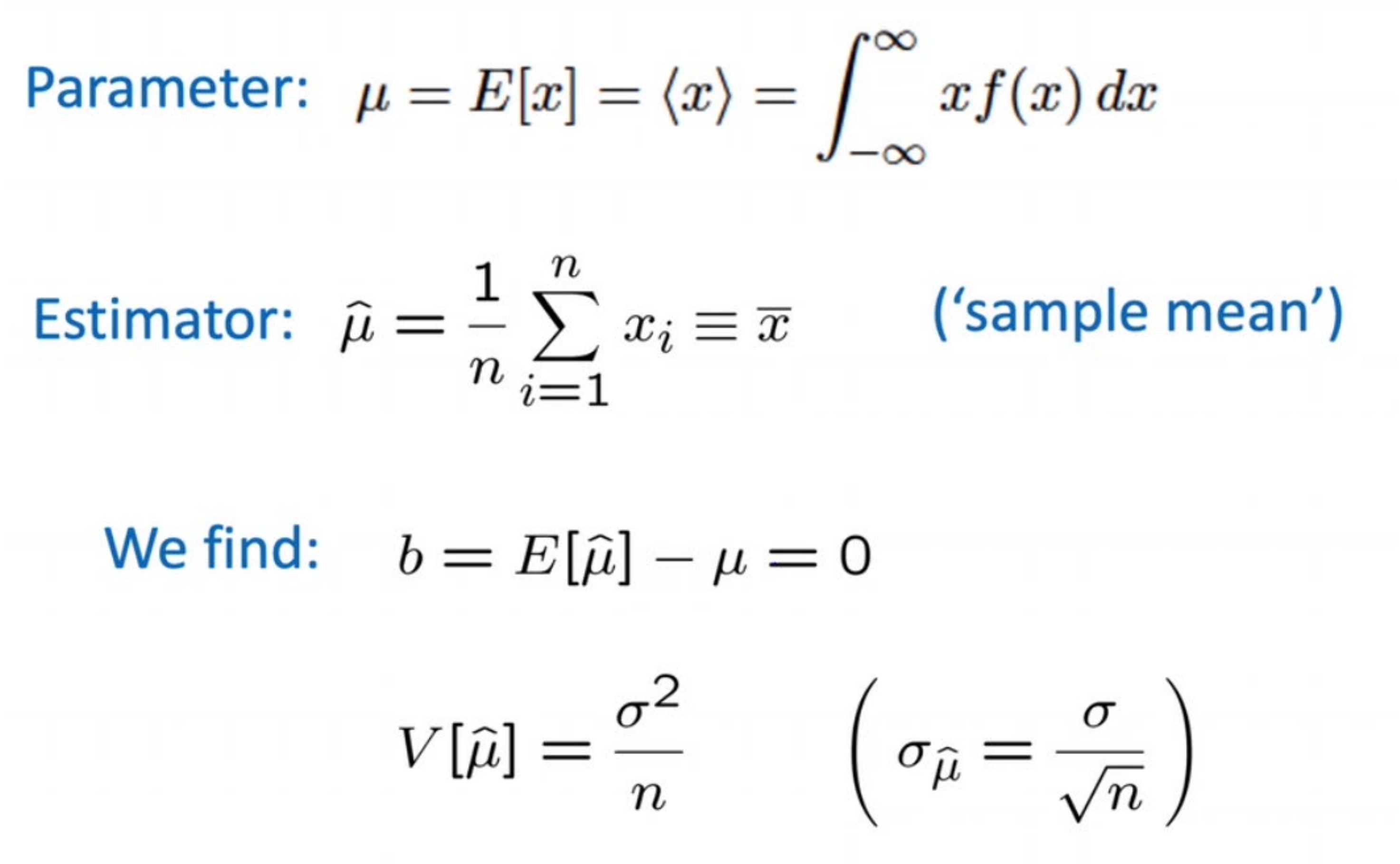
We want small (or zero) bias (systematic error): $b = E[\widehat{\theta}] - \theta$

$\rightarrow$ average of repeated measurements should tend to true value.

And we want a small variance (statistical error): $V[\widehat{\theta}]$

$\rightarrow$ small bias & variance are in general conflicting criteria

Estimator for the mean

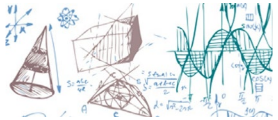


Parameter: $\mu = E[x] = \langle x \rangle = \int_{-\infty}^{\infty} x f(x)\, dx$

Estimator: $\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i \equiv \overline{x}$ ('sample mean')

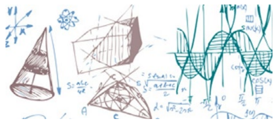
We find: $b = E[\hat{\mu}] - \mu = 0$

$$V[\hat{\mu}] = \frac{\sigma^2}{n} \qquad \left(\sigma_{\hat{\mu}} = \frac{\sigma}{\sqrt{n}} \right)$$



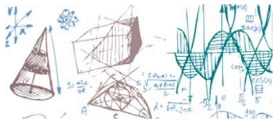
Summary so far

- A generic „algorithm” for point estimation task would be:
- **Collect the data** and understand it
- Come up with a **model**, this will specify a **parameter** or many parameters that we need to make an estimate
- For a given parameter(s) we need an **estimator(s)** (typically we will concentrate on the mean value or variance, however we also can tackle more ambitious cases – e.g., divorces)
- Work out the **estimate of the parameter** – this is a random variable and will be different for different data sets
- Finally, analyse the **sampling distribution of the estimator** to make a judgement of its usefulness
- We are looking for **unbiased** (expectation value) and **efficient** estimators (variance)



Estmator Wish List

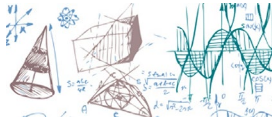
- We are looking for the best estimator (but what does „best” mean?
- In the best of all possible worlds, we could find an estimator $\hat{\mu}$ for which $\hat{\mu} = \mu$ in all samples. But this does not exist, sometimes $\hat{\mu}$ will be too small, fort other samples too big.
- Let’s write (in general): $\hat{\theta} = \theta + error\ of\ estmation$. Therefore the best estimator $\hat{\theta}$:
    - has small estimator errors: the mean squared error RMS $E\left[\left(\hat{\theta} - \theta\right)^2\right]$ shoud be the smallest
    - should be **unbiased** $E\left[(\hat{\theta})\right] = \theta$
    - should have small variance $VAR\left[(\hat{\theta})\right]$
- We are looking for **unbiased** (expectation value) and **efficient** estimators (variance).



Sampling distribution

- Any sample statistics is a function of R.Vs and is therefore itself a random variable – that is absolutely critical to remember!
- The probability distribution of a sample statistics is called **the sampling distribution** of this statistics (sorry for complicated circular sentences…)
  - A recipe to get such distribution would be as follow: we should draw all possible samples of size n from a population, next we should compute the statistics at hand, thus, obtaining the distribution of this statistics. We call it the **sampling distribution**
- It is perfectly ok to compute the mean, variance, standard deviation and other moments for the sampling distribution!
- To make it a bit more comprehensible, let's consider the sample mean. Let $X_1, X_2, \cdots, X_n$ be independent, identically distributed RVs. The mean of the sample is another R.V. defined as follow:

$$\boldsymbol{\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{\sum_{i/1}^{i/n} X_i}{n}}$$



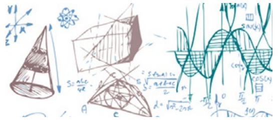
Sampling dist. of means

- **Theorem 3**. If the population is not infinite (of size N) or is the sampling is done without replacement, then the variance should be evaluated using:

$$\sigma'^2_{\bar{X}} = \frac{1}{n}\sigma^2\left(\frac{N-n}{N-1}\right), N \to \infty: \sigma'^2_{\bar{X}} \to \sigma^2_{\bar{X}}$$

- **Theorem 4**. If the population from which we draw samples is normally distributed with mean μ and variance σ^2, then the sample mean is also normally distributed with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\frac{\sigma^2}{n}}$
- **Theorem 5**. Let's assume that the population from which samples are drawn has mean μ and variance σ^2. The population **may or may not be normally distributed**. The standardised variable associated with $\bar{X}$ can be written as:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$



Sampling dist. of means

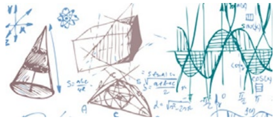
- **Theorem 1**. The mean of the sample means is a consistent etimator of μ:

$$\boxed{E[\bar{X}] = \mu_{\bar{X}} = \mu}$$

where μ is the mean of the population. So, we say, that the expected value of the sample mean is the population mean – **how interesting**!

- **Theorem 2**. If a population is infinite and the sampling is random, or if a population is finite and sampling is with replacement, then the variance of the distributions of the sample means, denoted by $\sigma_{\bar{X}}$, is:

$$\boxed{E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \frac{1}{n}\sigma^2}$$



Estimator for sample variance

- If $\{X_1, X_2, \cdots X_n\}$ denote R.Vs for a random sample of size n, the R.V. giving the variance of the sample (the sample variance) is defined as:

$$S^2 = \frac{\mathbf{1}}{\boldsymbol{n}}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]$$

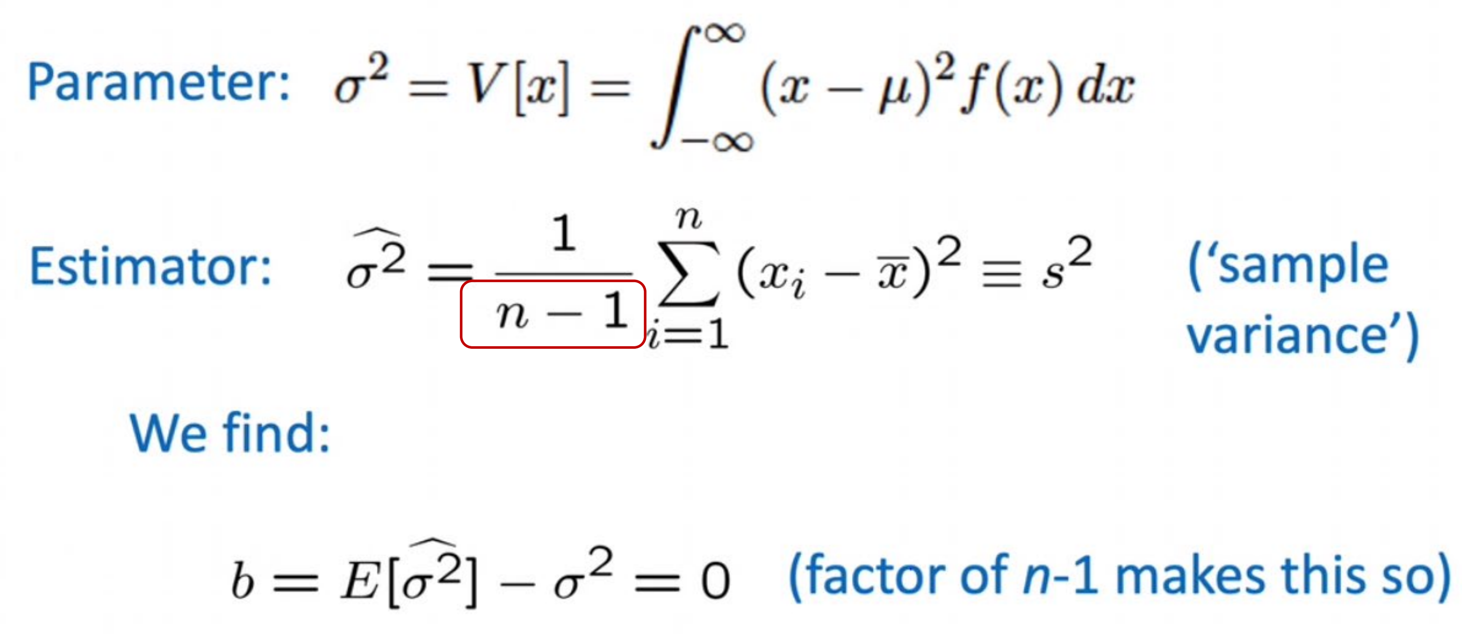
- We already know, that $E[\bar{X}] = \mu$, is this the same for $E[S^2] = \sigma^2$?
    - ✓ A little digression – whenever the expected value of a statistics **is equal** to the corresponding **population parameter**, we call this statistics **an unbiased estimator**. Its value is then an unbiased estimate of the respective parameter
- Unfortunately, it can be proved that for the sample variance, we have:

$$E[S^2] = \mu_{S^2} = \frac{n-1}{n}\sigma^2$$

- However, an unbiased variance estimator is easy to find:

$$\boxed{\hat{S}^2 = \frac{n}{n-1}S^2 = \frac{\mathbf{1}}{\boldsymbol{n}-\mathbf{1}}[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2]}$$

Estimator for the variance

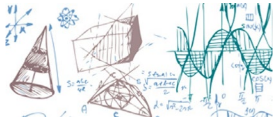


Parameter: $$\sigma^2 = V[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$$

Estimator: $$\widehat{\sigma^2} = \frac{1}{\boxed{n-1}} \sum_{i=1}^{n} (x_i - \overline{x})^2 \equiv s^2$$ ('sample variance')

We find:

$$b = E[\widehat{\sigma^2}] - \sigma^2 = 0$$ (factor of *n*-1 makes this so)



Point estimators - summary

- Sample mean $\bar{X}$ is the point estimator of parameter μ:

$$\boxed{\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} = \frac{1}{n}\sum_{i=1..n} X_i}$$

- The unbiased estimator for variance is:

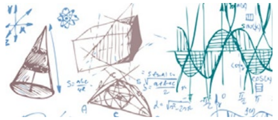
$$\boxed{\begin{gathered}\hat{S}^2 = \frac{1}{n-1}\left[(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2\right] = \\ \frac{1}{n-1}\sum (X_i - \bar{X})^2\end{gathered}}$$

- The estimator of the correlation (X, Y) is:

$$\boxed{r(X,Y) = \frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{yy}}}}$$

$$S_{XX} = \sum (X_i - \bar{X})^2 \qquad S_{YY} = \sum (Y_i - \bar{Y})^2$$

$$S_{XY} = \sum (X_i - \bar{X})(Y_i - \bar{Y})$$



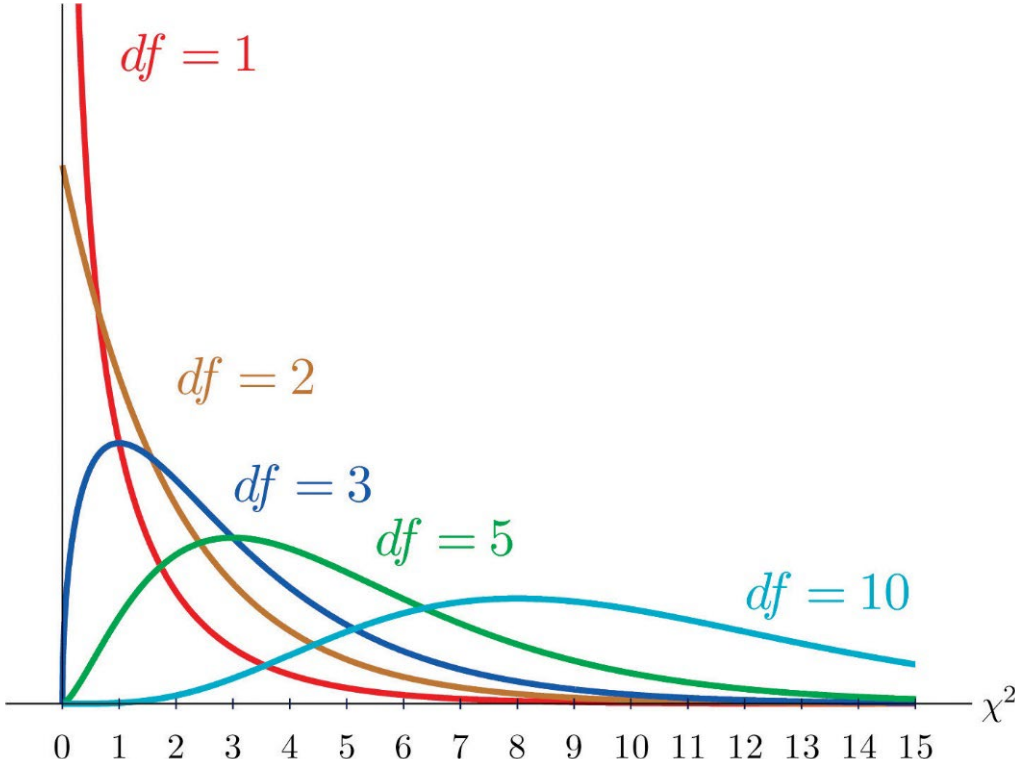
Sampling dist. of variances

- In order to create the sampling distribution of variances, we take all the possible samples of size n, that can be drawn from a population and calculate their variances
- One change is, that instead of looking directly at the distribution of the sample variance, we look at the R.V.:

$$\frac{nS^2}{\sigma^2} = \frac{(n-1)\hat{S}^2}{\sigma^2} = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{\sigma^2}$$

- **Theorem 6**. If a random samples of size n are taken from a population having a normal distribution, than the sampling variable $\frac{nS^2}{\sigma^2}$ has a χ^2 distribution with $n-1$ degrees of freedom

χ^2 distribution



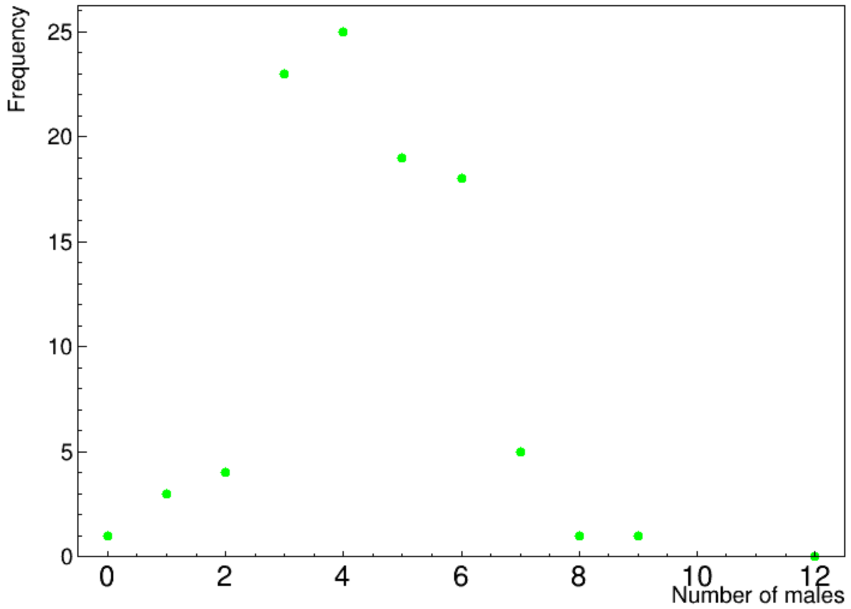
- This is another very popular distribution in Statistics!
- The mathematical formula describing it is quite complex, again we are going to use tabulated values when solving problems!

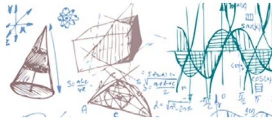
Number of males in a queue

- An experiment has been conducted in London Tube to check the number of males in each of 100 queues all of length 10. The results obtained were as follows

Counts	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Frequency	1	3	4	23	25	19	18	5	1	1	0

- And the plot



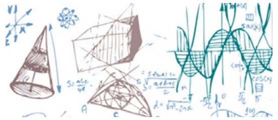


Number of males in a queue

- Can you tell what is the underlaying **parent distribution**?
- Well, one could prove that the **binominal** one fits quite good $\mathcal{B}(n, p)$, $n = 10$ being the length of the queue and $\boldsymbol{p}$ the proportion of males (check this on your own)
- We could estimate the p using the collected sample

$$\frac{\#males}{\#all\ passangers} = \frac{1 \cdot 0 + 3 \cdot 1 + \cdots + 1 \cdot 9 + 0 \cdot 10}{1000} = \frac{435}{1000} = \mathbf{0.435}$$

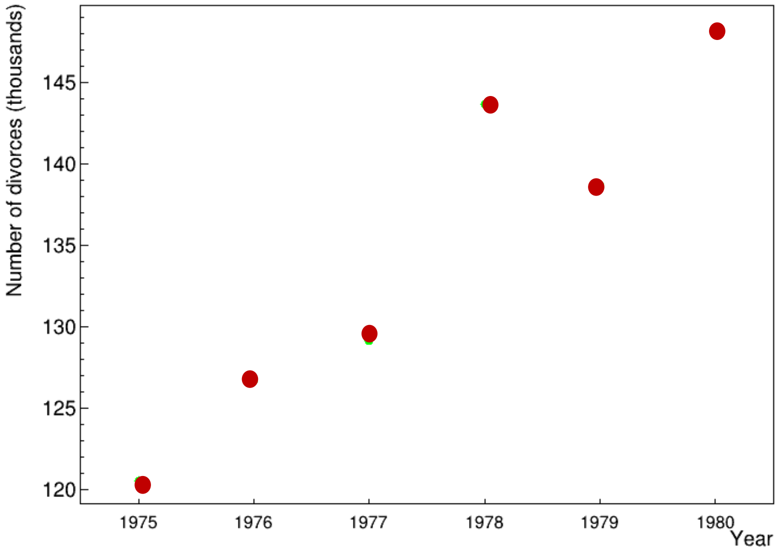
- **What would be the weak point of this assumption**?
- Can we actually come up with a generic strategy to say, the value of a parameter of interest is this and that?
- Yes! We can! We need to perform an experiment and run an analysis
- Another question would be how reliable this estimate is (but we leave it for the next lectures)



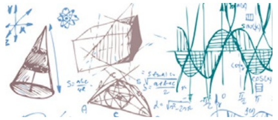
More than one way…

- Lets inspect the following data regarding the number of divorces in different years in some country in Europe

Year	1975	1976	1977	1978	1979	1980
# divorces (10^3)	120.5	126.7	129.1	143.7	138.7	148.3

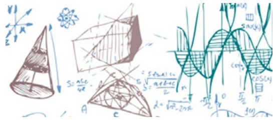


- Interesting…, very tempting to fit a model right away.



More than one way…

- From the plot we could conclude, that the **true underlaying distribution** describing the data can be represented by a **linear model**
- From the data we also conclude that **the slope** of the line is positive – ok, the task is then to **estimate this slope**, α, and then we could predict the annual rate of increase of divorces
- But how do we do that? It is not so obvious like the water example(??)
- Consider this:
  - $\hat{\alpha}_1$- join the first and the last point
  - $\hat{\alpha}_2$ - join the mid-points P_1P_2 and P_5P_6
  - $\hat{\alpha}_3$ - join the centroid of the first triplet and the second one
- Mind you, these are all sensible options!



The best parameters

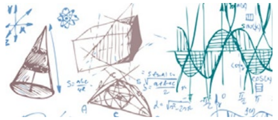
- Let’s do the calculations explicitly

$$\hat{\alpha}_1 = \frac{148.3 - 120.5}{80 - 75} = \frac{27.8}{5} = 5.6$$

$$\hat{\alpha}_2 = \frac{(138.7 + 148.3)/2 - (120.5 + 126.7)/2}{79.5 - 75.5} = \frac{19.9}{4} = 5.0$$

$$\hat{\alpha}_3 = \frac{(143.7 + 138.7 + 148.3)/3 - \cdots}{79 - 76} = \frac{18.13}{3} = 6.0$$

- So, is there a way to make a judgement on **which one of these estimates is the „best”**? **And what exactly the best means**?
- The second question actually pertains to **the estimator properties** and not the estimate (a number we obtained)
- So, we need to look at the properties of the **sampling distribution** of respective estimators!



Analysis

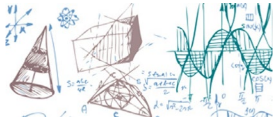
- Mind one thing. This example is not quite what we could call an experiment – we cannot repeat year 1976 and check the number of divorces again…
- However, we can still evaluate the deviations of data from the predicted model and treat them as random variable
- Say, the difference (**residual**) r is defined as follow:

$$r_i = y_i - (\beta + \alpha x_i)$$

- Next, we assume that the residual (random variable) will have a mean value and variance: μ_r and σ_r^2
- Further, we assume that the residual should have mean value equal to 0 (think about that!), so finally our model for the given data set is:

$$Y_i = \beta + \alpha x_i + r_i$$

- We have therefore three parameters $(\alpha, \beta, \sigma_r^2)$



Analysis

- Having formulated the model we can now start discussing the properties of the sampling distributions of our estimators
- So, we are going to treat **the estimate** (a number evaluated using the data):

$$\hat{\alpha}_1 = \frac{y_6 - y_1}{x_6 - x_1}$$

- … as a single measurement (observation) of the random variable (the estimator)

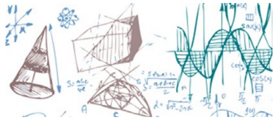
$$\hat{\alpha}_1 = \frac{Y_6 - Y_1}{x_6 - x_1}$$

- To come up with the answer regarding how good is such estimator we start from working out its mean and variance (we use the knowledge of these function of R.V. remembering that α, β, x_i are just constant numbers)

$$E[Y_i] = \beta + \alpha x_i$$

$$V[Y_i] = \sigma_r^2$$

$$E[r_i] = 0, V[r_i] = \sigma_r^2$$



Analysis

- Now, we can calculate the expected value of $\hat{\alpha}_1$:

$$E[\hat{\alpha}_1] = E\left[\frac{Y_6 - Y_1}{x_6 - x_1}\right] = \frac{1}{x_6 - x_1} E[Y_6 - Y_1] =$$

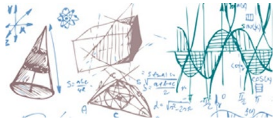
$$= \frac{1}{x_6 - x_1}\left(E[Y_6] - E[Y_1]\right) = \frac{1}{x_6 - x_1}\left((\beta + \alpha x_6) - (\beta + \alpha x_1)\right)$$

$$E[\hat{\alpha}_1] = \frac{1}{x_6 - x_1}(\alpha x_6 - \alpha x_1) = \alpha$$

- Neat! The expected value of the estimator is exactly equal to the unknown parameter. Good job
- What about the other estimators?

$$\hat{\alpha}_2 = \frac{\frac{1}{2}(Y_5 + Y_6) - \frac{1}{2}(Y_1 + Y_2)}{\frac{1}{2}(x_5 + x_6) - \frac{1}{2}(x_1 + x_2)} \qquad \hat{\alpha}_3 = ?$$

Analysis



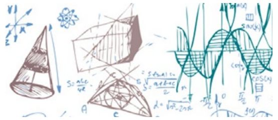
- Repeating the same calculations for remaining two estimators we conclude that their expected values are always the same and equal exactly the unknown parameter we estimating
- In general we say that an estimator $\hat{\theta}$, which we use to estimate an unknown parameter of a model, is **unbiased for parameter $\boldsymbol{\theta}$** if the following is true:

$$E\left[\hat{\theta}\right] = \theta$$

- So, it seems that all of them doing just fine. What next we can check…? The variance!
- In this case the best option would be to choose the one that features the least variability about its mean value, so:

$$V[\hat{\alpha}_1] = V\left[\frac{Y_6 - Y_1}{x_6 - x_1}\right] = \frac{1}{(x_6 - x_1)^2}(V[Y_6] + V[-Y_1]) =$$

$$= \frac{1}{(x_6 - x_1)^2}(\sigma_r^2 + \sigma_r^2) = \frac{2\sigma_r^2}{(x_6 - x_1)^2} = \frac{2\sigma_r^2}{25}$$

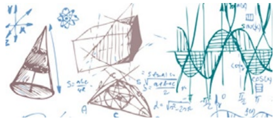


Analysis

- Again, we can repeat the calculations for the remaining two estimators (you are encuraged to do so!)
- We get the following:

$$V[\hat{\alpha}_2] = \frac{4\sigma_r^2}{64} \qquad V[\hat{\alpha}_3] = \frac{6\sigma_r^2}{81}$$

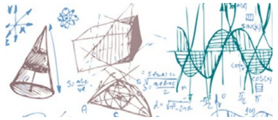
- So: $V[\hat{\alpha}_2] < V[\hat{\alpha}_3] < V[\hat{\alpha}_1]$
- Using the variance we say that the best (most efficient) estimator is the $\hat{\alpha}_2$ - thus we have the winner!



Summary so far

- A generic „algorithm” for point estimation task would be:

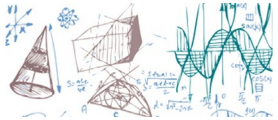
- **Collect the data** and understand it
- Come up with a **model**, this will specify a **parameter** of many parameters that we need to estimate
- For a given parameter we need an **estimator** (typically we will concentrate on the mean value or variance, however we also can tackle more ambitious cases – divorces)
- Work out the **estimate of the parameter** – this is a random variable and will be different for different data sets
- Finally, analyse the **sampling distribution of the estimator** to make a judgement of its usefulness
- We are looking for **unbiased** (expectation value) and **efficient** estimators (variance)



Can we do better?

- That was fun! And we learned a lot, however following such generic path each time we need to make an estimate **seems too much**
- We need a technique(s) that allows us **to define sensible estimators** (again, we could spend a lifetime on deriving estimators that are reasonable)
- So, such a technique would „automatically” **come up with a formula for best estimators**

- One thing to remember – there is no universal method to achieve the above task, in time a number of approaches have been proposed. <u>There is no „best” one</u>
- We concentrate on three techniques: the method **of least squares**, the method of **moments** and the method of **maximum likelihood**



The method of least squares

- Say, we are interested in estimating the mean value of some distribution which is θ. We take data sample: $\{X_1, X_2, \dots, X_n\}$
- Since the mean is „a typical" value, we conclude that the respective differences $X_i - \theta$ should be „small" (simultaneously)
- Also, the sum of squares of these differences should be as small as possible:

$$\boxed{S = \sum_i (X_i - \theta)^2 \rightarrow min}$$

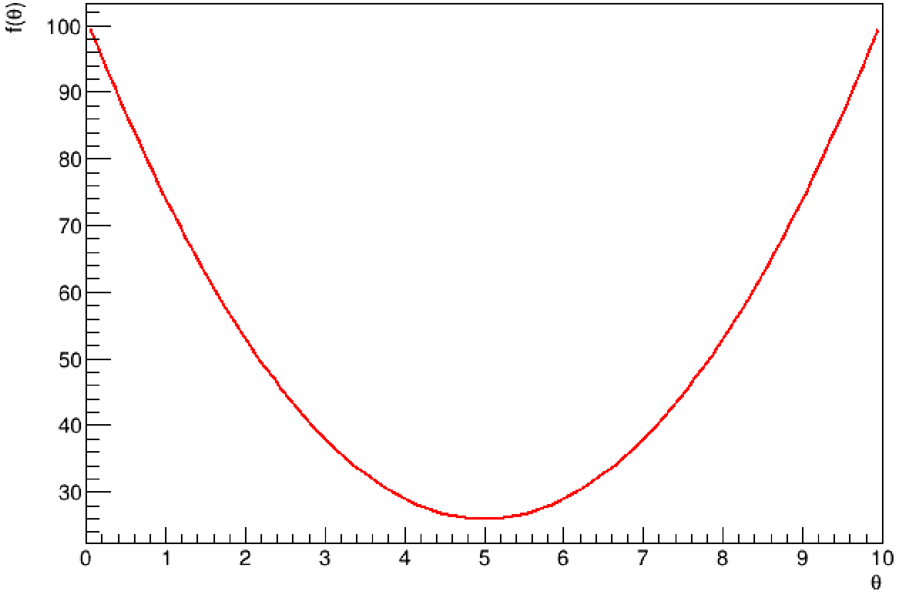
- This is the method of least squares (MLS)
- As usual, an **example** is in order! Say we collected sample: $X = \{2,4,9\}$ and we want to estimate the mean value of the parent distribution these numbers came from. Applying the MLS

$$S_3 = \sum_{i/1}^{i/3} (X_i - \theta)^2 = (2-\theta)^2 + (4-\theta)^2 + (9-\theta)^2 = \cdots =$$

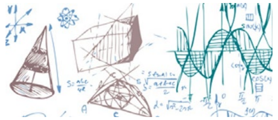
$$= 101 - 30\theta + 3\theta^2$$

The method of least squares

- This function will take different values for different θ – task is to identify the θ for which the sum is the smallest



- The minimum is attained at the point $\theta = 5$. We can also check that the sample mean, $\bar{x} = 5$. Reasonable!
- So, the method gives results compatible with the common sense, that is encouraging!



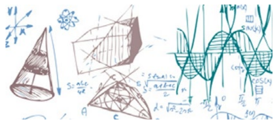
The method of least squares

- In a general case where we have a random sample of size n drawn from a population with an unknown mean value:

$$S_n = \sum_{i/1}^{i/n} (X_i - \theta)^2 = \sum_{i/1}^{i/n} \left(X_i^2 - 2\theta X_i + \theta^2\right) =$$

$$= \sum_{i/1}^{i/n} X_i^2 + 2\theta \sum_{i/1}^{i/n} X_i + n\theta^2$$

- This formula is minimised by: $\theta = \frac{1}{n}\sum_{i/1}^{i/n} X_i = \bar{X}$ - the sample mean
- So, the sample mean $\bar{x}$ of a random sample X taken from a population of an unknown mean value (we call is here in a generic way θ) is the **least squares estimator** $\hat{\theta}$ (or $\hat{\mu}$)
- The language here is important, so we are precise about what we mean!



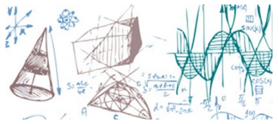
The method of moments

- We use the common sense when introducing this approach
- We introduced the notions of population moments and sample moments, the former are unknown and the latter are calculated using data samples
- The **method of moments** (MoM) uses the **sample moments** and match them to the analogous **population moments** to obtain **estimates for the unknown parameters**. Seems simple…
- For instance we have easy example: a normal distribution $\mathcal{N}(\mu, \sigma^2)$, we use sample mean and variance:

$$\hat{\mu} = \bar{x}, \hat{\sigma}^2 = s^2$$

What about the median?

- **Not all cases are so simple**, for instance what about the Poisson distribution? Both the population mean and its variance are equal μ. Shall we use the sample mean or variance as the best estimator $\hat{\mu}$?
- Need to understand the sampling distributions to answer this!

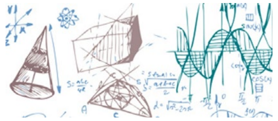


The method of moments

- Let’s write down explicitly moment estimators for a few most popular models we discussed so far

- The **Poisson**: one unknown parameter – the mean of the distribution. Matching sample and population moments gives the following estimate: $\hat{\mu} = \bar{x}$, the corresponding estimator we should use: $\hat{\mu} = \bar{X}$
- The **exponential**: the mean is $\frac{1}{\lambda}$, matching moments $\bar{X} = \frac{1}{\hat{\lambda}}$. So, we get: $\hat{\lambda} = \frac{1}{\bar{X}}$
- The binomial $\mathcal{B}(m, p)$: we have one unknown parameter p. The matching procedure gives: $\bar{X} = m\hat{p}$ (n is the sample size)

$$\hat{p} = \frac{\bar{X}}{m} = \frac{X_1 + X_2 + \cdots + X_n}{mn}$$



The method of moments

- In practice, we are going to observe some fluctuations, let’s consider the following: we used a generator of random numbers distributed according to the Poisson model. We draw two samples:

$$X_{(1)} = \{5,5,2,3,4,6,4,1\}, X_{(2)} = \{4,2,5,2,4,1,1,1\}$$

$$\hat{\mu} = \bar{x}_{(1)} = \frac{30}{8} = 3.75$$

$$\hat{\mu} = \bar{x}_{(2)} = \frac{20}{8} = 2.50$$

- These variations are „normal”, we are going to observe them
- The point is that (SV – sample variance):

$$E[\bar{X}] = \mu, SV[\bar{X}] = \sqrt{\frac{\sigma^2}{n}}$$

we will continue the discussion how to obtain the best estimators in a few weeks!