Probability and Statistics

DR. AHMED TAYEL

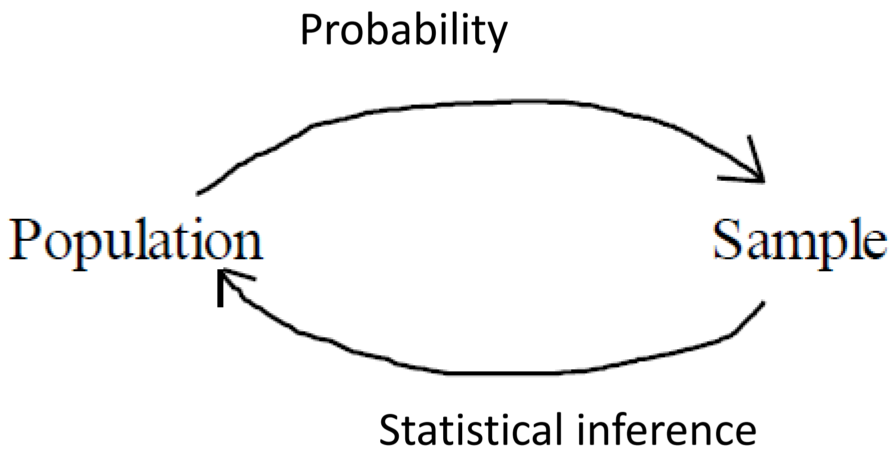
Department of Engineering Mathematics and Physics, Faculty of Engineering, Alexandria University

ahmed.tayel@alexu.edu.eg

Statistical Inference

Probability theory: the probability distribution of the population is known; we want to derive results about the probability of one or more values ("random sample") - *deduction*.

Statistics: the results of the random sample are known; we want to determine something about the probability distribution of the population - ***inference***.



Statistical Inference

In order to carry out **valid inference**, the sample must be **representative**, and preferably a **random** sample.

Random sample:

(i) **no bias** in the selection of the sample;

(ii) different members of the sample chosen **independently**.

(iii) large enough relative to the population size.

Statistical Inference

- The observed values are denoted as $\boldsymbol{x_1}, \boldsymbol{x_2}, \ldots, \boldsymbol{x_n}$.
- A "statistic" is any function of the observations in a random sample.
  - Sample mean.
  - Sample standard deviation.
  - Correlation functions.

All statistics help in *statistical inference* about the unknown population/parameters

Case study of
statistical inference

Fitting a distribution

Fitting a distribution

Steps for fitting a distribution

Hypothesizing Families of Distributions

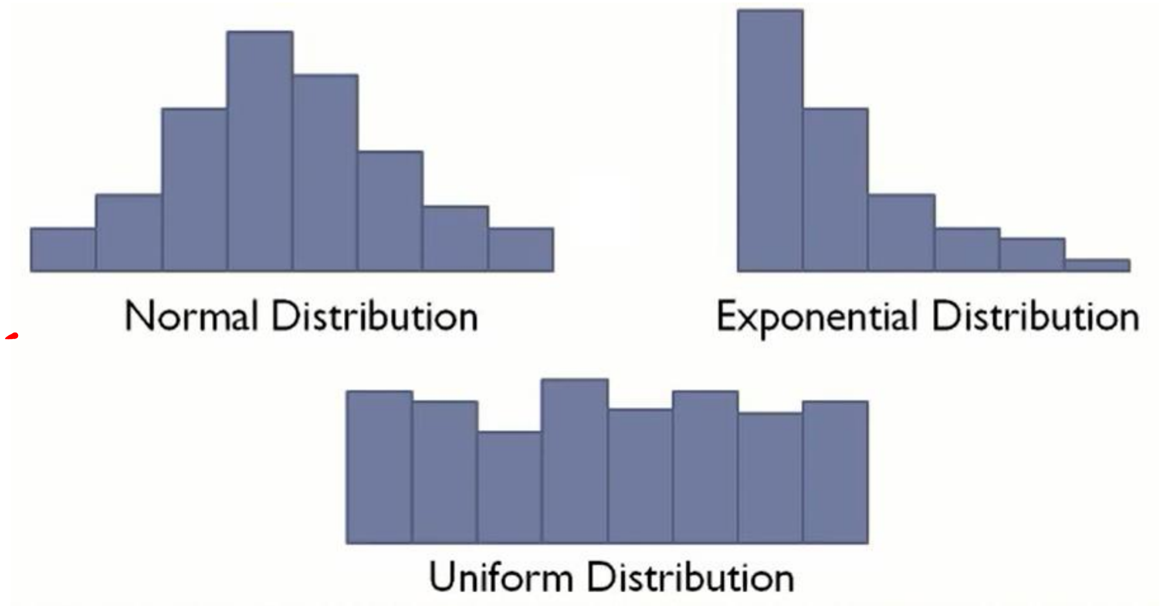
Parameters Estimation

Assessment of Goodness-of-fit

Fitting a distribution

Step 1/3: Hypothesizing Families of Distributions

- The first step in selecting a particular input distribution is to decide what general families (**exponential, normal, Uniform,…**) appear to be appropriate on the basis of their shapes, without yet worrying about the specific parameter values for these families.



* looks like a previously studied distn.

or * given a table of families of distns.

	Name	$E(X)$	$V(X)$

Fitting a distribution

Step 1/3: Hypothesizing Families of Distributions

Other useful indicators

- If **mean** $\approx$ **median** → Symmetric distribution (like the Normal distribution)
- If the **skewness factor** is $\boldsymbol{\upsilon}$ **= 0** → Symmetric distribution
- If $\boldsymbol{\upsilon}$ **> 0** → Positively skewed (like the **exponential** distribution)
- If the coefficient of variation $\boldsymbol{cv}$ **= 1** → **Exponential** (because mean = standard deviation = $1/\lambda$)

Recall: $\boldsymbol{cv = \frac{s(n)}{\bar{x}(n)}} = 1$

$s = \bar{x}$

$\sigma = \mu$

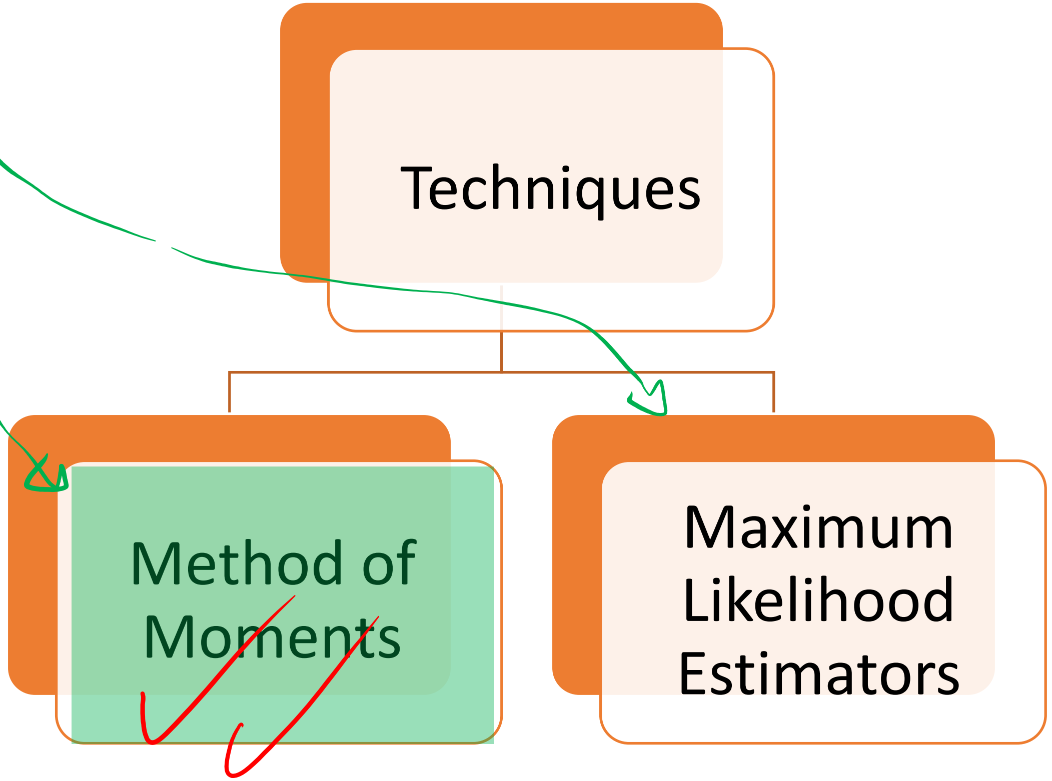
Fitting a distribution

Step 2/3: Parameter Estimation

- Once we've decided the type of **distribution**, we need to estimate its parameters.

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n}$$



The method of moments

in terms of the unknown parameters

ex: $\exp(\lambda) \dashrightarrow E(X) = 1/\lambda = \mu_1$ unknown

Definition:

Population and sample moments

Known

by equating $\mu_1 = m_1$

$\frac{1}{\lambda} = \bar{X} \Rightarrow \hat{\lambda} = \frac{1}{\bar{X}}$

hat "estimated"

The k^{th} **population moment** of a RV (about the origin) is

Given directly

previously studied distn.

Given pdf/pmf

Given MGF

Given E(X) & V(X)

$$\mu_k = E[X^k]$$

$$= \int_{-\infty}^{\infty} x^k\, f(x;\theta)\ dx \quad \text{"Continuous RV"}$$

$$= \sum_{x=-\infty}^{\infty} x^k\, f(x;\theta) \quad \text{"Discrete RV"}$$

in terms of unknown parameters

The k^{th} **sample moment** is

$$m_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

"known"

1st sample moment $m_1 = \frac{1}{n}\sum X_i = \bar{X}$

2nd sample moment $m_2 = \frac{1}{n}\sum X_i^2$

⋮

The method of moments

The **Method of Moments** (MoM) consists of equating sample moments and population moments. If a population has t parameters, the MOM consists of solving the system of equations

$$m_k = \mu_k, \; k = 1, 2, \ldots, t$$

for the t parameters.

If one parameter is unknown

$$\mu_1 = m_1$$

One equation in one unknown

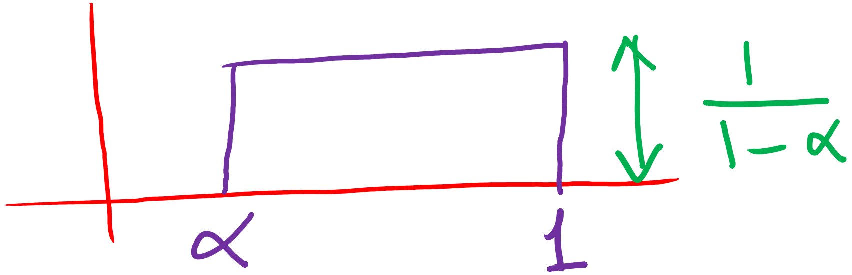
If two parameters are unknown

$$\mu_1 = m_1$$
$$\mu_2 = m_2$$

Two equations in two unknown

…

The method of moments



Example

Let **X** be uniformly distributed on the interval (α, 1). Given a random sample of size n, use the method of moments to obtain a formula for estimating the parameter α.

#

Solution

one unknown parameter

→ requires one eqn.

$\mu_1 = m_1$

The first population moment

$$\mu_1 = \frac{1+\alpha}{2}$$

The first sample moment

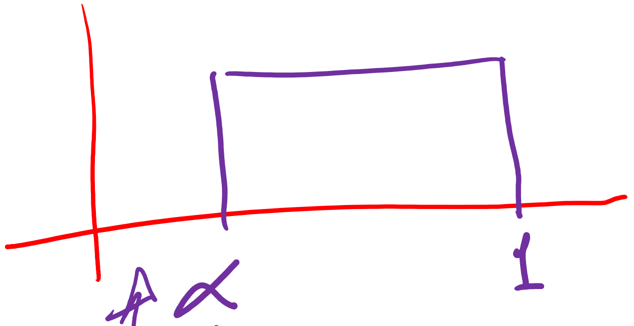
$$m_1 = \bar{X} = \frac{1}{n}\sum x_i$$

Using MME

$$m_1 = \mu_1 \Rightarrow \bar{X} = \frac{1+\hat{\alpha}}{2} \Rightarrow \hat{\alpha} = 2\bar{X} - 1$$

$$\hat{\alpha} = 2\bar{X} - 1$$

sub.



0.6

$$\hat{\alpha} = 2(0.8) - 1$$
$$= 1.6 - 1$$
$$= 0.6 \quad \leftarrow \text{MME}$$

$\bar{X} = 0.8$

0.97
0.95
0.91
0.88
0.85
0.83
0.7
0.4
0.1

⇒ Anothe better estimator $\hat{\alpha} = \min(X_i)$

using max. liklihood "MLE"

<u>Exercise</u> Assume Unif(α, β)

Both are unknowns

$\rightarrow E(X^2) = \int_{\alpha}^{\beta} x^2 \frac{1}{\beta - \alpha}\, dx$ or using $V(X) = 2^{nd}$ mom. $- (1^{st}\text{ mom})^2$

The method of moments

Example

Given a random sample of size n from a Poisson population, use the method of moments to obtain a formula for estimating the parameter λ. ← one unknown

→ one eqn.

Solution

The first population moment

$$\mu_1 = E(x) = \lambda$$

The first sample moment

$$m_1 = \bar{X}$$

Using MME

$$m_1 = \mu_1 \Rightarrow \quad \bar{X} = \hat{\lambda}$$

The method of moments

Example

Given a random sample of size n from a $\mathbf{N(\mu, \sigma^2)}$ population, use the method of moments to obtain formulas for estimating the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$.

Solution

Hint: $M_X(t) = e^{t\,\mu + \frac{1}{2} t^2\,\sigma^2}$

What is this function?

Definition: The moment generating function (MGF)

Definition

$$M_X(t) = E(e^{tX})$$

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0}$$

$$E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0}$$

$$E(X^n) = \left[\frac{d^n}{dt^n} M_X(t)\right]_{t=0}$$

MGF for important RV's

$Binomial(n,p)$	$(1-p+pe^t)^n$
$Geometric(p)$	$\dfrac{pe^t}{1-(1-p)e^t}$
$Poisson(\lambda)$	$e^{\lambda(e^t-1)}$
$Uniform(a,b)$	$\dfrac{e^{bt}-e^{at}}{t(b-a)}$
$Exponential(\lambda)$	$\dfrac{\lambda}{\lambda-t}$
$N(\mu,\sigma^2)$	$e^{t\mu+\frac{1}{2}t^2\sigma^2}$

The method of moments

2 unknowns → require 2 eqs. → up to 2nd moment

Example

Given a random sample of size n from a $\mathbf{N(\mu, \sigma^2)}$ population, use the method of moments to obtain formulas for estimating the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$.

Solution

Hint: $M_X(t) = e^{t\,\mu + \frac{1}{2} t^2 \sigma^2}$

The first two population moments

$$\boldsymbol{E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0}}$$

$$\boldsymbol{E(X) = \left[e^{t\,\mu + \frac{1}{2} t^2 \sigma^2}\left(\mu + t\,\sigma^2\right)\right]_{t=0}}$$

Product

$$\boldsymbol{E(X) = \mu}$$

d/dt

$$\boldsymbol{E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0}}$$

$$\boldsymbol{E(X) = \left[e^{t\,\mu + \frac{1}{2} t^2 \sigma^2}\sigma^2 + e^{t\,\mu + \frac{1}{2} t^2 \sigma^2}\left(\mu + t\,\sigma^2\right)^2\right]_{t=0}}$$

$$\boldsymbol{E(X^2) = \sigma^2 + \mu^2}$$

or using $E(X)$ & $V(X)$

For Normal $\boxed{E(X) = \mu}$ & $V(X) = \sigma^2$

$$= E(X^2) - [E(X)]^2$$

$$\sigma^2 = E(X^2) - \mu^2$$

$$\boxed{E(X^2) = \sigma^2 + \mu^2}$$

The method of moments

Example (Continued)

Given a random sample of size n from a $\mathbf{N(\mu, \sigma^2)}$ population, use the method of moments to obtain formulas for estimating the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\sigma^2}$.

Solution

Hint: $M_X(t) = e^{t\mu + \frac{1}{2}t^2\sigma^2}$

The first two population moments

$$\mu_1 = E(x) = \mu, \qquad \mu_2 = E(x^2) = \sigma^2 + \mu^2$$

$=$ $=$

The first two sample moments

$$m_1 = \overline{X} \quad \text{and} \quad m_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$s^2 = \frac{\Sigma(\ )^2}{n}$

Using MME

$$m_1 = \mu_1 \Rightarrow \overline{X} = \hat{\mu} \qquad \hat{\mu} = \overline{X}$$

$\overline{X}^2$

End of soln.

$$m_2 = \mu_2 \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \hat{\sigma}^2 + \hat{\mu}^2 \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum (X_i - \overline{X})^2$$

- <u>Example</u>: Let X be a random variable having a gamma distribution

$$f(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)} \quad , \quad x \geq 0 \, , \alpha > 0 \, , \beta > 0$$

with unknown parameters α and β . Find the moment estimators of the unknown parameters. 2 unknowns

(Hint: $\mathbf{E(X) = \alpha\,\beta}$ **and** $\mathbf{Var(X) = \alpha\,\beta^2}$)

<u>Note:</u> $\boldsymbol{Var(X) = E\left(X^2\right) - \mu^2}$

$$\alpha\beta^2 = E(x^2) - \alpha^2\beta^2$$

$$E(x^2) = \alpha\beta^2 + \alpha^2\beta^2$$

$$\mu_1 = E(X) = \alpha\beta$$

$=$

$$m_1 = \bar{X}$$

$$\Downarrow$$

$$\alpha\beta = \bar{X}$$

$$\Downarrow$$

$$\hat{\alpha} = \frac{\bar{X}}{\hat{\beta}}$$

$$\boxed{\hat{\alpha} = \frac{\bar{X}^2}{\frac{1}{n}\sum X_i^2 - \bar{X}^2}}$$

$$\mu_2 = E(X^2) = \alpha\beta^2 + \alpha^2\beta^2$$

$=$

$$m_2 = \frac{1}{n}\sum X_i^2$$

$$\Downarrow$$

$$\alpha\beta^2 + \alpha^2\beta^2 = \frac{1}{n}\sum X_i^2$$

sub.

$$\Downarrow$$

$$\beta(\alpha\beta) + (\alpha\beta)^2 = \frac{1}{n}\sum X_i^2$$

$$\hat{\beta}\bar{X} + \bar{X}^2 = \frac{1}{n}\sum X_i^2$$

sub.

$$\boxed{\hat{\beta} = \frac{\frac{1}{n}\sum X_i^2 - \bar{X}^2}{\bar{X}}}$$

Fitting a distribution

Step 3/3: Goodness of fit

- We have the **distribution** and its **parameters**.
- **How do we know if it's a good fit?**
- There are many methods, we will discuss:

→The P-P plot

"Probability-Probability" plot

- A Probability-Probability (P-P) plot is a graphical method for determining whether sampled data conform to a hypothesized distribution or not, based on a subjective **visual examination** of the data.
- It is **more reliable than the histogram** for small to moderate data sample sizes.

P – P plot

Reference CDF

does not change as the no. of observas. is fixed

Median = 50% percentile $= x_{0.5} = F^{-1}(0.5)$

Steps

1st percentiles 25%

= 2nd ~ 50%

median

3rd ~ 75%

- Arrange the observations (data) in ascending order.
- Calculate the observed (empirical) CDF as follows:

order of observations

Note difference betn. population & sample

ranges from 1 to n

$$\tilde{F}(x_{(i)}) = \frac{i - 0.5}{n}$$

Empirical CDF

$\frac{1-0.5}{n>>>}$ → almost zero

$\frac{n-0.5}{n>>}$ ∴ near 1

All based on order of obss. not their values

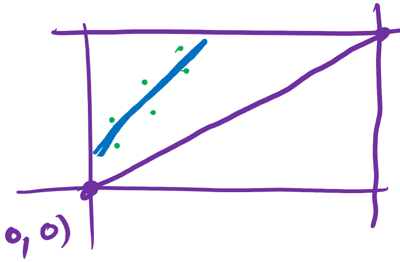
- Calculate the **theoretical CDF** using the **estimated parameters** and distribution.

↳ based on the observations values

≠ CDF

- Plot them together.
- The closer it is to a straight line (of **unit slope** and **zero intercept**), the better.

P – P plot



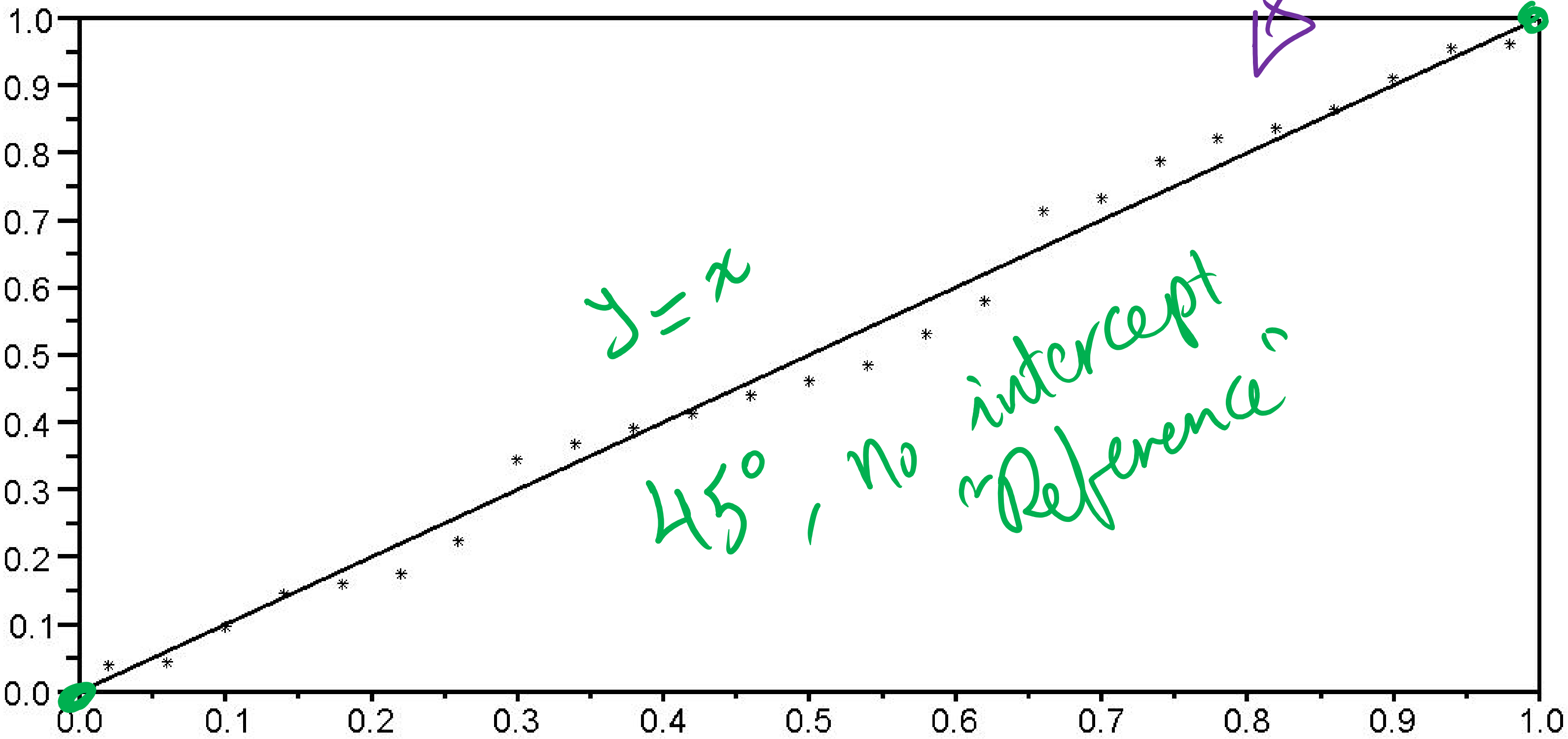
Not a good fit

Good fit

Theoretical cdf fitted

Based on observations values

$F(x_{(j)})$



$y = x$

45°, no intercept "Reference"

(1,1)

(0,0)

$\tilde{F}(x_{(j)})$

Empirical cdf

"Reference"

Based on the order of points

Example: Suppose the following data is given

$\hat{\lambda} = \frac{1}{\bar{x}} =$ ✓

(5.03 , 7.73 , 11.1 , 11.96 , 16.07 , 22.38 , 31.5 , 67.4)

Examine the goodness of an exponential fit to the given data set using a P-P plot.

Theritical CDF

Data in ascending order

Empirical CDF

i	x_i	Empirical cdf	Exponential cdf
1	5.03	0.0625	0.2074
2	7.73	0.1875	0.3003
3	11.1	0.3125	0.4012
4	11.96	0.4375	0.4245
5	16.07	0.5625	0.52404
6	22.38	0.6875	0.6444
7	31.5	0.8125	0.7667
8	67.4	0.9375	0.9556

x-axis   y-axis

Mode → 7: TABLE → Enter f(x)
→ Start, end, step
1, 8, 1

Empirical CDF

$$\tilde{F}(x_{(i)}) = \frac{i - 0.5}{n}$$

← 8

Theoretical CDF

$$\tilde{F}(x_{(i)}) = 1 - e^{-\hat{\lambda}\, x_{(i)}}$$

Mode → 1: COMP → Enter f(x)
→ CALC → x?

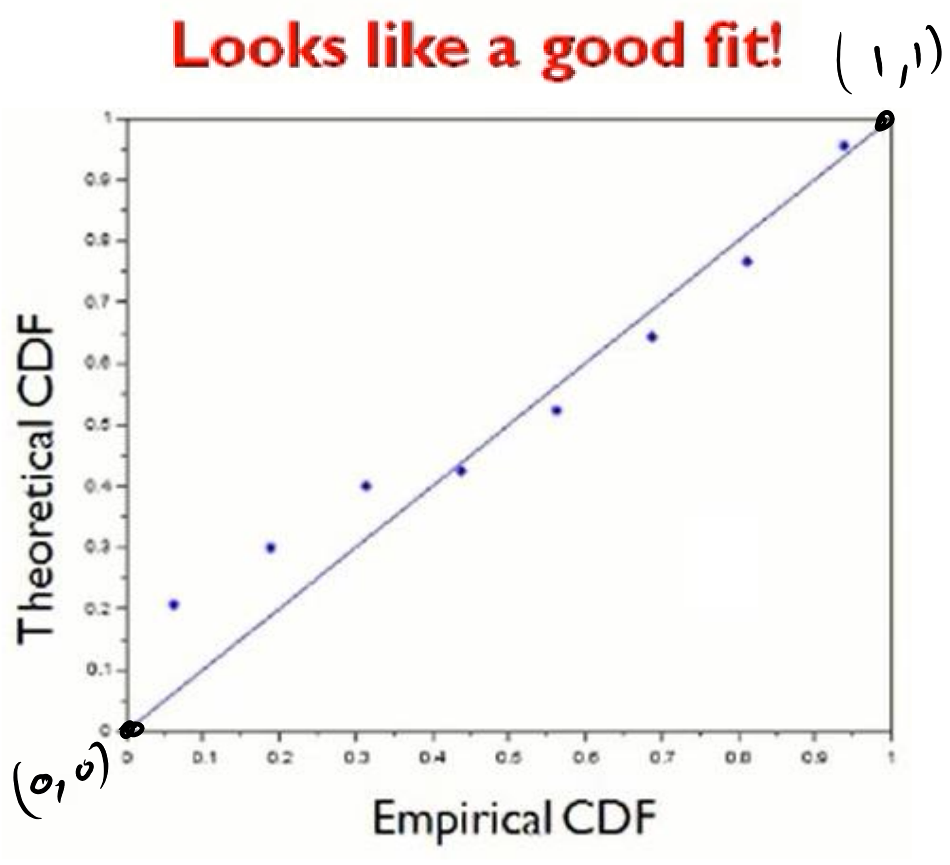
Using MME: $\hat{\lambda} = \frac{1}{\bar{X}} = \frac{8}{5.03 + 7.73 + 11.1 + 11.96 + 16.07 + 22.38 + 31.5 + 67.4} = 0.0462$

Example: Suppose the following data is given

(5.03 , 7.73 , 11.1 , 11.96 , 16.07 , 22.38 , 31.5 , 67.4)

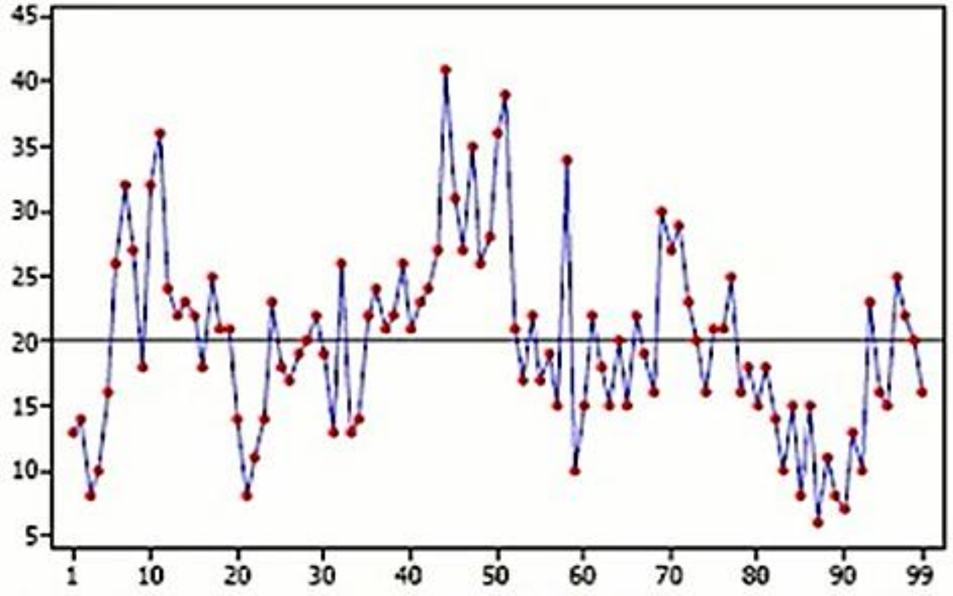
Examine the goodness of an exponential fit to the given data set using a P-P plot.

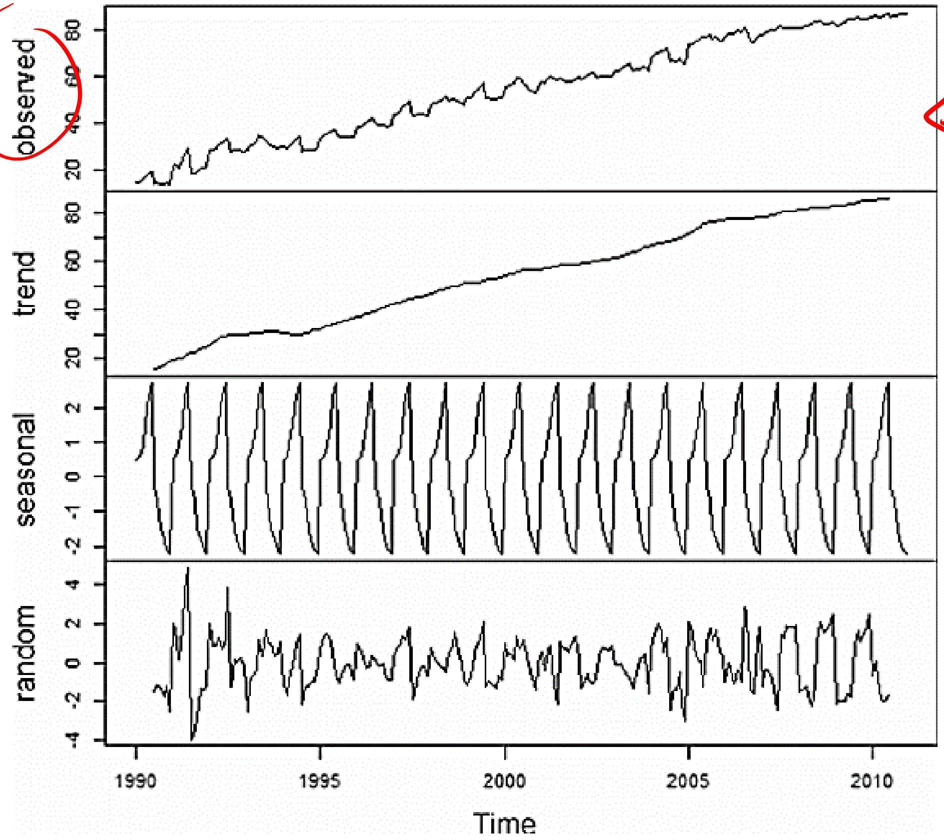
i	x_i	Empirical cdf	Exponential cdf
1	5.03	0.0625	0.2074
2	7.73	0.1875	0.3003
3	11.1	0.3125	0.4012
4	11.96	0.4375	0.4245
5	16.07	0.5625	0.52404
6	22.38	0.6875	0.6444
7	31.5	0.8125	0.7667
8	67.4	0.9375	0.9556



Time Series Plots

- A plot where
  - Horizontal axis (**x-axis**) → **Time**.
  - Vertical axis (**y-axis**) → **observed value**.





=

Σ

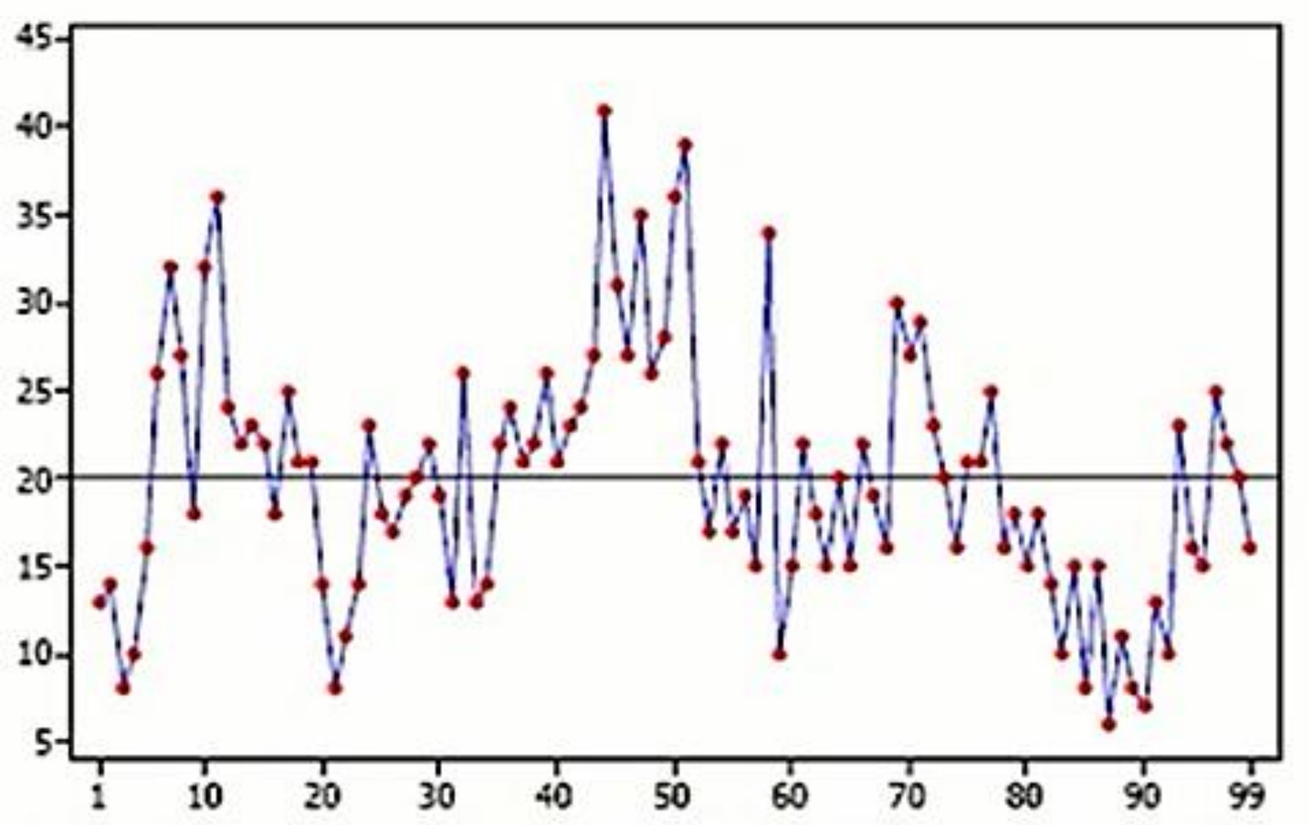
← original

- Time series help us notice **trends** and **cycles** in the observed data.

decomposition

Time Series Plots - Correlation

- Thus, a **time series plot** reveals whether there exists correlation between successive points or not.



Time Series Plots - Correlation

delay

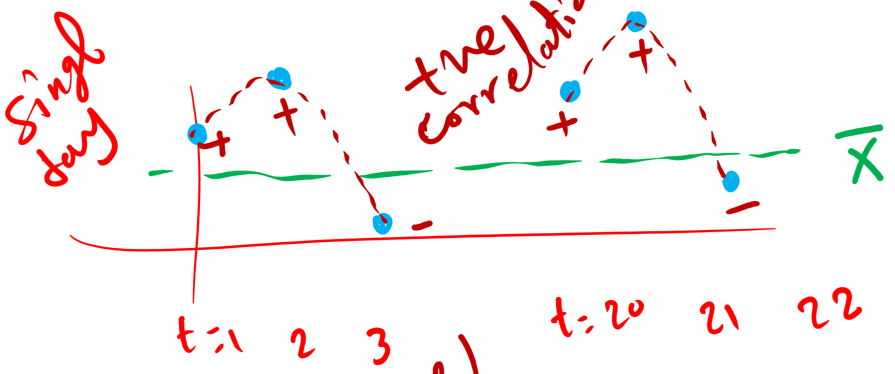
$x(\text{time}=1) \rightarrow x(\text{time } 20)$
$x(2) \rightarrow x(21)$
$\vdots \qquad \vdots$

lag 20

Auto-correlation Coefficient (at lag k)

- Correlation of a time series with a shifted version of itself.

$$\hat{\rho}_k = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})\ (x_{i+k} - \bar{x})}{n\ s^2}$$



increase/ decrease together

$\hat{\rho}_k$

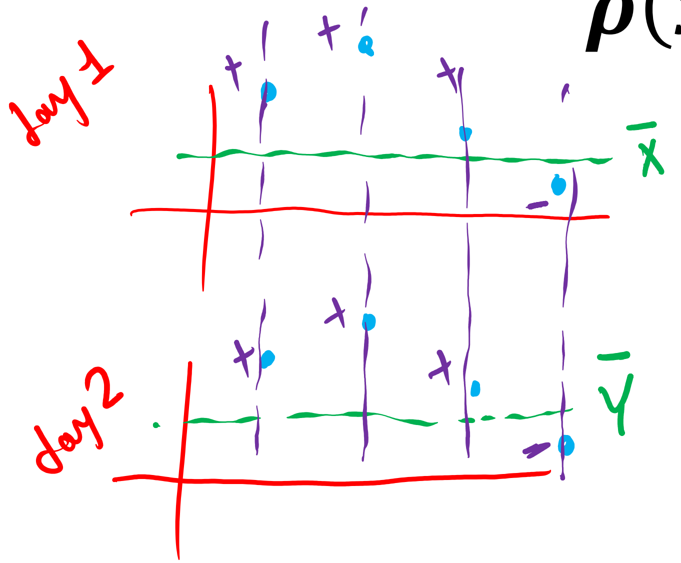
- > 0: increase/ decrease together beyond the mean
- $= 0$: Uncorrelated
- < 0: As one increases, the other decreases beyond the mean

Time Series Plots - Correlation

Cross-correlation Coefficient

- **Correlation** of a set of **observations X_i** with another set of **observations Y_i**

$$\hat{\rho}(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\,(y_i - \bar{y})}{n\,(s_x\, s_y)}$$



Time Series Plots - Correlation

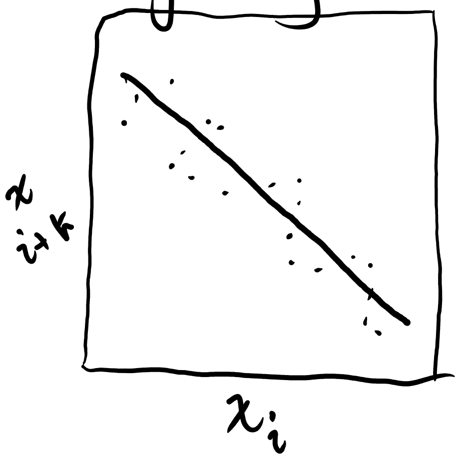
Scatter diagram

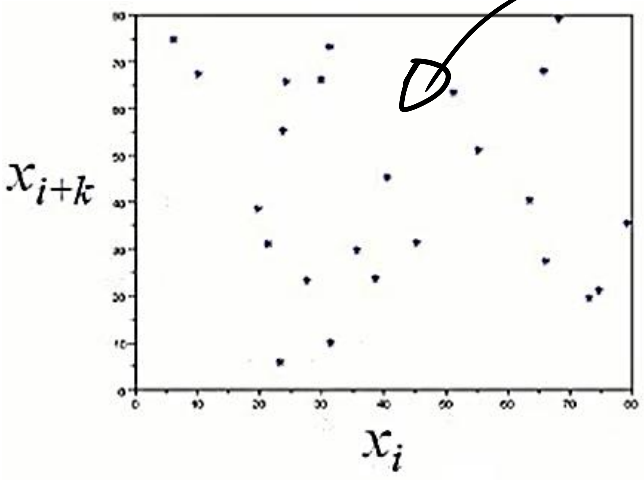
- A scatter diagram is another useful tool to **assess the correlation** between two workload parameters or successive points for the same parameter.
- It is a **visual display of the set of points** $(x_i\ , x_{i+k})$.

+ve slope straight line

Negatively correlate (-ve slop)

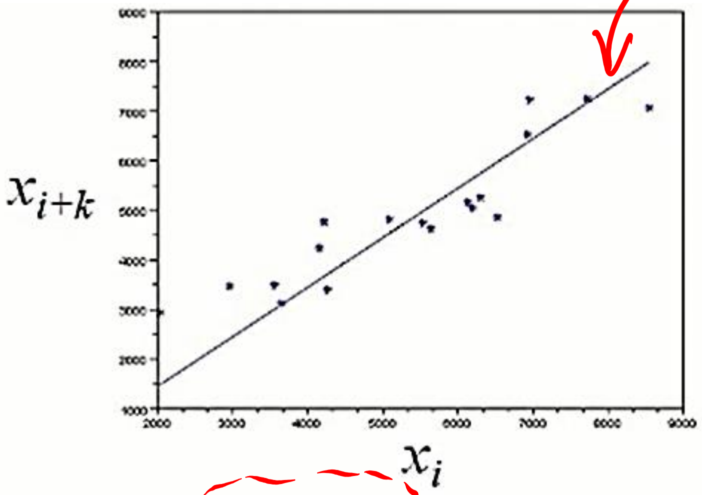
Scattered





Uncorrelated

"not necessary to be 45° or zero intercept"



Positively correlated

End of the course

الحمد لله الذي بنعمته تتم الصالحات