Elettra
Sincrotrone
Trieste

Introduction to Basics on radiation probing and imaging using x-ray detectors

Ralf Hendrik Menk
Elettra Sincrotrone Trieste
INFN Trieste

Part 1

Characterization of experimental data

- Data set: N independent measurements of the same physical quantity
  - $x_1, x_2, x_3, \ldots, x_i, \ldots, x_N$
  - every single value x_i can <u>only</u> assume <u>integer</u> values
- Basic properties of this data set
  - sum $$\Sigma = \sum_{i=1}^{N} x_i$$
  - experimental mean $$\overline{x_e} = \frac{\Sigma}{N}$$

Frequency distribution function F(x)

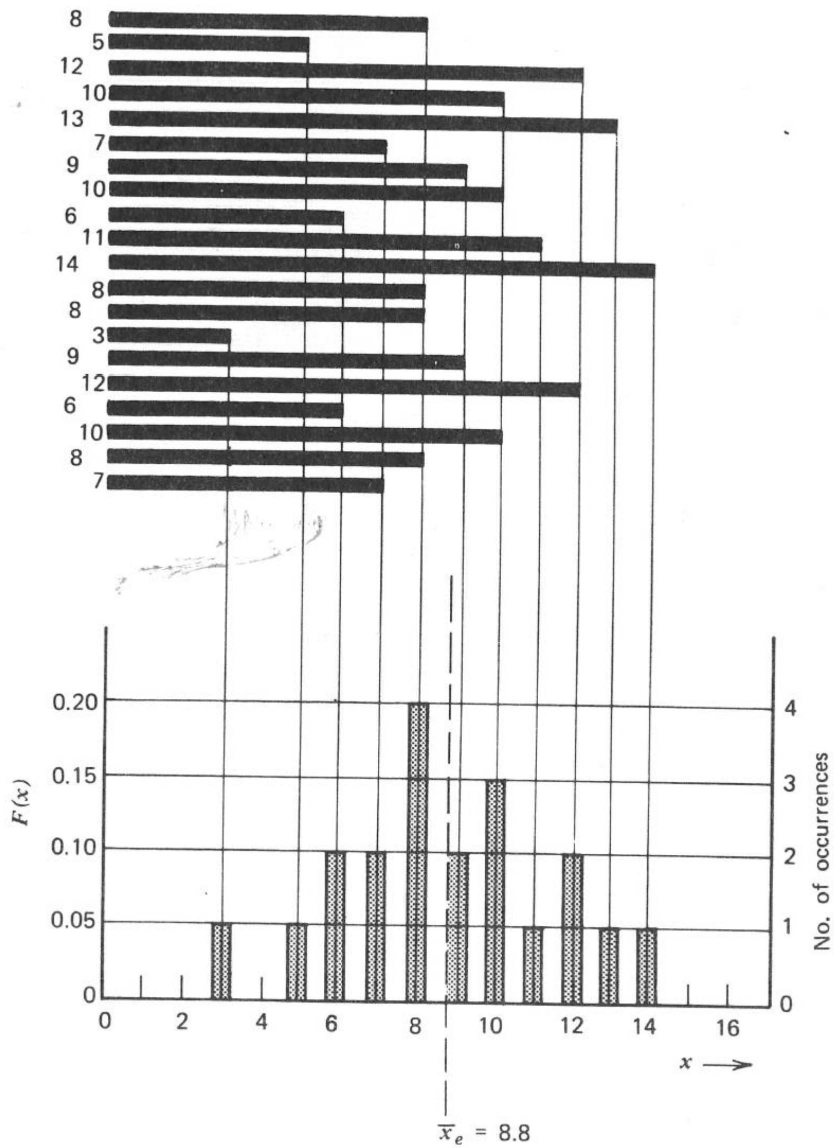
- The data set can be represented by means of a Frequency distribution function *F(x)*
- The value of *F(x)* is the relative frequency with which the number x appears in the collected data
  - $$F(x) \equiv \frac{number\ of\ occurences\ of\ the\ value\ x}{N}$$
- The frequency distribution is automatically normalized, i.e.
  - $$\sum_{x=0}^{\infty} F(x) = 1$$
- If we don't care about the sequence of the data *F(x)* represents all the information contained in the original data set

Data		Frequency Distribution Function	
8	14	$F(3) = 1/20$	$= 0.05$
5	8	$F(4)$	$= 0$
12	8	$F(5)$	$= 0.05$
10	3	$F(6)$	$= 0.10$
13	9	$F(7)$	$= 0.10$
7	12	$F(8)$	$= 0.20$
9	6	$F(9)$	$= 0.10$
10	10	$F(10)$	$= 0.15$
6	8	$F(11)$	$= 0.05$
11	7	$F(12)$	$= 0.10$
		$F(13)$	$= 0.05$
		$F(14)$	$= 0.05$
		$\sum_{x=0}^{\infty} F(x)$	$= 1.00$

Pictorial view

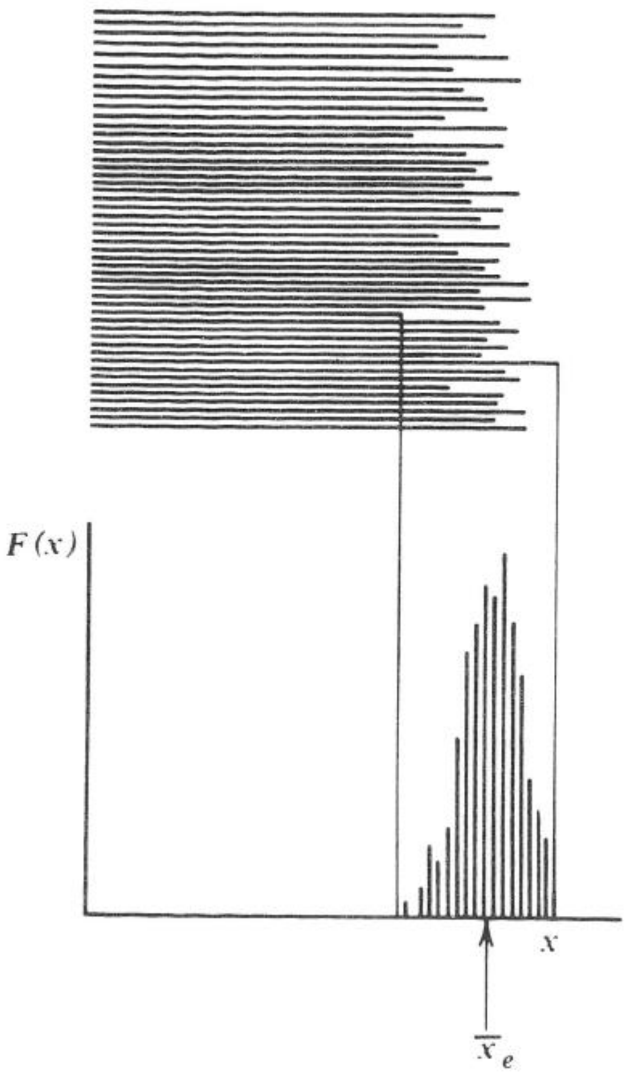
data

$F(x)$

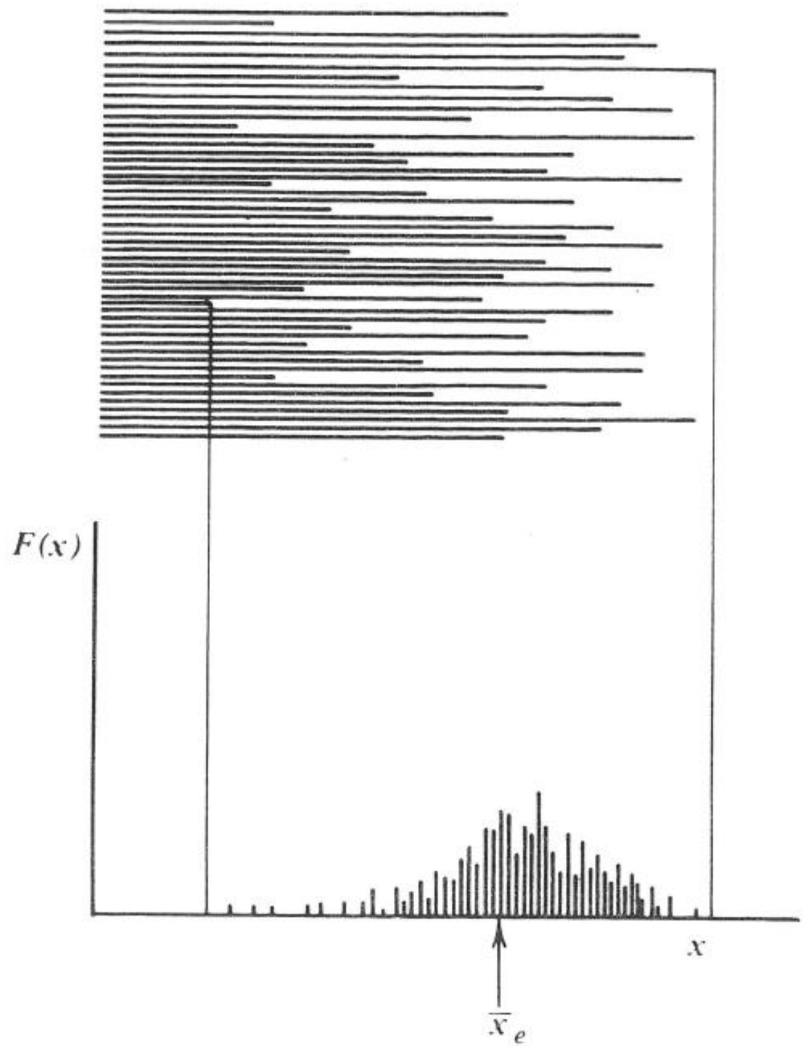


Properties of *F(x)*

- The experimental mean can be calculated as the first moment of the Frequency distribution function *F(x)*
  - $$\overline{x}_e = \sum_{x=0}^{\infty} x \cdot F(x)$$
- The width of the Frequency distribution function *F(x)* is a relative measure of the amount of fluctuation (or scattering) about the mean inherent in a given data set



F(x)
x
$\bar{x}_e$

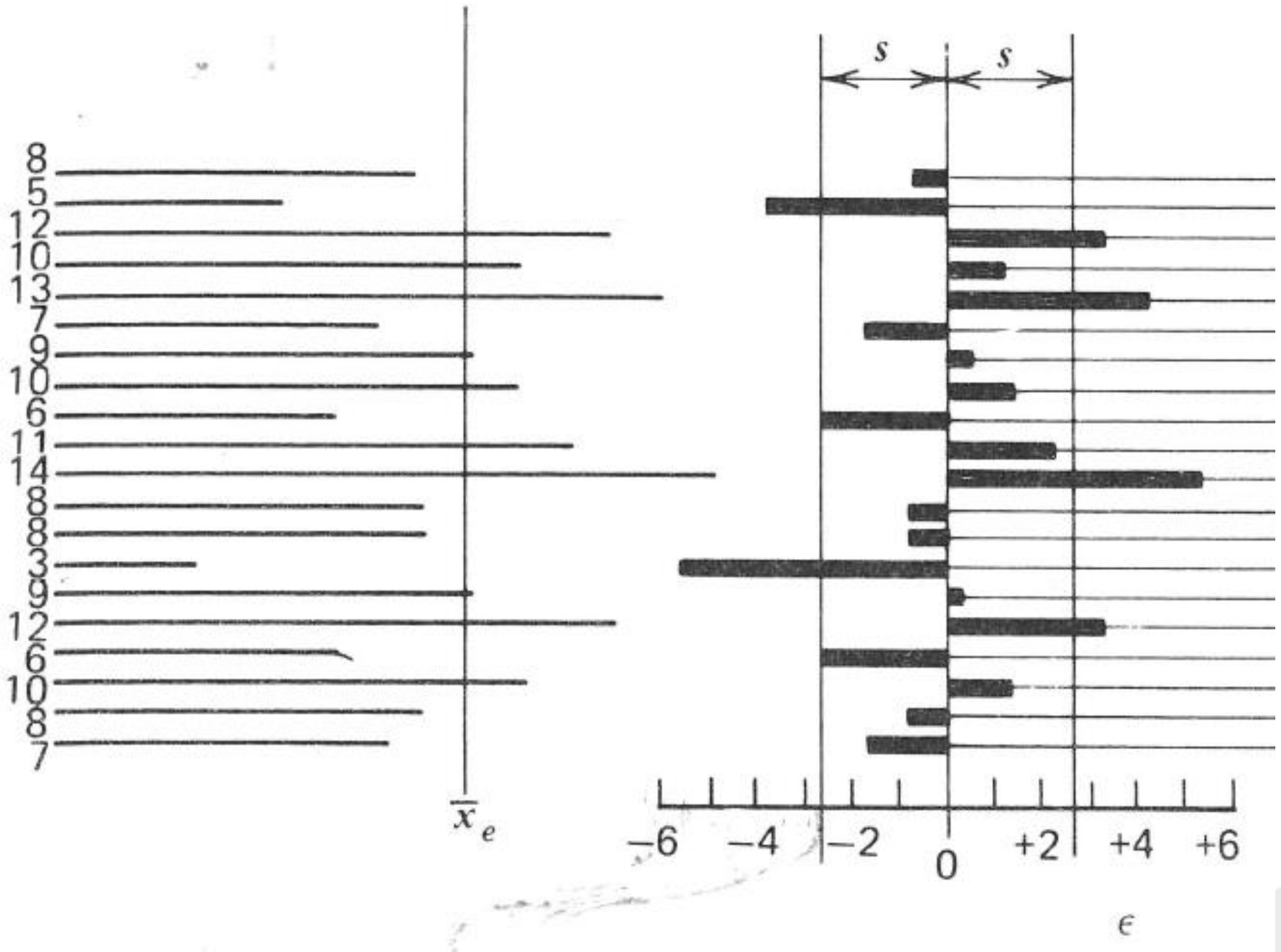


F(x)
x
$\bar{x}_e$

Deviations

- We define the deviation of any data point as the amount by which it differs from the mean value:
  - $$\varepsilon_i \equiv x_i - \overline{x}_e$$
- One could think to use the mean of the deviations to quantify the internal fluctuations of the data set, but actually:
  - $$\sum_{i=1}^{N} \varepsilon_i \equiv 0$$

Deviations: Pictorial View

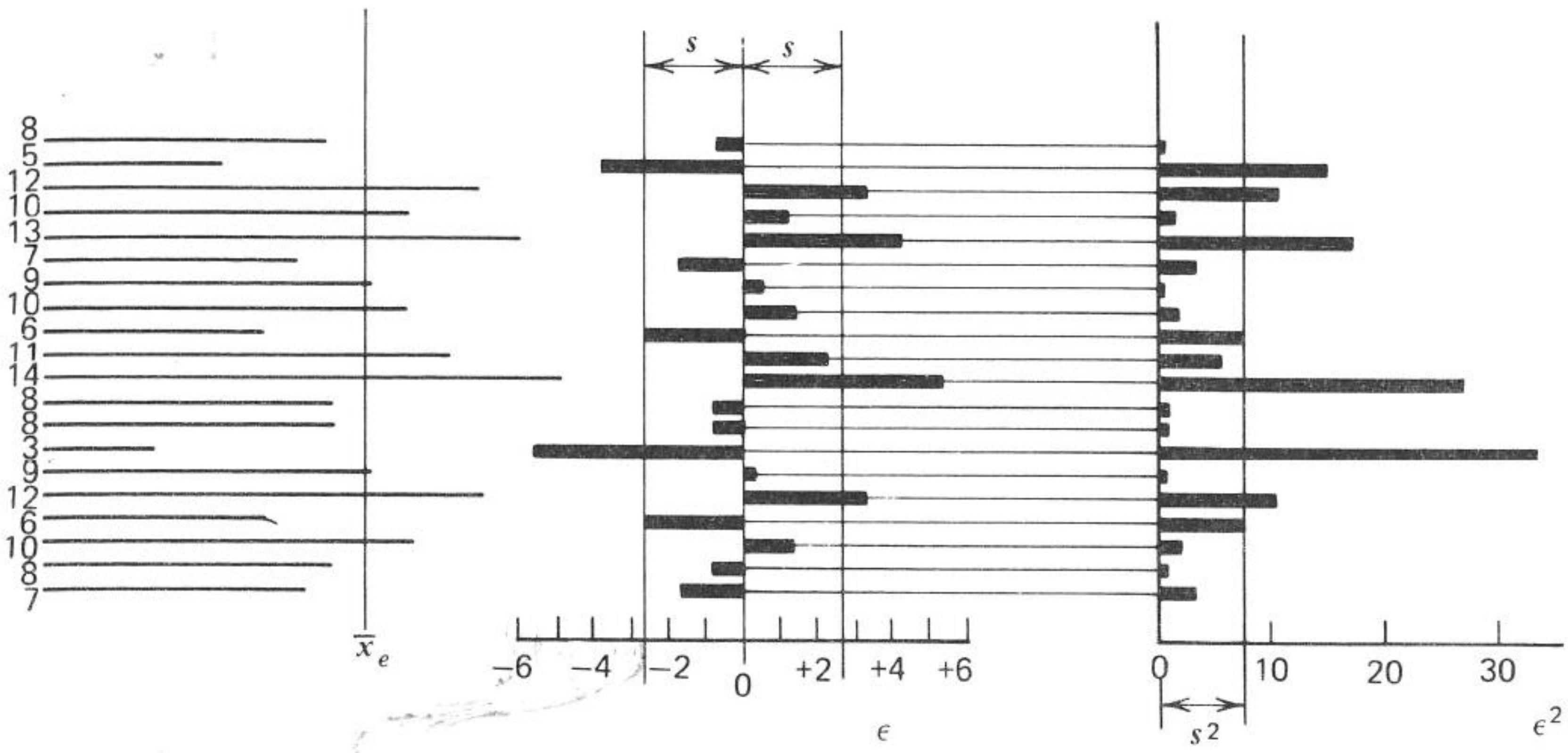


Elettra Sincrotrone Trieste

Sample variance

- A better idea is to take the square of each deviation
- The sample variance s is defined as
  - $$s^2 \equiv \frac{1}{N-1} \cdot \sum_{i=1}^{N} \varepsilon_i^2 = \frac{1}{N-1} \cdot \sum_{i=1}^{N} (x_i - \overline{x}_e)^2$$
- Or, more fundamentally, as the average value of the squared deviation of each data point from the "true" mean value (usually unknown)
  - $$s^2 \equiv \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2$$

Sample variance: Pictorial View



Sample variance

- The equation $s^2 \equiv \frac{1}{N} \cdot \sum_{i=1}^{N} (x_i - \overline{x})^2$ can be also rewritten in terms of *F(x)*, the data frequency distribution function as
  - $$s^2 \equiv \sum_{i=1}^{N} (x_i - \overline{x})^2 \cdot F(x)$$
- An expansion of the latter yields a well-known result
  - $$s^2 \equiv \overline{x^2} - (\overline{x})^2$$

  where $$\overline{x^2} = \sum_{x=0}^{\infty} x^2 \cdot F(x)$$

Trials

- We define a measurement as counting the number of successes resulting from a given number of trials
- The trial is assumed to be a binary process in which only two results are possible, either:
  - success or
  - not a success
- The probability of success is indicated as p and it is assumed to be constant for all trials
- The number of trials is usually indicated as n

Trials (examples)

Trial	Definition of success	Probability of success (*p*)
Tossing a coin	A Head	1/2
Rolling a die	A Six	1/6
Picking a card from a full deck	An Ace	4/52=1/13
Observing a given radioactive nucleus for a time *t*	The nucleus decays during the observation	$1-e^{-\lambda t}$

Elettra Sincrotrone Trieste

Statistical models

- Under certain circumstances, we can predict the distribution function that will describe the results of many repetitions of a given measurement
- Three specific statistical models are well-known
  - The Binomial Distribution
    - The most general but computationally cumbersome
  - The Poisson Distribution
    - A simplification of the above when *p* is small and *n* large
  - The Gaussian or Normal Distribution
    - A further simplification of the above if the average number of successes *pn* is relatively large (in the order of 20 or more)

The Binomial Distribution

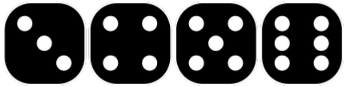
- The predicted probability of counting exactly *x* successes in *n* trials is:

$$P(x) = \frac{n!}{x!\,(n-x)!} \cdot p^x \cdot (1-p)^{n-x} = \binom{n}{x} \cdot p^x \cdot (1-p)^{n-x}$$

- Important properties
  - $\sum_{x=0}^{N} P(x) = 1$ P(x) is normalized
  - $\overline{x} = \sum_{x=0}^{N} x \cdot P(x) = n \cdot p$ expected average number of succ.
  - $\sigma^2 \equiv \sum_{x=0}^{n} (x-\overline{x})^2 \cdot P(x) = n \cdot p \cdot (1-p) = \overline{x} \cdot (1-p)$

    predicted variance

The Binomial Distribution (example)

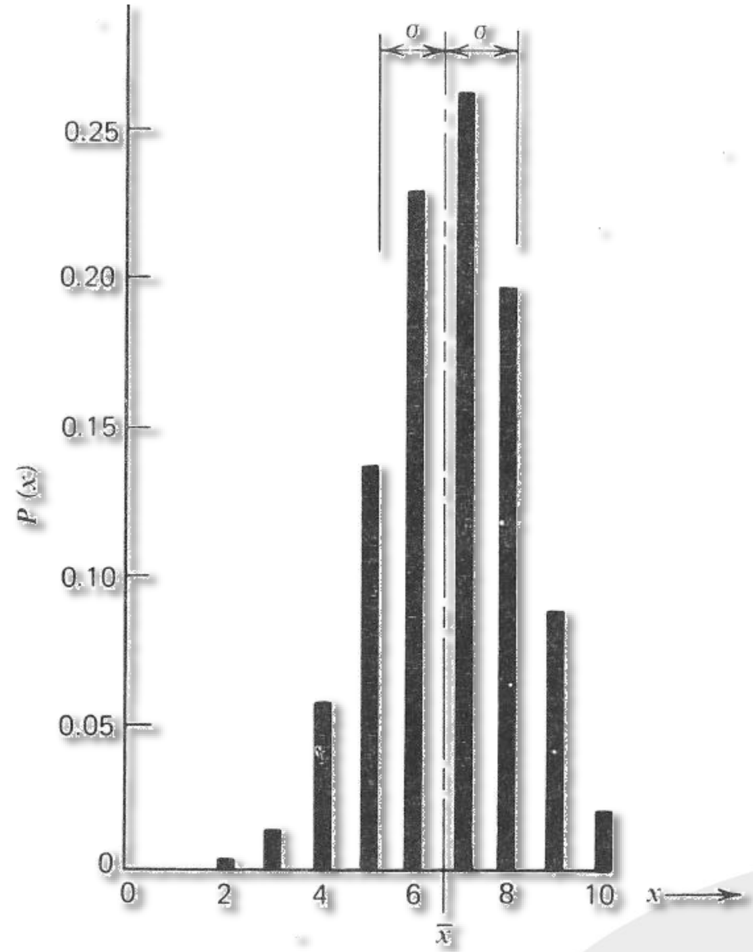
- Trial: rolling a die
- Success: any of 
- p = 4/6 = 2/3 = 0.667
- n = 10
- $\overline{x} = n \cdot p = \frac{2}{3} \cdot 10 = 6.67$
- $\sigma^2 = n \cdot p \cdot (1-p) = 10 \cdot 0.667 \cdot 0.333 = 2.22$
- $\sigma = \sqrt{\sigma^2} = \sqrt{2.22} = 1.49$

The Binomial Distribution (example)

Values of the Binomial Distribution for the Parameters $p = \frac{4}{6}$ or $\frac{2}{3}$, $n = 10$

x	$P(x)$
0	0.00002
1	0.00034
2	0.00305
3	0.01626
4	0.05690
5	0.13656
6	0.22761
7	0.26012
8	0.19509
9	0.08671
10	0.01734

$$\sum_{x=0}^{10} P(x) = 1.00000$$



A plot of the binomial distribution for $p = \frac{2}{3}$ and $n = 10$.

The Poisson Distribution

- When *p << 1* and *n* is reasonably large, so that *np=x* the binomial distribution reduces to the Poisson Distribution

$$P(x) = \frac{(\overline{x})^x \cdot e^{-\overline{x}}}{x!}$$

- Important properties
  - $\sum_{x=0}^{n} P(x) = 1$ P(x) is normalized
  - $\overline{x} = \sum_{x=0}^{n} x \cdot P(x) = n \cdot p$ expected average number of succ.
  - $\sigma^2 \equiv \sum_{x=0}^{n} (x-\overline{x})^2 \cdot P(x) = n \cdot p = \overline{x}$ predicted variance

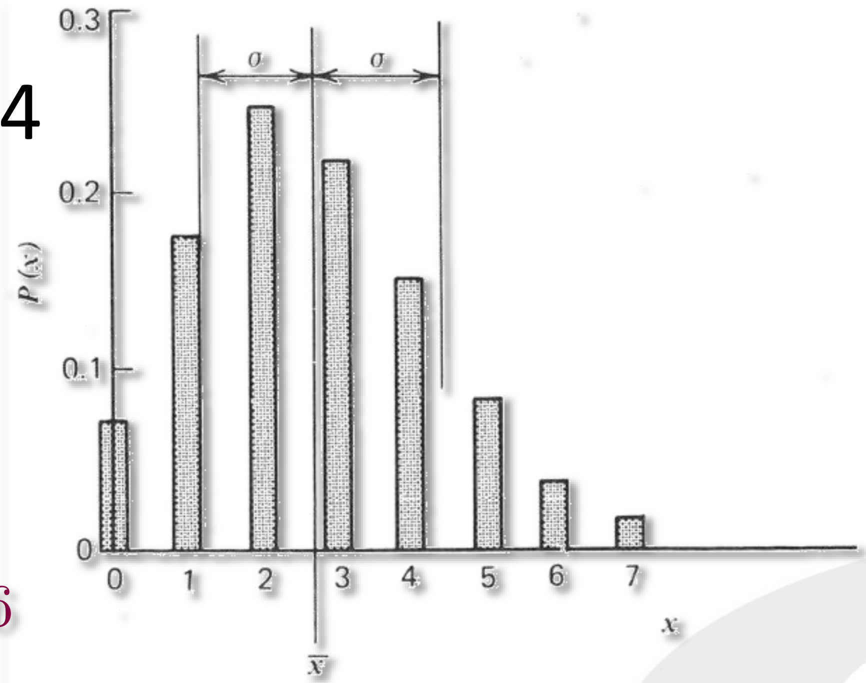
The Poisson Distribution

- Siméon Denis Poisson (21 June 1781 – 25 April 1840

- Ladislaus Josephovich Bortkiewicz (August 7, 1868 – July 15, 1931)
  - First practical application: investigating the number of soldiers in the Prussian army killed accidentally by horse kicks
  - Rare events!

The Poisson Distribution (example)

- Trial: birthdays in a group of 1000 people
- Success: if a person has his/her birthday today
- $p = 1/365 = 0.00274$
- $n = 1000$
- $\overline{x} = n \cdot p = 2.74$
- $\sigma^2 = \overline{x} = 2.74$
- $\sigma = \sqrt{\sigma^2} = \sqrt{2.74} = 1.66$



The Poisson Distribution with

- When $\bar{x} \gg 1$ the Poisson distribution can be approximated by a Gaussian (or Normal) distribution

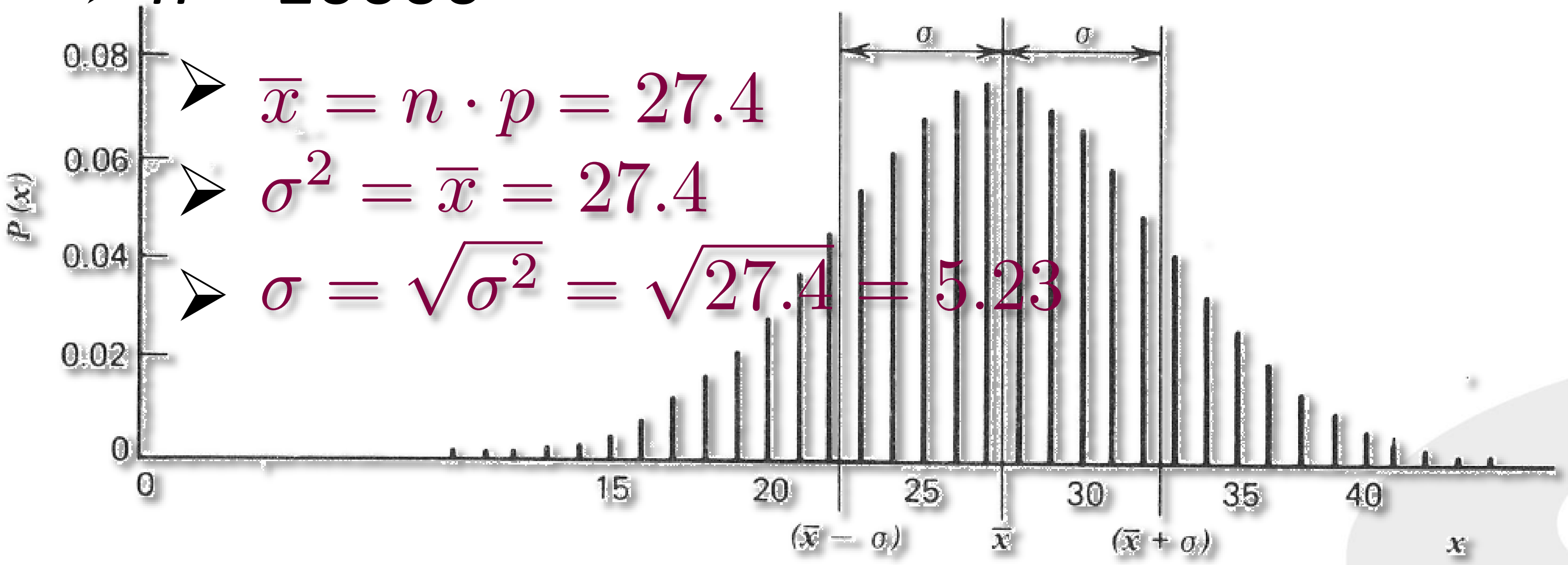
$$P(x) = \frac{1}{\sqrt{2 \cdot \pi \cdot \sigma^2}} \cdot e^{-\frac{(x-\overline{x})^2}{2 \cdot \sigma^2}}$$

  with the constraint $\sigma^2 = \overline{x}$

- As an example, we repeat the “birthday” experiment in a much larger group of 10000 people

The Poisson Distribution with *x >> 1*

- Trial: birthdays in a group of 10000 people
- Success: if a person has his/her birthday today
- p = 1/365 = 0.00274
- n = 10000
  - $\overline{x} = n \cdot p = 27.4$
  - $\sigma^2 = \overline{x} = 27.4$
  - $\sigma = \sqrt{\sigma^2} = \sqrt{27.4} = 5.23$



Variance of statistically independent trials

- N Trials : statistically independent
- variance in each trial σ_i^2
- Total variance

$$\sigma_{total} = \sqrt{\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2}$$

Elettra Sincrotrone Trieste

Distribution of time intervals between successive events

- We consider a random process characterized by a constant probability of occurrence per unit time
- Let r represent the average rate at which events are occurring
- Then $r\,dt$ is the (differential) probability that an event will take place in the differential time increment dt
- We assume that an event has occurred at time $t = 0$

Distribution of time intervals between successive events

- The (differential) probability $I_1(t) \cdot dt$ that the next event will take place <u>within a differential time</u> *dt* <u>after a time interval</u> of length *t* can be calculated as:

Probability of next event taking place in *dt* after delay of *t*	$=$	Probability of <u>no</u> events during time from 0 to *t*	$\times$	Probability of an event during *dt*
$I_1(t) \cdot dt$	$=$	$P(0)$	$\times$	$r \cdot dt$

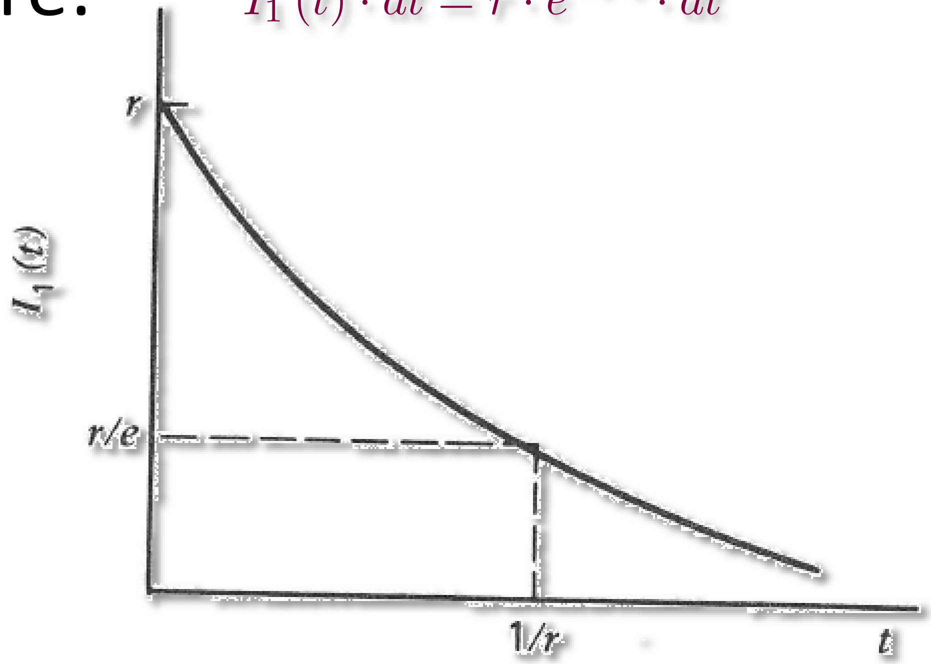
- Where *P(0)* is given by the Poisson distribution:

$$P(0) = \frac{(r \cdot t)^0 \cdot e^{-r \cdot t}}{0!}$$

$$P(0) = e^{-r \cdot t}$$

Distribution of time intervals between successive events

- The (differential) probability $I_1(t) \cdot dt$ that the next event will take place within a differential time *dt* after a time interval of length *t* is therefore: $I_1(t) \cdot dt = r \cdot e^{-r \cdot t} \cdot dt$



Elettra
Sincrotrone
Trieste

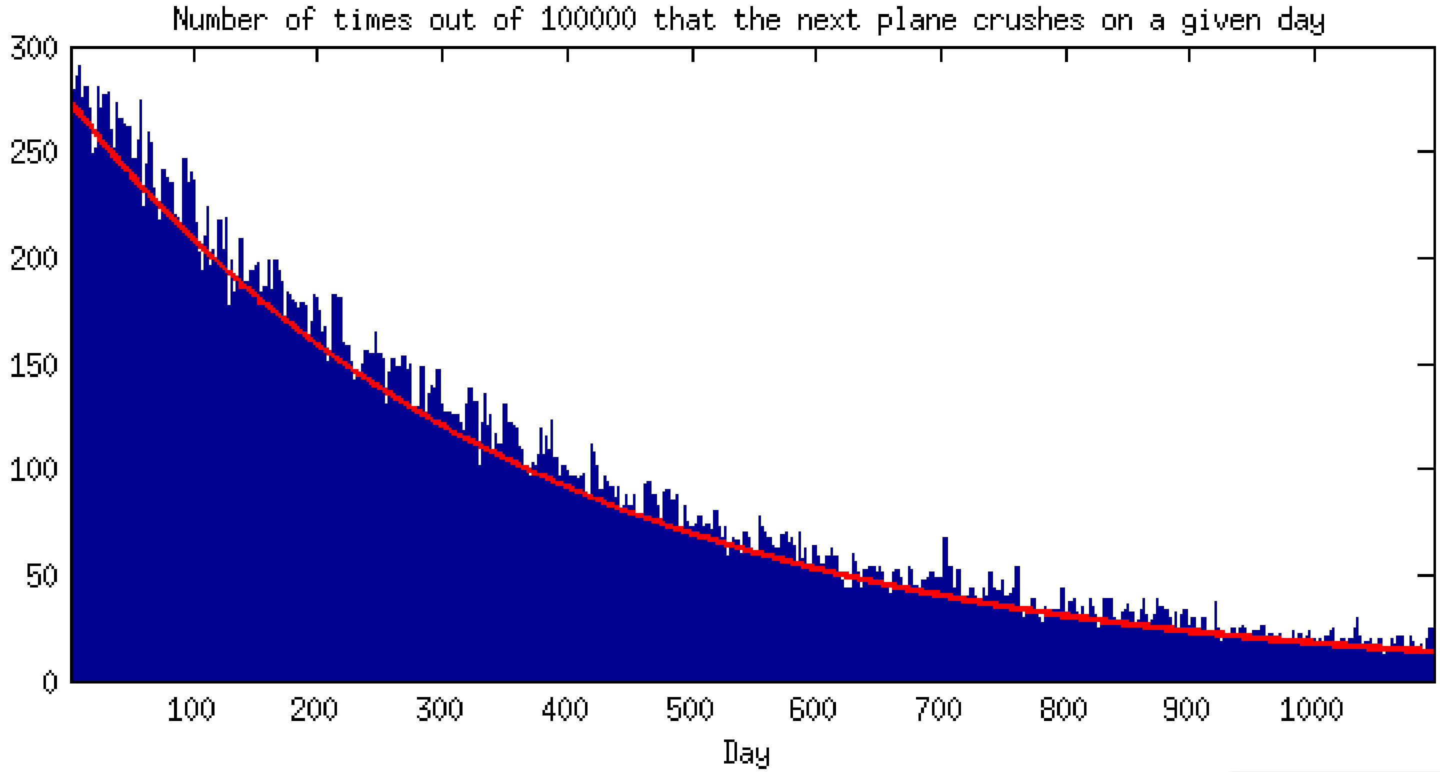
Distribution of time intervals between successive events

Elettra Sincrotrone Trieste

Distribution of time intervals between successive events

Event Date	Airline	Model (Age in Years)	Type of Operation	Accident Location	Phase of Flight	Event Description	Damage Category	Hull Loss	Injury Category	Onboard Fatalities/ Occupants (External Fatalities)	Major Accident
25-Jan-13	FedEx	MD-11-F (16)	Sched Cargo	Denver, USA	Landing	The airplane sustained damage due to a tail strike during landing. There were no injuries.	Substantial				
04-Feb-13	Asiana Airlines	767-300F (16)	Sched Cargo	Incheon, South Korea	Takeoff	During acceleration to takeoff power, both engines shut down due to ice/snow ingestion. Both engines sustained damage. There were no injuries.	Substantial				
06-Feb-13	Tunisair	A320 (22)	Sched Pax	Tunis, Tunisia	Landing	The airplane sustained damage when it veered off the runway while landing. The nose gear collapsed. There were no injuries.	Substantial	X			
11-Feb-13	Pakistan Int'l Airlines	737-300 (20)	Sched Pax	Muscat, Oman	Landing	The airplane sustained damage after landing when the left main landing gear collapsed during rollout. There were no injuries.	Substantial	X			
29-Mar-13	Air Méditerranée	A321 (16)	Sched Pax	Lyon, France	Landing	The airplane sustained damage when it overran the end of the runway and came to rest in soft ground. Both engines ingested rocks and mud. There were no injuries.	Substantial				
05-Apr-13	US Airways	A321 (1)	Sched Pax	Las Vegas, USA	Landing	The airplane sustained damage due to a tail strike during landing. There were no injuries.	Substantial				
13-Apr-13	Lion Air	737-800 (0)	Sched Pax	Denpasar, Indonesia	Approach	The airplane landed in the water short of the runway during a non-precision approach.	Destroyed	X	Serious		X
16-Apr-13	Aeromexico	767-200 (23)	Sched Pax	Madrid, Spain	Takeoff	The airplane sustained a tail strike during takeoff. The cabin did not pressurize, oxygen masks deployed automatically, and the airplane performed an air turnback. There were no injuries.	Substantial	X			
16-Apr-13	Asiana Airlines	A321 (9)	Sched Pax	Seoul, South Korea	Landing	The airplane sustained damage due to a tail strike during landing. There were no injuries.	Substantial				
29-Apr-13	National Air Cargo Group	747-400 (20)	Sched Cargo	Bagram, Afghanistan	Initial Climb	Shortly after takeoff, the airplane pitched up significantly and control of the airplane was lost. An uncontrolled ground impact and post-impact fire ensued.	Destroyed	X	Fatal	7/7(0)	X
24-May-13	Air Via	A320 (1)	Sched Pax	Varna, Bulgaria	Landing	The airplane sustained damage during landing when it overran the end of the runway and came to rest in soft ground. Injuries were sustained during evacuation.	Substantial		Serious		
24-May-13	British Airways	A319 (12)	Sched Pax	London, United Kingdom	Initial Climb	During or shortly after takeoff rotation, the fan cowl doors from both engines departed, damaging the airframe and some aircraft systems. The flight crew elected to turn back. During approach, a fire developed on the right engine. A successful landing ensued. There were no injuries.	Substantial				

Distribution of time intervals between successive events



Elettra
Sincrotrone
Trieste

Distribution of time intervals between successive events