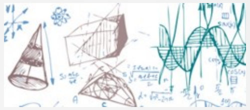
Introduction to probability, statistics and data handling



Tomasz Szumlak, Agnieszka Obłąkowska-Mucha
Faculty of Physics and Applied Computer Science

AGH UST Krakow

The Binomial Distribution

- We already know this distribution – it emerges when we consider an experiment such as tossing a coin, rolling a die or choosing a marble from a box repeatedly.
- So, we considering trials. Each outcome will have constant probability assigned (that should not change in time, and is the parameter of the Bernoulli prob. model family).
  - Sometimes we are also interested in processes where the probability is not constant (out of the scope of our lecture, however)
- We then say that p is a success and q is a failure (in a Bernoulli sense) and can compose the following P.D.F.

$$\boldsymbol{f(x) = B(n,p) = p(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!\,(n-x)!} p^x q^{n-x}}$$

- The RV denote the number of successes x in n trials, $x = 0, 1, \dots, n$

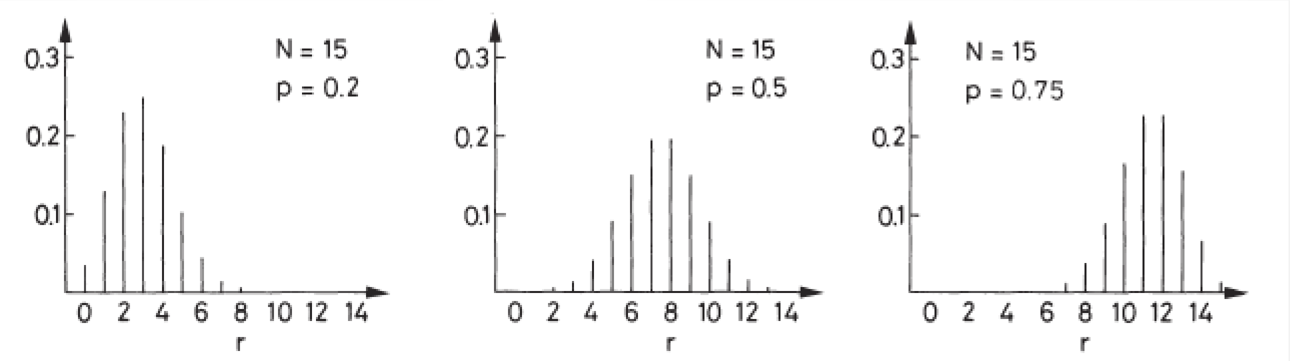
The Binomial Distribution

- The mean and variance can be fairly easy calculated:

$$\mu = \sum_x xP(x) = np$$

$$\sigma^2 = \sum_x (x-\mu)^2 P(x) = np(1-p)$$

- In the limit of „large” n and „no too small” p we can very accurately approximate the Binomial distribution with Gaussian one



The Binomial Distribution

- **Properties of the Binomial P.D.F.**

Mean	$\mu = np$
Variance	$\sigma^2 = npq$
Standard deviation	$\sigma = \sqrt{npq}$
Coefficient of skewness	$\alpha_3 = \dfrac{q - p}{\sqrt{npq}}$
Coefficient of kurtosis	$\alpha_4 = 3 + \dfrac{1 - 6pq}{npq}$

- We can mote something interesting here
- **Theorem 6**. Let X be the RV giving the number of successes in n Bernouli trials, so that $\frac{X}{n}$ is the proportion of successes. Then if p is the probability of success and ϵ is any positive number:

$$\lim_{n\to\infty} prob\left(\left|\frac{X}{n} - p\right| \geq \epsilon\right) = 0$$

Poisson distribution

- The RV of discrete type:

- the number of outcomes occurring, for instance, during a given time (e.g. number of radioactive decays in a sample of radioactive material) t: $X = X_t = 0, 1, 2, \ldots$

- number of events in a given region of space - e.g. number of typing errors per page

or:

- telephone calls arriving during a (short) period of time;
- light quanta (photons) arriving at detecting system;
- number of mutations on a strand of DNA (per unit length);
- number of customers arriving at a counter;
- number of cars arriving at a trac light;
- number of Losses/Claims;

- POISSON

Poisson distribution

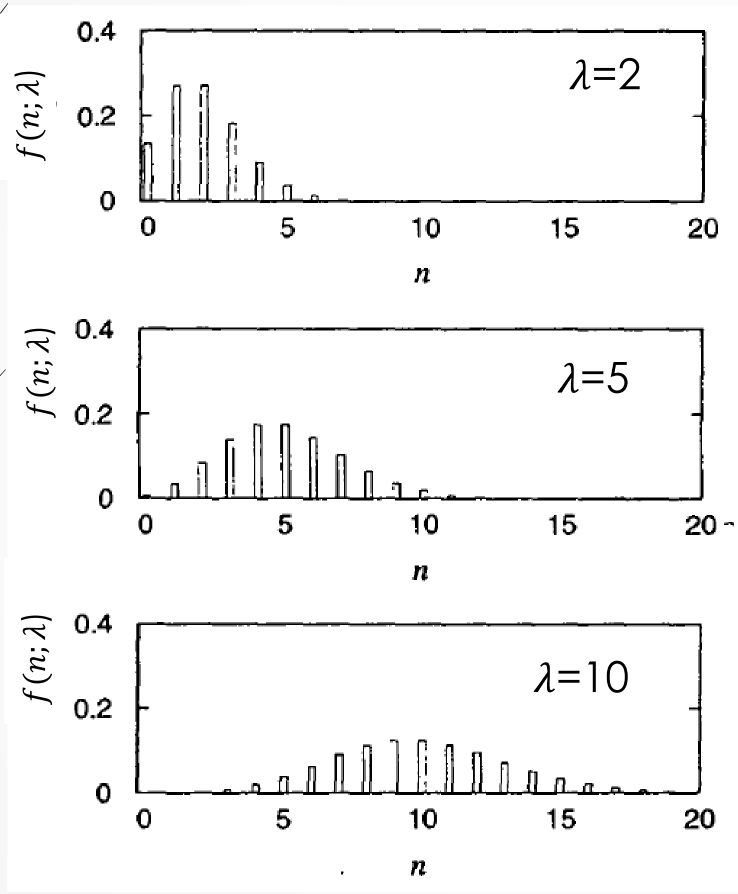
- Numbers of outcomes occurring in one time interval t are independent of each other, i.e. the number occurring in one time interval is independent of the number that occurs in any other disjoint time interval (Poisson process has no memory)
- The probability that a single outcome will occur during a very short time interval t is PROPORTIONAL to the length of interval:

$$P(X_{\Delta t} = 1) \sim \Delta t$$

- Let me to introduce the Poisson distribution:

$$P(X_t = k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Poisson distribution



- The number of decays of radioactive material in a fixed time period follows the Poisson distribution (given that the decay probability is constant over the time period)

Poisson distribution

- Let's take the binomial distribution, it can be shown that in the limit of large n and very small p (given that np is finite) we get a new distribution:

$$f(n; n \cdot p = \lambda) = \frac{\lambda^n}{n!} e^{-\lambda}$$

- This Poisson distribution is valid for integer variable n $(n = 0, 1, \ldots)$ and has a single parameter $n \cdot p = \nu$
- Its mean value and variance

$$E[n] = \sum_{n/0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} = \lambda \quad V[n] = \sum_{n/0}^{\infty} (n - \lambda)^2 \frac{\lambda^n}{n!} e^{-\lambda} = \lambda$$

- In many cases the n can be treated as a continuous variable. Also, if ν is large the Poisson random variable can be treated as a continuous variable similar to the **normal distribution**

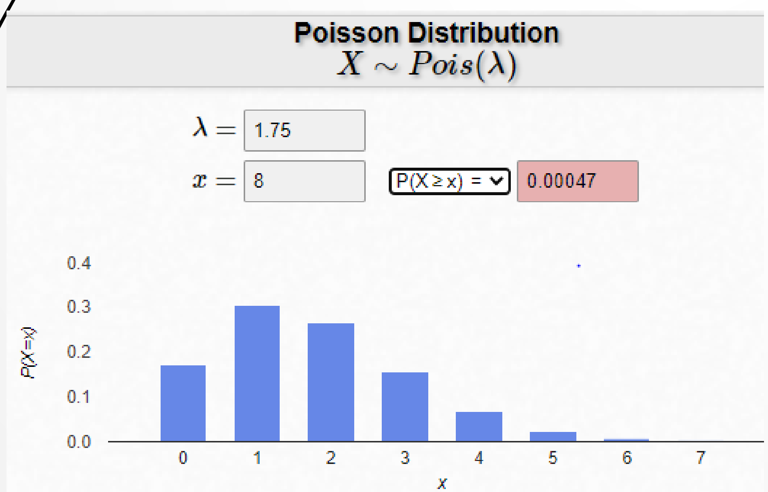
The Poisson Distribution

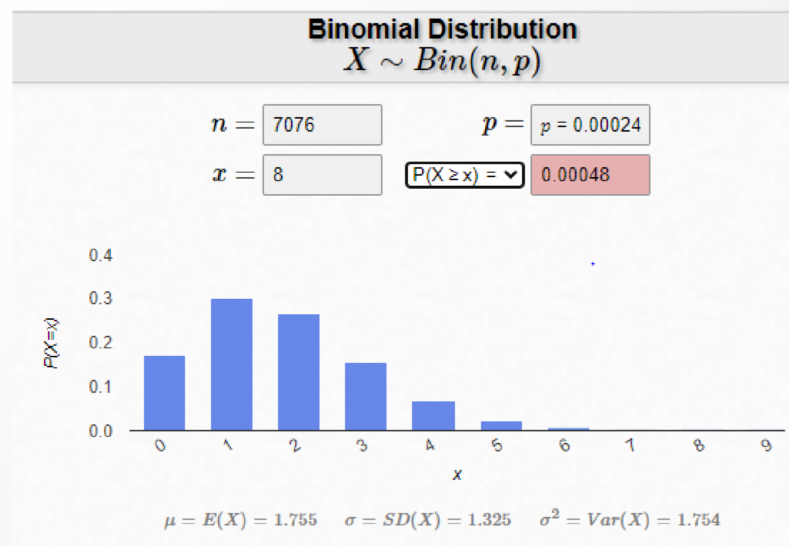
- The probability of getting leukemia is $p = 0.000248$. Using the approximation Bernoulli-to-Poisson find the P of eight or more leukemia cases in a population of size $n = 7076$.

$np = 7076 \times 0.000248 = 1.75 = \lambda.$

$$\mathcal{P}(X \leq 7) = \sum_{0 \leq x \leq 7} e^{-1.75} \frac{1.75^x}{x!} = 0.999518.$$

hence $\mathcal{P}(X \geq 8) \approx 1 - 0.999518 = 0.000482.$



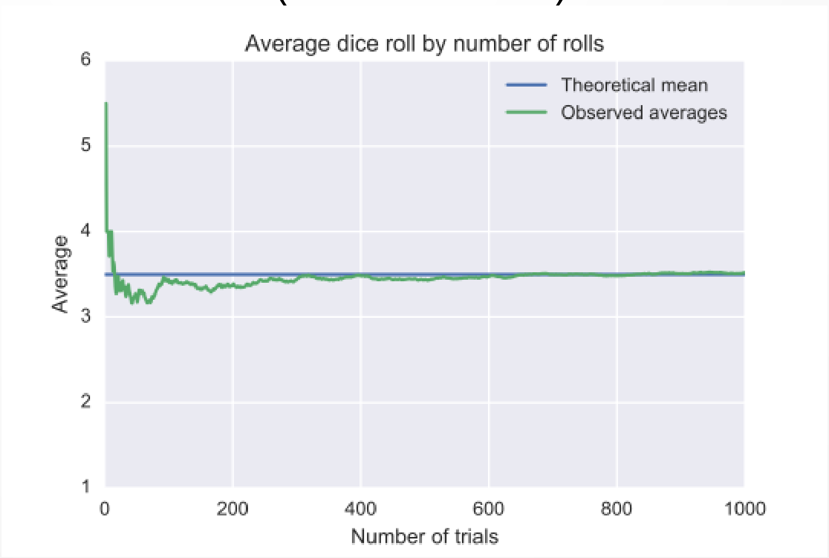


Law of large numbers

- Building on the knowledge we gained today, we can formulate a very advanced theorem, that is considered fundamental for statistics.
- **Theorem 7**. *Let $X_1, X_2, \cdots, X_n$ be mutually independent RV (discrete or continuous), each having finite mean μ and variance σ^2. Then if we take into consideration a new*

  RV: $\boldsymbol{S_n = X_1 + X_2 + \cdots + X_n}$, then:

$$\lim_{n\to\infty} p\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$



The Law of Large Numbers

- Imagine that our sample space can be divided into k events (or outcomes that we wish to study): $\{A_j\}, j = 1, \ldots, k$. What are the respective probabilities of such events p_j?
- Well, in principle we should conduct an experiment, collect a data sample and then calculate the frequency f_j (we assume that n below means the number of events of type j observed):

$$f_j = \frac{1}{n}\sum_{i/1}^{i/n} X_{ij} = \frac{1}{n} X_j$$

- So, note that X_j is a binomial R.V. that takes the following values:

$$X_j = \begin{cases} 1 \; if \, A_j \; occured \\ 0 \quad otherwise \end{cases}$$

- Now, how is f_j related to the probability p_j? Remember, the probability is just a number, whilst the frequency is a R.V.

The Law of Large Numbers

- It is essential to understand, the last point – remember frequency will always depend on a particular sample! Different sample will yield a different frequency.
- Having said that, we can however write (remember that X_j is a binomial R.V.!):

$$E[f_j] = E\left[\frac{X_j}{n}\right] = \boldsymbol{p_j}, E[X_j] = np_j$$

$$\sigma^2(f_j) = \sigma^2\left(\frac{X_j}{n}\right) = \frac{1}{n^2}\sigma^2(X_j) = \frac{1}{n}p_j q_j = \boldsymbol{\frac{1}{n}p_j(1-p_j)}$$

- In words: the expectation value of the frequency (event A_j) is equal to the probability of success. The variance of the frequency about its mean value can, in turn, be made arbitrarily small – just need to collect enough data! (large n).
- **This, actually, is the law of large numbers**!

The Law of Large Numbers

- Let's just think about the variance for a second. It is a product of these two elements: $1/n$ and $p_j(1-p_j)$. The latter is always less than unity (the max value is: $max\{p_j(1-p_j) = 1/4\}$), so the „smallness" of departure **will be governed by the number of observed events**.
- Using this argument and the result from previous slide we can justify that the approach, where **respective probabilities** of events that are estimated by **frequencies measured directly** in experiments, is **the right one**!
- The square of the error we make doing so is inversely proportional to the number of measurements in an experiment – this kind of error is called a **statistical** one
- This is essence of, so called, **counting experiments** such as: number of decaying particles, number of animals with a given traits, number of defective items, …

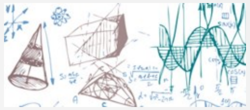
The Law of Large Numbers

- And finally to sum up the **LoLN**

Let $X_1, X_2, \dots, X_n$ be mutually independent random variables (no particular P.D.F. is assumed here), each of which have finite mean, μ, and variance, σ^2.
Now let's us define new R.V.: $S_n = X_1 + X_2 + \cdots + X_n, n = 1, 2, \dots$

The probability that the arithmetic mean of $X_1, X_2, \dots, X_n$ differs from its expected value more than ϵ approaches zero as $n \to \infty$:

$$\lim_{n\to\infty} p\left(\left|\frac{S_n}{n} - E\left[\frac{S_n}{n}\right]\right| \geq \epsilon\right) = \lim_{n\to\infty} p\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = 0$$



Uniform distribution

- Uniform P.D.F. is defined for the C.R.V. and given by

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a} & a\le x\le b\\ 0 & \text{otherwise}\end{cases}$$

- So, we say that x is equally likely to be found inside our interval of interest (a,b). The mean and variance:

$$E[x]=\int_a^b \frac{x}{b-a}\,dx=\frac{1}{2}(a+b)$$

$$V[x]=\int_a^b \left(x-\frac{1}{2}(a+b)\right)^2\frac{1}{b-a}\,dx=\frac{1}{12}(b-a)^2$$

- There is a very important application of the uniform model related to the fact that for any R.V. x with P.D.F. $f(x)$ we can easily find transformation to a new variable that is uniform.

Uniform distribution

- If we call the new transformed variable y, the transformation rule is simply related with calculating the C.D.F. Cool!

$$x \to y: y = F(x)$$

- Remember, x – any P.D.F., y uniform P.D.F. Next, for any C.D.F. the following is true

$$\frac{dy}{dx} = \frac{d}{dx}\int_{-\infty}^{x} f(x')dx' = f(x)$$

- And using the rule for change of variables

$$g(y) = f(x)\left|\frac{dx}{dy}\right| = f(x)\left|\frac{dy}{dx}\right|^{-1} = 1, \qquad (0 \le y \le 1)$$

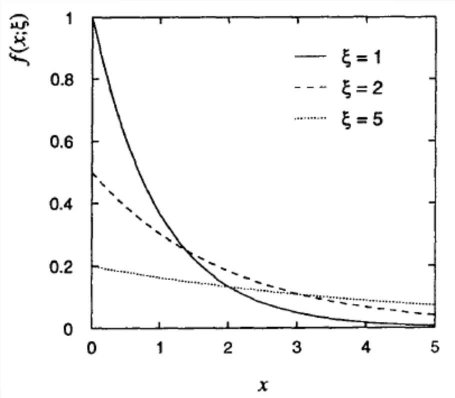
- This is the fundamental rule of random number generator programs that can give us a list of numbers with any distribution. We are going to look at this in more detail.

Exponential distribution

- Exp. distribution describes the amount of time between some (relatively rare) events
- This model is used for C.R.V. $x\colon 0 \le x \le \infty$ and is defined by

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

- The model depends on a single parameter θ. The mean and variance are as follow:



$$E[x] = \frac{1}{\xi}\int_0^\infty x e^{-\frac{x}{\xi}} dx = \xi$$

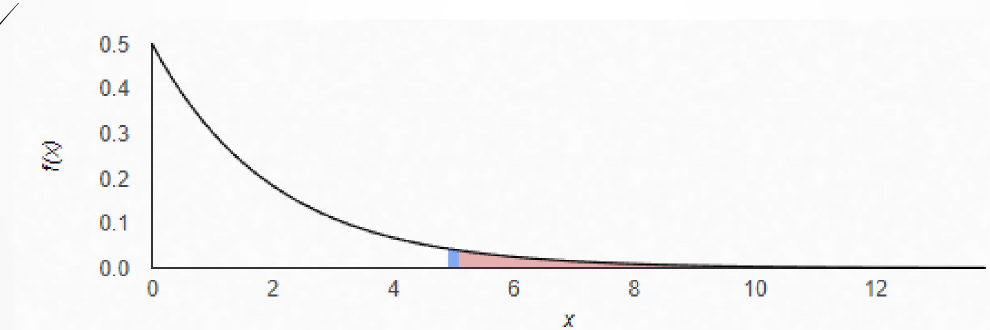
$$V[x] = \frac{1}{\xi}\int_0^\infty (x-\xi)^2 e^{-\frac{x}{\xi}} dx = \xi^2$$

- The decay time of an unstable particle measured in its rest frame follows the exponential distribution. The parameter of the distribution is then interpreted as the mean lifetime.

Exponential distribution has no memory

- Example: the amount of time between customers is exponentially distributed with a mean of two minutes $X \to \mathrm{Exp}(\xi = 2)$. Suppose that 5 min passed since the last customer arrived, is now more likely that a new will arrive within next 3 min?

$$P(X > 5\,min + 3min/X > 5min) = ?$$



Exponential distribution is memorylessness:

$$P(X > t + \Delta t/X > t) = P(X > \Delta t)$$

The memoryless property says that knowledge of what has occurred in the past has no effect on future probabilities.

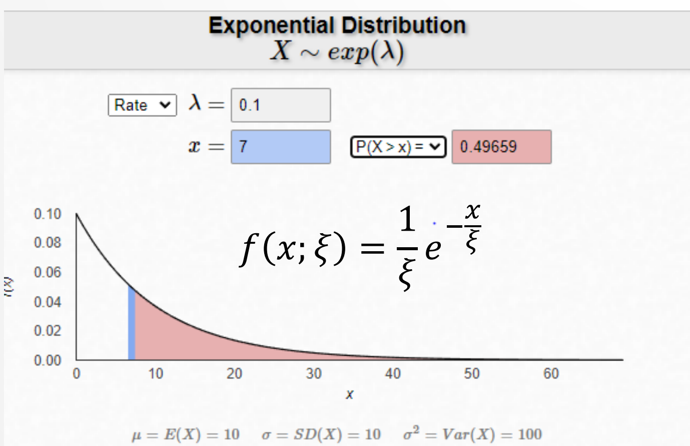
Exponential distribution has no memory

Exponential distribution is memorylessness:

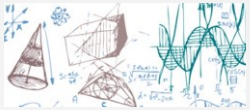
$$P(X > t + \Delta t / X > t) = P(X > \Delta t)$$

The memoryless property says that knowledge of what has occurred in the past has no effect on future probabilities.

Example: The exponential distribution is often used to model the longevity of an electrical or mechanical device.
he lifetime of a certain computer part has the exponential distribution with a mean of ten years $X \to \text{Exp}(\xi = 10)$. In this case it means that an old part is not any more likely to break down at any particular time than a brand new part. In other words, the part stays as good as new until it suddenly breaks. For example, if the part has already lasted ten years, then the probability that it lasts another seven years is:



$$P(X > 17 | X > 10) = P(X > 7) = 0.4966$$



The Poisson and Exponential Distribution

- Exp distribution: time between two events $X \rightarrow Exp\ (\mu)\ E[x] = \mu.$
- If the time between events is not affected by the times between previous events (times are independent) then number of events per unit time has Poisson distribution with $\lambda = 1/\mu$:

$$P(X = k) = \frac{\lambda^k e^\wedge - \lambda}{k!}$$

Conversely, if the number of events per unit time follows a Poisson distribution, then the amount of time between events follows the exponential distribution.

The Poisson and Exponential Distribution

- There is a direct connection between the exponential distribution and the Poisson distribution (process).
- We may remember that the unique parameter of **Poisson** distribution can be interpreted as a **mean number of events per unit time**.
- So, if we consider a time interval of the length t the number of events should be $\boldsymbol{N = \lambda \cdot t}$
- Consider now a RV described by the time required for the first event to occur, X.
- The probability that the length of time until the event will exceed x is equal to the probability that **no** $(k = 0)$ Poisson events will occur in x.

$$P(k = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!}$$ therefore: $$P(X \geq x) = e^{-\lambda t}$$

The Poisson and Exponential Distribution

CDF for X is: $F(x) = P(0 < X < x) = 1 - e^{-\lambda t}$

if we differentiate CDF: $f(x) = \lambda e^{-\lambda t}$

exponential distribution with $\xi = 1/\lambda$

example from W. Rosenkrantz book:

In a space shuttle a critical component of an experiment has an expected lifetime $\lambda = E(T) = 10$ days and it is exponentially distributed with $\xi = \frac{1}{\lambda} = 0.1$

The mission is scheduled for 10 days.

How many $(n - 1 = ?)$ spare parts should we put on the board if we want to have the probability of success at least equal to 0.99?

(All together we will be using n components).

The Poisson and Exponential Distribution

notation: T_i – the lifetime of the i-th component; W_n – the total lifetime for the n-th component:

$$W_n = T_1 + T_2 + \ldots + T_n.$$

Both T_i and W_n have the same probability density function.
$X(t)$ – number of events (break-downs) occuring in time t.

$$\mathcal{P}(W_n > t) = \mathcal{P}(X(t) < n) = \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}.$$

right-hand member: possible numbers of events (failures) during the time t.

The Poisson and Exponential Distribution

$$\mathcal{P}(W_n > t) = \sum_{k=0}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \ldots.$$

$$\mathcal{P}(W_n > t) = \boxed{\begin{array}{c} \lambda = 0.1 \\ t = 10 \end{array}} = e^{-1}\left[1 + 1 + \ldots + \frac{1}{(n-1)!}\right].$$

If we use two spares ($n - 1 = 2$; $n = 3$) the probability ia a meagre 92%!!
To have the probability of at least 99 percent we must put $n = 5$.
And this means four (!) spares (plus the component in the apparatus).

adapted from Andrzej Lenda

The Normal Distribution

- This is definitely one of the most fundamental PDF with great significance in statistics

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, -\infty < x < \infty$$

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-(v-\mu)^2/2\sigma^2}\, dv$$

- We can also introduce the standardised variable corresponding to X

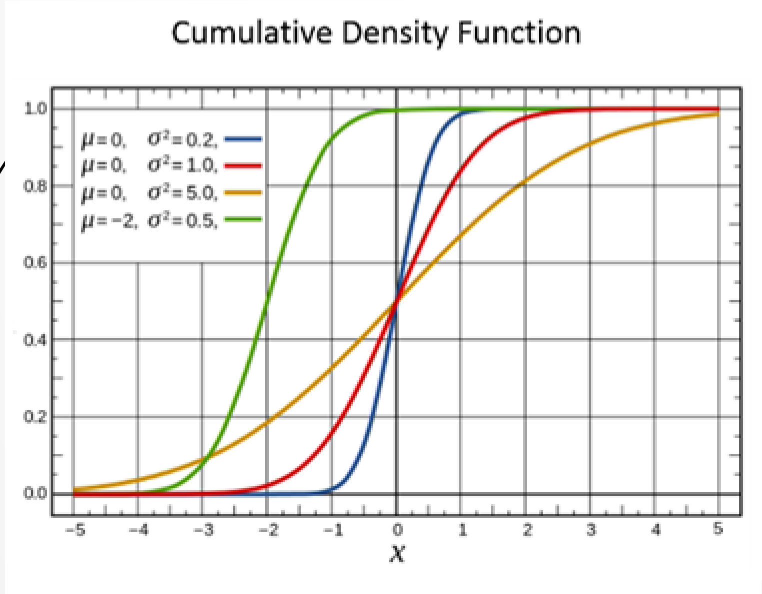
$$Z = \frac{X-\mu}{\sigma}, \mu_Z = 0, \sigma_Z = 1$$

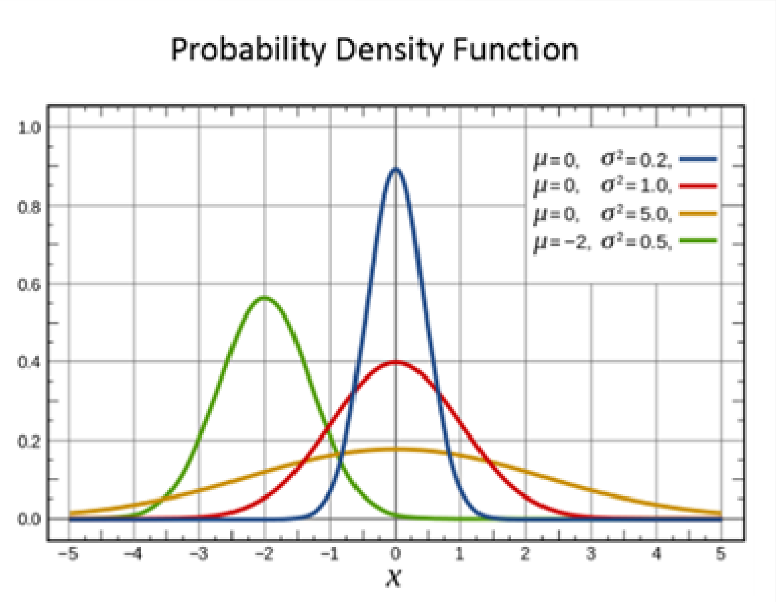
$$f(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty$$

$$F(z) = P(Z \leq z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{z} e^{-v^2/2}\, dv = \frac{1}{2} + \frac{1}{\sigma\sqrt{2\pi}} \int_{0}^{z} e^{-v^2/2}\, dv$$

The Normal Distribution -CDF

- In statistics we deal with two major players: probability density function PDF and cumulative density function CDF





The Normal Distribution

- We then call the Z the standard score and the distribution function $F(Z)$ can be related to error function (tabulated) $\mathrm{erf}(z)$

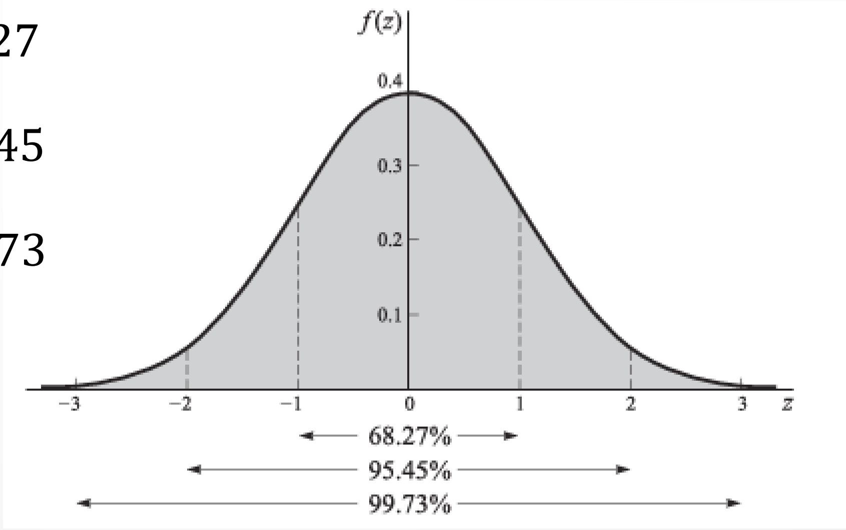
$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-u^2} dv$$

$$F(z) = \frac{1}{2}\left[1 + erf\left(\frac{z}{\sqrt{2}}\right)\right]$$

$$p(-1 < z < 1) = 0.6827$$

$$p(-2 < z < 2) = 0.9545$$

$$p(-3 < z < 3) = 0.9973$$



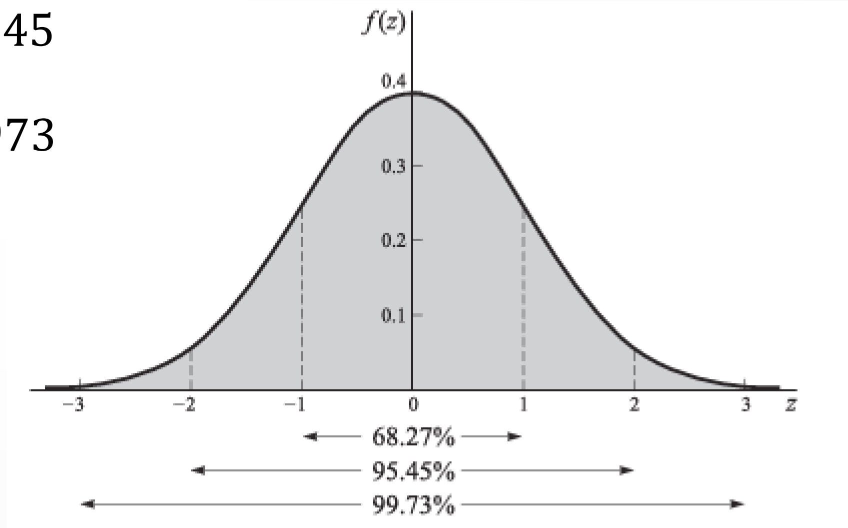
The Normal Distribution

- The Z score tells how many standard deviations the value x is above (to the right of) or below (to the left of) the mean, μ.

$$p(-1 < z < 1) = 0.6827$$

$$p(-2 < z < 2) = 0.9545$$

$$p(-3 < z < 3) = 0.9973$$



The empirical rule known as the **68-95-99.7 rule**.

The Normal Distribution

Exercise (use statistical tables and calculator only):
A citrus farmer who grows mandarin oranges finds that the diameters of mandarin oranges harvested on his farm follow a normal distribution with a mean diameter of 5.85 cm and a standard deviation of 0.24 cm.

a) Find the probability that a randomly selected mandarin orange from this farm has a diameter larger than 6.0 cm. Sketch the graph.
b) The middle 20% of mandarin oranges from this farm have diameters between ______ and ______.
c) Find the 90th percentile for the diameters of mandarin oranges, and interpret it in a complete sentence.
d) The middle 40% of mandarin oranges from this farm are between ______ and ______.
e) Find the 16th percentile and interpret it in a complete sentence.

The Normal Distribution

- The properties of the Gaussian distribution

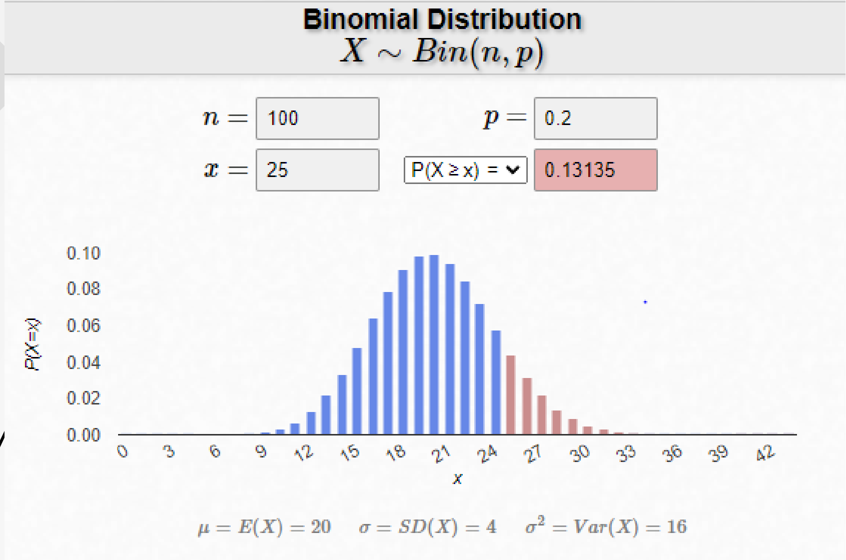
Mean	μ
Variance	σ^2
Standard deviation	σ
Coefficient of skewness	$\alpha_3 = 0$
Coefficient of kurtosis	$\alpha_4 = 3$

- Relation between binomial and normal distribution

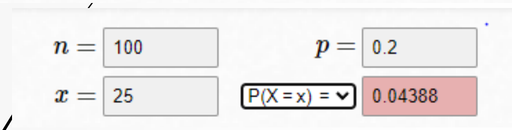
$$Z = \frac{X - np}{\sqrt{npq}}$$

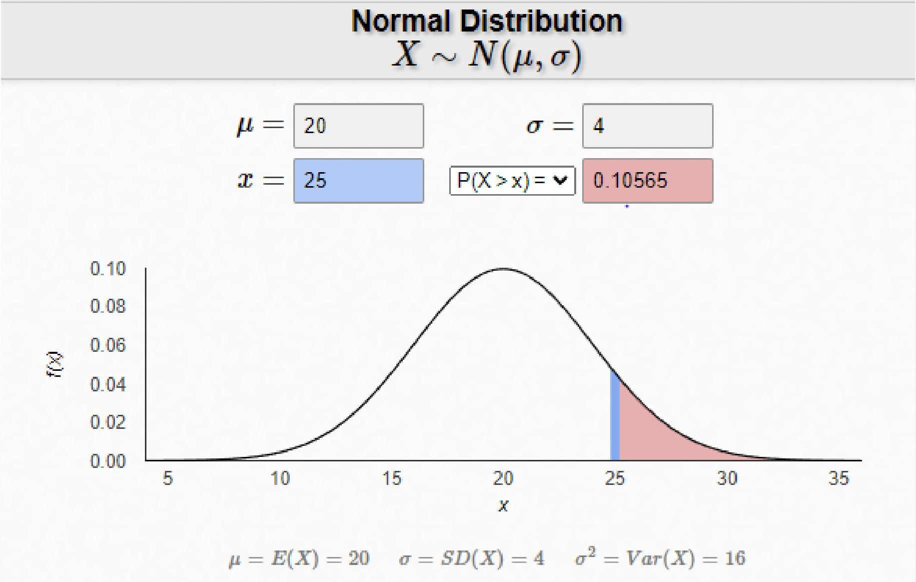
$$\lim_{n\to\infty} p\left(a \le \frac{X - np}{\sqrt{npq}} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-v^2/2}\, dv$$

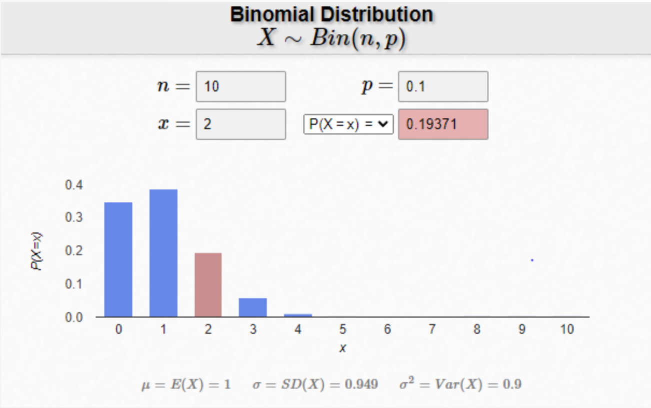
- We say, that the variable $\frac{X-np}{\sqrt{npq}}$ is **asymptotically normal**!


Binomial Distribution
X ~ Bin(n, p)
n = 100
p = 0.2
x = 25
P(X ≥ x) =
0.13135
P(X=x)
0.10
0.08
0.06
0.04
0.02
0.00
0 3 6 9 12 15 18 21 24 27 30 33 36 39 42
x
μ = E(X) = 20
σ = SD(X) = 4
σ² = Var(X) = 16


PROBLEMS
AHEAD


n = 100
p = 0.2
x = 25
P(X = x) =
0.04388


Normal Distribution
X ~ N(μ, σ)
μ = 20
σ = 4
x = 25
P(X > x) =
0.10565
f(x)
0.10
0.08
0.06
0.04
0.02
0.00
5 10 15 20 25 30 35
x
μ = E(X) = 20
σ = SD(X) = 4
σ² = Var(X) = 16


Binomial Distribution
X ~ Bin(n, p)
n = 10
p = 0.1
x = 2
P(X = x) =
0.19371
P(X=x)
0.4
0.3
0.2
0.1
0.0
0 1 2 3 4 5 6 7 8 9 10
x
μ = E(X) = 1
σ = SD(X) = 0.949
σ² = Var(X) = 0.9

More variables

- Especially important for the inference is operating with samples of measurements (data), and usually we have $\boldsymbol{n}$ of them
- Define the CDF for this case:

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1, X_2 < x_2, \dots, X_n < x_n)$$

- And the PDF in this case:

$$f(x_1, x_2, \dots, x_n) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(x_1, x_2, \dots, x_n)$$

- Any marginal PDF of RV x_k

$$g_k(x_k) = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_{k-1} dx_{k+1} \cdots dx_n$$

- … and the mean value for x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} x_k f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$$

More variables

- And the same using the marginal PDF of x_k

$$E[x_k] = \mu_k = \int_{-\infty}^{+\infty} x_k g_k(x_k) dx_k$$

NICE!

- Now, in this convention let's write out the mean, variance and covariance

$$E[x_i] = \boldsymbol{\mu_i}$$

$$E[(x_i - E[x_i])^2] = E[(x_i - \mu_i)^2] = \boldsymbol{\sigma_i^2}$$

$$Cov(x_i, x_j) = E\big[(x_i - E[x_i])\big(x_j - E\big[x_j\big]\big)\big] = E\big[(x_i - \mu_i)\big(x_j - \mu_j\big)\big] = \boldsymbol{c_{ij}}$$

- We can also introduce a pseudo-vector notation

$$\vec{x} = \{x_1, x_2, \dots, x_n\}, \vec{X} = \{X_1, X_2, \dots, X_n\}$$

$$f(\vec{x}) = \frac{\partial^n}{\partial x_1 \partial x_2 \cdots \partial x_n} F(\vec{x})$$

Nice and compact!

More variables

- We can also put all our variances and covariances in one structure that we call **covariance matrix**

$$C = \begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & \ddots & \vdots \\ c_{n1} & \cdots & c_{nn} \end{pmatrix} \quad c_{ii} = \sigma_i^2, c_{ij} = c_{ji}$$

- Also, we can do similar thing („vectorisation”) for the means

$$E[\vec{x}] = \vec{\mu}$$

$$c_{ij} = E\left[(x_i - \mu_i)\left(x_j - \mu_j\right)\right]$$

$$\boldsymbol{C} = \boldsymbol{E}[(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})(\vec{\boldsymbol{x}} - \vec{\boldsymbol{\mu}})^{\boldsymbol{T}}]$$

- The respective elements can be written explicitly (take 2 RV)

$$\vec{x} = \binom{x_1}{x_2}, \vec{x}^T = (x_1, x_2) \quad \vec{\mu} = \binom{\mu_1}{\mu_2}, \vec{\mu}^T = (\mu_1, \mu_2)$$

More variables

- Now make the complete calculations

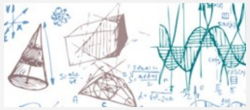
$$(\vec{x}-\vec{\mu})^T = (x_1-\mu_1, x_2-\mu_2), \vec{x}-\vec{\mu} = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$E[(\vec{x}-\vec{\mu})(\vec{x}-\vec{\mu})^T] = \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}(x_1-\mu_1, x_2-\mu_2) =$$

$$= \begin{pmatrix} (x_1-\mu_1)(x_1-\mu_1) & (x_1-\mu_1)(x_2-\mu_2) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)(x_2-\mu_2) \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & c_{12} \\ c_{21} & \sigma_2^2 \end{pmatrix}$$

- We can use our new and compact notation to derive one super important rule in statistics: **error propagation formula**
- It combines variable change and multivariate functions of RV
- Interested already? Go to the next page!

N-dim Gaussian



- Soon, we discuss the central limit theorem that states that the sum on n independent C.R.V. with finite means μ_i and variances σ_i becomes a Gaussian R.V. with mean $\mu = \sum_i \mu_i$ and variance $\sigma = \sum_i \sigma_i$ for n being large.
- Note, this is the justification that the random measurement errors should follow the Gaussian distribution!
- The N-dim generalisation of the Gaussian formula is given by:

$$f(\vec{x};\vec{\mu},C) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} exp\left[-\frac{1}{2}(\vec{x}-\vec{\mu})^T C^{-1}(\vec{x}-\vec{\mu})\right]$$

$|C|$ - determinant of covariance matrix

$$E[x_i] = \mu_i, V[x_i] = \sigma_i = C_{ii}, \mathrm{cov}[x_i, x_j] = C_{ij}$$

N-dim Gaussian

- For the pedagogical reasons let's write explicitly the 2-dim case of Gaussian distribution with $\rho = {}^{\mathrm{cov}[x_i, x_j]}/_{\sigma_1 \sigma_2}$

$$f(x_1, x_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho}} \times$$

$$\times\, exp\left[-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right)\right)\right]$$

- Which for the independent variables reduces to

$$f(x_1, x_2, \mu_1, \mu_2, \sigma_1, \sigma_2, \rho = 0) = \frac{1}{2\pi\sigma_1\sigma_2} \times exp\left[-\frac{1}{2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right]$$