Probability Theory & Statistics

Innopolis University, BS-I,II

Spring Semester 2016-17

Lecturer: Nikolay Shilov

Part I

RANDOM VARIABLES AND THEIR DISTRIBUTIONS

For sake of simplicity…

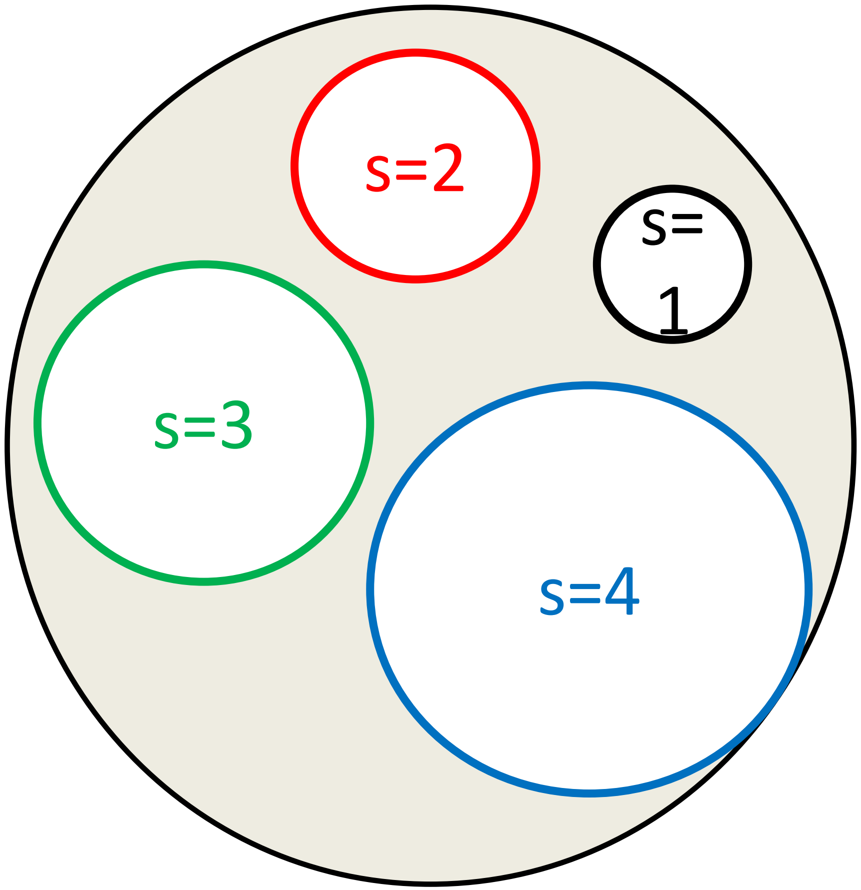
- Let us concentrate one-dimension case (R) but please keep in mind that 2D (R^2), 3D (R^3) and higher dimensional cases are very similar.
- Exercise: draw 1D, 2D, 3D and 4D cubes in R^2.
- Evening suggestion: read and enjoy *Flatland: A Romance of Many Dimensions*, a story (1884) by the English schoolmaster Edwin Abbott.
- Alternative: find and watch a movie inspired by the story!

Update on random variable (definition)

- In particular: a discrete random variable is any (total) real function on finite domain $X:\Omega\rightarrow R$. (Ref. Lecture for week 6.)
- In general: a *random variable* is any (total) function $X:\Omega\rightarrow R$ that range (co-domain) is an interval of real numbers (finite or infinite).

Example

- Consider all circles within a circle region of radius R on the Euclidean plane.
- A random variable S assigning area to each of these circles.



Discrete probability distribution (recall from lecture 6)

- Every discrete random variable defines *probability mass* function P_X and *cumulative distribution function*

$$F_X(x) = P_X(X \leq x) = \Sigma_{y \leq x} P_X(y).$$

Cumulative probability distribution

- Assume that
  - the domain of a random variable X is the sample space of some probability space with probability function P,
  - and for each $x \in R$ the pre-image of $(-\infty, x]$ is an event in this space;
- then (*cumulative probability*) *distribution function* (CDF) is the following function $\Phi: R^+ \to [0,1]$ defined as $\Phi(x) = P(-\infty < X < x)$.

Probability Space Definition (recall from week 11)

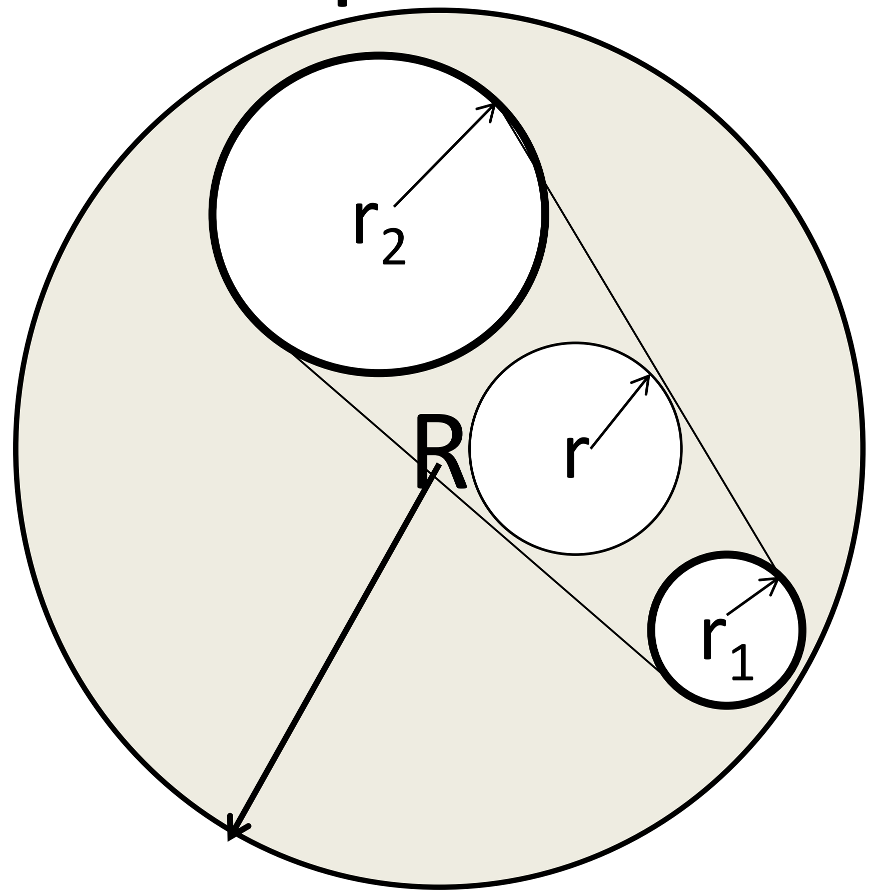
- A probability space is a triple $(\Omega, \mathcal{F}, \mathrm{P})$ where
  - Ω is a finite event/sample space,
  - $\mathcal{F} \subseteq 2^{\Omega}$ is the set of events,
  - and P: $\mathcal{F} \to [0,1]$ a (total) probability function satisfying *axioms*.

Probability Axioms (recall from week 11)

- Non-negativity: $0 \leq P(A)$ for every event;
- Normalization: $P(\Omega)=1$;
- Countable additivity: $P(\cup_{k \in N} A_k) = \Sigma_{k \in N} P(A_k)$ assuming all events are pair-wise exclusive.

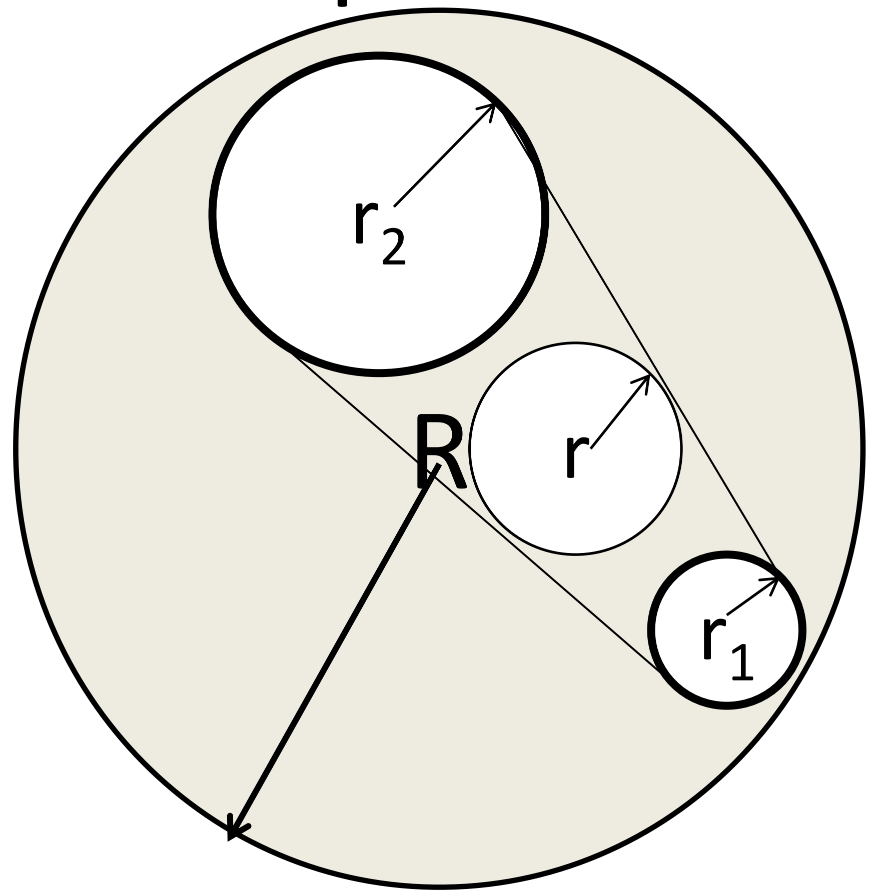
A variant of event space in the circle example

- Let *basic event* be a set of circles with radiuses r that fill some *range interval* ($[r_1,r_2]$).
- Let *event* be any countable set of basic events with disjoint range intervals.



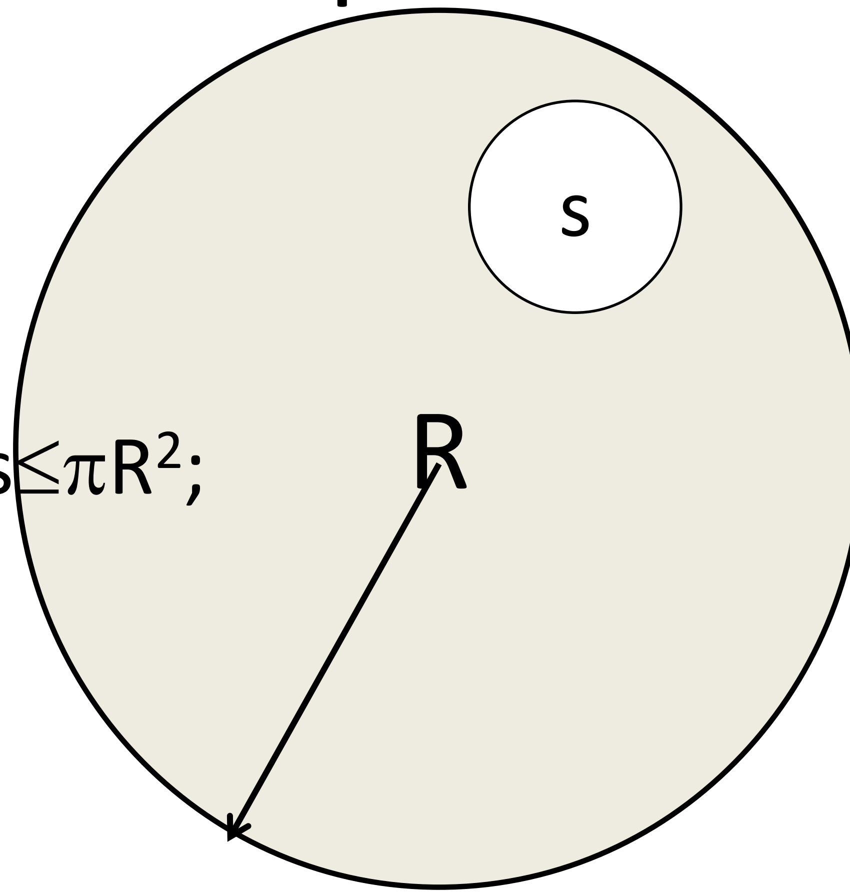
A variant of probability function in the circle example

- Let P for a *basic event* with *range interval* ($[r_1, r_2]$) be $|r_2 - r_1|/R$.
- Expand the probability function on events by countable additivity $P(\cup_{k\in N}A_k)=\Sigma_{k\in N}P(A_k)$.



Probability distribution function in the circle example

$$\Phi(s) = \begin{cases} 1, \text{ if } s>\pi R^2; \\ (s/\pi R^2)^{1/2}, \text{ if } 0<s\leq\pi R^2; \\ 0, \text{ if } s\leq 0. \end{cases}$$



Properties of distribution function of a random variable

- If Φ:R$^+$$\rightarrow$[0,1] is the distribution probability function for a random variable X then
  - Φ(a)$\leq$ Φ(b) for all a$\leq$b (monotonicity);
  - Φ(-∞) =0 and Φ(∞) =1;
  - P(a$\leq$X<b) = Φ(b) – Φ(a) for all a$\leq$b.

Part II

CONTINUOUS DISTRIBUTIONS AND PROBABILITY DENSITY

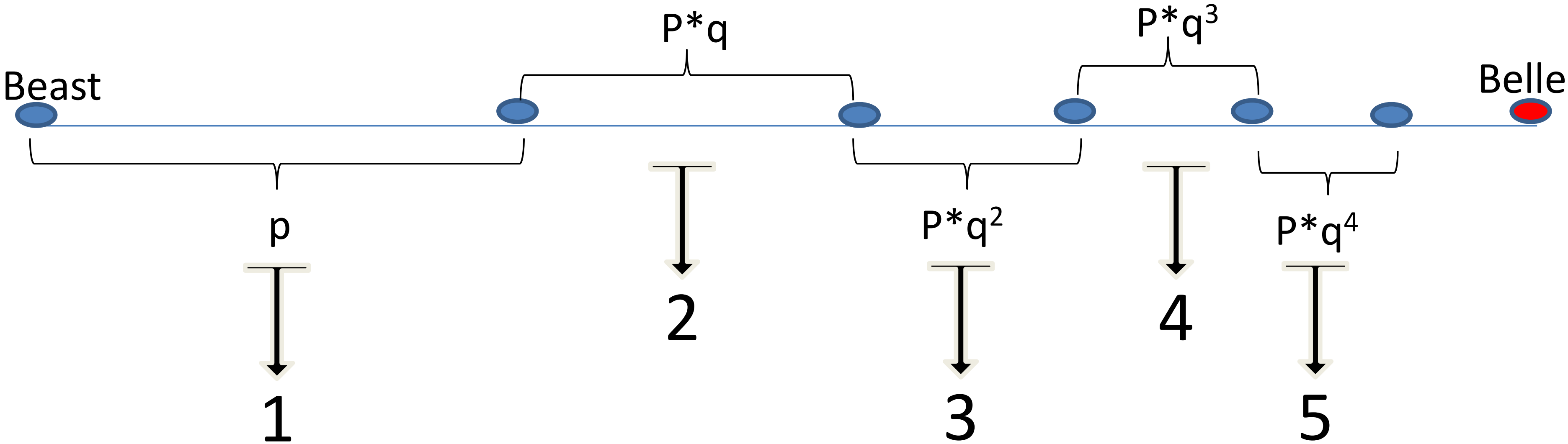
Continuous random variables and their distributions

- if distribution Φ of a random variable X is a continuous function then
  - the distribution Φ is called a *continuous distribution* and
  - the variable X is called a *continuous random variable*.
- If Φ is a continuous distribution than $\Phi(a)=0$ for all $a \in R$.

Example of non-continuous distribution

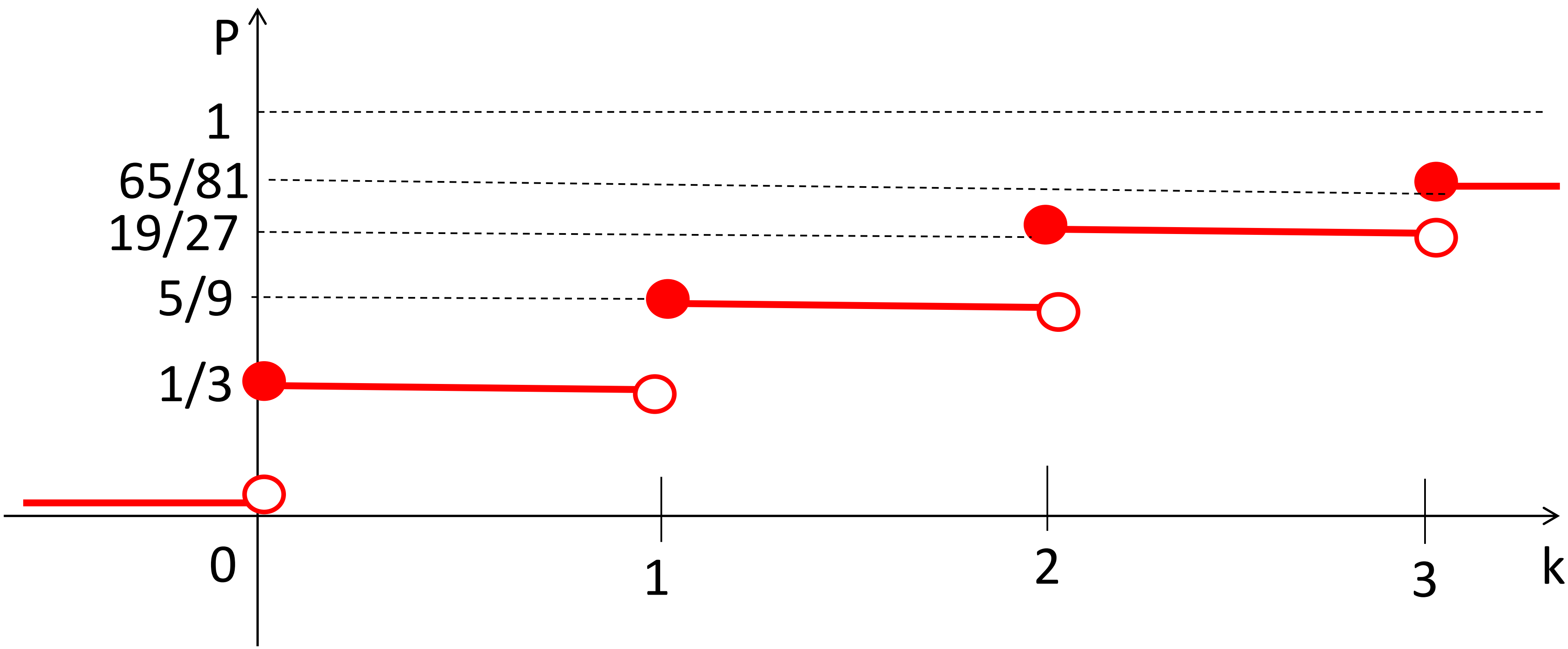
- Let Ω=[0,1) and X= Geom(p) be the staircase function (ref. lecture 7):

$$X(t)=k \text{ on } [(1-q^{(k-1)}p), (1-q^k p)).$$



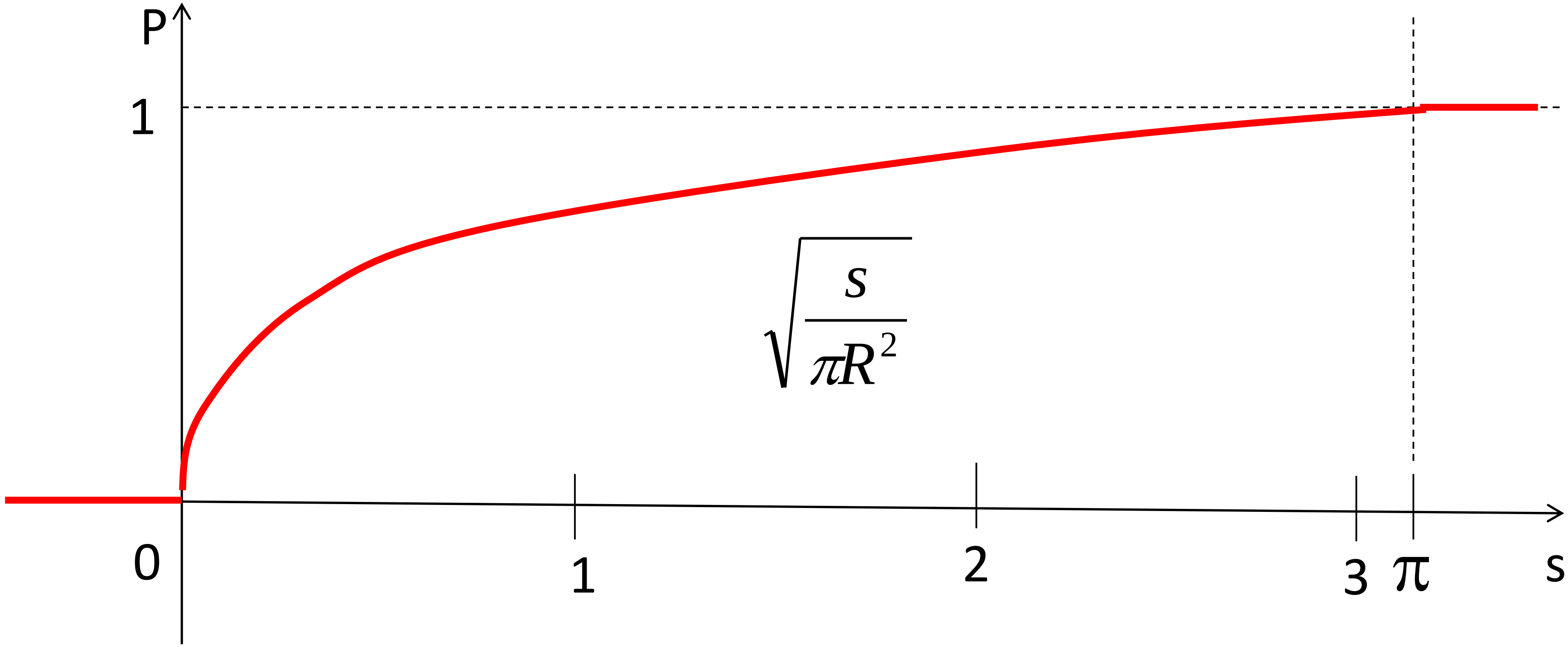
Example of non-continuous distribution (cont.)

- Then distribution Φ for X with p=1/3 is depicted below:



Back to probability distribution in the circle example

- Distribution Φ for circle example in case R=1:



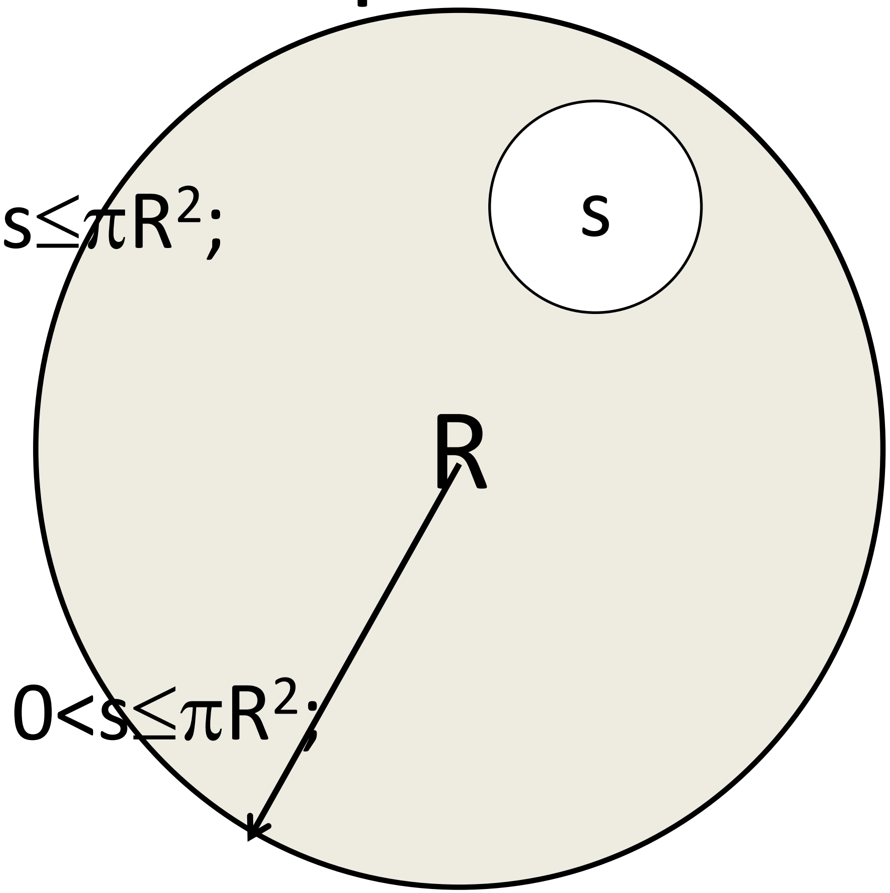
When distribution is differentiable…

- If $\Phi(x) = \int_{\infty}^{x} \varphi(t)dt$ then
  - $\varphi(x) = \Phi'(x)$ and $\varphi(x)$ is called *probability density function* (PDF);
  - $P(a \le X < b) = \Phi(b) - \Phi(a) = \int_{a}^{b} \varphi(t)dt$ for all $a \le b$.

Probability density function in the circle example

$$\Phi(s) = \begin{cases} 1, \text{ if } s>\pi R^2; \\ (s/\pi R^2)^{1/2}, \text{ if } 0<s\leq\pi R^2; \\ 0, \text{ if } s\leq 0; \end{cases}$$

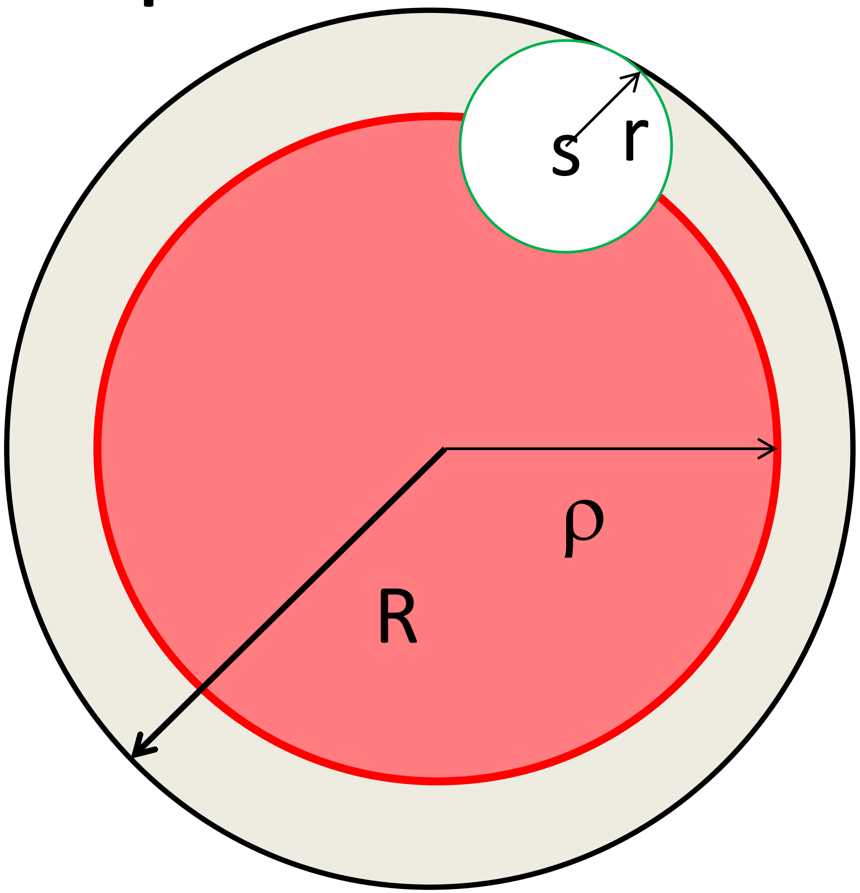
$$\varphi(s) = \begin{cases} 0, \text{ if } s>\pi R^2; \\ 1/[2R(\pi s)^{1/2}], \text{ if } 0<s\leq\pi R^2; \\ 0, \text{ if } s\leq 0. \end{cases}$$



Probability distribution in the circle example revised

Since the area of the red region where can be placed the center of a circle with area s is

$$\pi R^2 - 2R\sqrt{\pi s} + s$$



Probability distribution in the circle example revised (cont.)

hence $\varphi(s) = \dfrac{\pi R^2 - 2R\sqrt{\pi s} + s}{\pi R^2}$ may be adopted as a probability density function.

- Exercises:
  - draw graph of the probability density;
  - find the probability distribution;
  - draw graph of the probability distribution.

Moments of a random variable with probability density

- Let k>0 be an integer and X be a continuous random variable with distribution $\Phi(x)$ and density $\varphi(x)$.
- Hint: think interval $[x, x+\Delta x]$ as a “fat real value” (with intention $\Delta x \to 0$) and probability $\varphi(x)*\Delta x$ in the definition of k-th moment

$$M(X^k) = \Sigma_{x \in R} x^k *(\varphi(x)*\Delta x).$$

Moments of a random variable with probability density

- Then
  - k-th (initial) moment of X must be defined as
$$M(X^k) = \int_{-\infty}^{\infty} x^k \varphi(x)dx$$
  - and k-th central moment – as
$$M[(X - E(X))^k] = \int_{-\infty}^{\infty} [x - E(X)]^k \varphi(x)dx$$

Variance

- In particular, variance is the second central moment, i.e.

$$D(X) = M[(X - E(X))^2] = \int_{-\infty}^{\infty} [x - E(X)]^2 \varphi(x) dx$$

- Exercise: prove that D(X) = M(X^2) – [M(X)]^2.

Part III

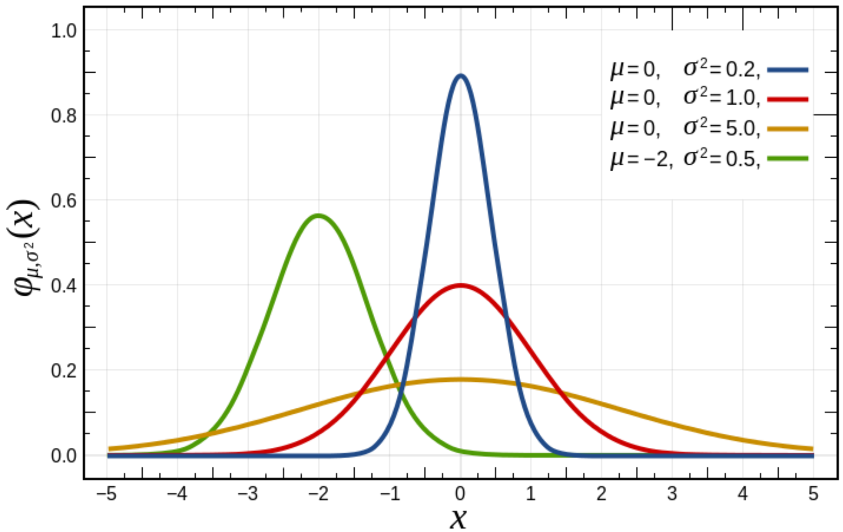
NORMAL DISTRIBUTION AND CENTRAL LIMIT THEOREM

The standard normal distribution: PDF

The standard normal distribution has the probability density function

$$\varphi_0(x) = e^{-x^2/2} \Big/ \sqrt{2\pi}$$

(please r lecture 9).



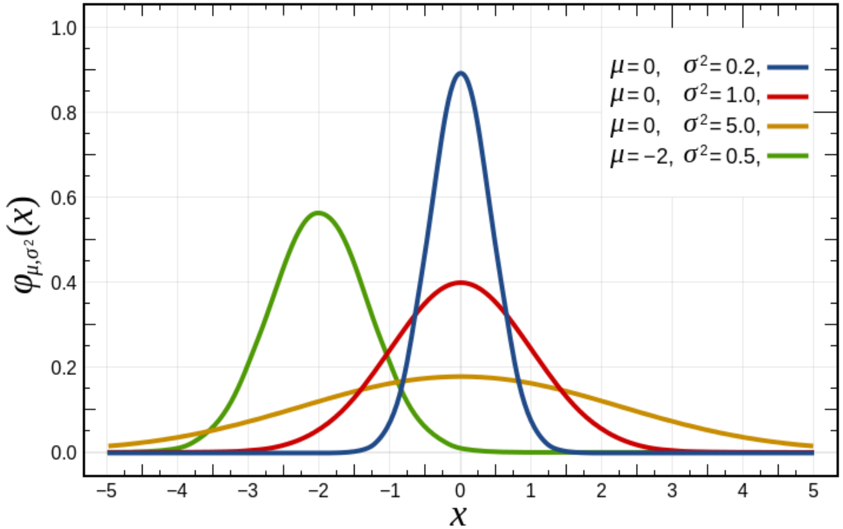
https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

The standard normal distribution: PDF

The standard normal distribution has the probability density function

$$\varphi_0(x) = e^{-x^2/2} \Big/ \sqrt{2\pi}$$

(please refer to Moivre-Laplace theorems in lecture 9).



https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

Exercises (with a non-trivial part)

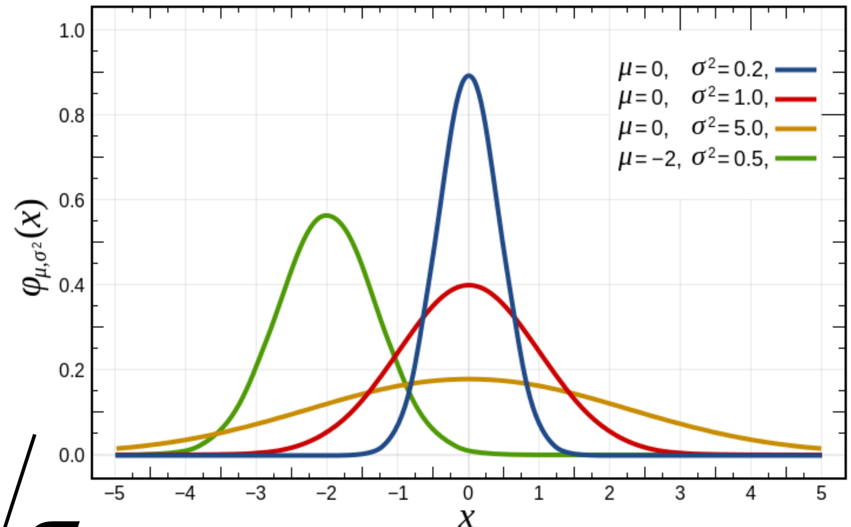
- Prove that:
  - $\int_{-\infty}^{\infty} \varphi_0(x)dx = 1$
  - $\int_{-\infty}^{\infty} x\varphi_0(x)dx = 0$
  - $\int_{-\infty}^{\infty} x^2\varphi_0(x)dx = 1$

What are these integrals for (in terms of) the standard normal distribution?

Normal (Gaussian) distribution: PDF

Every normal distribution is a version of the standard one with stretched domain with factor σ and then translated by μ:

$$f(x \mid \mu, \sigma^2) = \varphi_0\left(\frac{x-\mu}{\sigma}\right) \Big/ \sigma$$



https://en.wikipedia.org/wiki/File:Normal_Distribution_PDF.svg

Exercises (simple this time)

- Prove that:
  - $$\int_{-\infty}^{\infty} f(x \mid \mu, \sigma^2)dx = 1$$
  - $$\int_{-\infty}^{\infty} xf(x \mid \mu, \sigma^2)dx = \mu$$
  - $$\int_{-\infty}^{\infty} x^2 f(x \mid \mu, \sigma^2)dx = \sigma^2$$

What are these integrals for (in terms of) the normal distribution?

Normal (Gaussian) distribution: CDF

If

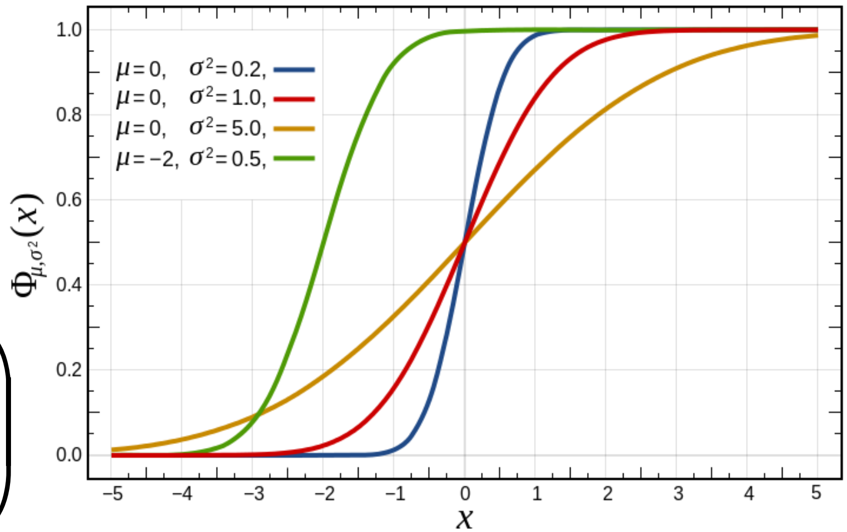
$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

then

$$\Phi(x) = \int_{-\infty}^{x} \varphi(t)dt = $$

$$= \frac{1}{2} + \Phi_0\left(\frac{x-\mu}{\sigma}\right)$$

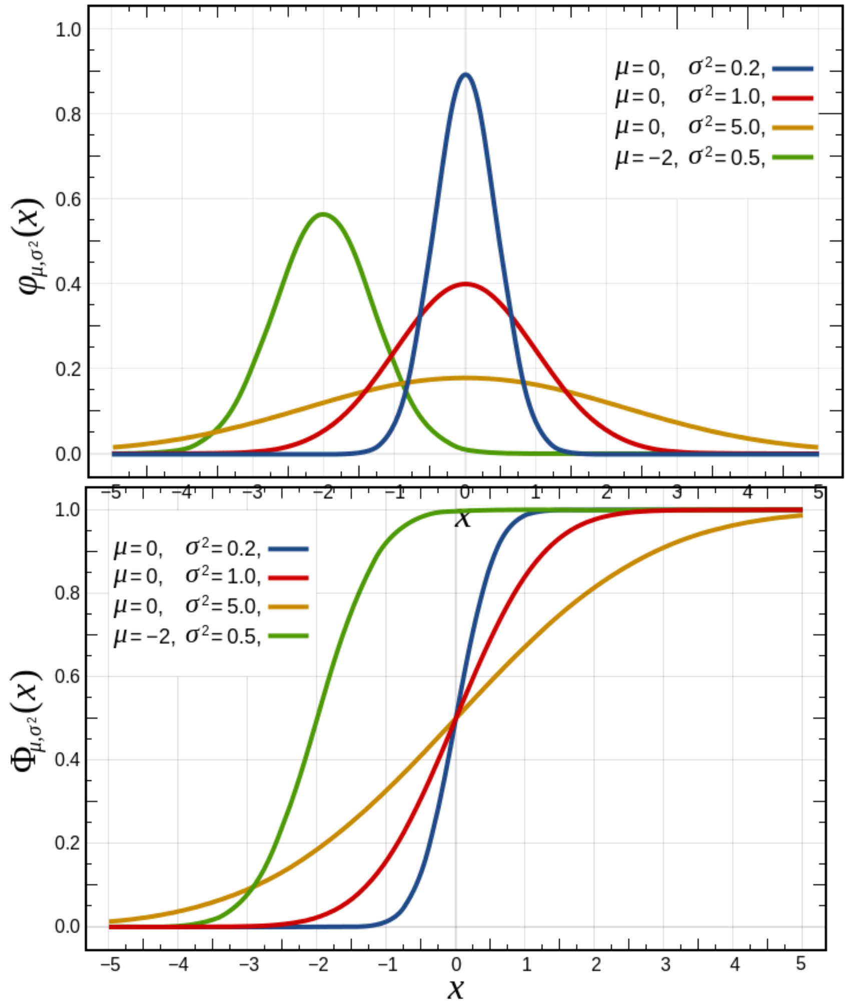
where function Φ_0 is defined in lecture 9.



https://en.wikipedia.org/wiki/File:Normal_Distribution_CDF.svg

Notation

- $X = N(\mu,\sigma^2)$ means that X is a random variable with normal distribution.
- In particular: $X = N(0,1)$ means that X has the standard normal distribution.



Central Limit Theorem

- Let X_1, X_2, … be an infinite sequence of IID (i.e. independent and identically distributed) random variables with a finite expectation μ and finite non-zero deviation σ.
- Let $S_{\mathbf{n}} = X_1 + \ldots + X_n$.
- Then $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ *converges in distribution* to N(0,1) (as $n \to \infty$).

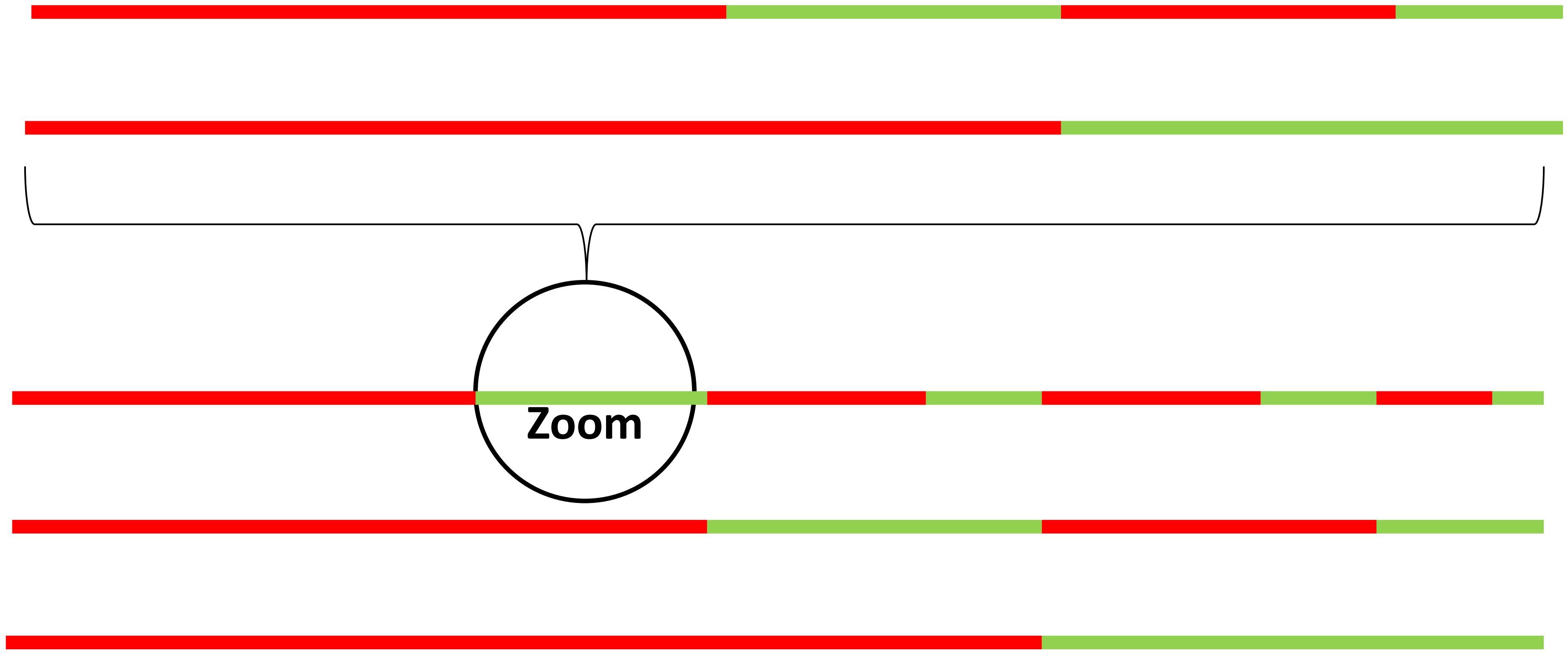
Why???

- De Moivre-Laplace theorems: for every series of Bernoulli trials X_1, … X_n, … where X_n= binomial(n, p) and 0<p<1

$$\frac{X_n - np}{\sqrt{npq}} \xrightarrow{n\to\infty} N(0,1)$$

- Khintchin's law of large numbers: for every X_1, X_2, … sequence of IID random variables with expectation μ and deviation $\sigma\neq 0$, for every $\varepsilon>0$ the probability $P(|\underline{\mathbf{X}}_{\mathbf{n}} - \mu|>\varepsilon) \to 0$.

IID's can't be "very random" but must form a *fractal*



68-95-99.7 rule

- The 68–95–99.7 rule is a shorthand to remember the percentage of values that lie within a band around the mean in a normal distribution:
  - 68.27%,
  - 95.45% and
  - 99.73%

of the values lie within distance σ, 2σ and 3σ of the mean μ.

Toward 68-95-99.7 rule

- If $X=N(\mu,\sigma^2)$ then

  $P(a\leq X\leq b) = \Phi(b) - \Phi(a) =$

  $= (1/2 + \Phi_0[(b-\mu)/\sigma]) - ([1/2 + \Phi_0[(a-\mu)/\sigma]) =$

  $= \Phi_0[(b-\mu)/\sigma] - \Phi_0[(a-\mu)/\sigma];$
- In particular, for any $k>0$:

  $P(|X - \mu|<k\sigma) =$

  $= P(\mu - k\sigma< X < \mu + k\sigma) =$

  $= \Phi_0(k) - \Phi_0(-k) = 2\Phi_0(k).$

68-95-99.7 rule

- According to lecture 9:

x	0	0.5	1	1.5	2	2.5	3
Φ_0	0	0.192	0.341	0.433	0.477	0.494	0.499

- In particular – правило трех сигм:

$$P(|X - \mu| < 3\sigma) \approx 0.998.$$