Università degli Studi di Pavia
Dipartimento di Ingegneria Industriale e dell'Informazione

Corso di Identificazione dei Modelli e Analisi dei Dati

Central Limit Theorem and Law of Large Numbers

Prof. Giuseppe De Nicolao, Federica Acerbi, Alessandro Incremona

Outline

- Rejection Sampling
- Law of Large Numbers
- Central Limit Theorem

Random number generator:

- **Empirical procedure** (e.g. flip of a real coin)

- **Inversion method** (when we can compute the inverse of the cumulative distribution function)

```
>> random
>> icdf
```

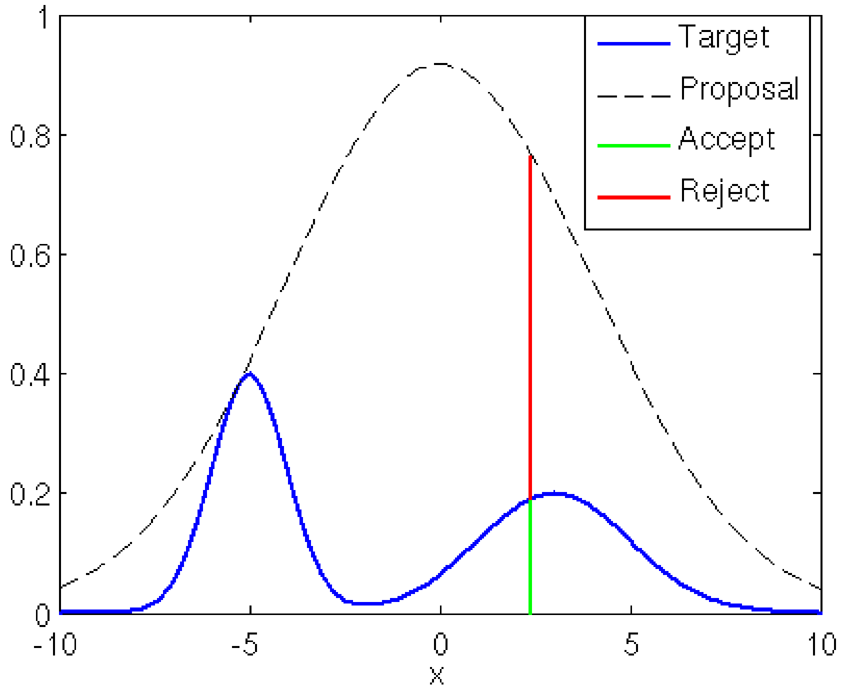
- **Monte Carlo methods:**
  - Rejection sampling
  - Metropolis –Hastings algorithm
  - etc

Rejection Sampling

Idea behind the rejection sampling method:

Sampling from a probability distribution $g(x)$ that is easy to sample from and applying some rejection criterion such that the samples that are accepted are distributed according to $f(x)$.

The probability distribution $g(x)$ has to have an important property, namely, $g(x)$ has to envelope the target distribution $f(x)$. That means, given a scaling factor k, it has to be $kg(x) > f(x)$ for all x.



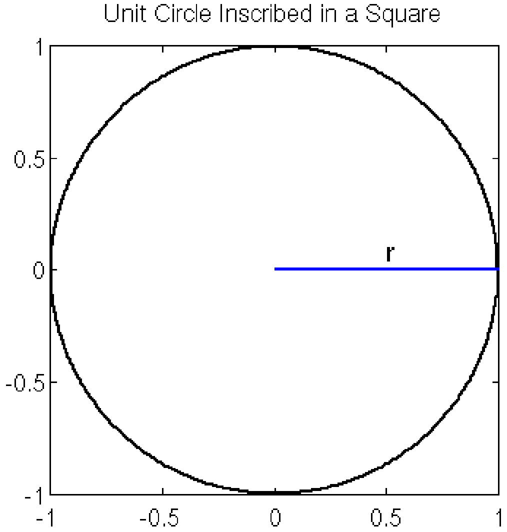
Exercise 1

1. Sample points uniformly from a circle centred in 0 with radius 1

   Idea:
   we could sample Cartesian spatial coordinates x and y uniformly from the interval (-1,1) and reject those points that lie outside of the radius $r = \sqrt{x^2 + y^2} = 1$

2. Approximate the value π, knowing that:

$$\pi = 4\frac{A_{circle}}{A_{square}} = 4\frac{NuberOfSamplesInsidetheCircle}{TotalNumberofSamples}$$



LLN (Law of Large Numbers)

DEF: Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\sigma^2 := Var[X_i] < +\infty$ and $m := E[X_i]$. Then,

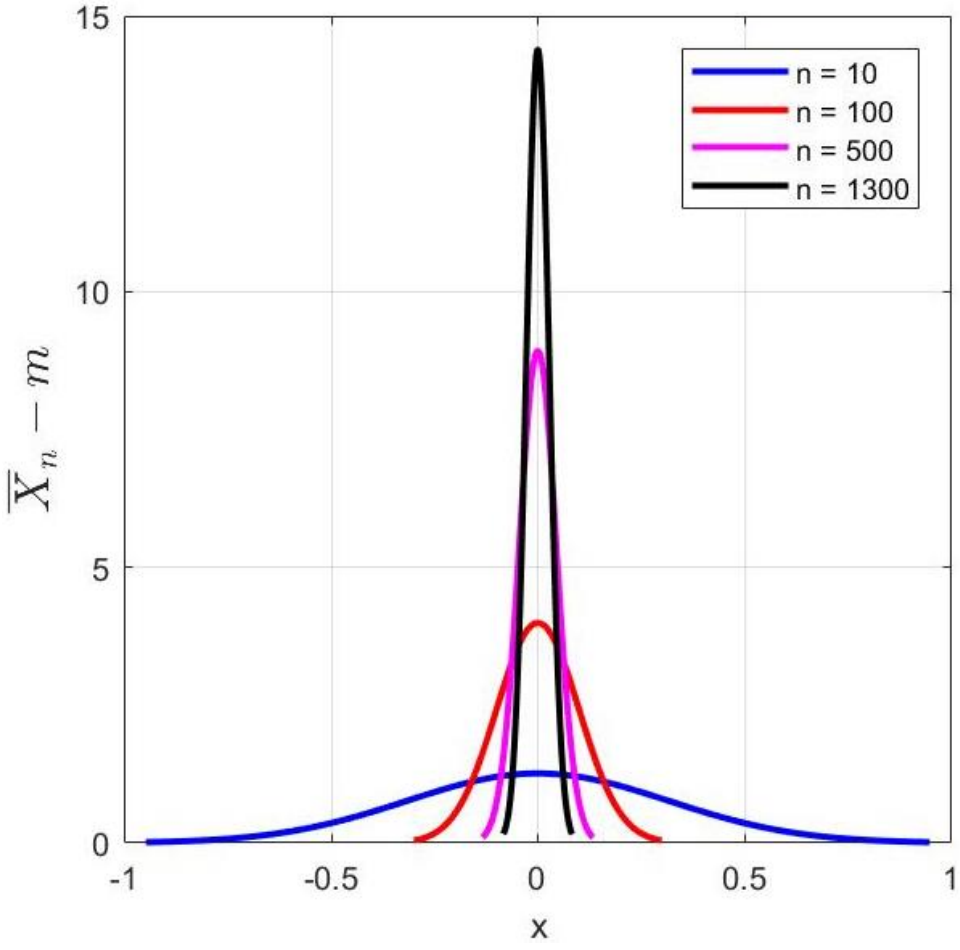
$$\lim_{n\to\infty} E\left[\left(\overline{X}_n - m\right)^2\right] = 0$$

where $\overline{X}_n := \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean.

LLN (Law of Large Numbers)

$$\lim_{n\to\infty} E\left[\overline{X}_n\right] = \lim_{n\to\infty} \frac{1}{n}\, n\, m = m$$

$$\lim_{n\to\infty} Var\left[\overline{X}_n\right] = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$$



Exercise 2

1. Generate n1 = 10 samples from a Normal distribution with m = 2 and sigma = 3 and compute the sample mean. Repeat the experiment 1000 times (**hint**: use a `for` loop) and store all the sample means in a vector.

2. Repeat point 1 with n2 = 100 and store the new sample means values in another vector.

3. Plot, in the same figure, the two histograms of the two sample means vectors and notice the differences.

4. Plot, in the same figure, the theoretical distributions of the two sample means vectors.

CLT (Central Limit Theorem)

<u>*DEF*</u>: Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\sigma^2 := Var[X_i] < +\infty$

and $m := E[X_i]$. Let

$$\widetilde{X}_n := \sum_{i=1}^{n} X_i$$ be the cumulative sum,

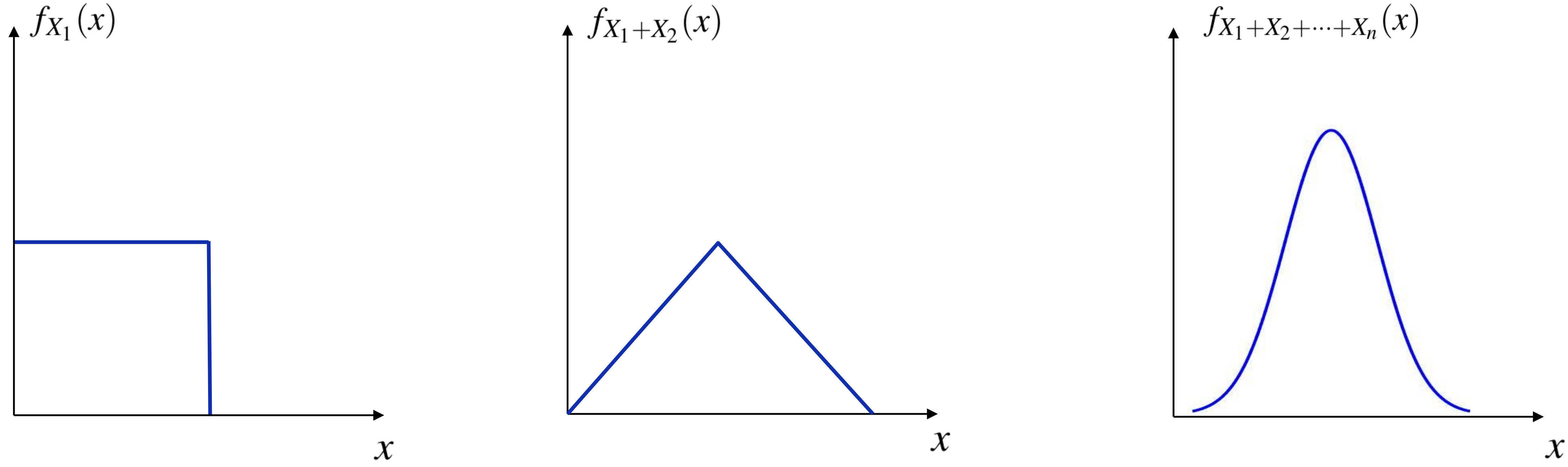
$$S_n := \frac{\widetilde{X}_n - n\,m}{\sqrt{n\,\sigma^2}}$$ be the standardized cumulative sum,

Then $\widetilde{X}_n$ converges in distribution to $\mathcal{N}(nm, n\sigma^2)$

and S_n converges in distribution to the standard Normal distribution $\mathcal{N}(0,1)$.

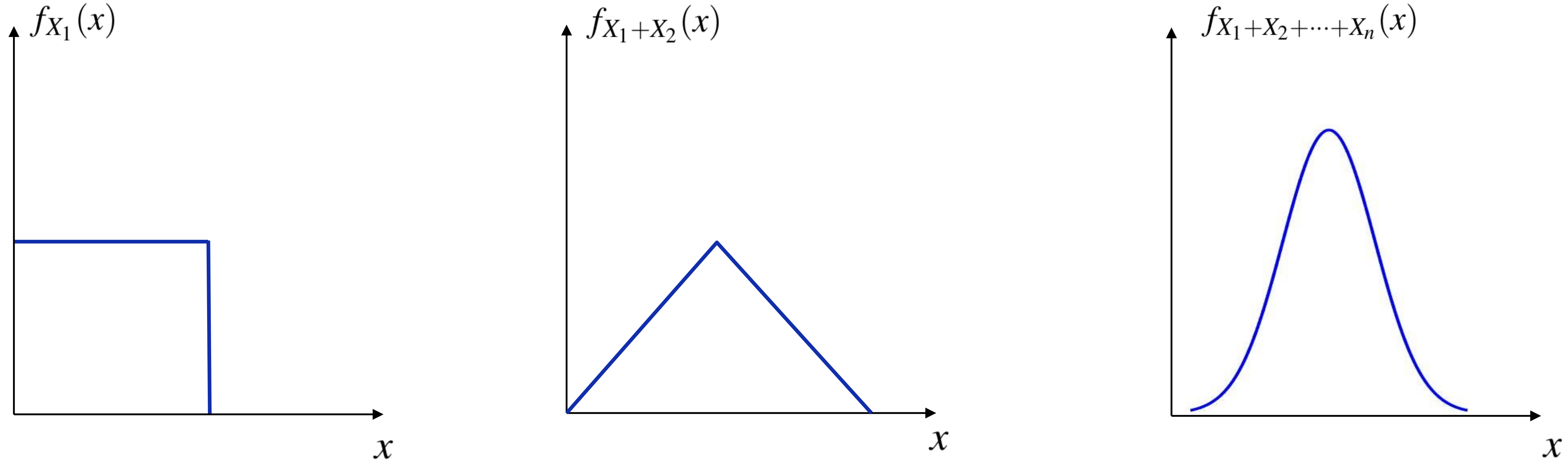
CLT (Central Limit Theorem)

Example: X_i uniform i.i.d.



CLT (Central Limit Theorem)

Convolution ⟶ $$f_Z(z) = \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x)\,dx = f_X(z) * f_Y(z)$$



Exercise 3

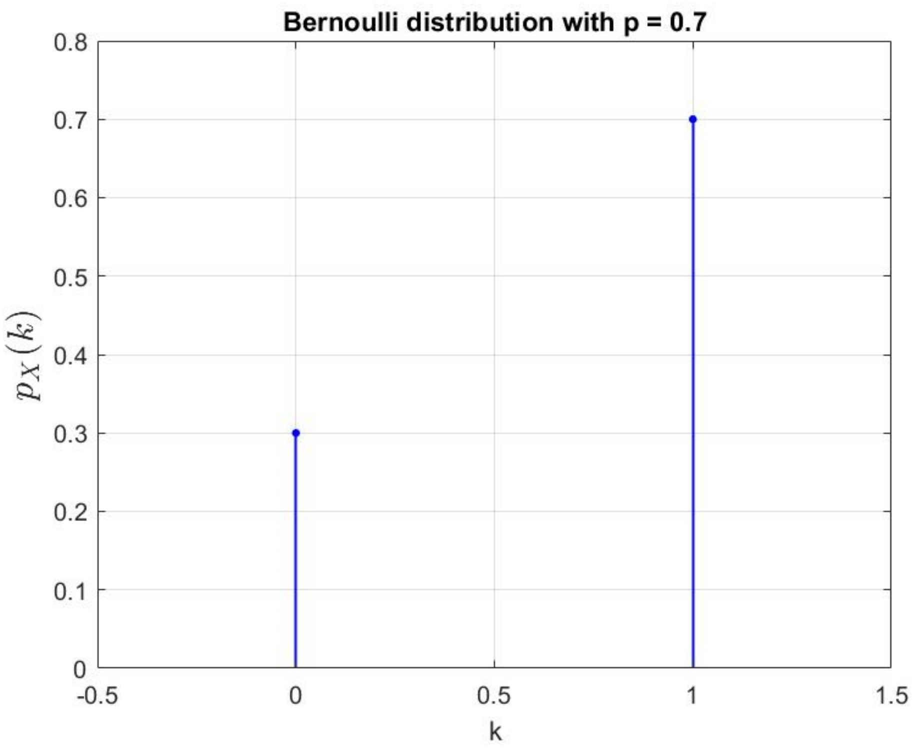
1. Generate n = 1000 samples from an Uniform distribution with a = -2 and b = 2 and compute the cumulative sum. Repeat the experiment 1000 times (**hint**: use a `for` loop) and store all the cumulative sums in a vector, then standardize the vector.

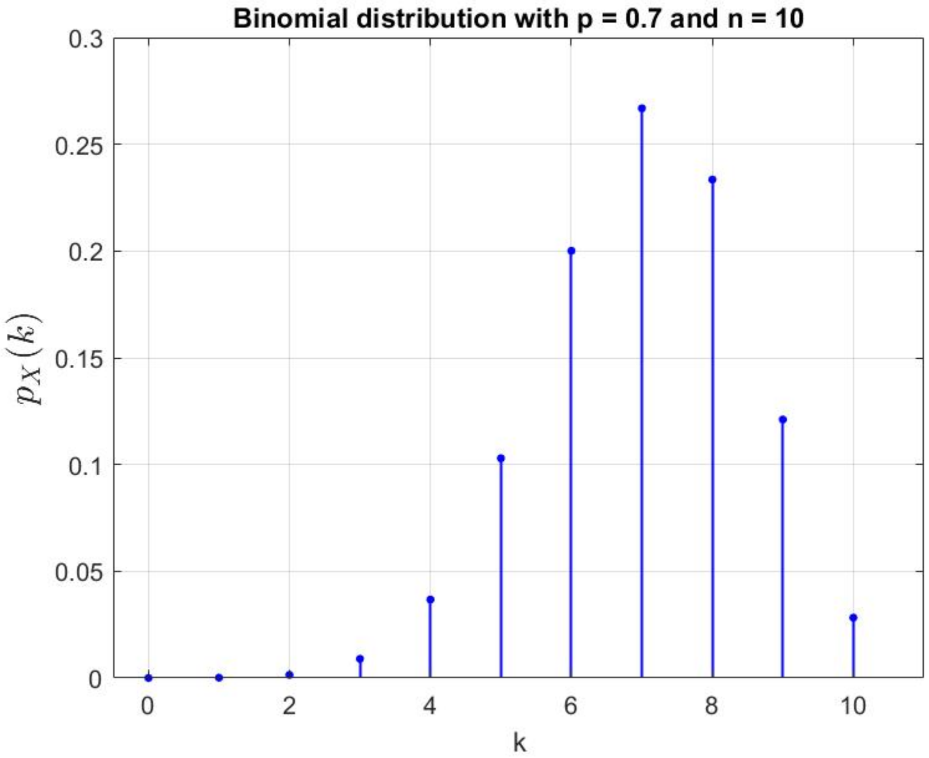
2. Plot, in same figure, the histogram of the standardized cumulative sums vector and the theoretical standard Normal distribution.

3. Repeat points 1 and 2 considering an Exponential distribution with lambda = 1.

4. Repeat points 1, 2 and 3 with n = 2. In the exponential case, which distribution does the histogram of the standardized cumulative sums approximate?

CLT (Central Limit Theorem): binomial distribution

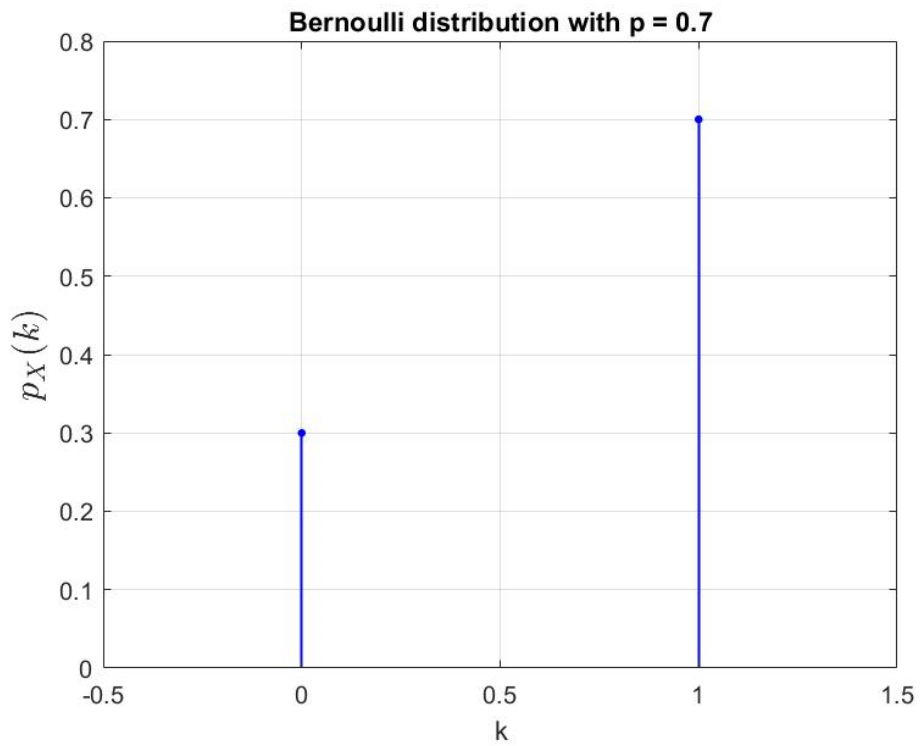


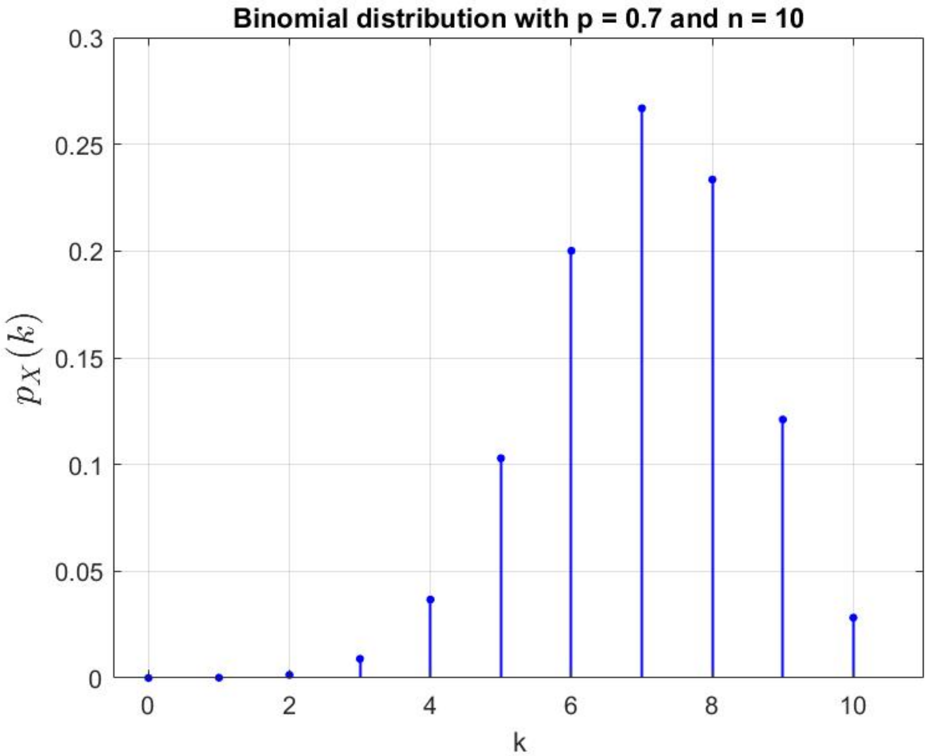


$$p_X(k) = \begin{cases} 1-p, & \text{if } k=0 \\ p, & \text{if } k=1 \end{cases}$$

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

CLT (Central Limit Theorem): binomial distribution





Convolution $\longrightarrow$ $$p_Z(z) = \sum_{x=-\infty}^{+\infty} p_X(x) p_Y(z-x) = p_X(z) * p_Y(z)$$

CLT (Central Limit Theorem): binomial distribution

```
p = 0.7; % prob. of success

f = [(1-p) p]; % prob. of k = 0, 1 success given 1 trial

f2 = conv(f,f); % prob. of k = 0, 1, 2 successes given 2 trials

f10 = f;
for i = 1:9 % prob. of k = 0, 1, ..., 10 successes given 10 trials
    f10 = conv(f10, f);
end

figure; stem([0 : 1], f); grid on;
figure; stem([0 : 2], f2); grid on;
figure; stem([0 : 10], f10); grid on;
```

EXERCISE: compute and plot ''f50''.

Exercise 4

1. Create a binomial probability distribution object with parameters n = 10 (number of trials) and p = 0.7 (probability of success).

2. Create a normal distribution object with parameters m and sigma equal respectively to the mean and the standard deviation of the binomial distribution created on point 1.

3. Compute the cumulative distribution functions (using the command `cdf`) of both distributions.

4. Plot, in the same figure, the two cumulative distribution functions. What happens if you increase the value of n?