وزارة الإتصـــــــالات
وتكنولوجيا المعلومـــات

Probability and Statistics for Machine Learning

AI TALENTS HUB

EPITA
ÉCOLE D'INGÉNIEURS EN INFORMATIQUE

Session 3

Probability and Statistics for Machine learning

Contents

Recap on Sessions 1 & 2

Recap and further discussion

Covariance

$$Cov(X,Y) = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \overline{X})(Y_i - \overline{Y})$$

where

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{n} X_i \text{ and } \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

are the means of X, Y

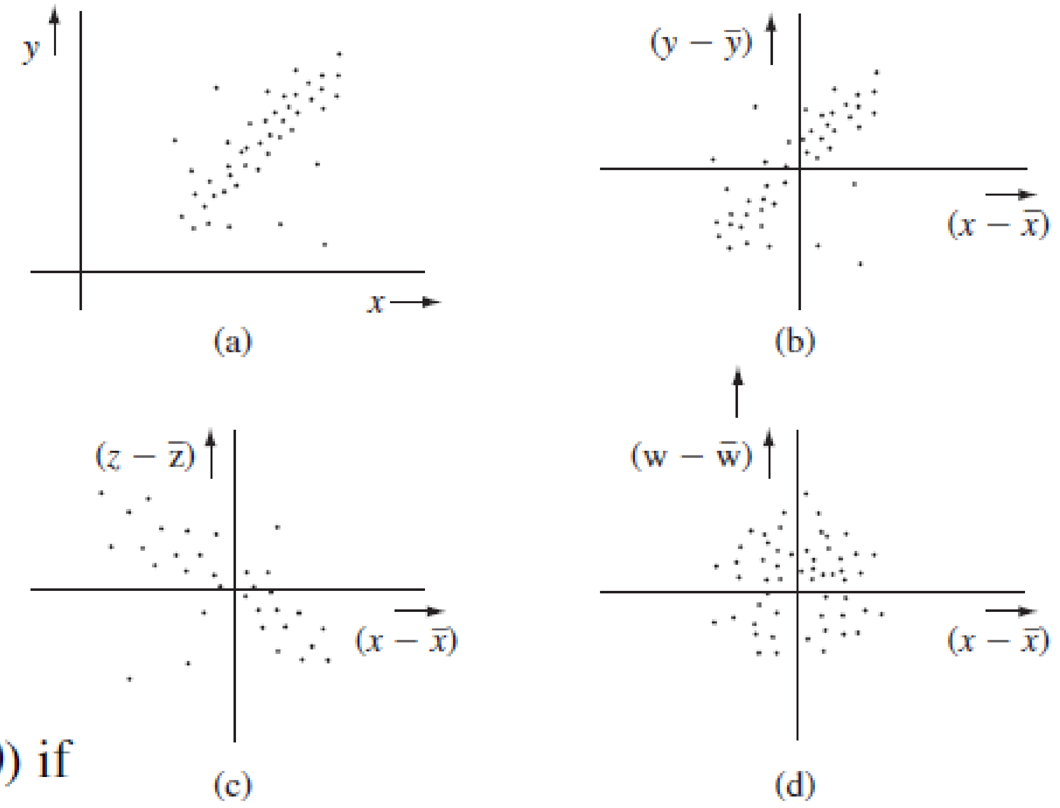
Covariance and Correlation are the tools that we have to measure if the two attributes are related to each other or not.

- **Covariance** measures how two variables vary in tandem to their means.

Covariance

$$
\begin{aligned}
\sigma_{xy} &= \overline{(x-\bar{x})(y-\bar{y})} \\
&= \overline{xy} - \overline{\bar{x}y} - \overline{x\bar{y}} + \overline{\bar{x}\bar{y}} \\
&= \overline{xy} - \bar{x}\bar{y} - \bar{x}\bar{y} + \bar{x}\bar{y} \\
&= \overline{xy} - \bar{x}\bar{y}
\end{aligned}
$$

Figure 7.18 Scatter diagrams: (a), (b) positive correlation; (c) negative correlation; (d) zero correlation.



it follows that the variables x and y are uncorrelated ($\sigma_{xy} = 0$) if

$$\overline{xy} = \bar{x}\bar{y}$$

The Oxford Series in Electrical and Computer Engineering) Ding, Zhi; Lathi, Bhagwandas Pannalal - Modern digital and analog communication systems-Oxford University Press, USA (2018).

Discrete Probability Distributions

Introducing discrete random variables and discrete distributions

Discrete Random Variables

- A random variable is discrete if there exists a denumerable sequence of distinct numbers x_i such that

$$\sum_{i} P_{\mathrm{x}}(x_i) = 1$$

Discrete RV

- A random variable that can only assume a finite number of values or an infinite sequence of values such as 0, 1, 2, ….
- **Probability Mass Function (PMF)**

- Discrete Random Variable
  - A random variable is discrete if there exists a denumerable sequence of distinct numbers x_i such that

$$\sum_i P_x(x_i) = 1$$

Probability Mass Function (PMF)

$$0 \le P_X(x) \le 1, \qquad (2.26a)$$

$$\sum_x P_X(x) = 1. \qquad (2.26b)$$

When developing the probability mass function for a random variable, it is useful to check that the PMF satisfies these properties.

Cumulative Distribution function (CDF)

The CDF of a random variable, X, is

$$F_X(x) = \Pr(X \le x).$$

These properties of CDFs are summarized as follows:

(1) $F_X(-\infty) = 0, F_X(\infty) = 1,$

(2) $0 \le F_X(x) \le 1,$

(3) For $x_1 < x_2, F_X(x_1) \le F_X(x_2),$

(4) For $x_1 < x_2, \Pr(x_1 < X \le x_2) = F_X(x_2) - F_X(x_1).$

CDF of a Uniformly Distributed RV

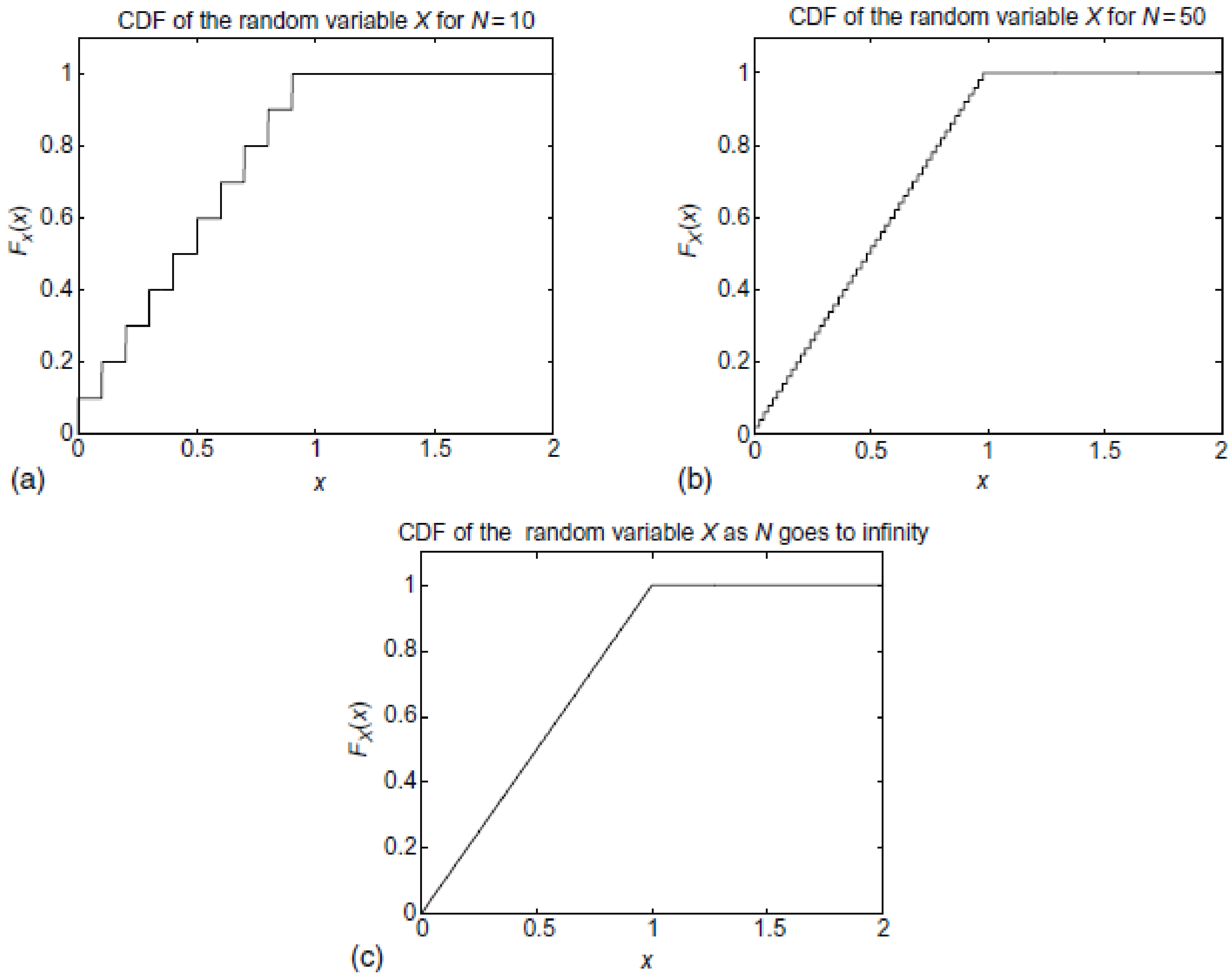


Figure 3.2
CDF of the random variable X for (a) $N = 10$, (b) $N = 50$, and (c) $N \to \infty$.

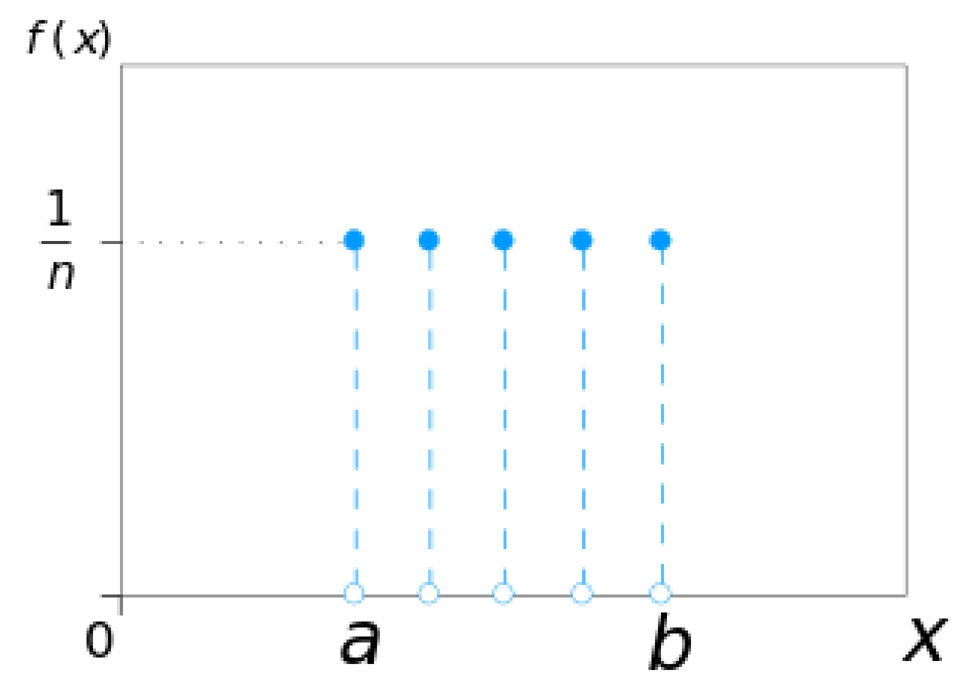
Discrete Probability Distributions

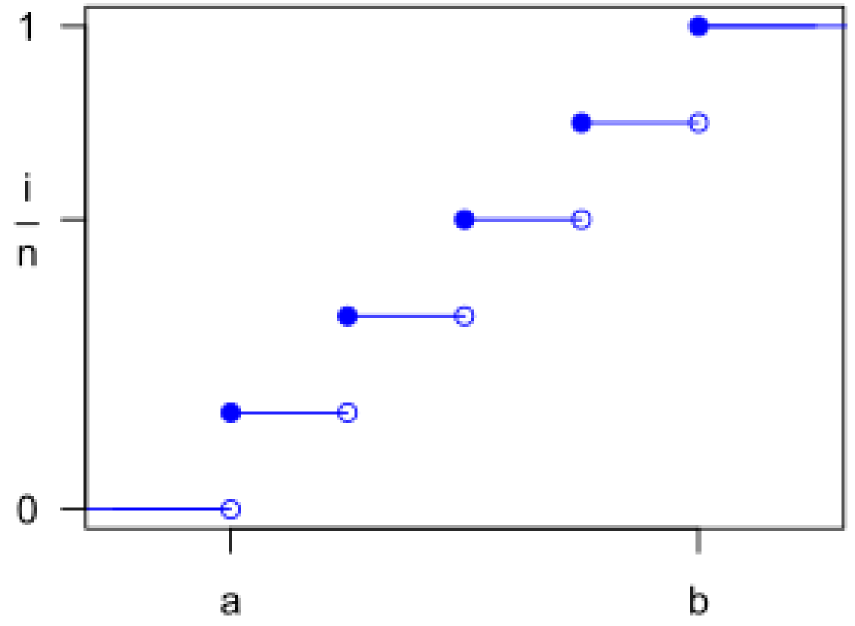
- **Uniform**
- **Binomial**
- **Bernoulli**
- **Poisson**
- **Geometric**
- **Hypergeometric**
- **…**

Discrete Uniform Distribution

PMF	$\frac{1}{n}$
CDF	$\frac{\lfloor k \rfloor - a + 1}{n}$

PMF and CDF





Bernoulli Trials

http://www.math.wichita.edu/history/topics/probability.html#bern-trials

- Boy? Girl? Heads? Tails? Win? Lose? Do any of these sound familiar? When there is the possibility of only two outcomes occuring during any single event, it is called a Bernoulli Trial. Jakob Bernoulli, a profound mathematician of the late 1600s, from a family of mathematicians, spent 20 years of his life studying probability. During this study, he arrived at an equation that calculates probability in a Bernoulli Trial. His proofs are published in his 1713 book *Ars Conjectandi* (Art of Conjecturing).

Some examples of Bernoulli Trials

http://en.wikipedia.org/wiki/Bernoulli_trial

- Flipping a coin. In this context, obverse ("heads") denotes success and reverse ("tails") denotes failure. A fair coin has the probability of success 0.5 by definition.
- Rolling a die, where for example we designate a six as "success" and everything else as a "failure".
- In conducting a political opinion poll, choosing a voter at random to ascertain whether that voter will vote "yes" in an upcoming referendum.
- Call the birth of a baby of one sex "success" and of the other sex "failure." (Take your pick.)

What constitutes a Bernoulli Trial?

- To be considered a **Bernoulli trial**, an experiment must meet each of three criteria:

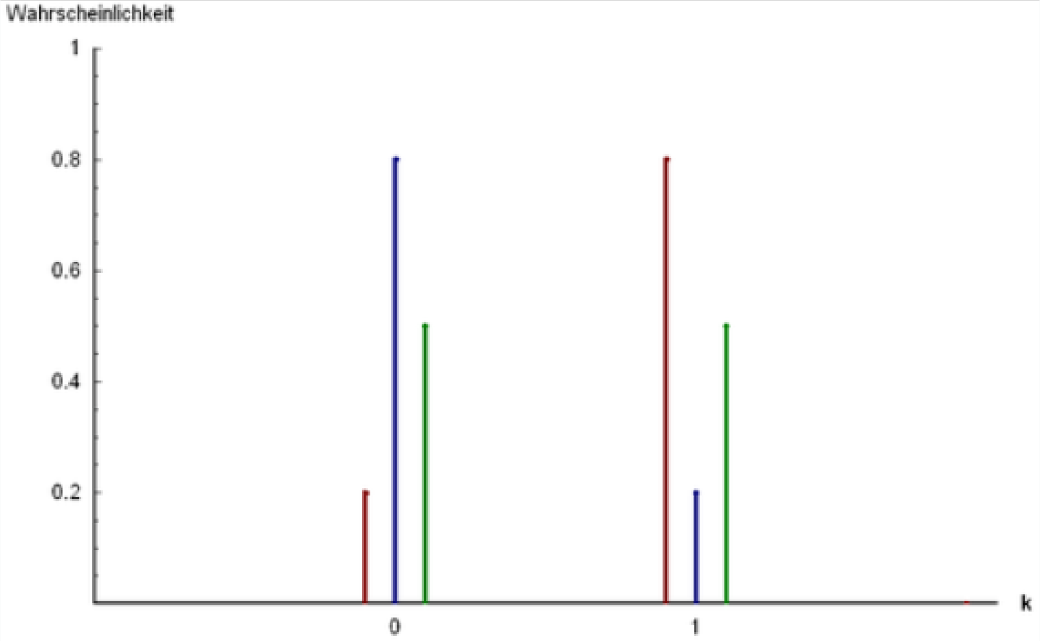
- There must be **only 2 possible outcomes**, such as: black or red, sweet or sour. One of these outcomes is called a **success**, and the other a **failure**. Successes and Failures are denoted as S and F, though the terms given do not mean one outcome is more desirable than the other.

- Each outcome has a **fixed probability** of occurring; a success has the probability of ***p***, and a failure has the probability of ***1 - p***.
- Each experiment and result are completely **independent** of all others.

Bernoulli Distribution

PMF	$\begin{cases} q = 1-p & \text{if } k = 0 \\ p & \text{if } k = 1 \end{cases}$ $p^k(1-p)^{1-k}$
CDF	$\begin{cases} 0 & \text{if } k < 0 \\ 1-p & \text{if } 0 \le k < 1 \\ 1 & \text{if } k \ge 1 \end{cases}$

Probability mass function



Three examples of Bernoulli distribution:

$P(x=0) = 0.2$ and $P(x=1) = 0.8$

$P(x=0) = 0.8$ and $P(x=1) = 0.2$

$P(x=0) = 0.5$ and $P(x=1) = 0.5$

Binomial Distribution

- Describes processes whose trials have only two possible outcomes.

- The binomial distribution gives the discrete probability distribution of obtaining exactly n successes out of N Bernoulli trials (where the result of each Bernoulli trial is true with probability p and false with probability 1-p). The binomial distribution is therefore given by

Binomial Distribution

Characteristics of the Binomial Distribution

-The binomial distribution requires that the experiment's trials be independent.

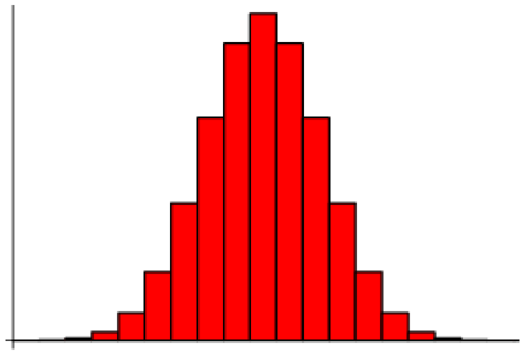
-Each trial has two outcome; a success outcome with probability P a failure out come with probability (1-p)

Binomial Formula

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \qquad \textbf{(4)}$$

where:

n = Random sample size
x = Number of successes (when a success is defined as what we are looking for)
$n - x$ = Number of failures
p = Probability of a success
$q = 1 - p$ = Probability of a failure
$n! = n(n-1)(n-2)(n-3)\ldots(2)(1)$
$0! = 1$ by definition



Plot of Binomial probabilities with n = 20 trials, p = 0.5

Example: Binomial Distribution

- **Studies show that 60 % of US families use physical aggression to resolve conflict. If 10 families are selected at random, find the probability that the number that use physical aggression to resolve conflict is:**

- **exactly 5**
- **Between 5 and 7 , inclusive**
- **over 80 % of those surveyed**
- **fewer than nine**
- **Solution: P(x = 5) =** $\binom{10}{5} \cdot 0.6^5 (1-0.6)^{(10-5)}$
- Probability (between 5 and 7) inclusive)=Prob(5) or prob(6) or prob(7) =

$$\binom{10}{5} 0.60^5 (0.40)^5 + \binom{10}{6} (0.6)^6 (0.4)^4 + \binom{10}{7} (0.6)^7 (0.4)^3$$

Example: Binomial Distribution

- A manager of a department store has determined that there is a probability of 0.30 that a particular customer will buy at least one product from his store. If three customers walk in a store, find the probability that two of three customers will buy at least one product.
- 1. Determine which two will buy at least one product.
- The outcomes are b b b′ (first two buy and third does not buy) or b b′ b , or b′ b b .
- There are three possible outcomes each consisting of two b's along with one not b (b'). Considering "buy" as a success, the probability of success is 0.30. Each customer is independent of the others and there are two possible outcomes, success or failure (not buy) .

Example: Binomial Distribution

A balanced coin was tossed three consecutive times, if the random variable X is defined as the number of heads appeared in the outcomes, find the probability distribution of X using:

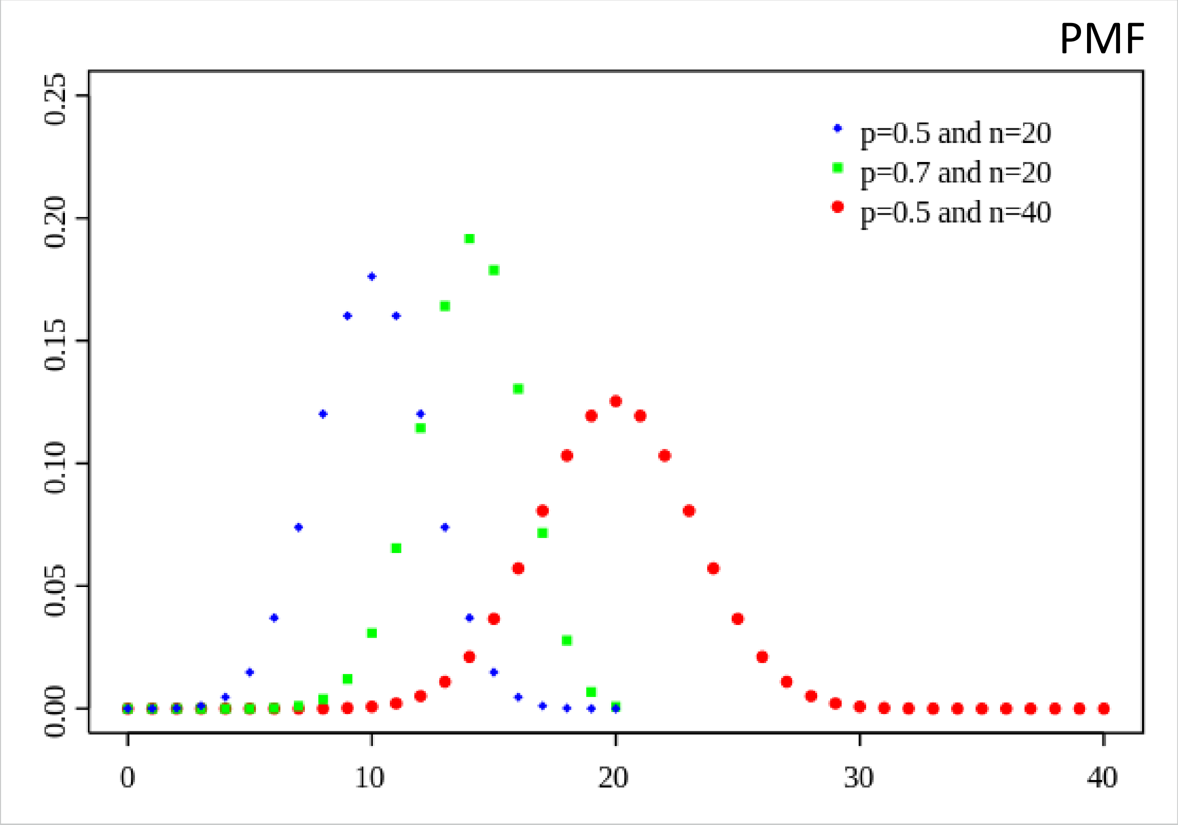
(a) The sample space of this experiment

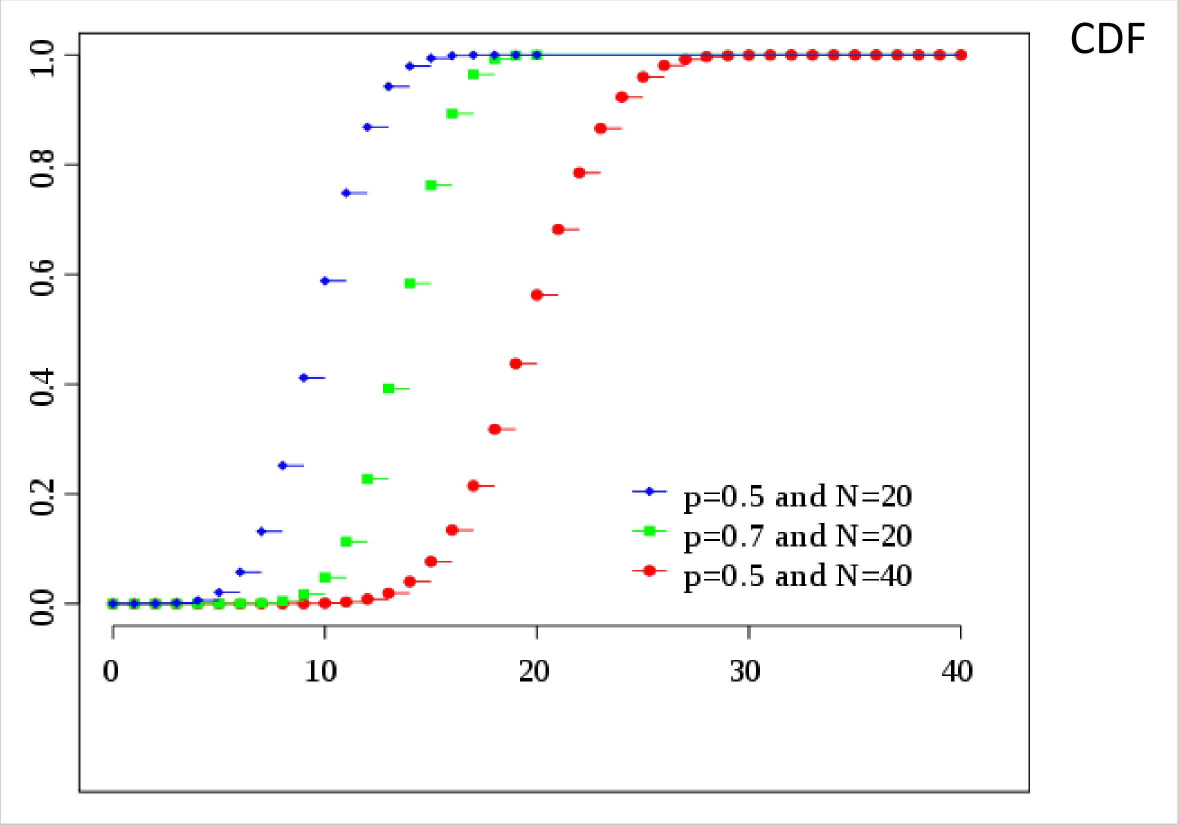
(b)The binomial distribution function

(c)Graph the probability distribution of X

Binomial Distribution

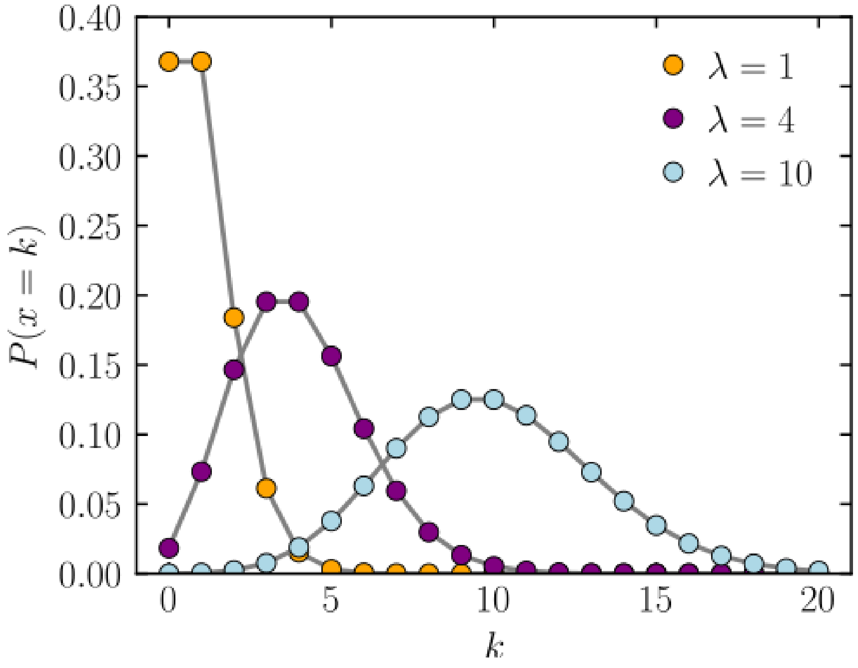
PMF	$\binom{n}{k} p^k q^{n-k}$
CDF	$I_q(n-k, 1+k)$





Poisson Random Variables

$$f(k;\lambda) = \Pr(X{=}k) = \frac{\lambda^k e^{-\lambda}}{k!},$$



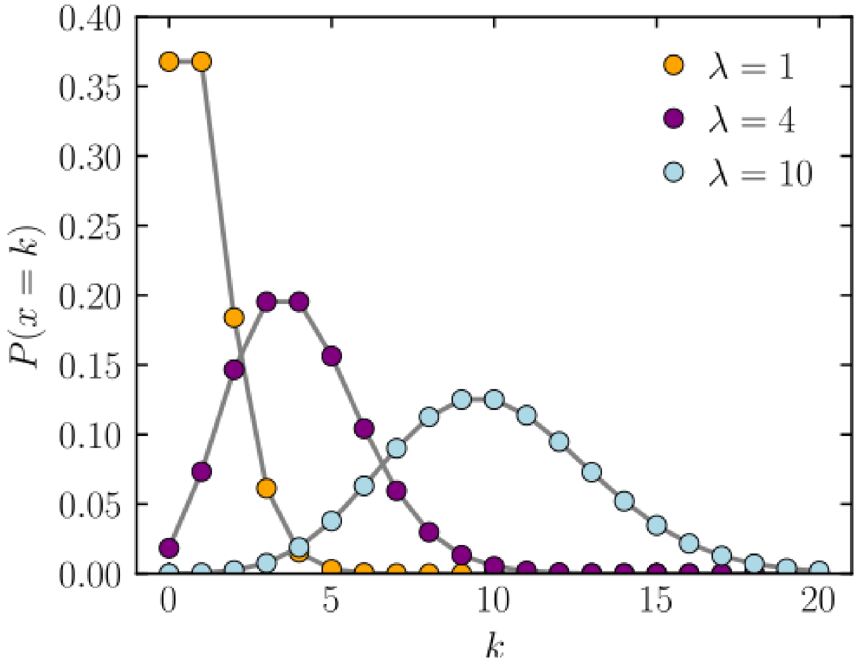
Expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate (lambda, λ) and independently of the time since the last event.

Can also be used for the number of events in other specified interval types such as distance, area or volume.

For example, a call center receives an average of 180 calls per hour, 24 hours a day. The calls are independent; receiving one does not change the probability of when the next one will arrive.

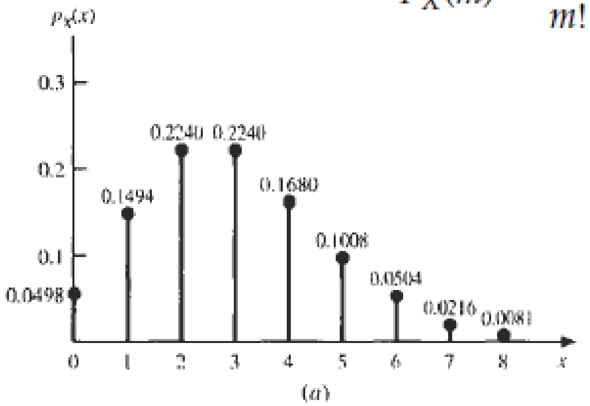
PMF	$\frac{\lambda^k e^{-\lambda}}{k!}$
CDF	$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !}$, or $e^{-\lambda} \sum_{j=0}^{\lfloor k \rfloor} \frac{\lambda^j}{j!}$

A Binomial random variable can be approximated by a Poisson random variable



C. *Poisson Random Variable* Consider a binomial random variable, X, where the number of repeated trials, n, is very large. In that case, evaluating the binomial coefficients can pose numerical problems. If the probability of success in each individual trial, p, is very small, then the binomial random variable can be well approximated by a *Poisson random variable*. That is, the Poisson random variable is a limiting case of the binomial random variable. Formally, let n approach infinity and p approach zero in such a way that $\lim_{n\to\infty} np = \alpha$. Then the binomial probability mass function converges to the form

$$P_X(m) = \frac{\alpha^m}{m!}e^{-\alpha}, \quad m = 0, 1, 2, \ldots, \tag{2.34}$$



Poisson distribution with $\lambda = 3$.

Statistical Tools for Discrete Probability Distributions

Statistical tools for Discrete Distributions

- **The mean of a discrete probability distribution** is also called the expected value of the random variable from an experiment.
- The expected value (Eq. 1) is a weighted average of the random variable values in which the weights are the probabilities assigned to the values.

Expected Value of a Discrete Probability Distribution

$$E(x) = \sum xP(x) \quad \textbf{(1)}$$

where:

$E(x)$ = Expected value of x

x = Values of the random variable

$P(x)$ = probability of the random variable taking on the value x

Statistical tools for Discrete Distributions

- **The standard deviation** measures the spread, or dispersion, in a set of data. The standard deviation also measures the spread in the values of a random variable. To calculate the standard deviation for a discrete probability distribution.

Standard Deviation of a Discrete Probability Distribution

$$\sigma_x = \sqrt{\Sigma[x - E(x)]^2 P(x)} \qquad (2)$$

where:

x = Values of the random variable

$E(x)$ = Expected value of x

$P(x)$ = Probability of the random variable taking on the value x

Example 1

Finding the mean & the standard deviation of a Bernoulli distribution

$E[x] = 0(1-p) + 1(p) = p$

Can you find the standard deviation?

Example 2

Finding the mean & the standard deviation of a Binomial distribution

See next slide.

Binomial Distribution (Mean, Std. dev.)

Expected Value of a Binomial Distribution

$$\mu_x = E(x) = np \tag{5}$$

where:

n = Sample size

p = Probability of a success

Standard Deviation of the Binomial Distribution

$$\sigma = \sqrt{npq} \tag{6}$$

where:

n = Sample size
p = Probability of a success
$q = 1 - p$ = Probability of a failure

Questions?

References

- B,P. Lathi and Z. Ding , “Modern Digital and Analog communication Systems” (Fourth Edition 2009).
- Schaum’s outline, “Probability, Random variables, and Stochastic Process”.
- Papoulis and Pillai, “Probability, Random Variables and Stochastic Process”.