Statistical Methods in Natural Language Processing (NLP)

Class 4: Distributions and Random Variables

Charalambos (Haris) Themistocleous

Department of Philosophy, Linguistics and Theory of Science, Centre for Linguistic Theory and Studies in Probability

GÖTEBORGS UNIVERSITET

Preceding Class

- the *Law of Total probability*
- *Conditional Probability* $P(A|B) = P(A, B)/P(B)$, this is simply the probability of A, given that B occurred and is fundamental for understanding probability theory.
- The conditional Probability *is* Probability $P(A|B)$ is a probability function for any fixed B.
- Any theorem that holds for probability also holds for conditional probability. We use this as the definition of conditional probability.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$$

- We introduced the *Bayes' Rule*, which unites marginal, joint, and conditional probabilities.

GÖTEBORGS UNIVERSITET

Random Variables

There are two main categories of random variables,

1. the discrete
2. the continuous.

GÖTEBORGS UNIVERSITET

Discrete Variables

- A random variable X is **discrete**, when there is a finite list of values $a_1, a_2, \ldots, a_n$ or an infinite list of values $a_1, a_2, \ldots$, such that $\mathrm{P}(\mathrm{X} = a_j \text{ for some } j) = 1$.
- The support of X is the finite or countably infinite set of values x, such that $P(X = x) > 0$ for discrete random variables.

GÖTEBORGS UNIVERSITET

Discrete Variables

- In essence, discrete variables are a set of integers ($\mathbb{Z}$).
- The set of integers consists of zero (0), the natural numbers: $(1, 2, 3, \ldots)$, also the negative integers, i.e., $-1, -2, -3, \ldots$).

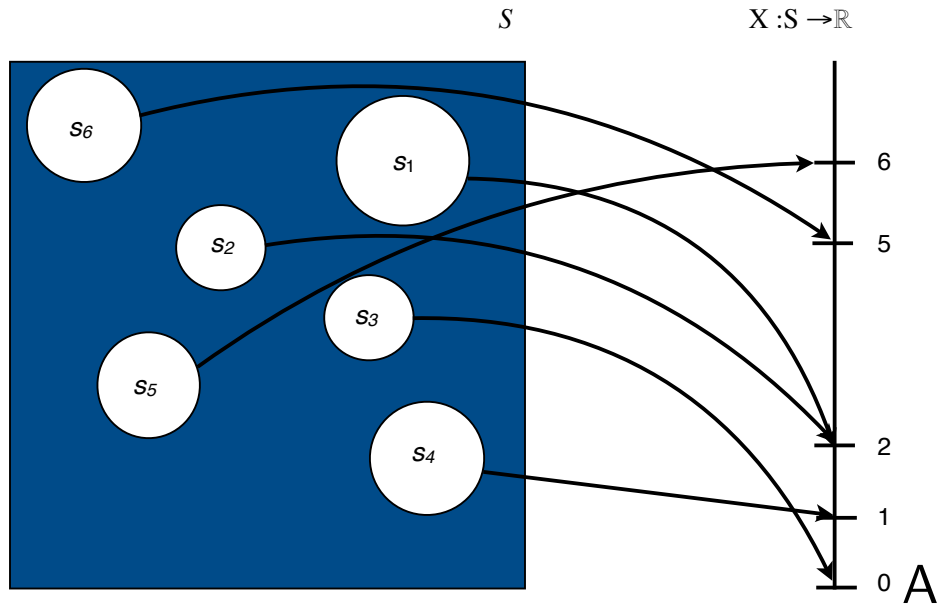
GÖTEBORGS UNIVERSITET

Continuous variables

- A **continuous variable** is a variable which can take on infinitely many, uncountable values, such as $\sqrt{2}, \pi$, etc.
- The reason is that any range of real numbers $a, b \in \mathbb{R}; a \neq b$ is infinite and uncountable.

Random variable

A **random variable** X is a function that maps the events of the sample space to the real line $X : S \rightarrow \mathbb{R}$ (*continuous*)/$\mathbb{Z}$ (*discrete*).



probability function (P) maps the events $s_1 \ldots s_6$ of a sample space S to the line of real numbers $\mathbb{R}$.

GÖTEBORGS UNIVERSITET

Example

Let us start with the coin example.

- If we toss two coins, the sample space will be $S = \{HH, HT, TT, TH\}$.
- An event s is a subset of S:
- So, the event **at least one head**, has an event space $\mathscr{B} = HH, HT, TH,$
- an event **at least one tail** has an event space $\mathscr{B} = TT, HT, TH,$
- an event **only one tail** has an event space $\mathscr{B} = HT, TH$.

GÖTEBORGS UNIVERSITET

Example

- A **random variable** X provides a number with respect to an event space.
- If X is the event **at least one head**, then X assigns a value 0, 1, 2, that corresponds to the possible occurrences of head in the sample space.

GÖTEBORGS UNIVERSITET

Example

So, that $X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$ (as selected from $S = \{HH, HT, TT, TH\}$).

- for the outcome HH, the X will assign the value 2,
- for the outcomes HT and TH, the X will assign the outcome 1, and
- for the outcome TT, the X, will assign the outcome 0.

GÖTEBORGS UNIVERSITET

Example

- What will be the selection of Y, if Y is the event **at least one tail**?

GÖTEBORGS UNIVERSITET

Example

- What will be the selection of Y, if Y is the event **at least one tail**?
- Y will again assign a value 0, 1, 2 to the events of the sample space S as follows: $Y(TT) = 2, Y(HT) = 1, Y(TH) = 1, Y(HH) = 0$.

GÖTEBORGS UNIVERSITET

Example

- What will be the selection of Z, if Z is the event **only one tail**?

GÖTEBORGS UNIVERSITET

Example

- What will be the selection of Z, if Z is the event **only one tail**?
- Z will assign 1, when there is only one tail and 0 in all other cases, $Z(TT) = 0, Z(HT) = 1, Z(TH) = 1, Z(HH) = 0$.

GÖTEBORGS UNIVERSITET

Probability Mass Function

- A **probability mass function** (PMF) is a function that gives the probability that a discrete random variable X is equal to a certain value. Specifically,
- The **probability mass function** (PMF) of a discrete random variable X is the function p_X given by $p_X(x) = P(X = x)$.

GÖTEBORGS UNIVERSITET

Probability Mass Function

- The **total probability** for all hypothetical outcomes x is therefore, the physical mass is the sum of all the probabilities of all events and it is equal to 1.

$$\sum_{x \in A} f_X(x) = 1$$

GÖTEBORGS UNIVERSITET

Probability Mass Function

- The PMF allows the definition of a **discrete probability distribution**, including discrete scalar or multivariate random variables.
- In contrast, **the probability density function** (PDF) associates with continuous rather than discrete random variables; the values of the latter are not probabilities so to provide probabilities a PDF should be integrated over an interval.

GÖTEBORGS UNIVERSITET

Probability Mass Function

In $P(X = x)$,

1. $X = x$ denotes the event, which consists of all outcomes s to which X assigns the number x.
2. an alternative denotation is $\{s \in S : X(s) = x\}$, also indicated more simply as $\{X = x\}$.

Probability Mass Function: Example

- Let us assume that we roll a fair dice with sample space: $S = \{1, 2, 3, 4, 5, 6\}$.
- If X is the number of possible outcomes from 1 to 6, then $X = 1$ ($P(X = 1)$, or in a more general form $P(X = x)$) consists of all the outcomes, since X assigns the number 1 to all 6 possible outcomes.
- So for a fair dice, all the possible outcomes have an equal chance, namely $1/6$.

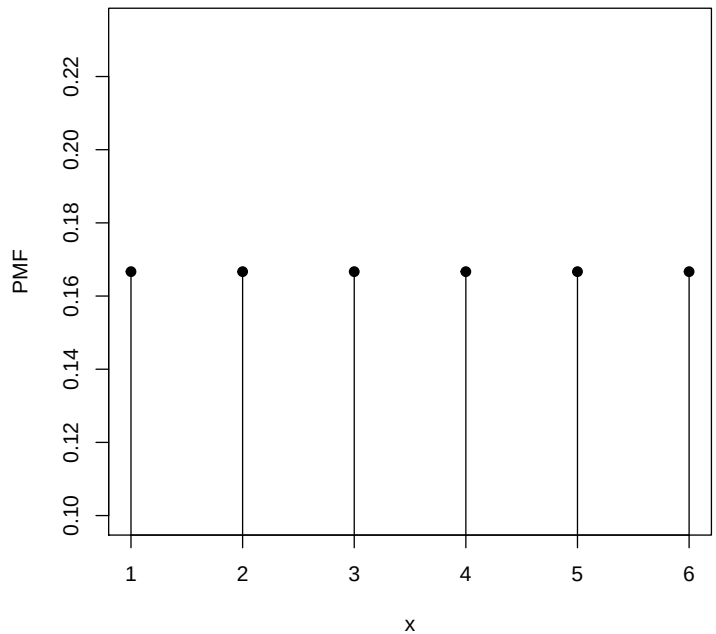


Figure: The PMF of fair dice, each outcome is equally possible, with possibility $1/6$.

GÖTEBORGS UNIVERSITET

Example

- Such fair outcomes are not always the case.
- In a fair coin roll, the sample space is $S = \{HH, HT, TT, TH\}$.
- For **At least one head**, the X takes the following values: $X(HH) = 2, X(HT) = X(TH) = 1, X(TT) = 0$.

GÖTEBORGS UNIVERSITET

Example

The PMF is in this case:

- $p_X(0) = P(X = 0) = 1/4$,
- $p_X(1) = P(X = 1) = 2/4$,
- $p_X(2) = P(X = 2) = 1/4$,

GÖTEBORGS UNIVERSITET

Example

- If Y is the event **at least one tail**, what will be the values of Y?

GÖTEBORGS UNIVERSITET

Example

- If Y is the event **at least one tail**, what will be the values of Y?
- Y will again assign a value 0, 1, 2 to the events of the sample space S as follows $Y(TT) = 2, Y(HT) = 1, Y(TH) = 1, Y(HH) = 0$.

GÖTEBORGS UNIVERSITET

Example

- $p_Y(0) = P(Y = 0) = 1/4,$
- $p_Y(1) = P(Y = 1) = 2/4,$
- $p_Y(2) = P(Y = 2) = 1/4,$

GÖTEBORGS UNIVERSITET

Example

- If Z is the event *only one tail*,

GÖTEBORGS UNIVERSITET

Example

- If Z is the event *only one tail*,
- then Z will assign 1 when there is only one tail and 0 in all other cases, $Z(TT) = 0, Z(HT) = 1, Z(TH) = 1, Z(HH) = 0$.

GÖTEBORGS UNIVERSITET

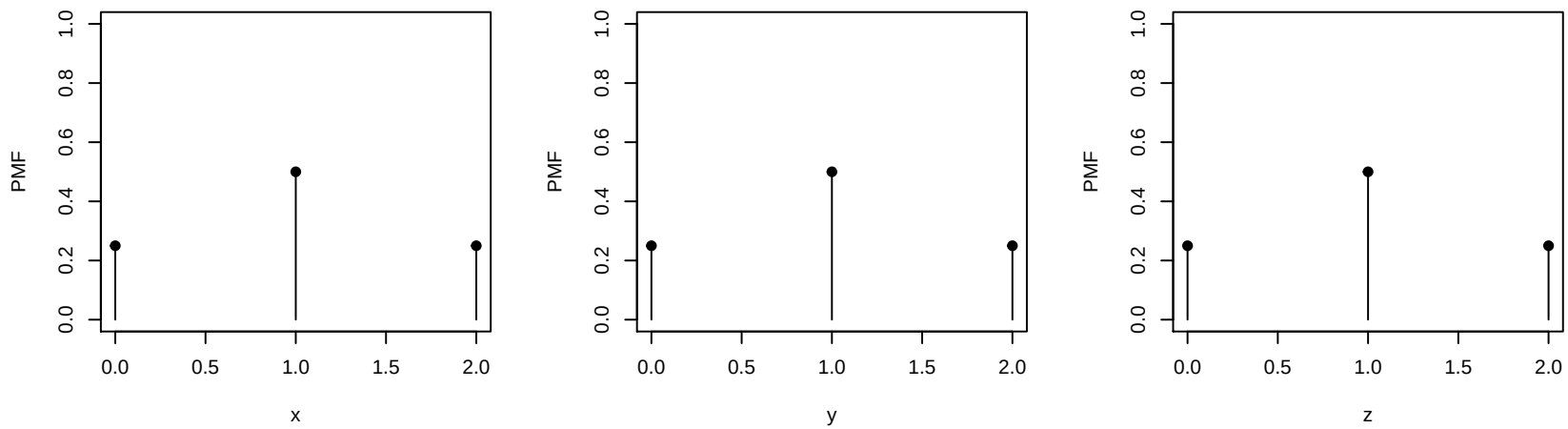
Example

- $p_Z(0) = P(Z = 0) = 2/4$,
- $p_Z(1) = P(Z = 1) = 2/4$,

GÖTEBORGS UNIVERSITET

Probability Mass Function: Example

The PMF for the examples can be plotted as follows:



GÖTEBORGS UNIVERSITET

Cumulative distribution function

- The **cumulative distribution function** (CDF)—unlike the PMF—describes the distribution of both discrete and continuous random variables.
- The **cumulative distribution function** (CDF) of an random variable X is the function FX given by $FX(x) = P(X \leq x)$.
- The **cumulative distribution function** (CDF) of a random variable X evaluated at x, is the probability that X will take a value less than or equal to x.

GÖTEBORGS UNIVERSITET

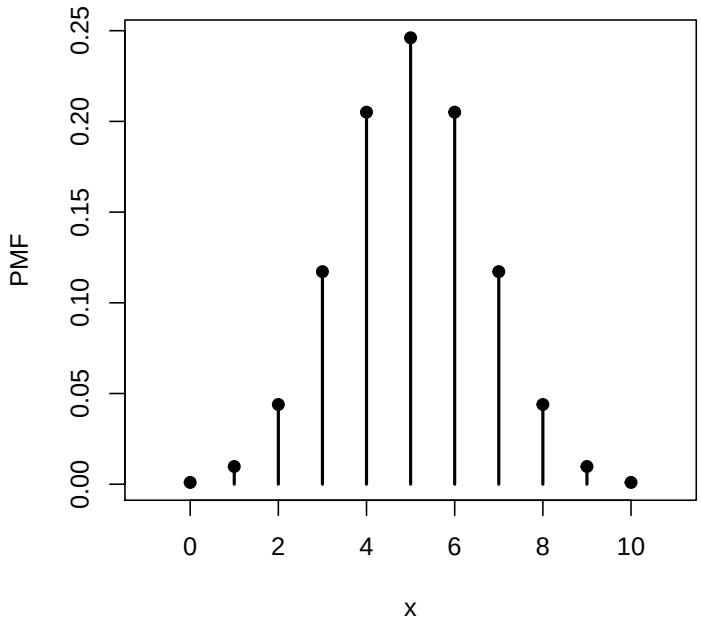
Cumulative distribution function

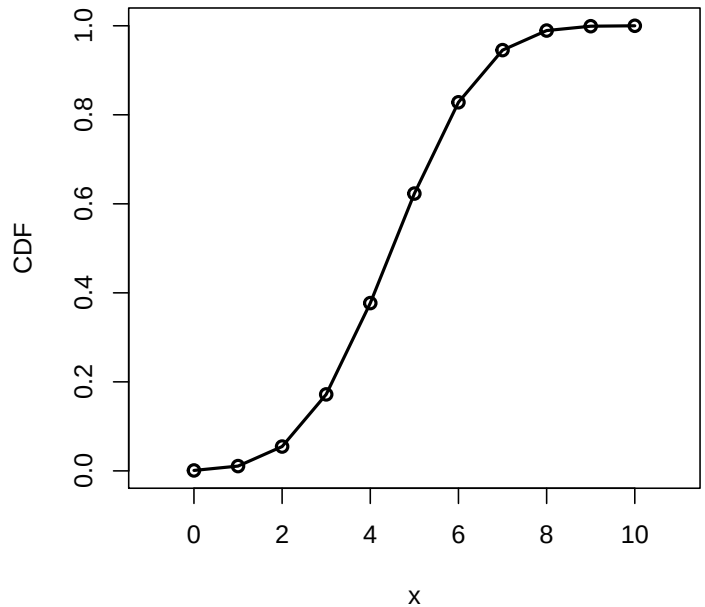
Every **cumulative distribution function** F—note that we write here F as a shorthand of FX—is **non-decreasing** and **right-continuous**:

1. Increasing:
   If $x1 \leq x2$, then $F(x_1) \leq F(x_2)$.
2. Right-continuous:
   $F(a) = \lim_{x \to -a^+}$
3. Converges to 0 and 1 in the limits:
   $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

GÖTEBORGS UNIVERSITET

Cumulative distribution function





Cumulative distribution function

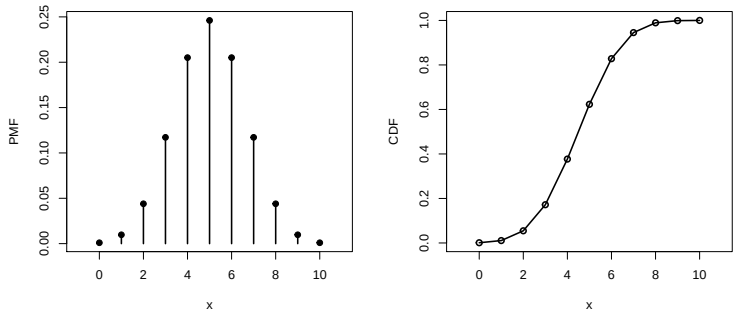


Figure: Probability Mass Function (left panel) and Cumulative Distribution Function for a Binomial distribution ($X = Bin(10, 1/2)$) (right panel).

From these two plots, it is possible to verify the properties of the CDF. Specifically,

1. it increases from almost 0 (the actual starting value is 0.00097) and goes up to 1.
2. it is right-continuous in that increases from left to right and there are no discontinuities.
3. it converges to 0 and 1 in these limits.

By contrast, the lines in the PMF are not connected, nor they approach specific limits.

GÖTEBORGS UNIVERSITET

Bernoulli Distribution

A random variable that can have 0 and 1 as possible values, is distributed as **Bernoulli Distribution**. Two relevant concepts with the Bernoulli distribution are the following:

1. **The Indicator Random Variable**. Any event that can be distributed as Bernoulli if we assign 0, when the event does not occur and 1 when the event occurs, this is the *indicator random variable*.
2. **Bernoulli trial**. An experiment that can be a **success** or a **failure** us know as *Bernoulli trial*.

GÖTEBORGS UNIVERSITET

Example

1. A coin experiment can be considered a **Bernoulli trial** when we agree that a *Head* is a *success* and *Tail* a *failure*.
2. Playing basketball can be considered a Bernoulli trial when we consider winning the game a success and losing the game a failure.
3. Requesting a deadline extension for an assignment can be Bernoulli trial with success when our Professor gives us an extension and failure when she does not give us an extension.
4. ...

GÖTEBORGS UNIVERSITET

Bernoulli random variables

A Bernoulli random variable X takes the values 0 and 1

$$P(X = 1) = p$$
$$P(X = 0) = 1 - p$$

where $0 < p < 1$. The $X \sim Bern(p)$ is read as the X variable is distributed as **Bernoulli distribution**. Let us define the Bernoulli Distribution more formally: If X is a random variable with this distribution, we have:

$$Pr(X = 1) = 1 - \Pr(X = 0) = 1 - q = p. \tag{1}$$

GÖTEBORGS UNIVERSITET

Probability mass function

The **probability mass function** f of Bernoulli distribution, over possible outcomes k, is

$$f(k;p) = p^k(1-p)^{1-k} \quad \text{for } k \in \{0,1\} \tag{2}$$

In a different notation, **the probability mass function** of f is defined as

$$f(k;p) = \begin{cases} p & \text{if } k = 1, \\ 1-p & \text{if } k = 0. \end{cases} \tag{3}$$

GÖTEBORGS UNIVERSITET

Cumulative Distribution Function

The **Cumulative Distribution Function** for the Bernoulli Distribution is shown in (4) (see however the discussion on the Binomial Distribution for more details).

$$\begin{cases} 0 & \text{for } k < 0 \\ 1-p & \text{for } 0 \leq k < 1 \\ 1 & \text{for } k \geq 1 \end{cases} \tag{4}$$

GÖTEBORGS UNIVERSITET

Random variable

- A random variable X that is distributed as Bernoulli is associated with a parameter of the distribution, denoted with p, that is equal to 1.
- The parameter of distribution p is an **index** of a family of Bernoulli distributions.
- To determine a distribution, we have to (1) name it and (2) determine the parameter value, we have a Bernoulli distribution with parameter value 2/5, which can be denoted as $X \sim Bern(2/5)$.

GÖTEBORGS UNIVERSITET

Problem

If there are 30 multiple choice questions in the exams and each question has 5 answers but only 1 is the correct answer. What is the probability of finding 4 or less correct answers if you decide to answer randomly the questionnaire?

GÖTEBORGS UNIVERSITET

Problem

The probability of answering a question correctly by random is 1/5=0.2. To calculate the probability of having 4 correct answers, we can do the following in R:

```
> dbinom(4, size=30, prob=0.2)\\
[1] 0.1325224
```

GÖTEBORGS UNIVERSITET

Problem

To calculate the probability of having 4 or less correct answers, we add the individual probabilities or we can use the **cumulative function**:

```
dbinom(0, size=30, prob=0.2) +
  + dbinom(1, size=30, prob=0.2) +
  + dbinom(2, size=30, prob=0.2) +
  + dbinom(3, size=30, prob=0.2) +
  + dbinom(4, size=30, prob=0.2)
```

or use the cumulative function
pbinom(4, size=30, prob=0.2)

GÖTEBORGS UNIVERSITET

Mean

The mean value of a Bernoulli random variable X is

$$\mathrm{E}(X) = p \tag{5}$$

GÖTEBORGS UNIVERSITET

Expected value for a Bernoulli distributed random variable

So the expected value for a Bernoulli distributed random variable X with $\Pr(X = 1) = p$ and $\Pr(X = 0) = q$ that is

$$\mathrm{E}[X] = \Pr(X = 1) \cdot 1 + \Pr(X = 0) \cdot 0 = p \cdot 1 + q \cdot 0 = p \qquad (6)$$

GÖTEBORGS UNIVERSITET

Variance

The variance of a Bernoulli distributed random variable X is

$$Var[X] = p(1-p) \tag{7}$$

This can be explained from the

$$\mathrm{E}[X^2] = \Pr(X=1)\cdot 1^2 + \Pr(X=0)\cdot 0^2 = p\cdot 1^2 + q\cdot 0^2 = p \tag{8}$$

So,

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = p - p^2 = p(1-p) = pq \tag{9}$$

GÖTEBORGS UNIVERSITET

Skewness and Kurtosis

The skewness is

$$\frac{q-p}{\sqrt{pq}} = \frac{1-2p}{\sqrt{pq}}. \tag{10}$$

When we take the standardized Bernoulli distributed random variable

$$\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}} \tag{11}$$

Skewness and Kurtosis

The random variable gets $\frac{q}{\sqrt{pq}}$ with probability p and $-\frac{p}{\sqrt{pq}}$ with probability q. Then we have,

$$\begin{aligned}
\gamma_1 &= \mathrm{E}\left[\left(\frac{X - \mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}\right)^3\right] \\
&= p \cdot \left(\frac{q}{\sqrt{pq}}\right)^3 + q \cdot \left(-\frac{p}{\sqrt{pq}}\right)^3 \\
&= \frac{1}{\sqrt{pq}^3}\left(pq^3 - qp^3\right) \\
&= \frac{pq}{\sqrt{pq}^3}(q - p) \\
&= \frac{q - p}{\sqrt{pq}}
\end{aligned}$$

The kurtosis goes to infinity for high and low values of p, but for $p = 1/2$ the distribution has a lower excess kurtosis 2.

GÖTEBORGS UNIVERSITET

Binomial Distribution

- **The binomial distribution** is a discrete probability distribution with parameters n and p of successes in a sequence of n independent yes/no experiments, each of which yields success with probability p.
- **The Bernoulli distribution** is a special case of the binomial distribution with $n = 1$.

GÖTEBORGS UNIVERSITET

Binomial Distribution: Example

- let us assume that we perform n independent Bernoulli trials.
- Each trial can have the same probability of success, p and the number of successes is X.
- The n is the number of trials and the p is the success probability.

GÖTEBORGS UNIVERSITET

Binomial Distribution: Example

To say that the random variable X has Binomial distribution with parameters n and p we write:

$$X \sim Bin(n, p) \tag{12}$$

Where n is a positive integer and $0 < p < 1$.

GÖTEBORGS UNIVERSITET

Probability mass function

- The probability mass function (Binomial PMF) provides the probability of getting exactly k successes in n trials.
- In Binomial distribution k successes occur with probability p^k; nk failures occur with probability $(1p)nk$.
- As, the k successes occur somewhere in the n trials, there are $\binom{n}{k}$ different ways of distributing k successes in a sequence of n trials.

GÖTEBORGS UNIVERSITET

Probability mass function

Therefore, the probability mass function is

$$f(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{13}$$

for $k = 0, 1, \ldots, n$ and

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{14}$$

is the binomial coefficient (read as n choose k).

GÖTEBORGS UNIVERSITET

Probability mass function

Note that R has this very nice function to perform the $\binom{n}{k}$ without calculating manually $\frac{n!}{k!(n-k)!}$, this function is choose(n, k). To calculate $\binom{6}{3}$

```
#
> choose(6,3)
[1] 20
```

GÖTEBORGS UNIVERSITET

Example

Suppose a biased coin is tossed 5 times and the probability to come on heads is 0.43 when tossed. X is the number of heads that result after each toss. Calculate:

1. $P(X = 3)$
2. $P(X = 1)$
3. $P(1 < X \leq 4)$

GÖTEBORGS UNIVERSITET

Example

If heads are considered a success, then the X has a binomial distribution with parameters $n = 5$ and $p = 0.4$.

$$Pr(X = 3) = \binom{5}{3} 0.4^3 (0.6)^2 \approx 0.23 \tag{15}$$

```
> choose(5,3) * 0.4^3 * 0.6^2
[1] 0.2304
```

GÖTEBORGS UNIVERSITET

Example

If heads are considered a success, then the X has a binomial distribution with parameters n $= 5$ and p $= 0.4$.

$$Pr(X = 1) = \binom{5}{1} 0.4^1 (0.6)^4 \approx 0.26 \tag{16}$$

```
> choose(5,1) * 0.4^1 * 0.6^4
[1] 0.2592
```

GÖTEBORGS UNIVERSITET

Example

3. Let us calculate $P(1 < X \leq 4)$

$$Pr(X = 2) = \binom{5}{2} 0.4^2 (0.6)^3 = 0.3456$$

$$Pr(X = 3) = \binom{5}{3} 0.4^3 (0.6)^2 = 0.2304$$

$$Pr(X = 4) = \binom{5}{4} 0.4^4 (0.6)^1 = 0.0768$$

```
> choose(5,2) * 0.4^2 * 0.6^3
[1] 0.3456
> choose(5,3) * 0.4^3 * 0.6^2
[1] 0.2304
> choose(5,4) * 0.4^4 * 0.6^1
[1] 0.0768
```

GÖTEBORGS UNIVERSITET

Mean

If X B(n, p), that is, X is a binomially distributed random variable, n being the total number of experiments and p the probability of each experiment yielding a successful result, then the expected value of X is:

$$\mathrm{E}[X] = np, \tag{17}$$

(For example, if n=100, and p=1/4, then the average number of successful results will be 25)

GÖTEBORGS UNIVERSITET

Mean: Proof

Proof: The mean μ can be directly calculated from its definition $\mu = \sum_{i=1}^{n} x_i p_i$ and the binomial theorem:

$$
\begin{aligned}
\mu &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\
&= np \sum_{k=0}^{n} k \frac{(n-1)!}{(n-k)!k!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{((n-1)-(k-1))!(k-1)!} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{\ell=0}^{n-1} \binom{n-1}{\ell} p^{\ell} (1-p)^{(n-1)-\ell} \quad \text{with } \ell := k-1 \\
&= np \sum_{\ell=0}^{m} \binom{m}{\ell} p^{\ell} (1-p)^{m-\ell} \qquad \text{with } m := n-1 \\
&= np\,(p + (1-p))^m = np 1^m = np
\end{aligned}
$$

GÖTEBORGS UNIVERSITET

Binomial Distribution in R

The probability mass function is calculated with the function dbinom(k,n,p), where where k is the number of success, n is the number of trials, and p is the probability of success. Take for instance the examples we examined above:

$$Pr(X = 2) = \binom{5}{2} 0.4^2 (0.6)^3 = 0.3456$$

$$Pr(X = 3) = \binom{5}{3} 0.4^3 (0.6)^2 = 0.2304$$

$$Pr(X = 4) = \binom{5}{4} 0.4^4 (0.6)^1 = 0.0768$$

are calculated as follows

```
> dbinom(2,5,0.4)
[1] 0.3456
> dbinom(3,5,0.4)
[1] 0.2304
> dbinom(4,5,0.4)
[1] 0.0768
```

GÖTEBORGS UNIVERSITET

Hypergeometric Distribution

- **The hypergeometric distribution** n is a discrete probability distribution that describes the probability of k successes in n draws from a finite population of size N that contains exactly K successes, wherein each draw is either a *success* or a *failure*.
- **The hypergeometric distribution** differs from the binomial distribution in that it is *without replacement*.
- In other words in a hypergeometric distribution the probability of success changes on each draw, because with every draw the population decreases.

GÖTEBORGS UNIVERSITET

Hypergeometric Distribution

The Hypergeometric distribution is denoted as follows:

$$HGeom \sim (K, N, n) \tag{18}$$

where

- K is the number of success states in the population
- k is the number of success states in the sample
- N is the size of the population
- n is the number sampled
- $\binom{k}{x}$ is a binomial coefficient, that stands for the number of combinations of k things taken x at time.

GÖTEBORGS UNIVERSITET

Hypergeometric Distribution

And the probability mass function is calculated as follows:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}, \tag{19}$$

GÖTEBORGS UNIVERSITET

Mean and Variance

The mean and variance of the hypergeometric distribution as calculated as follows: Mean:

$$n\frac{K}{N} \tag{20}$$

Variance

$$n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1} \tag{21}$$

GÖTEBORGS UNIVERSITET

What does *without replacement* mean? An Example

- Suppose we have an urn with 10 black and 10 white balls, with white to be a successful outcome and black to be non-successful outcome.
- If the distribution is **hypergeometric** it means that from the 10 white balls and 10 black balls in the urn and in the first trial we get a black ball, then in the urn will remain 1 black ball less,
- if in the following trial we get again a black ball then there will be only 8 black balls in the urn so the third time that we try to get a ball the probability of getting a white will be increased as there will be more white than black balls in the urn.

GÖTEBORGS UNIVERSITET

What does *without replacement* mean? An Example

- Suppose that we take from the urn 5 balls, what is the probability that three of them are white balls?
- Since, we do not draw one ball and then put it back again in the urn, draw another ball and but it back again, we cannot calculate this probability using binomial distribution. We have to employ the hypergeometric distribution.

GÖTEBORGS UNIVERSITET

What does *without replacement* mean? An Example

We have 40 balls (20 black and 20 white), we select 5 balls and we want to calculate the probability that 3 out of 5 are white balls. To do this let us put all this information in the table below:

	drawn	not drawn	total
white balls	$k = 3$	$K - k = 7$	$K = 10$
black balls	$n - k = 2$	$N + k - n - K = 28$	$N - K = 30$
total	$n = 5$	$N - n = 35$	$N = 40$

$$P(X = k) = \frac{\binom{K}{k} \cdot \binom{N-K}{n-k}}{\binom{N}{n}}$$

$$P(X = 3) = \frac{\binom{10}{3} \cdot \binom{30}{2}}{\binom{40}{5}}$$

$$P(X = 3) = \frac{120 \cdot 435}{658008} = \frac{52200}{658008} = 0.079$$

GÖTEBORGS UNIVERSITET

Calculating the example in R

In R the probability mass function can be calculated as follows:

```
p(x) = choose(K, k) choose(N–K, n–k) / choose(N, n)
```

GÖTEBORGS UNIVERSITET

Calculating the example in R

Alternatively, the probability mass function can be calculated in R by using the R function dhyper(k, K, x, n) where

- k the number of white balls drawn without replacement from the urn.
- K the number of white balls in the urn.
- x the number of black balls in the urn, namely the N-K.
- n the number of balls drawn from the urn.

We calculate the example in R as follows:

```
> (choose(10,3) * choose(30,2) ) / choose(40,5)
[1] 0.07933034
> dhyper(3,10,30,5)
[1] 0.07933034
```

GÖTEBORGS UNIVERSITET

Dependent vs. Independent Random Variables

- Consider a **binomial distribution** and **a hypergeometric distribution**, in the **binomial distribution** every trial does not affect the other.
- For instance, in the case that we have 20 balls in an urn every time we get a ball we put it back so that the probabilities do not change in a second trial.

GÖTEBORGS UNIVERSITET

Dependent vs. Independent Random Variables

- By contrast, in the **hypergeometric distribution** every trial changes the probabilities, since we remove the ball from the urn and we do not put it back.
- So, if there is urn with 8 black and 2 white balls in the urn the probability to get a white is 2/10, but we take two white balls out of the urn the probability of getting a white becomes 0 where the probability of getting a black ball becomes 100%.
- So, the probabilities of each new trial depend on the outcome of the preceding trial(s).

GÖTEBORGS UNIVERSITET

Dependent vs. Independent Random Variables

In more formal terms, two random variables X and Y are said to be independent iff (if and only if) the subsets P of the set S generated by them are independent; that is to say, for every a and b, the events $X \leq a$ and $Y \leq b$ are independent events. That is, X and Y are independent iff

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \tag{22}$$

GÖTEBORGS UNIVERSITET

Next

1. Series and DataFrames in Python
2. Basic Plots
3. Descriptive Statistics in Python