Quick review on Discrete Random Variables

Notations.
- \( Z = \{ \ldots, -2, -1, 0, 1, 2, \ldots \} \), set of all integers;
- \( \mathbb{N} = \{0, 1, 2, \ldots\} \), set of natural numbers, i.e. all nonnegative integers;
- \( \mathbb{R} = (-\infty, \infty) \), set of all real numbers.
- For a set \( S \), \(|S|\) denotes the size of \( S \), i.e. the number of elements in \( S \).

1 Probability Space

Example 1. Assume a pizza restaurant proposes its customers to create their own customized pizzas using exactly 5 toppings from a total of 15 (pepperoni, mozzarella, mushrooms, bacon, sausage, green peppers, \ldots).

- A possible outcome of that random experiment is for instance a pizza composed of pepperoni, mushrooms, bacon, onions and green peppers.
- The sample space \( \Omega \) of the experiment is the set of all possible outcomes. Here,
  \[
  \Omega = \{ \text{all possible combinations of 5 toppings among 15} \}
  \]
  The size of the sample space is: \( \binom{15}{5} \).
- An event is a subset of \( \Omega \). For example, let \( E \) be the event that a customer picks pepperoni and mushrooms. The number of outcomes that compose \( E \) is: \( \binom{13}{3} \).
- The complement of event \( E \), denoted \( E^c \), consists of all outcomes in \( \Omega \) that are not in \( E : E^c = \Omega \setminus E \)
- Let \( F \) be the event that the customer picks bacon. The union of \( E \) and \( F \), denoted \( E \cup F \), consists of all outcomes that are either in \( E \) or \( F \).
- The intersection of \( E \) and \( F \), denoted \( E \cap F \), consists of all outcomes that are both in \( E \) and \( F \).

Definition 1.1. A probability space \( (\Omega, \mathcal{A}, \mathbb{P}) \) consists of three parts:
• A sample space, $\Omega$, which is the set of all possible outcomes of a random experiment.

• A set of events $A$, where each event $E \in A$ is a subset of $\Omega$. Formally, $A$ is a $\sigma$-algebra$^1$ on $\Omega$.

• A function mapping $P : A \to [0, 1]$, called probability measure that assigns probabilities to the events.

Example 1 (continued). Assuming all toppings are equally likely to be picked, the probability of event $E$ that a customer picks pepperoni and mushrooms is:

$$P(E) = \frac{|E|}{|\Omega|} = \frac{13}{15} \cdot \frac{5}{5}.$$

2 Discrete Random Variables

2.1 Real-valued Random Variables

Definition 2.1. Let $(\Omega, A)$ be a measurable space of events on the sample space $\Omega$. A real-valued random variable (r rv) $X$ is a function mapping with domain $\Omega$, i.e. $X : \Omega \to \mathbb{R}$ such that for any subset $B \subset \mathbb{R}$:

$$\{\omega \in \Omega \mid X(\omega) \in B\} \in A \quad (1)$$

The event $\{\omega \in \Omega \mid X(\omega) \in B\}$ is simply denoted $\{X \in B\}$.

Notation: We denote $X(\Omega)$ all the possible values that $X$ can take on.

A real-valued random variable (r rv) serves to describe some numerical property that outcomes in $\Omega$ may have.

Example 1 (continued). Suppose that each topping has some cost. Let $X$ be the price in dollars of a pizza composed of 5 toppings.

• Let $\omega$ be the outcome that a customer picks pepperoni ($2), mushrooms ($1.75), bacon ($2), onions ($1.50) and green peppers ($1). Then $X(\omega) = 8.25$

• $\{X = 10\}$ is the event that the customer picks a combination of 5 toppings that cost $10.

• $\{8 \leq X \leq 10.5\}$ is the event that the customer picks a combination of 5 toppings that cost between $8 and $10.5.

$^1$Modeling the set of events as a $\sigma$-algebra is convenient because it allows us to define formally the unions, intersections and complements of events.

$^2$Formally, the set of values that a real-valued random variable $X$ can take on should be endowed with a $\sigma$-algebra. In this course, the $\sigma$-algebra that we will work with is called the Borel $\sigma$-algebra on $\mathbb{R}$.
2.2 Discrete random variables

**Definition 2.2.** A real-valued random variable $X$ is said to be **discrete** if $X$ can take:

- either a finite number of values: $X(\Omega) = \{x_i \in \mathbb{R}, i = 1, \ldots, n\}$ for a given $n \in \mathbb{N}$, $n \geq 1$

- or a countably infinite number of values: $X(\Omega) = \{x_i \in \mathbb{R}, i \in I\}$ for a given subset $I \subset \mathbb{N}$.

**Example 1 (continued).** Suppose that each topping has some cost. Let $X$ be the price in dollars of a pizza composed of 5 toppings. $X$ is discrete since the number of possible values for $X$ is finite. To be more precise, $|X(\Omega)| \leq \binom{15}{5}$.

**Example 2.** Consider the dimensions of a standard tennis court: 23.77 metres long and 10.97 metres wide. We are interested in the location of Federer’s landing ball in the $xy$-plane. Assuming Federer’s ball cannot land outside of the court, the sample space of this experiment is:

$\Omega = \{(x, y) | 0 \leq x \leq 10.97, 0 \leq y \leq 23.77\}$

Let $X$ be the distance between Federer and the location where the ball hits the court. $X$ can take an uncountably infinite number of values. Therefore, $X$ is not discrete.

**Definition 2.3** (Probability mass function). Let $X$ be a discrete rrv on probability space $(\Omega, \mathcal{A}, P)$. The probability mass function (pmf) $p_X$ of $X$ is a function with domain $X(\Omega)$ and is defined by:

$$p_X(x) = \mathbb{P}(X = x), \text{ for } x \in X(\Omega)$$

2.3 Expectation and Variance

**Definition 2.4** (Expected Value). Let $X$ be a discrete rrv on probability space $(\Omega, \mathcal{A}, P)$ with pmf $p_X$. If $\sum_{x \in X(\Omega)} |x|p_X(x) < \infty^3$, then the expectation (or expected value) of $X$ exists and is defined as follows:

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} xp_X(x)$$

The expected value of a discrete rrv $X$ is the weighted average of the values that $X$ can take on, where each possible value is weighted by its respective probability. In practice, if we observe $n$ independent realizations of $X$, denoted $x_1, \ldots, x_n$, then the sample mean $\sum_{i=1}^n x_i/n$ gets close to $\mathbb{E}[X]$, for sufficiently

---

$^3$We then say that $X$ is integrable.
large values of $n^4$. For that reason, $E[X]$ is sometimes referred as the theoretical mean.

**Definition 2.5** (Variance–Standard Deviation). Let $X$ be a discrete rrv on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with pmf $p_X$. If $E[X^2]$ exists, the variance of $X$ is defined as follows:

$$Var(X) = E[(X - E[X])^2]$$  \hfill (4)

$Var(X)$ is sometimes denoted $\sigma_X^2$. The positive square root of the variance is called the standard deviation of $X$, and is denoted $\sigma_X$. That is:

$$\sigma_X = \sqrt{Var(X)}$$  \hfill (5)

Variance is a measure of the dispersion of a random variable around its mean. The larger are the possible deviations of $X$ from its expected value $E[X]$, the larger the variance of $X$ is.

**Theorem 2.1** (König-Huygens formula). Let $X$ be a rrv on probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Upon existence, the variance of $X$ is also given by:

$$Var(X) = E[X^2] - (E[X])^2$$  \hfill (6)

2.4 Common Discrete Distributions

**Example 3.** An instructor recklessly assigns a random grade (integer between 1 and 4) to his students. Let $X$ be the grade of a student of his class. $X$ follows a discrete uniform on $\{1, 2, 3, 4\}$.

**Definition 2.6** (Discrete Uniform Distribution). Let $X$ be a discrete rrv on probability space $(\Omega, \mathcal{A}, \mathbb{P})$ with $|X(\Omega)| = n \in \mathbb{N}$, $n \geq 1$. $X$ is said to follow a discrete uniform distribution $U_n$ if its probability mass function is given by:

$$p_X(x) = \mathbb{P}(X = x) = \frac{1}{n}, \text{ for } x \in X(\Omega)$$  \hfill (7)

**Property 2.1** (Mean and Variance for a Discrete Uniform Distribution). If $X$ follows a discrete uniform distribution $U_n$ on $\{1, \ldots, n\}$, then

- its expected value is given by:

$$E[X] = \frac{n + 1}{2}$$  \hfill (8)

\footnote{This intuitive interpretation of the expected value is justified by one of the most important results in Probability called “The Strong Law of Large Numbers” and will be discussed later this quarter.}

\footnote{We then say that $X$ is square integrable.}
• its variance is given by:

\[ \text{Var}(X) = \frac{n^2 - 1}{12} \]  \hspace{1cm} (9)

**Example 4.** Ten people attending a match between Roger Federer and Novak Djokovic are randomly selected. A person is either a Federer fan (F) or a Djokovic fan (D). The sample space of this experiment is:

\[ \Omega = \{ (a_1, \ldots, a_{10}) \mid a_i \in \{F, D\} \} = \{F, D\}^{10} \]

The size of \( \Omega \) is \( 2^{10} \).

**Definition 2.7.** A Bernoulli or binomial process has the following features:

1. We repeat \( n \in \mathbb{N}, \; n \geq 1 \) identical trials

2. A trial can result in only two possible outcomes, that is, a certain event \( E \), called success, occurs with probability \( p \), thus event \( E^c \), called failure, occurs with probability \( 1 - p \)

3. The probability of success \( p \) remains constant trial after trial. In this case, the process is said to be stationary.

4. The \( n \) trials are mutually independent.

**Example 4 (continued).** Assume that a person has a likelihood of \( p = 80\% \) to be a Federer fan. Compute the probabilities of the following events:

- \( A \): “exactly two of the picked people like Federer”
- \( B \): “at least three of the picked people like Federer”

**Answer.** Let \( X \) be the number of Federer fans among the 10 picked people. Then \( X \) follows a binomial distribution with parameters 10 (sample size) and 0.8 (probability of success). Hence,

\[ \mathbb{P}(A) = \mathbb{P}(X = 2) = \binom{10}{2} 0.8^2 0.2^{10-2} \]

\[ \mathbb{P}(B) = \mathbb{P}(X \geq 3) = \sum_{x=3}^{10} \binom{10}{x} 0.8^x 0.2^{10-x} \]

\[ = 1 - \mathbb{P}(X \leq 2) = 1 - \sum_{x=0}^{2} \binom{10}{x} 0.8^x 0.2^{10-x} \]
**Definition 2.8** (Binomial Distribution). Let \((\Omega, \mathcal{A}, \mathbb{P})\) be a probability space. Let \(E \in \mathcal{A}\) be an event labeled as success, that occurs with probability \(p\). If \(n \in \mathbb{N}, n \geq 1\) trials are performed according to a Bernoulli process, then the random variable \(X\) defined as the number of successes among the \(n\) trials, is said to have a **binomial distribution** Bin\((n, p)\) and its probability mass function is given by:

\[
p_X(x) = \mathbb{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for} \quad x = 0, \ldots, n
\]  

(10)

**Property 2.2** (Mean and Variance for a Binomial Distribution). If \(X\) follows a binomial distribution Bin\((n, p)\), then

- its expected value is given by:
  \[
  \mathbb{E}[X] = np
  \]  
  (11)

- its variance is given by:
  \[
  \text{Var}(X) = np(1-p)
  \]  
  (12)

**Example 5.** Let \(X\) be the number of people joining the line of a movie theater in an interval of one hour. Assume that the mean number of people arriving in an interval of one hour is 50. \(X\) follows a Poisson distribution with parameter 50.

**Definition 2.9** (Poisson Distribution). Let \(X\) be a discrete rrv on probability space \((\Omega, \mathcal{A}, \mathbb{P})\). \(X\) is said to have a **Poisson distribution** \(\mathcal{P}(\lambda)\), with \(\lambda > 0\) if its probability mass function is given by:

\[
p_X(x) = \mathbb{P}(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad \text{for} \quad x \in \mathbb{N}
\]  

(13)

**Property 2.3** (Mean and Variance for a Poisson Distribution). If \(X\) follows a Poisson distribution \(\mathcal{P}(\lambda)\), then

- its expected value is given by:
  \[
  \mathbb{E}[X] = \lambda
  \]  
  (14)

- its variance is given by:
  \[
  \text{Var}(X) = \lambda
  \]  
  (15)

---

6If \(n = 1\), we talk about Bernoulli distribution Ber\((p)\).
**Definition 2.10 (Approximate Poisson process).** Let $X$ denote the number of events in a given continuous interval of length 1. Assume there are $\lambda > 0$ events on average in that interval. Then:

1. The probability of exactly two or more events in a “short” interval is essentially zero.
2. The number of events occurring in non-overlapping intervals are independent.
3. $X$ follows a Poisson distribution with parameter $\lambda$.
4. The number of events in an interval of length $h$ follows a Poisson distribution with parameter $\lambda h$.

**Example 5 (continued).**

- Let $E$ be the event that between 35 and 60 people join the line between 7:00pm and 8:00pm. Then,

$$
P(E) = P(35 \leq X \leq 60) = \sum_{x=35}^{60} \frac{e^{-50} 50^x}{x!}
$$

- Let $F$ be the event that 55 people arrive between 8:00pm and 9:00pm. According to Assumption (2) of a Poisson process,

$$
P(F \mid E) = P(F) = e^{-50} \frac{50^{55}}{55!}
$$

- Let $G$ be the event that at least 30 people arrive between 8:00pm and 8:30pm. Let $Y$ be the number of people joining the line of a movie theater in an interval of 30 minutes. According to Assumption (4) of a Poisson process, $Y$ follows a Poisson distribution of parameter $50 \cdot 1/2 = 25$.

$$
P(G) = P(Y \geq 30) = \sum_{x=30}^{\infty} \frac{e^{-25} 25^x}{x!}
$$

- Let $H$ be the event that 55 people arrive between 7:00pm and 8:00pm.

$$
P(H \mid E) = \frac{P(H \cap E)}{P(E)} = \frac{P(X = 55)}{P(35 \leq X \leq 60)} = \frac{e^{-50} \frac{50^{55}}{55!}}{\sum_{x=35}^{60} \frac{e^{-50} 50^x}{x!}}
$$

**Example 6.** In France, the “galette des rois” (King cake) contains a figurine, the “fève”, hidden in the cake and the person who finds the trinket in his or her slice becomes king/queen for the day. Assume that galettes are cut into 6 identical slices. Let $X$ be the number of galettes you eat until you find the fève. Then $X$ follows a geometric distribution with parameter $1/6$.
Definition 2.11 (Geometric Distribution). Let \((\Omega, A, \mathbb{P})\) be a probability space. Let \(E \in A\) be an event labeled as success, that occurs with probability \(p\). If all the assumptions of a Bernoulli process are satisfied, except that the number of trials is not preset, then the random variable \(X\) defined as the number of trials until the first success is said to have a geometric distribution \(G(p)\) and its probability mass function is given by:

\[
p_X(x) = \mathbb{P}(X = x) = (1 - p)^{x-1}p, \quad \text{for } x \in \mathbb{N}, x \geq 1
\] (16)

Property 2.4 (Mean and Variance for a Geometric Distribution). If \(X\) follows a geometric distribution \(G(p)\), then

- its expected value is given by:

\[
\mathbb{E}[X] = \frac{1}{p}
\] (17)

- its variance is given by:

\[
\text{Var}(X) = \frac{1 - p}{p^2}
\] (18)