Quick review on Discrete Random Variables

Néhémy Lim

University of Washington

Winter 2017
Example

Pick 5 toppings from a total of 15. Give the sample space $\Omega$ of the experiment and examples of events on $\Omega$. 

- Pepperoni
- Mozzarella
- Mushrooms
- Bacon
- Sausage
- Green peppers
- ...
Definition

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 \(\mathcal{A}\), where each event \(E \in \mathcal{A}\) is a subset of \(\Omega\). Formally, \(\mathcal{A}\) is a \(\sigma\)-algebra on \(\Omega\).
- A function mapping \(\mathbb{P} : \mathcal{A} \rightarrow [0, 1]\), called probability measure that assigns probabilities to the events.
Real-valued Random Variables

Definition

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

\[
\{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{A}
\]  

(1)

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

Definition

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

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

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

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.
Example

Dimensions of a tennis court: 23.77 metres long and 10.97 metres wide. Location of Federer’s landing ball. Assuming Federer’s ball can’t land outside of the court, give the sample space of this experiment.

Let $X$ be the distance between Federer and the location where the ball hits the court.
Definition (Probability mass function)

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

$$p_X(x) = \mathbb{P}(X = x), \text{ for } x \in X(\Omega) \quad (2)$$

Definition (Expected Value)

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

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x p_X(x) \quad (3)$$
Definition (Variance–Standard Deviation)

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

$$\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

(4)

The standard deviation of $X$ is:

$$\sigma_X = \sqrt{\text{Var}(X)}$$

(5)

Theorem (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:

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

(6)
Ten people attending a match between Roger Federer and Novak Djokovic are randomly selected. A person is either a Federer fan or a Djokovic fan. Give the sample space of this experiment.
Bernoulli process

Definition

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.
Ten people attending a match between Roger Federer and Novak Djokovic are randomly selected. A person is either a Federer fan or a Djokovic fan.

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”
Example

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.
Approximate Poisson process

Definition
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

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.

Compute the probabilities of the following events:

- $E$: “between 35 and 60 people join the line between 7:00pm and 8:00pm”
- $F$: “55 people arrive between 8:00pm and 9:00pm”
- $F \mid E$
- $G$: “at least 30 people arrive between 8:00pm and 8:30pm”
Example

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.