Lecture 5 - Random Variables

Intuition

- A random variable is a function from Ω to $\mathbb{R}$, which assigns a number to every outcome.
- Motivation: we are often interested with converting the outcomes of a random experiment to numbers:
  - We toss a coin n times, and we are only interested in the number of obtained heads
  - We roll dice until we obtain "six", but we are only interested in the number of rolls
- In probability, we usually denote random variables with large Latin letters, e.g. X, Y, Z.
- Example: we toss a coin 3 times, and we are only interested in the number of heads:

$$
\begin{gathered}
\Omega=\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\} \\
X: \Omega \rightarrow\{0,1,2,3\} \\
X(HHH)=3, \quad X(HHT)=2, \quad X(HTH)=2, \quad X(HTT)=1, \\
X(THH)=2, \quad X(THT)=1, \quad X(TTH)=1, \quad X(TTT)=0
\end{gathered}
$$

- We can assign probabilities to every such outcome. E.g., $P(\{X=2\})=\frac{3}{8}$
- Similarly, we can ask for any set of outcomes, e.g.: $P(\{X\leq 1\})=\frac{4}{8}$
- Disclaimer. We often drop the curly brackets, writing $P(X=2)=\frac{3}{8}$ instead of $P(\{X=2\})=\frac{3}{8}$.

Probability and preimage

- Sometimes we have a subset of real numbers A, and we are interested in an event E (a subset of Ω) that for some random variable X (a function from Ω to $\mathbb{R}$) corresponds to this subset. In other words:

$$E = \{\omega : X(\omega) \in A\} \text{ where } A \subseteq \mathbb{R}$$

- We can then write the probability of the event E as:

$$P(E) = P(X \in A) = P(\{\omega : X(\omega) \in A\}) = P(X^{-1}(A))$$

- Where $X^{-1}(A)$ is the preimage of A.

Definition

- A random variable is any measurable function $X : \Omega \to \mathbb{R}$.
- The probability distribution of X is a probability measure P_X defined as:

$$P_X(A) = P(X \in A) = P(X^{-1}(A))$$

- Remark: we will use $X \sim P_X$ to denote that a random variable X is distributed according to P_X.

Notations

$$\begin{aligned}
P_X(A) \quad &\Longleftrightarrow \quad P(X \in A) \\
P_X(\{a\}) \quad &\Longleftrightarrow \quad P(X = a) \\
P_X((-\infty, a]) \quad &\Longleftrightarrow \quad P(X \leqslant a) \\
P_X([a, b)) \quad &\Longleftrightarrow \quad P(a \leqslant X < b)
\end{aligned}$$

- Both the notation in blue and in red are correct, but we will use the red notation, as it seems more convenient.

Cumulative distribution function

- A cumulative distribution function (c.d.f.) of a random variable X is a function $F : \mathbb{R} \to \mathbb{R}$ defined as:

$$F(x) = P_X((-\infty, x]) = P(X \leq x)$$

- $F(x)$ is nondecreasing
- We have $P(X \in (a, b]) = F(b) - F(a)$
- $F(-\infty) = 0$
- $F(\infty) = 1$
- $F(x)$ is right-continuous

Discrete random variables

- A random variable X taking at most countable number of values $\mathcal{X} = \{x_1, x_2, \ldots\}$ is called a discrete random variable.
- For any $A \in \mathcal{B}$ it holds:

$$P(X \in A) = \sum_{i: x_i \in A} P(X = x_i)$$

- or in the "event" notation:

$$P_X(A) = \sum_{i: x_i \in A} P_X(\{x_i\})$$

- A function $P_X(\{x_i\})$ is sometimes called a probability mass function.
- Convention: for discrete random variables we omit their values taken with probability zero (i.e., we treat such values as not belonging to the codomain of X)

Examples of probability distributions

Degenerate distribution

- A random variable X has a degenerate probability distribution, if there exists $x \in R$, such that $P(X = x) = 1$.

Uniform distribution

- A random variable X has a uniform distribution, if $X \in \{x_1, x_2, \ldots, x_n\}$ and:

$$P(X = x_i) = \frac{1}{n}, \qquad i = 1, \ldots, n$$

Two-point and Bernoulli distribution

- A random variable X has a two-point distribution if $X \in \{x_0, x_1\}$.
- Denote:

$$p = P(X = x_1), \text{so } P(x = x_0) = 1 - p$$

- When $x_0 = 0$ and $x_1 = 1$, we say X has Bernoulli distribution; we then call p the probability of a success.
- The Bernoulli distribution is denoted by $B(p)$

Binomial distribution

- A random variable $X \in \{0, 1, 2, \ldots, n\}$ has a binomial distribution with parameter p if:

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

- For $k = 0, 1, \ldots, n$
- So X denotes the number of successes in n independent trials, if the probability of a success in a single trial is p (Bernoulli scheme).

- The binomial distribution is denoted by $B(n, p)$

Geometric distribution

- A random variable $X \in \{1, 2, \ldots\}$ has a geometric distribution with parameter p if:

$$P(X = k) = (1-p)^{k-1}p$$

- For $k = 1, 2, \ldots$
- So X denotes the number of (independent) trials until the first success in an infinite Bernoulli scheme.
- Geometric distribution is denoted by $G_1(p)$

Negative binomial (Pascal) distribution

- A random variable $X \in \{0, 1, \ldots\}$ has a negative binomial distribution with parameters r and p if:

$$P(X = k) = \binom{r+k-1}{r-1}(1-p)^r p^k$$

- The probability of k successes before r failures.
- The negative binomial distribution is denoted by $NB(r, p)$.

Poisson distribution

- A random variable $X \in \{0, 1, \ldots\}$ has Poisson distribution with parameter λ if:

$$P(X = k) = \frac{\lambda^k}{k!}e^{-\lambda}$$

- The Poisson distribution is denoted by $\mathrm{Pois}(\lambda)$.
- The Poisson distribution is a limit of Binomial distribution where n is very large, p is very small, but $n \cdot p = \lambda$ is moderate.

Memoryless property

- Let $X \sim G_1(p)$. Calculate $P(X \geq k+\ell | X \geq k+1)$ for $\ell \geq 1$.
- "The first success will not occur after at least $k+\ell$ trials if it did not occur within first k trials"
- By some calculations we can get:

$$P(X \geq k+\ell | X \geq k+1) = P(X \geq \ell)$$

- (Using the formula for geometric distribution)
- Memorylessness: the distribution of waiting time for the first success does not depend on the past (if the first success did not occur within first k trials, we can forget about the past and calculate the probabilities as if we have just started to draw)