IEOR 151 – Lecture 2
Probability Review

1 Definitions in Probability and Their Consequences

1.1 Defining Probability

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ consists of three elements:

- A sample space Ω is the set of all possible outcomes.
- The σ-algebra $\mathcal{F}$ is a set of events, where an event is a set of outcomes.
- The measure $\mathbb{P}$ is a function that gives the probability of an event. This function $\mathbb{P}$ satisfies certain properties, including: $\mathbb{P}(A) \geq 0$ for an event A, $\mathbb{P}(\Omega) = 1$, and $\mathbb{P}(A_1 \cup A_2 \cup \ldots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \ldots$ for any countable collection $A_1, A_2, \ldots$ of mutually exclusive events.

Some useful consequences of this definition are:

- For a sample space $\Omega = \{o_1, \ldots, o_n\}$ in which each outcome o_i is equally likely, it holds that $\mathbb{P}(o_i) = 1/n$ for all $i = 1, \ldots, n$.
- $\mathbb{P}(\overline{A}) = 1 - \mathbb{P}(A)$, where $\overline{A}$ denotes the complement of event A.
- For any two events A and B, $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- If $A \subseteq B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- Consider a finite collection of mutually exclusive events $B_1, \ldots, B_m$ such that $B_1 \cup \ldots \cup B_m = \Omega$ and $\mathbb{P}(B_i) > 0$. For any event A, we have $\mathbb{P}(A) = \sum_{k=1}^{m} \mathbb{P}(A \cap B_k)$.

1.2 Conditional Probability

The conditional probability of A given B is defined as

$$\mathbb{P}[A|B] = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Some useful consequences of this definition are:

- Law of Total Probability: Consider a finite collection of mutually exclusive events $B_1, \ldots, B_m$ such that $B_1 \cup \ldots \cup B_m = \Omega$ and $\mathbb{P}(B_i) > 0$. For any event A, we have

$$\mathbb{P}(A) = \textstyle\sum_{k=1}^{m} \mathbb{P}[A|B_k]\mathbb{P}(B_k).$$

- Bayes' Theorem: It holds that

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\mathbb{P}(B)}{\mathbb{P}(A)}.$$

1.3 Independence

Two events A_1 and A_2 are defined to be independent if and only if $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2)$. Multiple events $A_1, A_2, \ldots, A_m$ are mutually independent if and only if for every subset of events

$$\{A_{i_1}, \ldots, A_{i_n}\} \subseteq \{A_1, \ldots, A_m\},$$

the following holds:

$$\mathbb{P}(\cap_{k=1}^{n} A_{i_k}) = \Pi_{k=1}^{n} \mathbb{P}(A_{i_k}).$$

Multiple events $A_1, A_2, \ldots, A_m$ are pairwise independent if and only if every pair of events is independent, meaning $\mathbb{P}(A_n \cap A_k) = \mathbb{P}(A_n)\mathbb{P}(A_k)$ for all distinct pairs of indices n, k. Note that pairwise independence does not always imply mutual independence! Lastly, an important property is that if A and B are independent and $\mathbb{P}(B) > 0$, then $\mathbb{P}[A|B] = \mathbb{P}(A)$.

1.4 Random Variables

A random variable is a function $X(\omega) : \Omega \to \mathcal{B}$ that maps the sample space Ω to a subset of the real numbers $\mathcal{B} \subseteq \mathbb{R}$, with the property that the set $\{w : X(\omega) \in b\} = X^{-1}(b)$ is an event for every $b \in \mathcal{B}$. The cumulative distribution function (cdf) of a random variable X is defined by

$$F_X(u) = \mathbb{P}(\omega : X(\omega) \leq u).$$

The probability density function (pdf) of a random variable X is any function $f_X(u)$ such that

$$\mathbb{P}(X \in A) = \int_A f_X(u)du,$$

for any well-behaved set A.

1.5 Expectation

The expectation of $g(X)$, where X is a random variable and $g(\cdot)$ is a function, is given by

$$\mathbb{E}(g(X)) = \int g(u) f_X(u)du.$$

Two important cases are the mean

$$\mu(X) = \mathbb{E}(X) = \int u f_X(u) du,$$

and variance

$$\sigma^2(X) = \mathbb{E}((X-\mu)^2) = \int (u-\mu)^2 f_X(u) du.$$

Two useful properties are that if λ is a constant then

$$\mathbb{E}(\lambda X) = \lambda \mathbb{E}(X)$$
$$\sigma^2(\lambda X) = \lambda^2 \sigma^2(X).$$

2 Common Distributions

2.1 Uniform Distribution

A random variable X with uniform distribution over support $[a, b]$ is denoted by $X \sim \mathcal{U}(a, b)$, and it is the distribution with pdf

$$f_X(u) = \begin{cases} \frac{1}{b-a}, & \text{if } u \in [a, b] \\ 0, & \text{otherwise} \end{cases}.$$

The mean is $\mu = (a + b)/2$, and the variance is $\sigma^2 = (b - a)^2/12$.

2.2 Gaussian/Normal Distribution

A random variable X with Guassian/normal distribution and mean μ and variance σ^2 is denoted by $X \sim \mathcal{N}(\mu, \sigma^2)$, and it is the distribution with pdf

$$f_X(u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(u-\mu)^2}{2\sigma^2}\right).$$

For a set of iid (mutually independent and identically distributed) Gaussian random variables $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, consider any linear combination of the random variables.

$$S = \lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_n X_n.$$

The mean of the linear combination is

$$\mathbb{E}(S) = \mu \cdot \sum_{i=1}^{n} \lambda_i,$$

and the variance of the linear combination is

$$\sigma^2(S) = \sigma^2 \cdot \sum_{i=1}^{n} \lambda_i^2.$$

Note that in the special case where $\lambda_i = 1/n$ (which is also called a sample average):

$$\overline{X} = 1/n \cdot \sum_{i=1} nX_i$$

we have that $\mathbb{E}(\overline{X}) = \mathbb{E}(X)$ and $\sigma^2(\overline{X}) = \sigma^2/n$ (which also implies that $\lim_{n\to\infty} \sigma^2(\overline{X}) = 0$).

2.3 Chi-Squared Distribution

A random variable X with chi-squared distribution and k-degrees of freedom is denoted by $X \sim \chi^2(k)$, and it is the distribution of the random variable defined by

$$\sum_{i=1}^{n} Z_i^2,$$

where $Z_i \sim \mathcal{N}(0, 1)$. The mean is $\mathbb{E}(X) = k$, and the variance is $\sigma^2(X) = 2k$.

2.4 Exponential Distribution

A random variable X with exponential distribution is denoted by $X \sim \mathcal{E}(\lambda)$, where $\lambda > 0$ is the *rate*, and it is the distribution with pdf

$$f_X(u) = \begin{cases} \lambda \exp(-\lambda u), & \text{if } u \geq 0, \\ \text{otherwise} \end{cases}.$$

The cdf is given by

$$F_X(u) = \begin{cases} 1 - \exp(-\lambda u), & \text{if } u \geq 0, \\ \text{otherwise} \end{cases}$$

and so $\mathbb{P}(X > u) = \exp(-\lambda u)$ for $u \geq 0$. The mean is $\mu = \frac{1}{\lambda}$, and the variance is $\sigma^2 = \frac{1}{\lambda^2}$. One of the most important aspects of an exponential distribution is that is satisfies the memoryless property:

$$\mathbb{P}[X > s + t | X > t] = \mathbb{P}(X > s), \text{for all values of } s, t \geq 0.$$