Probability Review

Yutian Li

Stanford University

January 18, 2018

Outline

Outline

Elements of probability

Definition (Sample space Ω)

The set of all the outcomes of a random experiment.

Elements of probability

Definition (Event space $\mathcal{F}$)

A set whose elements $A \in \mathcal{F}$ (called *events*) are subsets of Ω.

Elements of probability

Definition (Probability measure)

A function $P : \mathcal{F} \to \mathbb{R}$ that satisfies the following properties.

- $P(A) \geq 0$, for all $A \in \mathcal{F}$.
- $P(\Omega) = 1$.
- If $A_1, A_2, \dots$ are disjoint events, then

$$P(\cup_i A_i) = \sum_i P(A_i).$$

These three properties are called the *the Axioms of Probability*.

Elements of probability

Definition (Probability measure)

A function $P : \mathcal{F} \rightarrow \mathbb{R}$ that satisfies the following properties.

- $P(A) \geq 0$, for all $A \in \mathcal{F}$.
- $P(\Omega) = 1$.
- If $A_1, A_2, \ldots$ are disjoint events, then

$$P(\cup_i A_i) = \sum_i P(A_i).$$

These three properties are called the *the Axioms of Probability*.

Elements of probability

Properties:

1. If $A \subset B \implies P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A), P(B))$.
3. Union bound: $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \setminus A) = 1 - P(A)$.
5. Law of total probability: If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^{k} A_i = \Omega$, then

$$\sum_{i=1}^{k} P(A_i) = 1.$$

Elements of probability

Properties:

1. If $A \subset B \implies P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A), P(B))$.
3. Union bound: $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \setminus A) = 1 - P(A)$.
5. Law of total probability: If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^{k} A_i = \Omega$, then

$$\sum_{i=1}^{k} P(A_i) = 1.$$

Elements of probability

Properties:

1. If $A \subset B \implies P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A), P(B))$.
3. Union bound: $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \setminus A) = 1 - P(A)$.
5. Law of total probability: If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^k A_i = \Omega$, then

$$\sum_{i=1}^{k} P(A_i) = 1.$$

Elements of probability

Properties:

1. If $A \subset B \implies P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A), P(B))$.
3. Union bound: $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \setminus A) = 1 - P(A)$.
5. Law of total probability: If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^{k} A_i = \Omega$, then
$$\sum_{i=1}^{k} P(A_i) = 1.$$

Elements of probability

Properties:

1. If $A \subset B \implies P(A) \leq P(B)$.
2. $P(A \cap B) \leq \min(P(A), P(B))$.
3. Union bound: $P(A \cup B) \leq P(A) + P(B)$.
4. $P(\Omega \setminus A) = 1 - P(A)$.
5. Law of total probability: If $A_1, \ldots, A_k$ are a set of disjoint events such that $\cup_{i=1}^k A_i = \Omega$, then
$$\sum_{i=1}^{k} P(A_i) = 1.$$

Elements of probability

Definition (Conditional probability)

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

Elements of probability

Theorem (Chain rule)

Let $S_1, \ldots, S_k$ be events, $P(S_i) > 0$. Then

$$
\begin{aligned}
&P(S_1 \cap S_2 \cap \cdots \cap S_k) \\
=&P(S_1)P(S_2 \mid S_1)P(S_3 \mid S_2 \cap S_1) \cdots P(S_k \mid S_1 \cap S_2 \cap \cdots \cap S_{k-1}).
\end{aligned}
$$

Elements of probability

Definition (Independence)

Two events are called independent if and only if $P(A \cap B) = P(A)P(B)$.

Outline

Random variables

Definition (Random variable)

A random variable X is a function $X : \Omega \to \mathbb{R}$.

Typically we denote random variables using upper case letters $X(\omega)$ or more simply X. We denote the value that a random variable may take on using lower case letters x.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0,1]$ defined as
$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Cumulative distribution function)

A cumulative distribution function (CDF) is a function $F_X : \mathbb{R} \to [0, 1]$ defined as

$$F_X(x) = P(X \leq x).$$

Properties:

1. $0 \leq F_X(x) \leq 1$.
2. $\lim_{x \to -\infty} F_X(x) = 0$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $x \leq y \implies F_X(x) \leq F_X(y)$.
5. Right-continuous: $\lim_{x \to +a} F_X(x) = F_x(a)$.

Random variables

Definition (Probability density function)

If the cumulative distribution function F_X is differentiable everywhere, we define the probability density function (PDF) as the derivative of the CDF,

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

Note that the PDF for a continuous random variable may not always exist.

Properties:

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$.
3. $\int_{x \in A} f_X(x)\mathrm{d}x = P(X \in A)$.

Random variables

Definition (Probability density function)

If the cumulative distribution function F_X is differentiable everywhere, we define the probability density function (PDF) as the derivative of the CDF,

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

Note that the PDF for a continuous random variable may not always exist.
Properties:

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$.
3. $\int_{x \in A} f_X(x)\mathrm{d}x = P(X \in A)$.

Random variables

Definition (Probability density function)

If the cumulative distribution function F_X is differentiable everywhere, we define the probability density function (PDF) as the derivative of the CDF,

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

Note that the PDF for a continuous random variable may not always exist.
Properties:

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$.
3. $\int_{x\in A} f_X(x)\mathrm{d}x = P(X \in A)$.

Random variables

Definition (Probability density function)

If the cumulative distribution function F_X is differentiable everywhere, we define the probability density function (PDF) as the derivative of the CDF,

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}.$$

Note that the PDF for a continuous random variable may not always exist.

Properties:

1. $f_X(x) \geq 0$.
2. $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$.
3. $\int_{x \in A} f_X(x)\mathrm{d}x = P(X \in A)$.

Random variables

Definition (Expectation)

If X is a continuous random variable with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. The expectation or expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x.$$

Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[ag(X)] = aE[g(X)]$ for any constant $a \in \mathbb{R}$.
3. Linearity: $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable X, $E[\mathbf{1}\{X = k\}] = P(X = k)$.

Random variables

Definition (Expectation)

If X is a continuous random variable with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. The expectation or expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x.$$

Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[ag(X)] = aE[g(X)]$ for any constant $a \in \mathbb{R}$.
3. Linearity: $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable X, $E[\mathbf{1}\{X = k\}] = P(X = k)$.

Random variables

Definition (Expectation)

If X is a continuous random variable with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. The expectation or expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x.$$

Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[ag(X)] = aE[g(X)]$ for any constant $a \in \mathbb{R}$.
3. Linearity: $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable X, $E[\mathbf{1}\{X = k\}] = P(X = k)$.

Random variables

Definition (Expectation)

If X is a continuous random variable with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. The expectation or expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x.$$

Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[ag(X)] = aE[g(X)]$ for any constant $a \in \mathbb{R}$.
3. Linearity: $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable X, $E[\mathbb{1}\{X = k\}] = P(X = k)$.

Random variables

Definition (Expectation)

If X is a continuous random variable with PDF $f_X(x)$ and $g : \mathbb{R} \to \mathbb{R}$ is an arbitrary function. The expectation or expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x.$$

Properties:

1. $E[a] = a$ for any constant $a \in \mathbb{R}$.
2. $E[ag(X)] = aE[g(X)]$ for any constant $a \in \mathbb{R}$.
3. Linearity: $E[f(X) + g(X)] = E[f(X)] + E[g(X)]$.
4. For a discrete random variable X, $E[\mathbf{1}\{X = k\}] = P(X = k)$.

Random variables

Definition (Variance)

The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean. Formally, the variance of a random variable X is defined as

$$Var[X] = E[(X - E[X])^2].$$

Alternatively, $Var[X] = E[X^2] - E[X]^2$.

Properties:

1. $Var[a] = 0$ for any constant $a \in \mathbb{R}$.
2. $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Random variables

Definition (Variance)

The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean. Formally, the variance of a random variable X is defined as

$$Var[X] = E[(X - E[X])^2].$$

Alternatively, $Var[X] = E[X^2] - E[X]^2$.

Properties:

1. $Var[a] = 0$ for any constant $a \in \mathbb{R}$.
2. $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Random variables

Definition (Variance)

The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean. Formally, the variance of a random variable X is defined as

$$Var[X] = E[(X - E[X])^2].$$

Alternatively, $Var[X] = E[X^2] - E[X]^2$.

Properties:

1. $Var[a] = 0$ for any constant $a \in \mathbb{R}$.
2. $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Random variables

Definition (Variance)

The variance of a random variable X is a measure of how concentrated the distribution of a random variable X is around its mean. Formally, the variance of a random variable X is defined as

$$Var[X] = E[(X - E[X])^2].$$

Alternatively, $Var[X] = E[X^2] - E[X]^2$.

Properties:

1. $Var[a] = 0$ for any constant $a \in \mathbb{R}$.
2. $Var[af(X)] = a^2 Var[f(X)]$ for any constant $a \in \mathbb{R}$.

Random variables

Discrete random variables:

- $X \sim \text{Bernoulli}(p)$ (where $0 \leq p \leq 1$): one if a coin with heads probability p comes up heads, zero otherwise.

$$p(x) = \begin{cases} p & \text{if } p = 1 \\ 1 - p & \text{if } p = 0 \end{cases}$$

- $X \sim \text{Binomial}(n, p)$ (where $0 \leq p \leq 1$): the number of heads in n independent flips of a coin with heads probability p.

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Random variables

Discrete random variables:

- $X \sim \text{Geometric}(p)$ (where $p > 0$): the number of flips of a coin with heads probability p until the first heads.

$$p(x) = p(1-p)^{x-1}$$

- $X \sim \text{Poisson}(\lambda)$ (where $\lambda > 0$): a probability distribution over the nonnegative integers used for modeling the frequency of rare events.

$$p(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

Random variables

Continuous random variables:

- $X \sim \text{Uniform}(a, b)$ (where $a < b$): equal probability density to every value between a and b on the real line.

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- $X \sim \text{Exponential}(\lambda)$ (where $\lambda > 0$): decaying probability density over the negative reals.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Random variables

Continuous random variables:

- $X \sim \text{Normal}(\mu, \sigma^2)$: also known as the Gaussian distribution.

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Outline

Multiple random variables

Definition (Joint cumulative distribution function)

Suppose that we have two random variables X and Y,

$$F_{XY}(x, y) = P(X \leq x, Y \leq y).$$

The joint CDF $F_{XY}(x, y)$ and the marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$ of each variable separately are related by

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y),$$

$$F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y).$$

Multiple random variables

Definition (Joint cumulative distribution function)

Suppose that we have two random variables X and Y,

$$F_{XY}(x, y) = P(X \leq x, Y \leq y).$$

The joint CDF $F_{XY}(x, y)$ and the marginal cumulative distribution functions $F_X(x)$ and $F_Y(y)$ of each variable separately are related by

$$F_X(x) = \lim_{y \to \infty} F_{XY}(x, y),$$

$$F_Y(y) = \lim_{x \to \infty} F_{XY}(x, y).$$

Multiple random variables

In the discrete case, the conditional probability mass function of X given Y is simply

$$p_{Y|X}(y \mid x) = \frac{p_{XY}(x, y)}{p_X(x)},$$

assuming that $p_X(x) \neq 0$.

A useful formula that often arises when trying to derive expression for the conditional probability of one variable given another, is Bayes' rule:

$$p_{Y|X}(y \mid x) = \frac{p_{XY}(x, y)}{p_X(x)} = \frac{p_{X|Y}(x \mid y) p_Y(y)}{\sum_{y'} p_{X|Y}(x \mid y') p_Y(y')}.$$

Multiple random variables

In the discrete case, the conditional probability mass function of X given Y is simply

$$p_{Y|X}(y \mid x) = \frac{p_{XY}(x, y)}{p_X(x)},$$

assuming that $p_X(x) \neq 0$.
A useful formula that often arises when trying to derive expression for the conditional probability of one variable given another, is Bayes' rule:

$$p_{Y|X}(y \mid x) = \frac{p_{XY}(x, y)}{p_X(x)} = \frac{p_{X|Y}(x \mid y) p_Y(y)}{\sum_{y'} p_{X|Y}(x \mid y') p_Y(y')}.$$

Outline

Common inequalities

Theorem (Markov inequality)

Let X be a non-negative random variable, then

$$P(X \geq a) \leq \frac{E[X]}{a}.$$

Common inequalities

Theorem (Chebyshev inequality)

Let X be a random variable, then

$$P(|X - E[X]| \geq a) \leq \frac{Var(x)}{a^2}.$$

Common inequalities

Theorem (Jensen inequality)

Let ϕ be a convex function and X be a random variable, then

$$E[\phi(X)] \geq \phi(E[X]).$$