

Lecture 2

Basic Concepts

Instructor: Shibo Li

shiboli@cs.fsu.edu



Department of Computer Science
Florida State University

Definition

A **sample space** is a set Ω consisting of all possible outcomes of a random experiment.

► Discrete Examples

- Tossing a coin: $\Omega = \{H, T\}$
- Rolling a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Radioactive decay, number of particles emitted per minute: $\Omega = \mathbb{N} = \{0, 1, 2, \dots\}$

► Continuous Examples

- Measuring height of spruce trees: $\Omega = [0, \infty)$
- Image pixel values: $\Omega = [0, M]$

Definition

An **event** in a sample space Ω is a subset $A \subseteq \Omega$.

Examples:

- ▶ In the die rolling sample space, consider the event “An even number is rolled”. This is the event $A = \{2, 4, 6\}$.
- ▶ In the spruce tree example, consider the event “The tree is taller than 80 feet”. This is the event $A = (80, \infty)$.

Given two events A, B of a sample space Ω .

- ▶ Union: $A \cup B$ “or” operation
- ▶ Intersection: $A \cap B$ “and” operation
- ▶ Complement: \bar{A} “negation” operation
- ▶ Subtraction: $A - B$ A happens, B does not

Given a sample space Ω , the space of all possible events \mathcal{F} must satisfy several rules:

- ▶ $\emptyset \in \mathcal{F}$
- ▶ If $A_1, A_2, \dots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.
- ▶ If $A \in \mathcal{F}$, then $\bar{A} \in \mathcal{F}$.

Definition

A set $\mathcal{F} \subseteq 2^{\Omega}$ that satisfies the above rules is called a **σ -algebra**.

Definition

A **measure** on a σ -algebra \mathcal{F} is a function

$\mu : \mathcal{F} \rightarrow [0, \infty)$ satisfying

- ▶ $\mu(\emptyset) = 0$
- ▶ For pairwise disjoint sets $A_1, A_2, \dots \in \mathcal{F}$,
$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

Definition

A measure P on (Ω, \mathcal{F}) is a **probability measure** if
 $P(\Omega) = 1$.

Definition

A **probability space** is a triple (Ω, \mathcal{F}, P) , where

1. Ω is a set, called the **sample space**,
2. \mathcal{F} is a σ -algebra, called the **event space**,
3. and P is a measure on (Ω, \mathcal{F}) with $P(\Omega) = 1$,
called the **probability measure**.

For any probability measure P and events A, B :

- ▶ $P(\bar{A}) = 1 - P(A)$
- ▶ $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Definition

Given a probability space (Ω, \mathcal{F}, P) , the **conditional probability** of an event A given the event B is defined as

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

Die Example:

Let $A = \{2\}$ and $B = \{2, 4, 6\}$. $P(A) = \frac{1}{6}$, but

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}.$$

Definition

Let A and B be two events in a sample space. We say A and B are **independent** given that

$$P(A \cap B) = P(A)P(B).$$

Two events that are not independent are called **dependent**.

Consider two events A and B in a sample space.

If the probability of A doesn't depend on B , then

$$P(A|B) = P(A).$$

Notice, $P(A) = P(A|B) = P(A \cap B)/P(B)$. Multiplying by $P(B)$ gives us

$$P(A \cap B) = P(A)P(B)$$

We get the same result if we start with $P(B|A) = P(B)$.

Theorem

Let A and B be two events in a probability space (Ω, \mathcal{F}, P) . The following conditions are equivalent:

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

Definition

A **random variable** is a function defined on a probability space. In other words, if (Ω, \mathcal{F}, P) is a probability space, then a random variable is a function $X : \Omega \rightarrow V$ for some set V .

Note:

- ▶ A random variable is neither random nor a variable.
- ▶ We will deal with integer-valued ($V = \mathbb{Z}$) or real-valued ($V = \mathbb{R}$) random variables.
- ▶ Technically, random variables are *measurable* functions.

Let (Ω, \mathcal{F}, P) be the probability space for rolling a pair of dice, and let $X : \Omega \rightarrow \mathbb{Z}$ be the random variable that gives the sum of the numbers on the two dice. So,

$$X[(1, 2)] = 3, \quad X[(4, 4)] = 8, \quad X[(6, 5)] = 11$$

Most of the time the random variable X will just be the identity function. For example, if the sample space is the real line, $\Omega = \mathbb{R}$, the identity function

$$X : \mathbb{R} \rightarrow \mathbb{R},$$

$$X(s) = s$$

is a random variable.

Setting a real-valued random variable to a value or range of values defines an event.

$$[X = x] = \{s \in \Omega : X(s) = x\}$$

$$[X < x] = \{s \in \Omega : X(s) < x\}$$

$$[a < X < b] = \{s \in \Omega : a < X(s) < b\}$$

Definition

Let X be a real-valued random variable on the probability space (Ω, \mathcal{F}, P) . Then the **cumulative distribution function** (cdf) of X is defined as

$$F(x) = P(X \leq x)$$

Let X be a real-valued random variable with cdf F . Then F has the following properties:

1. F is monotonic increasing.
2. F is right-continuous, that is,

$$\lim_{\epsilon \rightarrow 0^+} F(x + \epsilon) = F(x), \quad \text{for all } x \in \mathbb{R}.$$

3. $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow \infty} F(x) = 1$.

Definition

The **probability mass function** (pmf) for a discrete real-valued random variable X , denoted p , is defined as

$$p(x) = P(X = x).$$

The cdf can be defined in terms of the pmf as

$$F(x) = P(X \leq x) = \sum_{k \leq x} p(k).$$

Definition

The **probability density function** (pdf) for a continuous real-valued random variable X , denoted p , is defined as

$$p(x) = \frac{d}{dx}F(x),$$

when this derivative exists.

The cdf can be defined in terms of the pdf as

$$F(x) = P(X \leq x) = \int_{-\infty}^x p(t) dt.$$

$$X \sim \text{Unif}(0, 1)$$

" X is uniformly distributed between 0 and 1."

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

$$F(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & x > 1 \end{cases}$$

Consider a differentiable function $f : \mathbb{R} \rightarrow \mathbb{R}$ that transforms a random variable X into a random variable Y by $Y = f(X)$. Then the pdf of Y is given by

$$p(y) = \left| \frac{d}{dy}(f^{-1}(y)) \right| p(f^{-1}(y))$$

Definition

The **expectation** of a continuous random variable X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x p(x) dx.$$

The **expectation** of a discrete random variable X is

$$\mathbb{E}[X] = \sum_i x_i P(X = x_i)$$

This is the “mean” value of X , also denoted $\mu_X = \mathbb{E}[X]$.

If X and Y are random variables, and $a, b \in \mathbb{R}$, then

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$

This extends to several random variables X_i and constants a_i :

$$\mathbb{E}\left[\sum_{i=1}^N a_i X_i\right] = \sum_{i=1}^N a_i \mathbb{E}[X_i].$$

We can also take the expectation of any continuous function of a random variable. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous function and X a random variable, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) p(x) dx.$$

Or, in the discrete case,

$$E[g(X)] = \sum_i g(x_i) P(X = x_i).$$

Definition

The **variance** of a random variable X is defined as

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

- ▶ This formula is equivalent to
$$\text{Var}(X) = E[X^2] - \mu_X^2.$$
- ▶ The variance is a measure of the “spread” of the distribution.
- ▶ The **standard deviation** is the sqrt of variance:
$$\sigma_X = \sqrt{\text{Var}(X)}.$$

$$X \sim N(\mu, \sigma)$$

“ X is normally distributed with mean μ and standard deviation σ .”

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

$$\mathbb{E}[X] = \mu$$

$$\text{Var}(X) = \sigma^2$$

Recall that given two events A, B , we can talk about the intersection of the two events $A \cap B$ and the probability $P(A \cap B)$ of both events happening.

Given two random variables, X, Y , we can also talk about the intersection of the events these variables define. The distribution defined this way is called the **joint distribution**:

$$F(x, y) = P(X \leq x, Y \leq y) = P([X \leq x] \cap [Y \leq y]).$$

Just like the univariate case, we take derivatives to get the joint pdf of X and Y :

$$p(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y).$$

And just like before, we can recover the cdf by integrating the pdf,

$$F(x, y) = \int_{-\infty}^y \int_{-\infty}^x p(s, t) ds dt.$$

Definition

Given a joint probability density $p(x, y)$, the **marginal densities** of X and Y are given by

$$p(x) = \int_{-\infty}^{\infty} p(x, y) dy, \quad \text{and}$$
$$p(y) = \int_{-\infty}^{\infty} p(x, y) dx.$$

The discrete case just replaces integrals with sums:

$$p(x) = \sum_j p(x, y_j), \quad p(y) = \sum_i p(x_i, y).$$

Two Bernoulli random variables:

$$C = \text{cold / no cold} = (1/0)$$

$$R = \text{runny nose / no runny nose} = (1/0)$$

Joint pmf:

		C	
		0	1
R	0	0.40	0.05
	1	0.30	0.25

		<i>C</i>	
		0	1
<i>R</i>	0	0.50	0.05
	1	0.20	0.25

Marginals:

$$P(R = 0) = 0.55, \quad P(R = 1) = 0.45$$

$$P(C = 0) = 0.70, \quad P(C = 1) = 0.30$$

Definition

If X, Y are random variables with joint density $p(x, y)$, then the **conditional density** of X given $Y = y$ is

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

		C		0.55
		0	1	
R	0	0.50	0.05	0.55
	1	0.20	0.25	
		0.7	0.3	

Conditional Probabilities:

$$P(C = 0|R = 0) = \frac{0.50}{0.55} \approx 0.91$$

$$P(C = 1|R = 1) = \frac{0.25}{0.45} \approx 0.56$$

Definition

Two random variables X, Y are called **independent** if

$$p(x, y) = p(x)p(y).$$

If we integrate (or sum) both sides, we see this is equivalent to

$$F(x, y) = F(x)F(y).$$

Definition

Given two random variables X, Y , the **conditional expectation** of X given $Y = y$ is

Continuous case:

$$E[X|Y = y] = \int_{-\infty}^{\infty} x p(x|y) dx$$

Discrete case:

$$E[X|Y = y] = \sum_i x_i P(X = x_i | Y = y)$$

We can take the expected value of the product of two random variables, X and Y :

$$\mathbb{E}[XY] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy.$$

Definition

The **covariance** of two random variables X and Y is

$$\begin{aligned}\text{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\ &= \mathbb{E}[XY] - \mu_X\mu_Y.\end{aligned}$$

This is a measure of how much the variables X and Y “change together”.

We'll also write $\sigma_{XY} = \text{Cov}(X, Y)$.

Definition

The **correlation** of two random variables X and Y is

$$\rho(X, Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}, \quad \text{or}$$

$$\rho(X, Y) = E \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right].$$

Correlation normalizes the covariance between $[-1, 1]$.

If X and Y are two independent RVs, then

$$\begin{aligned}\mathbb{E}[XY] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x, y) dx dy \\&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy p(x)p(y) dx dy \\&= \int_{-\infty}^{\infty} xp(x) dx \int_{-\infty}^{\infty} yp(y) dy \\&= \mathbb{E}[X] \mathbb{E}[Y] = \mu_X \mu_Y\end{aligned}$$

So, $\sigma_{XY} = \mathbb{E}[XY] - \mu_X \mu_Y = 0$.

Warning: Independence implies uncorrelation, but uncorrelated variables are not necessarily independent!

Independence \Rightarrow Uncorrelated

Uncorrelated $\not\Rightarrow$ Independence

OR

Correlated \Rightarrow Dependent

Dependent $\not\Rightarrow$ Correlated