Probability Crash Course

CS 6957: Probabilistic Modeling

January 7, 2014

Sample Spaces

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: $\Omega = \mathbb{N} = \{0, 1, 2, \ldots\}$
- Continuous Examples
  - Measuring height of spruce trees: $\Omega = [0, \infty)$
  - Image pixel values: $\Omega = [0, M]$

Events

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 $A = (80, \infty)$.

Operations on Events

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

Event Spaces

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

- $\emptyset \in \mathcal{F}$
- If $A_1, A_2, \ldots \in \mathcal{F}$, 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**.

Probability Measures

Definition

A **measure** on a σ-algebra $\mathcal{F}$ is a function $\mu : \mathcal{F} \to [0, \infty)$ satisfying

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

Definition

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

Probability Spaces

Definition

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$, 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 $(\Omega, \mathcal{F})$ with $P(\Omega) = 1$, called the **probability measure**.

Some Properties of Probability Measures

For any probability measure P and events A, B:

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

Conditional Probability

Definition

Given a probability space $(\Omega, \mathcal{F}, P)$, 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 $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{1/2} = \frac{1}{3}$.

Independence

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**.

Independence

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

Independence

Theorem

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

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

Random Variables

Definition

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

Dice Example

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

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

Even Simpler Example

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

$$X : \mathbb{R} \to \mathbb{R},$$
$$X(s) = s$$

is a random variable.

Defining Events via Random Variables

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

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

Cumulative Distribution Functions

Definition

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

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

Properties of CDFs

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,

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

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

Probability Mass Functions (Discrete)

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

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

Probability Density Functions (Continuous)

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

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

Example: Uniform Distribution

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

"X is uniformly distributed between 0 and 1."

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

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

Transforming a Random Variable

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

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

Expectation

Definition

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

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

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

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

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

Linearity of Expectation

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

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

This extends the several random variables X_i and constants a_i:

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

Expectation of a Function of a RV

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

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

Or, in the discrete case,

$$\mathrm{E}[g(X)] = \sum_{i} g(x_i)\, P(X = x_i).$$

Variance

Definition

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

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

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

Example: Normal Distribution

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

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

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

$$\mathrm{E}[X] = \mu$$
$$\mathrm{Var}(X) = \sigma^2$$

Joint Distributions

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**:

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

Joint Densities

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,

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

Marginal Distributions

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:

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

Cold Example: Probability Tables

Two Bernoulli random variables:
$C =$ cold / no cold $= (1/0)$
$R =$ runny nose / no runny nose $= (1/0)$

Joint pmf:

		C = 0	C = 1
R	0	0.40	0.05
	1	0.30	0.25

Cold Example: Marginals

		C = 0	C = 1
R	0	0.50	0.05
	1	0.20	0.25

Marginals:

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

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

Conditional Densities

Definition

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

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

Cold Example: Conditional Probabilities

R \ C	0	1	
0	0.50	0.05	0.55
1	0.20	0.25	0.45
	0.7	0.3	

Conditional Probabilities:

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

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

Independent Random Variables

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

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

Conditional Expectation

Definition

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

Continuous case:

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

Discrete case:

$$\mathrm{E}[X|Y = y] = \sum_i x_i\, P(X = x_i|Y = y)$$

Expectation of the Product of Two RVs

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

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

Covariance

Definition

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

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

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

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

Correlation

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 $[-1, 1]$.

Independent RVs are Uncorrelated

If X and Y are two independent RVs, then

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

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

More on Independence and Correlation

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

$$\text{Independence} \Rightarrow \text{Uncorrelated}$$
$$\text{Uncorrelated} \not\Rightarrow \text{Independence}$$

OR

$$\text{Correlated} \Rightarrow \text{Dependent}$$
$$\text{Dependent} \not\Rightarrow \text{Correlated}$$