Prof. Lorenzo Naranjo

Probability Basics

In these notes, I review important probability concepts that will be used throughout the course. For clarity, I present the results using a discrete probability space. However, all results extend to the general case, where $(\Omega, \mathcal{F}, \mathrm{P})$ consists of an arbitrary sample space Ω, a σ-algebra $\mathcal{F}$ of subsets of Ω, and a probability measure P defined on $\mathcal{F}$.

Probability Measure

Consider a *probability space* $(\Omega, \mathcal{F}, \mathrm{P})$, where $\Omega = \{\omega_1, \omega_2, \ldots\}$ is a countable set of outcomes, and $\mathcal{F}$ is the collection of all subsets of Ω (the power set 2^Ω). For any event $A \subseteq \Omega$, its probability is given by

$$\mathrm{P}(A) = \sum_{\omega \in A} \mathrm{P}(\omega),$$

where $\mathrm{P}(\omega)$ is the probability assigned to outcome ω. We require that $\mathrm{P}(\Omega) = 1$, so the total probability sums to one. In practice, we focus on outcomes with positive probability, since those with zero probability do not affect any calculations.

If $\{A_i : i \in I\}$ is a collection of pairwise disjoint subsets of Ω, then no outcome ω belongs to more than one A_i. In this case, the probability of their union is simply the sum of their probabilities:

$$\mathrm{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathrm{P}(A_i).$$

This property is called countable additivity, and with $\mathrm{P}(\Omega) = 1$ is enough to define a probability measure on $\mathcal{F}$.

Random Variables

A *random variable* X is a function that assigns a real value to each outcome: $X(\omega)$ for $\omega \in \Omega$. Several outcomes may have the same value of X. For any real number x, the set

$$\{X = x\} = \{\omega \in \Omega : X(\omega) = x\}$$

collects all outcomes where X takes the value x. The probability that X equals x is given by the *probability mass function*:

$$p_X(x) = \mathsf{P}(X = x).$$

For most $x \in \mathbb{R}$, $p_X(x) = 0$; only a countable set of values have positive probability. The *support* of X is the set

$$R_X = \{x \in \mathbb{R} : p_X(x) > 0\},$$

which is countable because Ω is countable.

Expectation and Variance

The *expectation* (mean) of X is

$$\mathsf{E}(X) = \sum_{x \in R_X} x\, p_X(x).$$

The expectation captures the average value of X over all possible outcomes, weighted by their probabilities. The *variance* of X measures the spread of X around its mean:

$$\mathsf{V}(X) = \sum_{x \in R_X} (x - \mathsf{E}(X))^2\, p_X(x).$$

Note that we have

$$
\begin{aligned}
\mathsf{V}(X) &= \sum_{x \in R_X} (x - \mathsf{E}(X))^2 \, p_X(x) \\
&= \sum_{x \in R_X} (x^2 - 2x \, \mathsf{E}(X) + \mathsf{E}(X)^2) \, p_X(x) \\
&= \sum_{x \in R_X} x^2 \, p_X(x) - 2 \, \mathsf{E}(X) \sum_{x \in R_X} x \, p_X(x) + \mathsf{E}(X)^2 \\
&= \mathsf{E}(X^2) - 2 \, \mathsf{E}(X)^2 + \mathsf{E}(X)^2 \\
&= \mathsf{E}(X^2) - \mathsf{E}(X)^2.
\end{aligned}
$$

Joint Probability Mass Function

Suppose X and Y are two random variables. Their joint probability mass function is

$$p_{X,Y}(x, y) = \mathsf{P}(X = x, Y = y),$$

which gives the probability that X takes value x and Y takes value y simultaneously.

To find the probability that Y equals a specific value y, we sum over all possible values of X:

$$p_Y(y) = \sum_{x \in R_X} p_{X,Y}(x, y).$$

This works because the events $\{X = x, Y = y\}$ for different x are disjoint and together cover all ways Y can be y. Similarly,

$$p_X(x) = \sum_{y \in R_Y} p_{X,Y}(x, y).$$

Therefore, we can marginalize out Y to obtain the probability mass function of X, in the same way that we can marginalize out X to obtain the probability mass function of Y.

Conditional Probability

For two events A and B, the *conditional probability* of A given B is

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

Similarly, the conditional probability mass function of Y given $X = x$ is

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

The *conditional expectation* of Y given $X = x$ is

$$\mathsf{E}(Y \mid X = x) = \sum_{y \in R_Y} y\, p_{Y|X}(y \mid x).$$

This is a function of x, and we can define the random variable $\mathsf{E}(Y \mid X)$ by assigning to each outcome ω the value $\mathsf{E}(Y \mid X = X(\omega))$.

A key result in probability theory is the *law of iterated expectations*:

$$\begin{aligned}
\mathsf{E}(\mathsf{E}(Y \mid X)) &= \sum_{x \in R_X} \mathsf{E}(Y \mid X = x)\, p_X(x) \\
&= \sum_{x \in R_X} \sum_{y \in R_Y} y\, p_{Y|X}(y \mid x)\, p_X(x) \\
&= \sum_{x \in R_X} \sum_{y \in R_Y} y\, p_{X,Y}(x, y) \\
&= \sum_{y \in R_Y} y\, p_Y(y) \\
&= \mathsf{E}(Y).
\end{aligned}$$

This means that the expected value of the conditional expectation equals the expected value of Y itself.

Covariance and Correlation

Suppose we have a function $g(X, Y)$ that depends on two random variables X and Y. To compute its expected value, we take a weighted average of $g(x, y)$ over all possible pairs (x, y), using the joint probability mass function:

$$\mathsf{E}(g(X, Y)) = \sum_{x \in R_X} \sum_{y \in R_Y} g(x, y)\, p_{X,Y}(x, y).$$

Alternatively, this can be written as

$$\mathsf{E}(g(X, Y)) = \sum_{(x,y) \in R_{X,Y}} g(x, y)\, p_{X,Y}(x, y),$$

where $R_{X,Y}$ is the set of all pairs (x, y) with $p_{X,Y}(x, y) > 0$. This formula generalizes the expectation to any function of X and Y.

The *covariance* between X and Y measures how much the two variables move together. It is defined as

$$\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\ \mathsf{E}(Y),$$

where $\mathsf{E}(XY)$ is the expected value of the product XY. If X and Y tend to be above or below their means at the same time, the covariance is positive; if one tends to be above its mean when the other is below, the covariance is negative.

The *correlation* between X and Y is a normalized measure of their linear relationship:

$$\rho_{X,Y} = \frac{\mathsf{Cov}(X, Y)}{\sigma_X\, \sigma_Y},$$

where σ_X and σ_Y are the standard deviations of X and Y. Correlation ranges from -1 (perfect negative linear relationship) to 1 (perfect positive linear relationship), with 0 meaning no linear relationship.

Independence

We say that two events A and B are independent if

$$\mathsf{P}(A \cap B) = \mathsf{P}(A)\,\mathsf{P}(B).$$

This definition is easily extended to random variables. Two random variables X and Y are independent if

$$\mathsf{P}(X = x, Y = y) = \mathsf{P}(X = x)\,\mathsf{P}(Y = y),$$

or in more compact notation if $f_{X,Y}(x, y) = f_X(x) f_Y(y)$. We then have that

$$\begin{aligned}
\mathsf{E}(XY) &= \sum_{x \in R_X} \sum_{y \in R_Y} xy\, f_{X,Y}(x, y) \\
&= \sum_{x \in R_X} \sum_{y \in R_Y} xy\, f_X(x) f_Y(y) \\
&= \sum_{y \in R_Y} y\, f_Y(y) \sum_{x \in R_X} x\, f_X(x) \\
&= \sum_{y \in R_Y} y\, f_Y(y)\, \mathsf{E}(X) \\
&= \mathsf{E}(X) \sum_{y \in R_Y} y\, f_Y(y) \\
&= \mathsf{E}(X)\, \mathsf{E}(Y).
\end{aligned}$$

Thus, if two random variables are independent the expectation of their product is equal the product of their expectations. An immediate consequence of this observation is that if X and Y are independent then

$$\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\,\mathsf{E}(Y) = 0.$$

Note that the conversion of this statement is not true. Zero covariance only implies no linear relationship, but X and Y may still be dependent in a nonlinear way.

Example 1. Consider a random variable X that takes the values $\{1, 0, -1\}$, each with probability $1/3$. We compute:

$$\mathsf{E}(X) = \frac{1}{3}(1) + \frac{1}{3}(0) + \frac{1}{3}(-1) = 0,$$

and

$$\mathsf{E}(X^3) = \frac{1}{3}(1^3) + \frac{1}{3}(0^3) + \frac{1}{3}((-1)^3) = 0.$$

Now, define $Y = X^2$. The covariance between X and Y is

$$\mathsf{Cov}(X, Y) = \mathsf{E}(XY) - \mathsf{E}(X)\,\mathsf{E}(Y) = \mathsf{E}(X^3) - \mathsf{E}(X)\,\mathsf{E}(X^2) = 0.$$

This example shows that X and Y are uncorrelated (zero covariance), but they are not independent—Y is completely determined by X.