2002.1 Probability

Part of @Stats Index

Foundations and notation

Basic idea is *sample space* Ω - all the things that can happen. This can be pretty abstract e.g. all the outcomes an experiment could have had.

An *event* is a subset of the sample space eg. $A, B \subseteq \Omega$
A probability function is a mapping $A \to P$ where $0 \leq P \leq 1$ that is defined for all $A \subseteq \Omega$.

Probability of one event $P(A)$ is a *marginal probability*.

The probability of two events both happening - the intersection of the events - is a *joint probability* $P(A \cap B)$.

Probability of one of A or B happening - the union of the events - is $P(A \cup B)$.

Probability functions observe the Kolmogorov axioms:

- Non-negativity $P(A) > 0 \; \forall A$
- Normalization $P(\Omega) = 1$
- Additivity - if $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$
- We could construct something more rigorous with measure theory, but I don't know how, and I don't think I need to

Multiple Events

Given $P(A \cap B)$ and $P(B)$ we can create a new idea - $P(A|B)$, a *conditional probability*.

This can be interpreted as the probability of event A, given the knowledge that event B has occurred. It is defined as

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

This imples that $P(A \cap B) = P(B)P(A|B)$.

Given this definition, we can create:

Law of Total Probability

Let $A_1, \ldots, A_n$ be a partition of Ω - that is, $\bigcup A_i = \Omega$ and $A_i \cap A_j = \emptyset \; \forall i, j$.

Then, for any event $B \subseteq \Omega$,

$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_n \cap B)$$

Equivalently,

$$P(B) = P(A_1)P(B|A_1) + \ldots + P(A_n)P(B|A_n)$$

Bayes' Rule

We know that $P(A|B) = P(A \cap B)/P(B)$, and $P(A \cap B) = P(B \cap A) = P(A)P(B|A)$. Substituting:

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

Independence

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

It follows that $B \indep A$, and $P(A|B) = P(A)$ and vice versa.

Conditional Independence means that $P(A \cap B|C) = P(A|C)P(B|C)$. Can be written $A \indep B|C$.

Random Variables

A random variable is a mapping from a subset of Ω to $\mathbb{R}$.

The *probability law* of a variable $P_X(X)$ is a function that maps each $x \subseteq X$ to $\mathbb{R}$.

The behavior of the random variable is determined by a *distribution function*. This function can be characterized by parameters e.g. the expected value, the variance.

Basics: Discrete variables have a mass function or PMF, continuous ones have a PDF. The PDF/PMF corresponds to a *Cumulative* Density/Mass function CDF/CMF. E.g.

$$F_X(x) = \int_{-\infty}^{x} f_X(z)dz$$

Expected Value

The expected value of a variable is a measure of the location of a distribution. It is given by a weighted average of sorts:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx$$

The expectation operator $E()$ is linear: $E(aX + b) = aE(X) + b$

Variance

Variance is a measure of a random variable's dispersion. It is the expected squared deviation from the expected value:

$$V(X) = E[(X - E(X))^2]$$

Expanding this, we get that

$$\begin{aligned} V(X) &= E[X^2 - 2XE(X) + E(X)^2] \\ &= E(X^2) - 2E(X)E(X) + E(X)^2 \\ &= E(X^2) - E(X)^2 \end{aligned}$$

I like to remember this as "the expectation of the square minus the square of the expectation".

A more easily interpretable quantity is the *standard deviation* of a random variable, defined as

$$\sigma(X) = \sqrt{V(X)}$$

Combinations of Random Variables

There are marginal, joint, and conditional distributions for random variables.

A joint distribution is a function

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

A *marginal* distribution is a function representing $P(X = x)$:

$$f_X(a) = \int_{-\infty}^{\infty} f_{X,Y}(a,y)dy$$

A *conditional* distribution is a function of y, $P(Y = y|X = x)$

$$f_{Y|X}(y|x = a) = \int_{-\infty}^{\infty} f_{X,Y}(a,y)dy$$

Conditional Expectation

The conditional expectation is often what we're actually interested in. For example, what's the distribution of income conditional on kindergarten class sizes?

We can write

$$E(Y|X = a) = \int_{-\infty}^{\infty} y f_{Y|X}(y|a)dy$$

Define a conditional variance

$$V(Y|X = a) = \int [y - E(Y = y|x = a)]^2 f_{Y|X}(y|a)dy$$

Law of Iterated Expectation

For RVs X, Y,

$$E[Y] = E[E[Y|X]] = \int_{-\infty}^{\infty} E(Y|X = a) f_X(a)da$$

Law of Total Variance

$$V(Y) = E(V(Y|X)) + V(E(Y|X))$$

Suppose a variable is measured for several groups - eg income by race. Then the total variance of income is the expected variance of income within each race, plus the variance between the expected values of each race.

Independence

RVs are independent iff $f_{X,Y}(x,y) = f_X(x)f_Y(y)$

Covariance

Covariance of two variables measures whether they vary from their respective means together. Specifically

$$\text{Cov}(X,Y) = E[(X - E(X))(Y - E(Y))]$$

Equivalently

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

If $X \indep Y$ then $\text{Cov}(X,Y) = 0$. But can't infer the converse - e.g. many nonlinear relationships.

'Linearity' of Variance

Variance is not quite linear: $V(aX+b) = a^2V(X)$

For two variables, $V(X+Y) = V(X) + V(Y) + 2\text{Cov}(X,Y)$

If $X \indep Y$ then $V(X \pm Y) = V(X) + V(Y)$

Correlation

A normalized covariance concept.

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}$$

$\rho(X,Y)$ is 0 iff $Y = aX + b$ where $a \neq 0$. A measure of *linear* relationship.