Part A Probability

Michaelmas Term 2020

Matthias Winkel

winkel@stats.ox.ac.uk
or
winkel@maths.ox.ac.uk
Themes of the course

- Convergence of random variables
- Probabilistic limit laws:
  - Laws of large numbers
  - Central limit theorem
- Joint distributions
- Random processes:
  - Markov chains
  - Poisson processes
Review

Probability spaces and random variables

A probability space is a collection \((\Omega, \mathcal{F}, \mathbb{P})\) where:

- \(\Omega\) is a set, called the sample space.
- \(\mathcal{F}\) is a collection of subsets of \(\Omega\). An element of \(\mathcal{F}\) is called an event.
- \(\mathbb{P}\) is a function from \(\mathcal{F}\) to \([0, 1]\), called the probability measure. It assigns a probability to each event in \(\mathcal{F}\).

If we think of the probability space as modelling some “experiment”, then \(\Omega\) represents the “set of outcomes” of the experiment.
Events

The set of events $\mathcal{F}$ should satisfy the following natural conditions:

1. $\Omega \in \mathcal{F}$
2. If $\mathcal{F}$ contains some set $A$ then $\mathcal{F}$ also contains its complement $A^c$ (i.e. $\Omega \setminus A$).
3. If $(A_i, i \in \mathcal{I})$ is a finite or countably infinite collection of events in $\mathcal{F}$, then their union $\bigcup_{i \in \mathcal{I}} A_i$ is also in $\mathcal{F}$.

By combining (2) and (3), we can also get finite or countable intersections as well as unions.
Probability axioms

The probability measure $P$ should satisfy the following conditions:

1. $P(\Omega) = 1$

2. If $(A_i, i \in I)$ is a finite or countably infinite collection of disjoint events, then

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

The second condition is known as countable additivity.
Random variables

A random variable is a function from $\Omega$, for example to $\mathbb{R}$. A random variable represents an observable in our experiment; something we can measure.

Formally, for a function $X : \Omega \mapsto \mathbb{R}$ to be a random variable, we require that the events

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

are contained in $\mathcal{F}$, for every $x$. (Then by taking complements and unions, we will in fact have that the event $\{\omega \in \Omega: X(\omega) \in B\}$ is in $\mathcal{F}$ for a very large class of sets $B$).

We normally write just $\{X \in B\}$ for $\{\omega \in \Omega: X(\omega) \in B\}$. We write $\mathbb{P}(X \in B)$ for the probability of the event $\{X \in B\}$. 
Within one experiment, there will be many observables! That is, on the same probability space we can consider many different random variables.

We generally do not work with the sample space $\Omega$ directly. Instead we work directly with the events and random variables (the “observables”) in the experiment.
Examples of systems (or “experiments”) that we might model using a probability space.

- Throw two dice, one red, one blue. Random variables: the score on the red die; the score on the blue die; the sum of the two; the maximum of the two; the indicator function of the event that the blue score exceeds the red score....

- A Geiger counter detecting particles emitted by a radioactive source. Random variables: the time of the $k$th particle detected, for $k = 1, 2, \ldots$; the number of particles detected in the time interval $[0, t]$ for $t \in \mathbb{R}_+$, ...

- A model for the evolution of a financial market. Random variables: the prices of various stocks at various times; interest rates at various times; exchange rates at various times....

- The growth of a colony of bacteria. Random variables: the number of bacteria present at a given time; the diameter of the colonised region at a given time....

- A call-centre. The time of arrival of the $k$th call; the length of service required by the $k$th caller; the wait-time of the $k$th caller in the queue before receiving service....
Distribution

The distribution of a random variable $X$ is summarised by its (cumulative) distribution function:

$$F_X(x) = \mathbb{P}(X \leq x).$$

Once we know $F$ we can obtain $\mathbb{P}(X \in B)$ for a large class of sets $B$ by taking complements and unions.

$F$ obeys the following properties:

1. $F$ is non-decreasing
2. $F$ is right-continuous
3. $F(x) \to 0$ as $x \to -\infty$
4. $F(x) \to 1$ as $x \to \infty$.

Note that two different random variables (two different “observables” within the same experiment) can have the same distribution. If $X$ and $Y$ have the same distribution we write $X \overset{d}{=} Y$. 
Discrete random variables

A random variable $X$ is **discrete** if there is a finite or countably infinite set $B$ such that $\mathbb{P}(X \in B) = 1$.

We can represent its distribution by the **probability mass function**

$$p_X(x) = \mathbb{P}(X = x), \text{ for } x \in \mathbb{R}$$

We have

- $\sum_x p_X(x) = 1$
- $\mathbb{P}(X \in A) = \sum_{x \in A} p_X(x)$ for any set $A \subseteq \mathbb{R}$. 
Continuous random variables

A random variable $X$ is continuous if its distribution function $F$ can be written as an integral; i.e. there is a function $f$ such that

$$
P(X \leq x) = F(x) = \int_{-\infty}^{x} f(u)du.
$$

$f$ is the (probability) density function of $X$.

$f$ is not unique; for example we can change the value at any single point without affecting the integral. At points where $F$ is differentiable, it's natural to take $f(x) = F'(x)$.

For general (well-behaved) sets $B$,

$$
P(X \in B) = \int_{x \in B} f(x)dx.
$$
Expectation

If $X$ is discrete, its expectation (or mean) is given by

$$
\mathbb{E}(X) = \sum_x x p_X(x).
$$

For $X$ continuous, instead

$$
\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) \, dx.
$$

We could unify these definitions (and extend to random variables which are neither discrete nor continuous). For example, consider approximations of a general random variable by discrete random variables (analogous to the construction of an integral of a general function by defining the integral of a step function using sums, and then extending to general functions using approximation by step functions).
Properties of expectation

(1) $\mathbb{E} I_A = \mathbb{P}(A)$ for any event $A$.
(2) If $\mathbb{P}(X \geq 0) = 1$ then $\mathbb{E} X \geq 0$.
(3) (Linearity 1): $\mathbb{E}(aX) = a\mathbb{E} X$ for any constant $a$.
(4) (Linearity 2): $\mathbb{E}(X + Y) = \mathbb{E} X + \mathbb{E} Y$.

Expectation of a function of a random variable:

For a discrete random variable $X$ with probability mass function $p_X(x)$:

$$\mathbb{E} g(X) = \sum_x g(x)p_X(x) \quad \text{(discrete case)}$$

For a continuous random variable $X$ with probability density function $f(x)$:

$$\mathbb{E} g(X) = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{(continuous case)}$$
Variance and covariance

The variance of a random variable $X$ is defined by

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

The covariance of two random variables $X$ and $Y$ is defined by

$$
\text{Cov}(X, Y) = \mathbb{E} \left[ (X - \mathbb{E} X)(Y - \mathbb{E} Y) \right]
= \mathbb{E} (XY) - (\mathbb{E} X)(\mathbb{E} Y).
$$

Properties:

$$
\text{Var}(aX + b) = a^2 \text{Var} X
$$

$$
\text{Cov}(aX + b, cY + d) = ac \text{Cov}(X, Y)
$$

$$
\text{Var}(X + Y) = \text{Var} X + \text{Var} Y + 2 \text{Cov}(X, Y)
$$

$$
\text{Var}(X_1 + X_2 + \cdots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).
$$
Independence

Events $A$ and $B$ are independent if

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

More generally, a collection of events $\{A_i, i \in \mathcal{I}\}$ are independent if

$$P\left( \bigcap_{i \in J} A_i \right) = \prod_{i \in J} P(A_i)$$

for all finite subsets $J$ of $\mathcal{I}$. 
Random variables $X_1, \ldots, X_n$ are independent if for all $B_1, \ldots, B_n \subset \mathbb{R}$, the events $\{X_1 \in B_1\}, \ldots, \{X_n \in B_n\}$ are independent.

In fact, it’s sufficient that for all $x_1, \ldots, x_n$,

$$
P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(X_1 \leq x_1) \ldots P(X_n \leq x_n).
$$

If $X$ and $Y$ are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, i.e. $\text{Cov}(X, Y) = 0$. The converse is not true!
Examples of probability distributions

- **Continuous:**
  - Uniform, exponential, normal, gamma...

- **Discrete:**
  - Discrete uniform, Bernoulli, binomial, geometric, Poisson...