Fundamental Tools - Probability Theory II

MSc Financial Mathematics

The University of Warwick

September 29, 2015

Informal introduction to measurable random variables

In an example of rolling a die with $\Omega = \{1, 2, 3, 4, 5, 6\}$:

- A random variable maps each outcome in Ω to a real number. Eg

$$X_1(\omega) = \omega, \quad X_2(\omega) = \begin{cases} 1, & \omega \in \{1, 3, 5\}; \\ -1, & \omega \in \{2, 4, 6\}, \end{cases}$$

  are both random variables on Ω.

- X_1 gives the exact outcome of the roll, and X_2 is a binary variable whose value depends on whether the roll is odd or even.
- If we only have information on whether the roll is odd/even (represented by a σ-algebra $\mathcal{F} = \{\emptyset, \Omega, \{1, 3, 5\}, \{2, 4, 6\}\}$), we can determine the value of X_2 but not X_1.
- We say X_2 is measurable w.r.t $\mathcal{F}$, but X_1 is NOT measurable w.r.t $\mathcal{F}$.

Formal definition of random variables

We wrap the formal definition of a random variable and measurability as follows:

Definition (Measurable random variables)

A random variable is a function $X : \Omega \to \mathbb{R}$. It is said to be measurable w.r.t $\mathcal{F}$ (or we say that X is a random variable w.r.t $\mathcal{F}$) if for every Borel set $B \in \mathcal{B}(\mathbb{R})$

$$X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

Informally, X is measurable w.r.t $\mathcal{F}$ if all possible inverses of X can be found in $\mathcal{F}$.

Examples

1. Back to our first example of rolling a die:
   - The possible sets of inverse of X_2 are $\{1,3,5\}$, $\{2,4,6\}$, Ω and $\emptyset$. They are all in $\mathcal{F}$ so X_2 is $\mathcal{F}$-measurable.
   - For X_1, note for example that $X_1^{-1}(6) = \{6\} \notin \mathcal{F}$. X_1 is hence not $\mathcal{F}$-measurable.
2. Let $\Omega = \{-1, 0, 1\}$ and $\mathcal{F} = \{\emptyset, \Omega, \{-1, 1\}, \{0\}\}$:
   - $X_1(\omega) := \omega$ is NOT $\mathcal{F}$-measurable. Eg $X_1^{-1}(1) = \{1\} \notin \mathcal{F}$.
   - $X_2(\omega) := \omega^2$ is $\mathcal{F}$-measurable.

σ-algebra generated by a random variable

- With a given information set (or a σ-algebra), we check if we can determine the value of a random variable (i.e. if it is $\mathcal{F}$-measurable).
- Conversely, given a random variable we want to extract the information contained therein.

Revisiting the example of rolling a die:

- The value of X_1 gives the information on the exact outcome.
- The value of X_2 gives the information on odd/even.

Definition (σ-algebra generated by a r.v)

The σ-algebra generated by a random variable X, denoted by $\sigma(X)$, is the smallest σ-algebra which X is measurable with respect to.

Example

It is hard (or too tedious) to write down precisely the set of $\sigma(X)$ apart from few simple examples.

Example
In an experiment of flipping a coin twice, let $\Omega = \{HH, HT, TH, TT\}$ and consider the random variables

$$X_1(\omega) = \begin{cases} 1, & \omega \in \{HH, HT\}; \\ -1, & \omega \in \{TH, TT\}, \end{cases} \quad X_2(\omega) = \begin{cases} 2, & \omega \in \{HH\}; \\ 1, & \omega \in \{HT\}; \\ -1, & \omega \in \{TH\}; \\ -2, & \omega \in \{TT\}. \end{cases}$$

Here, $\sigma(X_1) = \{\Omega, \emptyset, \{HT, HT\}, \{TH, TT\}\}$ and $\sigma(X_2) = 2^{\Omega}$. In particular, $\sigma(X_1) \subset \sigma(X_2)$ so X_2 is "more informative" than X_1.

From probability space to distribution functions

In practice, we seldom bother working with the abstract concept of a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but rather just focusing on the distributional properties of a random variable X representing the random phenomenon.

For example, suppose we want to model the number of coin flips required to get the first head:

- Formally, we would write $\Omega = \{1, 2, 3, ...\}$, $\mathcal{F} = 2^{\Omega}$ and let $\mathbb{P}$ be a probability measure satisfying $\mathbb{P}(\{\omega : \omega = k\}) = (1-p)^{k-1}p$ for $k = 1, 2, 3...$
- In practice, we would simply let X be the number of flips required, and consider $\mathbb{P}(X = k) = (1-p)^{k-1}p$ for $k = 1, 2, 3...$

From now on whenever we write expression like $\mathbb{P}(X \in B)$, imagine there is a probability space "in the background", and $\mathbb{P}(X \in B)$ actually means $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$.

Cumulative distribution function

For a random variable X, its cumulative distribution function (CDF) is defined as

$$F(x) = \mathbb{P}(X \leqslant x), \quad -\infty < x < \infty.$$

One can check that F has the following properties:

1. F is non-decreasing and right-continuous;
2. $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

Conversely, if a given function F satisfies the above properties, then it is a CDF of some random variable.

Classes of random variables

We can talk about CDF of general variables. But for random variables belonging to two important subclasses, it is more informative to consider their

- *probability mass functions* for *discrete* random variables;
- *probability density functions* for *continuous* random variables.

Warning: there are random variables which are neither discrete nor continuous!

Discrete random variables

- A random variable X is discrete if it only takes value on a countable set $S = \{x_1, x_2, x_3, ...\}$, which is called the support of X.
- A discrete random variable is fully characterised by its probability mass function (p.m.f).

Definition (Probability mass function)

A probability mass function of a discrete random variable X is defined as

$$p_X(x) = \mathbb{P}(X = x), \quad x \in S.$$

Conversely, if a function $p(x)$ satisfies:

1. $p(x) > 0$ *for all* $x \in S$, *and* $p(x) = 0$ *for all* $x \notin S$;
2. $\sum_{x \in S} p(x) = 1$.

where S is some countable set. Then $p(\cdot)$ defines a probability mass function for some discrete random variable with support S.

Expected value and variance

For a discrete r.v X supported on S, its expected value is defined as

$$\mathbb{E}(X) = \sum_{x \in S} x \mathbb{P}(X = x).$$

More generally, for a given function $g(\cdot)$ we define

$$\mathbb{E}(g(X)) = \sum_{x \in S} g(x) \mathbb{P}(X = x).$$

The variance of X is defined as

$$\text{var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Some useful identities for computing $\mathbb{E}(X)$ and $\text{var}(X)$

- Geometric series

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad \sum_{k=1}^{\infty} k x^{k-1} = \frac{1}{(1-x)^2}$$

for $|x| < 1$.

- Binomial series

$$\sum_{k=0}^{n} C_k^n a^k b^{n-k} = (a+b)^n, \quad \sum_{k=1}^{n} k C_k^n a^{k-1} b^{n-k} = n(a+b)^{n-1}.$$

- Taylor's expansion of e^x

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x.$$

Some popular discrete random variables

Bernoulli $Ber(p)$

- A binary outcome of success (1) or failure (0) where the probability of success is p.

Binomial $Bin(n, p)$

- Sum of n independent and identically distributed (i.i.d) $Ber(p)$ r.v's.

Poisson $Poi(\lambda)$

- Limiting case of $Bin(n, p)$ on setting $p = \lambda / n$ and then let $n \to \infty$.

Geometric $Geo(p)$

- Number of trails required to get the first success in a series of i.i.d $Ber(p)$ experiments.

Some popular discrete random variables: a summary

Distribution	Support	Pmf $\mathbb{P}(X=k)$	$\mathbb{E}(X)$	var(X)
$Ber(p)$	$\{0,1\}$	$p1_{(k=1)}+(1-p)1_{(k=0)}$	p	$p(1-p)$
$Bin(n,p)$	$\{0,1,...,n\}$	$C_k^n p^k(1-p)^{n-k}$	np	$np(1-p)$
$Poi(\lambda)$	$\{0,1,2,...\}$	$\frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ
$Geo(p)$	$\{1,2,3,...\}$	$(1-p)^{k-1}p$	$1/p$	$(1-p)/p^2$

Exercises: For each distribution shown above, verify its $\mathbb{E}(X)$ and var(X)

Continuous r.v's and probability density functions

Definition (Continuous r.v and probability density function)

X is a continuous random variable if there exists a non-negative function f such that

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

for any x. f is called the probability density function (p.d.f) of X.
Conversely, if a given function f satisfies:

1. $f(x) \geqslant 0$ *for all* x;
2. $\int_{-\infty}^{\infty} f(x)dx = 1$.

Then f is the probability density function for some continuous random variable.

Remarks:

- $f(x) = F'(x)$. Thus the pdf is uniquely determined by the CDF of X.
- The set $\{x : f(x) > 0\}$ is called the support of X. This is the range which X can take values on.

Probability density function

Probabilities can be computed by integration: for any set A,

$$\mathbb{P}(X \in A) = \int_A f(u)du.$$

Now let's fix x and consider $A = [x, x + \delta x]$. Then

$$\mathbb{P}(x \leqslant X \leqslant x + \delta x) = \int_x^{x+\delta x} f(u)du \approx f(x)\delta x$$

for small δx. Hence $f(x)$ can be interpreted as the probability of X lying in $[x, x + \delta x]$ normalised by δx.

The second observation is that on setting $\delta x = 0$, we have $\mathbb{P}(X = x) = 0$. Thus a continuous random variable has zero probability of taking a *particular* value.

Expected value and variance of a continuous r.v

For discrete random variables, we work out expected value by summing over the countable possible outcomes. In the continuous case, the analogue is to use integration.

For a continuous r.v X, its expected value is defined as

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

More generally, for a given function $g(\cdot)$ we define

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

The variance of X again is defined as

$$\text{var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2.$$

Remarks on expectation and variance

- We have seen how to define $\mathbb{E}(X)$ (and more generally $\mathbb{E}(g(X))$) when X is either discrete or continuous.
- For more general random variables, it is still possible to define $\mathbb{E}(g(X))$ using notions from measure theory (which we won't discuss here).

Some fundamental properties of expectation and variance:

- $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$ for any two random variables X, Y and constants a, b.
- $\text{var}(aX + b) = a^2 \text{var}(X)$ for any two constants a and b.
- $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$ for any two independent random variables X and Y.

Popular examples of continuous r.v's

Distribution	Support	Pdf	$\mathbb{E}(X)$	$\text{var}(X)$
Uniform $U[0,1]$	$[0,1]$	1	$1/2$	$1/12$
Exponential $Exp(\lambda)$	$[0,\infty)$	$\lambda e^{-\lambda x}$	$1/\lambda$	$1/\lambda^2$
Normal $N(\mu,\sigma^2)$	$\mathbb{R}$	$\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

Exercises: For each distribution shown above, verify its $\mathbb{E}(X)$ and $\text{var}(X)$

Transformation of random variables

Suppose X is a random variable with known distribution. Let $g(\cdot)$ be a function and define a new random variable via $Y = g(X)$. How to find the distribution function of Y?

We start with the first principle: the cumulative distribution function of Y is defined as $F_Y(y) = \mathbb{P}(Y \leqslant y)$. Then

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(g(X) \leqslant y).$$

- If g has an inverse, then we can write

  $$F_Y(y) = \mathbb{P}(X \leqslant g^{-1}(y)) = F_X(g^{-1}(y))$$

  where F_X is the cdf of X.
- If g does not have an inverse (eg $g(x) = x^2$), then special care has to be taken to work out $\mathbb{P}(g(X) \leqslant y)$.

Example

Let $X \sim U[0,1]$. Find the distribution and density function of $Y = \sqrt{X}$.

Sketch of answer: For $X \sim U[0,1]$,

$$F_X(x) = \begin{cases} 0, & x < 0; \\ x, & 0 \leqslant x \leqslant 1; \\ 1, & x > 1. \end{cases}$$

Thus for $y \geqslant 0$, $F_Y(y) = \mathbb{P}(\sqrt{X} \leqslant y) = \mathbb{P}(X \leqslant y^2) = F_X(y^2)$, i.e

$$F_Y(y) = \begin{cases} 0, & y < 0; \\ y^2, & 0 \leqslant y \leqslant 1; \\ 1, & y > 1. \end{cases}$$

Differentiating F_Y gives the density function $f_Y(y) = 2y$ for $y \in [0,1]$ (and 0 elsewhere).

Applications of transformation of random variables

- Log-normal random variable:

  Defined via $Y = \exp(X)$ where $X \sim N(\mu, \sigma)$. It could serve as a simple model of stock price (see problem sheet as well).

- Simulation of random variables:

  Given F a CDF of a random variable X. Define the right-continuous inverse as $F^{-1}(y) = \min\{x : F(x) \geqslant y\}$. Then for $U \sim U[0,1]$, the random variable $F^{-1}(U)$ has the same distribution as X.

  Consequence: If we want to simulate X on a computer and if F^{-1} has an easy expression, we just need to simulate a U from $U[0,1]$ (which is very easy) and then $F^{-1}(U)$ is our sample of X.