Probability Review II

Harvard Math Camp - Econometrics

Ashesh Rambachan

Summer 2018

Outline

Outline

Outline

Random Variables: Borel σ-algebra

Going to present a measure-theoretic definition of random variables.

First building block: **Borel σ-algebra**

- $\Omega = \mathbb{R}$, $\mathcal{A} =$ collection of all open intervals. The "smallest" σ-algebra containing all open sets is the **Borel σ-algebra**, denoted as $\mathcal{B}$.
  - Rigorous definition: $\mathcal{B} =$ collection of all Borel sets - any set in $\mathbb{R}$ that can be formed by countable union, countable intersection and relative complement.
- $\mathcal{B}$ contains all closed intervals. Why?
- Higher-dimensions: $\mathcal{B} =$ smallest σ-algebra containing all open balls.

Random Variables: Measurable functions

Second building block

Definition: Let $(\Omega, \mathcal{A}, \mu)$ and $(\Omega', \mathcal{A}', \mu')$ be two measure spaces. Let $f : \Omega \to \Omega'$ be a function. f is **measurable** if and only if $f^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$.

What in the world...

- For a given set of values in the function's range, we can "measure" the subset of the function's domain upon which these values occur.
- $\mu(f^{-1}(A'))$ is well-defined.

Random variables: Measurable functions

Important case:

$$(\Omega', \mathcal{A}', \mu') = (\mathcal{R}, \mathcal{B}, \lambda).$$

λ is the lebesgue measure on $\mathbb{R}$. In this case, f will be real-valued.

f is μ-**measurable** iff

$$f^{-1}((-\infty, c)) = \{\omega \in \Omega : f(\omega) < c\} \in \mathcal{A} \quad \forall c \in \mathbb{R}.$$

Random Variables

Consider a probability space $(\Omega, \mathcal{A}, P)$. A **random variable** is simply a measurable function from the sample space Ω to the real-line.

Formal definition: Let $(\Omega, \mathcal{A}, P)$ be a probability space and $X : \Omega \to \mathbb{R}$ is a function. X is a **random variable** if and only if X is P-measurable. That is, $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$ where $\mathcal{B}$ is the Borel σ-algebra.

Whew... done with that now.

Outline

Cumulative Distribution Function

Let X be a random variable. The **cumulative distribution function** (cdf) $F : \mathbb{R} \to [0, 1]$ of X is defined as

$$F_X(x) = P(X^{-1}(x)) = P(\{\omega \in \Omega : X(\omega) \leq x\}).$$

- We write
$$F_X(x) = P(X \leq x).$$
- $(\mathbb{R}, \mathbb{B}, F_X)$ form a probability space.

Cumulative Distribution Function

The cumulative distribution function F_X has the following properties:

1. For $x_1 \leq x_2$,
$$F_X(x_2) - F_X(x_1) = P(x_1 < X < x_2).$$
2. $\lim_{x \to -\infty} F_X(x) = 0$, $\lim_{x \to \infty} F_X(x) = 1$.
3. $F_X(x)$ is non-decreasing.
4. $F_X(x)$ is right-continuous: $\lim_{x \to x_0^+} F_X(x) = F_X(x_0)$.

Cumulative Distribution Function

The **quantiles** of a random variable X are given by the inverse of its cumulative distribution function.

- The **quantile function** is

$$Q(u) = \inf\{x : F_X(x) \geq u\}.$$

If F_X is invertible, then

$$Q(u) = F_X^{-1}(u).$$

For any function F that satisfies the properties of a cdf listed above, we can construct a random variable whose cdf is F.

- $U \sim U[0,1]$ and $F_U(u) = u$ for all $u \in [0,1]$. Define

$$Y = Q(U),$$

where Q is the quantile function associated with F. When F is invertible, we have

$$F_Y(y) = P(F^{-1}(U) \leq y) = P(U \leq F(y)) = F(y)$$

Discrete Random Variables

If F_X is constant except at a countable number of points (i.e. F_X is a step function), then we say that X is a **discrete random variable**.

$$p_i = P(X = x_i) = F_X(x_i) - \lim_{x \to x_i^-} F_X(x)$$

Use this to define the **probability mass function** (pmf) of X.

$$f_X(x) = \begin{cases} p_i & \text{if} \quad x = x_i, \quad i = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

We can write

$$P(x_1 < X \leq x_2) = \sum_{x_1 < x \leq x_2} f_X(x).$$

Continuous Random Variables

If F_X can be written as

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$$

where $f_X(x)$ satisfies

$$f_X(x) \geq 0 \quad \forall x \in \mathbb{R}$$
$$\int_{-\infty}^{\infty} f_X(t)dt = 1,$$

we say that X is a **continuous random variable**.

At the points where f_X is continuous,

$$f_X(x) = \frac{dF_X(x)}{dx}.$$

We call $f_X(x)$ the **probability density function** (pdf) of X. We call

$$S_X = \{x : f_X(x) > 0\}$$

the **support** of X.

Continuous Random Variables

Note that for $x_2 \geq x_1$,

$$
\begin{aligned}
P(x_1 < X \leq x_2) &= F_X(x_2) - F_X(x_1) \\
&= \int_{x_1}^{x_2} f_X(t)dt
\end{aligned}
$$

and that

$$P(X = x) = 0$$

for a continuous random variable.

Outline

Joint Distributions

Let X, Y be two scalar random variables. A **random vector** (X, Y) is a measurable mapping from Ω to $\mathbb{R}^2$.

The **joint cumulative distribution function** of X, Y is

$$\begin{aligned} F_{X,Y}(x, y) &= P(X \leq x, Y \leq y) \\ &= P(\{\omega : X(\omega) \leq x\} \cap \{\omega : Y(\omega) \leq y\}) \end{aligned}$$

(X, Y) is a **discrete random vector** if

$$F_{X,Y}(x, y) = \sum_{u \leq x} \sum_{v \leq y} f_{X,Y}(u, v),$$

where $f_{X,Y}(x, y) = P(X = x, Y = y)$ is the **joint probability mass function** of (X, Y).

Joint Distributions

(X, Y) is a **continuous random vector** if

$$F_{X,Y}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(u, v) dv du,$$

where $f_{X,Y}(x, y)$ is the **joint probability density function** of (X, Y). As before,

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}$$

at the points of continuity of $F_{X,Y}$.

Joint to Marginal

From the joint cdf of (X, Y), we can recover the **marginal cdfs**.

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P(X \leq x, Y \leq \infty) \\ &= \lim_{y \to \infty} F_{X,Y}(x, y). \end{aligned}$$

We can also recover the **marginal pdfs** from the joint pdf:

$$f_X(x) = \sum_y f_{X,Y}(x, y) \quad \text{if discrete}$$

and

$$f_X(x) = \int_{S_y} f_{X,Y}(x, y) dy \quad \text{if continuous.}$$

Outline

Conditioning & Random Variables: Discrete case

Consider x with $f_X(x) > 0$. The **conditional pmf of Y given $X = x$** is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

This satisfies

$$f_{Y|X}(y|x) \geq 0$$
$$\sum_y f_{Y|X}(y|x) = 1.$$

- $f_{Y|X}(y|x)$ is a well-defined pmf.

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

$$F_{Y|X}(y|x) = P(Y \leq y|X = x) = \sum_{v \leq y} f_{Y|X}(v|x)$$

.

Conditioning & Random Variables: Continuous case

Consider x with $f_X(x) > 0$, the **conditional pdf** of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

- This is a well-defined pdf for a continuous random variable.

The **conditional cdf** is

$$F_{Y|X}(y|x) = \int_{-\infty}^{y} f_{Y|X}(v|x)dv.$$

Independence

The random variables X, Y are **independent** if

$$F_{Y|X}(y|x) = F_Y(y)$$

Equivalently, if

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

.

Also can be defined in terms of the densities. X, Y are independent if $f_{Y|X}(y|x) = f_Y(y)$ or equivalently, if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$.

Outline

Transformations of random variables

Let X be a random variable with cdf F_X. Define the random variable $Y = h(X)$, where h is a one-to-one function whose inverse h^{-1} exists. What is the distribution of Y?

Suppose that X is discrete with values $x_1, \ldots, x_n$. Y is also discrete with the values

$$y_i = h(x_i), \quad \text{for} \quad i = 1, \ldots, n.$$

The pmf of Y is given by

$$P(Y = y_i) = P(X = h^{-1}(x_i))$$

$$f_Y(y) = f_X(h^{-1}(y_i))$$

Transformations of random variables

Suppose that X is continuous. Suppose h is increasing

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X \leq h^{-1}(y)) = F_X(h^{-1}(y)). \end{aligned}$$

So,

$$\begin{aligned} f_Y(y) &= \frac{dF_Y(y)}{dy} \\ &= f_X(h^{-1}(y))\frac{dh^{-1}(y)}{dy} \end{aligned}$$

Suppose h is decreasing.

$$f_Y(y) = -f_X(h^{-1}(y))\frac{dh^{-1}(y)}{dy}$$

Combining these two cases, we have that, in general,

$$f_Y(y) = f_x(h^{-1}(y))|\frac{dh^{-1}(y)}{dy}|$$

Example

$X \sim U[0,1]$ and $Y = X^2$. What is the density of Y?

Outline

Outline

Definition: Discrete Random Variables

X is a discrete random variable. Its **expectation** or **expected value** is defined as

$$E[X] = \sum_x x f_X(x).$$

if $\sum_x |x| f_X(x) < \infty$. Otherwise, its expectation does not exist.

Let $g : \mathbb{R} \to \mathbb{R}$. Then,

$$E[g(X)] = \sum_x g(x) f_X(x)$$

Definition: Continuous Random Variables

Suppose X is a continuous random variable. Its expectation is defined as

$$E[X] = \int_{S_X} x f_X(x) dx$$

if $\int_{S_X} |x| f_X(x) dx < \infty$. Otherwise, its expectation does not exist.

Let $g : \mathbb{R} \to \mathbb{R}$. Then,

$$E[g(X)] = \int_{S_X} g(x) f_X(x) dx$$

Outline

Expectation is a linear operator

Suppose $a, b \in \mathbb{R}$ and $g_1(\cdot), g_2(\cdot)$ are real-valued functions.

1. $E[a] = a$.
2. $E[ag_1(X)] = aE[g_1(X)]$.
3. $E[g_1(X) + g_2(X)] = E[g_1(X)] + E[g_2(X)]$.

Multivariate Expectations

X, Y are random variables with joint density $f_{X,Y}(x,y)$. Let $g(x,y) : \mathbb{R}^2 \to \mathbb{R}$.

$$E[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) f_{X,Y}(x,y)\,dydx.$$

By linearity of the expectation, for $a, b \in \mathbb{R}$,

$$E[aX + bY] = aE[X] + bE[Y].$$

If X, Y are independent, then for any functions $h_1(\cdot), h_2(\cdot)$,

$$E[h_1(X)h_2(Y)] = E[h_1(X)]E[h_2(Y)].$$

Indicator Functions

An **indicator function** $1(A)$ is a function that is equal to one if condition A is true and zero otherwise.

- E.g. if X is a random variable, then

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

Note that (for the continuous case)

$$\begin{aligned} E[1(X \leq x)] &= \int_{-\infty}^{\infty} 1(X \leq x) f_X(x) dx \\ &= \int_{-\infty}^{x} f_X(x) dx \\ &= F_X(x) = P(X \leq x). \end{aligned}$$

More generally, if $A_X \subseteq \mathbb{R}$, we have that

$$E[1(X \in A_X)] = P(X \in A_X)$$

Outline

Moments

Consider a random variable X. The k**-th moment of** X is defined as $E[X^k]$.

- The first moment of X is its **mean**, $E[X]$.

The k**-th centered moment of** X is $E[(X - E[X])^k]$.

- he second centered moment of X is its **variance**, $V(X) = E[(X - E[X])^2]$.

Moment Generating Function (MGF)

The **moment generating function** (MGF) of a random variable X is defined as

$$\mu_X(t) = E[e^{tX}] = \int e^{tx} f_X(x) dx.$$

The MGF of X allows us to easily compute all of the moments of a random variable.

Moment Generating Function (MGF)

We have that

$$\mu'_X(t) = \int xe^{tx} f_X(x)dx, \quad \mu'_X(0) = \int xf_X(x)dx = E[X],$$
$$\mu''_X(t) = \int x^2 e^{tx} f_X(x)dx, \quad \mu''_X(0) = \int x^2 f_X(x)dx = E[X^2].$$

In general, we can show that

$$\mu_X^{(j)}(0) = E[X^j] \quad \text{for} \quad j = 1, 2, \dots$$

The MGF of a random variable completely characterizes the distribution of a random variable. If X, Y are two random variables with the same MGF, then they have the same distribution.

Outline

Covariance

X, Y are two random variables with joint density $f_{X,Y}(x, y)$. The **covariance** between X, Y is

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

The covariance is a linear operator

$$Cov(X, aY + bW) = aCov(X, Y) + bCov(X, W).$$

Moreover, suppose $Z = aX + bY$ for $a, b \in \mathbb{R}$. Then,

$$V(Z) = a^2 V(X) + b^2 V(Y) + 2abCov(X, Y).$$

Moments for Random Vectors

X is an n-dimensional random vector with $X = (X_1, \ldots, X_n)$.

- Its **mean vector** is

$$E[X] = \begin{pmatrix} E[X_1] \\ \vdots \\ E[X_n] \end{pmatrix}$$

- Its **covariance matrix** is

$$V(X) = \Sigma$$

where Σ is an $n \times n$ matrix whose ij-th entry is $\Sigma_{ij} = Cov(X_i, X_j)$.

Σ is a positive semi-definite matrix. Why? $\alpha \in \mathbb{R}^n$ and $Y = \alpha^T X$. Then,

$$V(Y) = \alpha^T \Sigma \alpha \geq 0.$$

This must hold for all $\alpha \in \mathbb{R}^n$.

Outline

Outline

Conditional Expectations

(X, Y) is a pair of random variables with a joint density $f_{X,Y}(x, y)$. The **conditional expectation** of Y given $X = x$ is

$$E[Y|X = x] = \int_{S_Y} y f_{Y|X}(y|x) dy.$$

Note that this is a function of x. It is sometimes denote $\mu_Y(x)$ and called the **regression function**.

Outline

Iterated Expectations

Law of Iterated Expectations:

$$E_Y[Y] = E_X E_{Y|X}[Y],$$

- E_X denotes the expectation taken with respect to the marginal density of X.
- $E_{Y|X}$ denotes the expectation taken with respect to the conditional density of Y given X.

Proof

$$
\begin{aligned}
E_X E_{Y|X}[Y] &= \int \left(\int y f_{Y|X}(y) dy \right) f_X(x) dx \\
&= \int \int y f_{Y|X}(y) f_X(x) dy dx \\
&= \int y \left(\int f_{X,Y}(x,y) dx \right) dy \\
&= \int y f_Y(y) dy = E[Y]
\end{aligned}
$$

Outline

Optimal Forecasting

What are some ways to interpret the conditional expectation?

- The conditional expectation is the solution to an *optimal forecasting* problem.

Suppose you wish to forecast the value of a random variable Y. Pick $h \in \mathbb{R}$ that minimizes the expected mean-square error

$$E[(Y-h)^2] = \int (y-h)^2 f_Y(y)dy.$$

The first-order condition is

$$\int y f_Y(y)dy = \int h f_Y(y)dy \implies h^* = E[Y].$$

Optimal Forecasting

Suppose that we observe another random variable X and see that $X = x$. We wish to forecast Y as a function of x. We minimize

$$E[(Y - h(X))^2].$$

<u>Claim 1</u>: We can write any function of X as

$$h(x) = \mu_Y(x) + g(x)$$

Why?

Choosing h is equivalent to choosing g. Then write

$$(Y - h(X))^2 = (Y - \mu_Y(X))^2 - 2g(X)(Y - \mu_Y(x)) + g(X)^2.$$

Optimal Forecasting

Claim 2:

$$E_{Y|X}[g(X)(Y - \mu_Y(x))] = 0$$

Why?

So,

$$E[(Y - h(X))^2] = E[(Y - \mu_Y(X))^2 + g(X)^2]$$

. and $g^*(x) = 0$ with

$$h^*(x) = \mu_Y(x).$$

L^2 Projection

We can also interpret the conditional expectation of Y given X as the orthogonal projection of Y onto the space of functions of the random variable X i.e. L^2 space.

- This is the focus of the first several lectures of Econ 2120.

Provides a unifying perspective on much of econometrics and this is really the through line of Econ 2120.

Outline

Jensen's Inequality

Jensen's Inequality: Let $h(\cdot)$ be a convex function and X be a random variable. Then,

$$E[h(X)] \geq h(E[X]).$$

If $h(\cdot)$ is concave, then

$$E[h(X)] \leq h(E[X]).$$

Jensen's Inequality Proof

If $h\cdot$ is a convex function, then $\forall x_0$, there exists some constant a such that

$$h(x) \geq h(x_0) + a(x - x_0) \quad \forall x$$

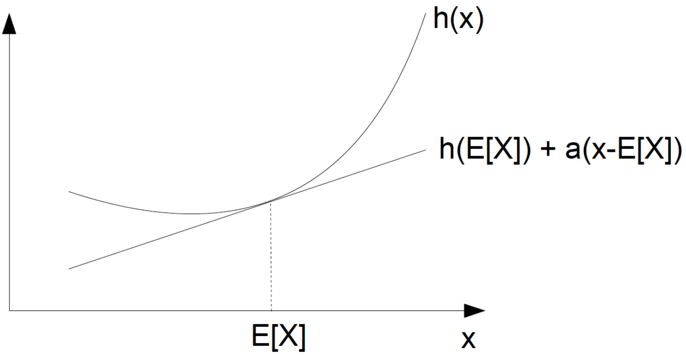
Set $x_0 = E[x]$. It follows that

$$h(X) \geq h(E[X]) + a(x - E[X])$$

holds for all x. Taking expectations, we have that

$$E[h(X)] \geq h(E[X]).$$

Jensen's Inequality Picture Proof



Markov's Inequality

Markov's Inequality: Suppose X is a random variable with $X \geq 0$ with $E[X] < \infty$. Then, for all $M > 0$,

$$P(X \geq M) \leq \frac{E[X]}{M}.$$

- $X \geq 0 \iff P(\{\omega : X(\omega) < 0\}) = 0.$

Application: Suppose that household income is non-negative. No more than $1/5$ of households can have an income that is greater than five times the average household income.

Markov's Inequality Proof

Note

$$X \geq M1(X \geq M).$$

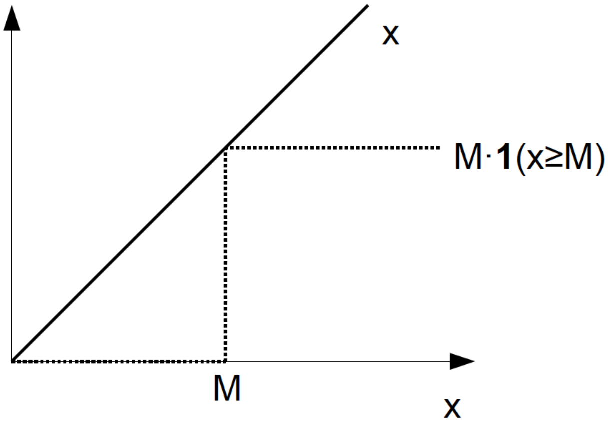
Taking expectations of both sides, we have that

$$E[X] \geq ME[1(X \geq M)] = MP(X \geq M)$$

and re-arrange.

Markov's Inequality Picture Proof

Figure: Proof of Markov's inequality



Chebyshev's Inequality

Chebyshev's Inequality: Suppose that X is a random variable such that $\sigma^2 = Var[X] < \infty$. Then, for all $M > 0$,

$$P(|X - E[X]| > M) \leq \frac{\sigma^2}{M^2}.$$

Chebyshev's Inequality Proof

Let $Y = (X - E[X])^2$. Apply Markov's inequality to Y and the cutoff M^2 to get

$$P(Y \geq M^2) \leq \frac{E[Y]}{M^2}.$$

Rewrite to get that

$$P(|X - E[X]| \geq M) \leq \frac{\sigma^2}{M^2}$$