ECE 730 Notes

John A. Gubner
University of Wisconsin–Madison

September 12, 2018
1

Random variables

*** Read Section 1.4 of the textbook before studying the following material. ***

1.1. The real numbers and extended real numbers

We denote the set of all real numbers by $\mathbb{R} := (-\infty, \infty)$, and we denote the extended real numbers by $\overline{\mathbb{R}} := [-\infty, \infty] = \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$. The extended real numbers behave pretty much as you would expect, but there are two things to watch out for. First, we always take $\pm \infty \cdot 0 = 0$. Second, the expressions $\infty - \infty$ and $-\infty + \infty$ are not defined, and we must guard against them in our calculations.

1.2. Random variables and cumulative distribution functions

Suppose $X$ is an extended-real-valued function defined on the sample space $\Omega$. (Since $\mathbb{R} \subset \overline{\mathbb{R}}$, a real-valued function is a special case of an extended-real-valued function.) We say that $X$ is a random variable if

$$\{ \omega \in \Omega : X(\omega) \leq x \} \in \mathcal{F}, \quad \text{for all real } x.$$  \hfill (1.1)

**Example 1.1.** Let $A$ be any subset of $\Omega$ and put $X := 1_A$, where $1_A$ denotes the indicator function of the set $A$; i.e., $1_A(\omega) := 1$ for $\omega \in A$ and $1_A(\omega) := 0$ for $\omega \notin A$. How do the sets in (1.1) vary with $x$? Determine conditions under which $X$ is a random variable.

**Solution.** To begin, first note that $1_A(\omega) \leq 1$ for all $\omega \in \Omega$. Second, $0 \leq 1_A(\omega) < 1$ if and only if $1_A(\omega) = 0$, which happens if and only if $\omega \notin A$. Third, $1_A(\omega)$ is never negative. Hence,

$$\{ \omega \in \Omega : 1_A(\omega) \leq x \} = \begin{cases} \Omega, & x \geq 1, \\ A^c, & 0 \leq x < 1, \\ \emptyset, & x < 0. \end{cases}$$

This shows that $1_A$ is a random variable if and only all of the sets $\Omega$, $A^c$, and $\emptyset$ belong to $\mathcal{F}$. Since $\Omega$ and $\emptyset$ always belong to $\mathcal{F}$, and since $A^c \in \mathcal{F}$ if and only if $A \in \mathcal{F}$, we conclude that $1_A$ is an random variable if and only if $A \in \mathcal{F}$.

It is common practice in probability theory to drop the $\omega$s and use the shorthand $\{X \leq x\}$ in place of $\{\omega \in \Omega : X(\omega) \leq x\}$. However, it is important to keep in mind that $\{X \leq x\}$ always denotes a subset of the sample space $\Omega$. 

---
If $X$ is a random variable; i.e., the probabilities of the sets in (1.1) are defined, then we further define the **cumulative distribution function** (cdf) of $X$ by

$$F(x) := P(\{\omega \in \Omega : X(\omega) \leq x\}), \quad x \in \mathbb{R}.$$  

Using our just-defined shorthand, we could write $F(x) = P(\{X \leq x\})$. However, we use the further shorthand $F(x) = P(X \leq x)$, always keeping in mind that this refers to the probability of a subset of $\Omega$.

Observe that using one of the limit properties of probability, we can write $P(X < \infty) = P(\bigcup_{n=1}^{\infty} \{X \leq n\}) = \lim_{N \to \infty} P(X \leq N) = \lim_{N \to \infty} F(N)$.

Hence, $P(X = \infty) = 1 - \lim_{N \to \infty} F(N)$. Similarly,

$$P(X > -\infty) = P(\bigcup_{n=1}^{\infty} \{X > -n\}) = \lim_{N \to \infty} P(X > -N) = \lim_{N \to \infty} 1 - F(-N),$$

and so $P(X = -\infty) = \lim_{N \to \infty} F(-N)$. We can now write

$$1 = P(\Omega) = P(X \in \mathbb{R}) = P(X = -\infty) + P(X \in \mathbb{R}) + P(X = \infty).$$

We say $X$ is real valued **with probability one** (w.p.1.) if $P(X \in \mathbb{R}) = 1$. This happens if and only if $F(N) \to 1$ and $F(-N) \to 0$, which corresponds to $P(X = \infty) = 0$ and $P(X = -\infty) = 0$, respectively.

**Remark.** Even though we do not define $F(\pm \infty)$, the values of $P(X = \pm \infty)$ are determined by the values of $F(x)$ for finite values of $x$ using the limits above.

### 1.2.1. Equivalent characterizations of random variables

Since $\{X > x\} = \{X \leq x\}^c$, and since $\{X \leq x\} = \{X > x\}^c$, it makes no difference if we replace $\leq$ by $>$ in the definition of a random variable in (1.1). In fact, we can replace $\leq$ by $<$. To see this, write

$$\{X < x\} = \bigcup_{n=1}^{\infty} \{X \leq x - 1/n\}. \quad (1.2)$$

From this equation, we see that if (1.1) holds, then $\{X < x\} \in \mathcal{A}$ for all real $x$. Conversely, it can be shown that if $\{X < x\} \in \mathcal{A}$ for all real $x$, then (1.1) holds.
2

Expectation

2.1. Definitions

If a random variable takes only finitely many distinct real values, we say it is a **simple random variable**. For a nonnegative simple random variable $X$ taking distinct values $x_1, \ldots, x_m$, we define its *expectation*

$$E[X] := \sum_{k=1}^{m} x_k P\left(\{\omega \in \Omega : X(\omega) = x_k\}\right), \quad \text{(2.1)}$$

or using our shorthand,

$$E[X] := \sum_{k=1}^{m} x_k P(X = x_k).$$

In order for this expression to make sense, it is necessary that the sets $\{X = x_k\} \in \mathcal{A}$. Fortunately, the assumption that $X$ is a random variable implies each

$$\{X = x_k\} = \{X \leq x_k\} \setminus \{X < x_k\}$$

is an event.

**Example 2.1.** Consider the function $a1_A(\omega)$, where $a$ is a positive constant and $A \in \mathcal{A}$. This function takes the two distinct values zero and $a$. Hence, $E[a1_A] = aP(A)$.

Given an arbitrary nonnegative random variable $Y$, we put

$$E[Y] := \sup\{E[X] : X \text{ is simple with } 0 \leq X \leq Y\}.$$  

Alternative notation for $E[Y]$ includes

$$\int Y \, dP, \quad \int Y(\omega) \, dP(\omega), \quad \text{and} \quad \int Y(\omega) \, P(d\omega). \quad \text{(2.2)}$$

We also write $\int_A Y \, dP$ for $E[Y1_A]$.

**Example 2.2** (approximation by simple random variables). Here is an easy way to construct a sequence of simple random variables $X_n$ with $0 \leq X_n \leq Y$ and $X_n \to Y$. Fix any positive integer $n$. Start by breaking the interval $[0, \infty]$ into two parts, $[0, n)$ and $[n, \infty]$. Then break $[0, n)$ into subintervals of length $1/2^n$. Notice that there are $n2^n$ such subintervals. When $Y(\omega) \in [n, \infty]$, put $X_n(\omega) := n$, and when $Y(\omega) \in [(k-1)/2^n, k/2^n)$, put $X_n(\omega) := (k-1)/2^n$. In each case, the value of $X_n(\omega)$ is taken as the left endpoint of the
2.2. Properties of expectation

To see this, observe that since $E$ is monotonic implies $X$ is simple with $0 \leq X \leq Y$, and so $X_n(\omega) = n \rightarrow \infty = Y(\omega)$. If $Y(\omega)$ is finite, then for each $n \geq Y(\omega)$, there will be some $k$, depending on $n$, with $(k-1)/2^n \leq Y(\omega) < k/2^n$. Since $X_n(\omega) = (k-1)/2^n$, the error between $Y(\omega)$ and $X_n(\omega)$ is at most $1/2^n$, which tends to zero as $n \rightarrow \infty$.

Remark. Because $Y$ is assumed to be a random variable, the sets $B_n = \{Y \geq n\}$ and $A_k = \{(k-1)/2^n \leq Y < k/2^n\}$ are events; i.e., belong to the $\sigma$-field $\mathcal{F}$.

Example 2.3. Show that if $Y \geq 0$ and $P(Y > 0) = 0$, then $E[X_n] = 0$ for each $n$.

Solution. The term with $k = 1$ in (2.3) is zero. For $k = 2, \ldots, n2^n$, we have $A_k \subset \{Y > 0\}$, and so $0 \leq P(A_k) \leq P(Y > 0) = 0$. Similarly, $B_n \subset \{Y > 0\}$ implies $P(B_n) = 0$.

2.2. Properties of expectation

Monotonicity. For arbitrary random variables $Y_1$ and $Y_2$ with $0 \leq Y_1 \leq Y_2$, $E[Y_1] \leq E[Y_2]$. To see this, first note that

$$\{X : X \text{ is simple with } 0 \leq X \leq Y_1\} \subset \{X : X \text{ is simple with } 0 \leq X \leq Y_2\}.$$ 

It then follows that

$$\sup\{E[X] : X \text{ is simple with } 0 \leq X \leq Y_1\} \leq \sup\{E[X] : X \text{ is simple with } 0 \leq X \leq Y_2\}.$$ 

Markov inequality. For a nonnegative random variable $Y$ and a constant $a > 0$,

$$P(Y \geq a) \leq \frac{E[Y]}{a}.$$ 

To see this, observe that since

$$a1\{y \geq a\} \leq Y1\{y \geq a\} \leq Y,$$

monotonicity implies $E[a1\{y \geq a\}] \leq E[Y]$, or $aP(Y \geq a) \leq E[Y]$. To conclude, divide by $a$. 

September 12, 2018
Example 2.4 (zero expectation). If a random variable $X$ satisfies $P(X = 0) = 1$, then we say that $X = 0$ with probability one (w.p.1.). Equivalently, if $P(X \neq 0) = 0$, we say that $X = 0$ almost surely (a.s.). Show that if $X$ is nonnegative and $E[X] = 0$, then $X = 0$ w.p.1.

Solution. Since $X \geq 0$, \( \{X \neq 0\} = \{X > 0\} \), and it suffices to show that $P(X > 0) = 0$. Next, observe that

\[
\{X > 0\} = \bigcup_{n=1}^{\infty} \{X > 1/n\}.
\]

Then by a limit property of probability,

\[
P(X > 0) = \lim_{N \to \infty} P(X > 1/N).
\]

By the Markov inequality, $P(X > 1/N) \leq P(X \geq 1/N) \leq N E[X] = 0$.

Example 2.5 (finite expectation). Show that if $X$ is a nonnegative random variable and $E[X] < \infty$, then $X$ is finite w.p.1. In other words, show that $P(X < \infty) = 1$.

Solution. We show that $P(X = \infty) = 0$. To see this, observe that

\[
\{X = \infty\} = \bigcap_{n=1}^{\infty} \{X > n\}.
\]

Then by a limit property of probability,

\[
P(X = \infty) = \lim_{N \to \infty} P(X > N).
\]

By the Markov inequality, $P(X > N) \leq E[X]/N$, which tends to zero as $N \to \infty$ on account of the assumption that $E[X] < \infty$.

Signed random variables

For a signed random variable $Y$, we define the positive and negative parts of $Y$ by

\[
Y^+(\omega) := \begin{cases} 
Y(\omega), & \text{if } Y(\omega) \geq 0, \\
0, & \text{if } Y(\omega) < 0,
\end{cases}
\]

and

\[
Y^-(\omega) := \begin{cases} 
-Y(\omega), & \text{if } Y(\omega) < 0, \\
0, & \text{if } Y(\omega) \geq 0.
\end{cases}
\]

Notice that both $Y^+$ and $Y^-$ are nonnegative random variables and satisfy

\[
Y(\omega) = Y^+(\omega) - Y^-(\omega) \quad \text{and} \quad |Y(\omega)| = Y^+(\omega) + Y^-(\omega).
\]

Hence,

\[
E[|Y|] = E[Y^+] + E[Y^-],
\]

by the additivity property for nonnegative random variables. We then define

\[
\]

provided at least one of the terms on the right is finite. Notice that both terms on the right are finite if and only if $E[|Y|] < \infty$. In this case, we say that $Y$ is integrable.
Expectation

For a signed random variable there are three possibilities:
- \( E[Y] \) may not be defined because the above formula would involve the prohibited expression \( \infty - \infty \);
- \( E[Y] \) may be \( \pm \infty \);
- \( E[Y] \) may be a finite number. This happens if and only if \( Y \) is integrable.

**General monotonicity.** For random variables \( Y_1 \leq Y_2 \), \( E[Y_1] \leq E[Y_2] \), provided both expectations exist. (no proof)

**General additivity.** For random variables \( Y_1 \) and \( Y_2 \), \( E[Y_1 + Y_2] = E[Y_1] + E[Y_2] \), provided the right-hand side is not of the form \( \infty - \infty \) or \( -\infty + \infty \). (no proof)

**Jensen inequality.** Let \( \varphi \) be a convex function defined on an interval of the real line, and let \( Y \) be a random variable taking values in that interval. If \( E[|Y|] < \infty \), then
\[
E[\varphi(Y)] \geq \varphi(E[Y]),
\]
where the left-hand side may be \( +\infty \), but cannot be \( -\infty \). (no proof)

### 2.3. Distributions on the real line

So far we have been discussing extended real valued random variables. Now we restrict attention to real-valued random variables. Fortunately, a real-valued random variable \( X \) has the property that \( \{X \in B\} \in \mathcal{A} \) whenever \( B \) is a Borel subset of \( IR \). It is then easy to show that
\[
\mu(B) := P(\{X \in B\}) = P(X^{-1}(B)), \quad B \in \mathcal{B},
\]
defines a probability measure on \( IR \). This measure is called the distribution of the random variable.

Recall that a real-valued function \( g \) on \( IR \) is said to be measurable if \( \{x \in IR : g(x) \leq t\} \in \mathcal{B} \) for all real \( t \). We can develop integrals for functions \( g \) with respect to \( \mu \) just as we developed expectation for random variables. Common notation for the integral of \( g \) with respect to \( \mu \) includes (cf. (2.2))
\[
\int g \, d\mu, \quad \int g(x) \, d\mu(x), \quad \text{and} \quad \int g(x) \, \mu(dx).
\]
We also write \( \int_B g \, d\mu \) for \( \int g1_B \, d\mu \). All of the results for \( E \) have analogs for \( \int \cdots d\mu \).

**Law of the unconscious statistician (LOTUS).**
\[
E[g(X)] = \int g \, d\mu
\]
in the sense that if either side is defined, then so is the other and both sides are equal.

\(^a\)Recall that the collection of Borel sets, denoted by \( \mathcal{B} \), is the smallest \( \sigma \)-algebra containing all the open subsets of \( IR \).
2.3 Distributions on the real line

2.3.1. Probability mass functions

A probability distribution \( \mu \) is said to have a **probability mass function** if there is a sequence of real numbers \( p_k \) and a sequence of distinct real numbers \( x_k \) such that

\[
\mu(B) = \sum_k 1_B(x_k)p_k, \quad B \in \mathcal{B}.
\]

Taking \( B = \{x_i\} \) shows that \( \mu(\{x_i\}) = p_i \). Since \( \mu \) is a probability measure, this implies \( 0 \leq p_i \leq 1 \). It can be shown that

\[
\int g \, d\mu = \sum_k g(x_k)p_k.
\]

It follows that if \( \mu \) is the distribution of a random variable \( X \), then we have the more familiar version of LOTUS,

\[
E[g(X)] = \sum_k g(x_k)p_k.
\]

2.3.2. Probability densities

A probability distribution \( \mu \) is said to have a **density** if there is a measurable function \( f \) such that

\[
\mu(B) = \int_B f(x) \, dx = \int_{-\infty}^{\infty} 1_B(x)f(x) \, dx, \quad B \in \mathcal{B}.
\]

Integrals \( dx \) are integrals with respect to Lebesgue measure. Lebesgue measure satisfies the first three properties of a probability measure on the Borel subsets of \( \mathbb{R} \). However, since the Lebesgue measure of an interval is defined to be its length, the measure of \( \mathbb{R} \) is infinite rather than one as would be the case for a probability measure on \( \mathbb{R} \). It can be shown that

\[
\int g \, d\mu = \int g(x)f(x) \, dx.
\]

In other words, if \( \mu \) is the distribution of a random variable \( X \), then we have the more familiar version of LOTUS,

\[
E[g(X)] = \int g(x)f(x) \, dx.
\]

\[b\] Consider the set \( N := \{x : f(x) < 0\} \). Then since \( \mu \) is a probability measure, \( \mu(N) \geq 0 \). On the other hand, from the integral formula, \( \mu(N) \leq 0 \). Therefore, \( \mu(N) = 0 \), and it follows that the Lebesgue measure of \( N \) is zero. We describe this condition by saying \( f \geq 0 \) almost everywhere (a.e.).
Index

a.e., see almost everywhere
a.s., see almost surely
almost everywhere. 7
almost surely. 5

Borel set. 6
cumulative distribution function. 2
distribution. 6
expectation. 3
extended real numbers. 1
indicator function. 1
integrable random variable. 5
Jensen inequality. 6
law of the unconscious statistician. 6
LOTUS, see law of the unconscious statistician
measurable function. 6
negative part of a function. 5
positive part of a function. 5
probability density. 7
probability mass function. 7
random variable. 1
  integrable. 5
  simple. 3
simple random variable. 3
w.p.1., see with probability one
with probability one. 2 5