September 1 **Semester 2018-II**

Tutorial 5: Random Processes

Lecturer: Utpal Mukherji/ Parimal Parag *Prepared by: Prathamesh and Karthik*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

5.1 Agenda

- Functions of random variables: sum, min of two random variables
- Jensen's inequality
- Cauchy-Schwartz inequality

5.2 Functions of Random Variables

In this section, we show that well-designed functions of random variables are also random variables. Specifically, let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable. Suppose $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. Then, we would like to show that $f(X) : \Omega \to \mathbb{R}$ is also a random variable, provided f is designed suitably.

Towards this, as a first example, we show from first principles that the sum of two random variables is a random variable.

Theorem 5.2.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If X and Y are two $\mathcal{F}$-measurable random variables, then $X + Y$ is also an $\mathcal{F}$-measurable random variable.*

Proof. Since X and Y are given to be random variables, by definition,

$$\{\omega \in \Omega : X(\omega) < y\} \in \mathcal{F} \text{ for all } y \in \mathbb{R}, \tag{5.1}$$

$$\{\omega \in \Omega : Y(\omega) < y\} \in \mathcal{F} \text{ for all } y \in \mathbb{R}. \tag{5.2}$$

In order to show that $X + Y$ is a random variable, it suffices to show that

$$\{\omega \in \Omega : X(\omega) + Y(\omega) < x\} \in \mathcal{F} \text{ for all } x \in \mathbb{R}. \tag{5.3}$$

Fix an arbitrary $x \in \mathbb{R}$. Then, $X(\omega) + Y(\omega) < x$ implies that <u>there exists</u> a rational number $q \in \mathbb{Q}$ such that $X(\omega) < q$ <u>and</u> $Y(\omega) < x - q$. Conversely, if there exists a rational number $q \in \mathbb{Q}$ such that $X(\omega) < q$ and $Y(\omega) < x - q$, then this implies that $X(\omega) + Y(\omega) < x$. By translating the words "there exists" and "and" into union and intersection of sets respectively, we get that

$$\{\omega \in \Omega : X(\omega) + Y(\omega) < x\} = \bigcup_{q \in \mathbb{Q}} \underbrace{\left(\underbrace{\{\omega \in \Omega : X(\omega) < q\}}_{\in \mathcal{F} \text{ from (5.1) with } y=q} \cap \underbrace{\{\omega \in \Omega : Y(\omega) < x - q\}}_{\in \mathcal{F} \text{ from (5.2) with } y=x-q} \right)}_{\in \mathcal{F} \text{ since intersection of two events in } \mathcal{F} \text{ belongs to } \mathcal{F}}$$

belongs to $\mathcal{F}$ since the union over $q \in \mathbb{Q}$ is a countable union, and countable union of events in $\mathcal{F}$ belongs to $\mathcal{F}$ by the property that $\mathcal{F}$ is a σ-algebra. Thus, $X + Y$ is a random variable. $\square$

On similar lines, as a second example, we show that min of two random variables is a random variable.

Theorem 5.2.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let X and Y be two $\mathcal{F}$-measurable random variables. Define a new function $Z : \Omega \to \mathbb{R}$ as*

$$Z(\omega) = \min\{X(\omega), Y(\omega)\}, \quad \omega \in \Omega.$$

Then, Z is also an $\mathcal{F}$-measurable random variable.

Proof. Since X and Y are given to be random variables, by definition,

$$\{\omega \in \Omega : X(\omega) \le y\} \in \mathcal{F} \text{ for all } y \in \mathbb{R}, \tag{5.4}$$

$$\{\omega \in \Omega : Y(\omega) \le y\} \in \mathcal{F} \text{ for all } y \in \mathbb{R}. \tag{5.5}$$

We need to show that

$$\{\omega \in \Omega : \min\{X(\omega), Y(\omega)\} \le x\} \in \mathcal{F} \text{ for all } x \in \mathbb{R}. \tag{5.6}$$

Fix an arbitrary $x \in \mathbb{R}$. Then, $\min\{X(\omega), Y(\omega)\} \le x$ implies that either $X(\omega) \le x$ or $Y(\omega) \le x$ or both are true, and the converse also holds. Translating "either, or, or both" to set theory by using union, we get

$$\{\omega \in \Omega : \min\{X(\omega), Y(\omega)\} \le x\} = \underbrace{\{\omega \in \Omega : X(\omega) \le x\}}_{\in \mathcal{F} \text{ from (5.4) with } y=x} \cup \underbrace{\{\omega \in \Omega : Y(\omega) \le x\}}_{\in \mathcal{F} \text{ from (5.5) with } y=x} \tag{5.7}$$

belongs to $\mathcal{F}$ since the union of two events in $\mathcal{F}$ belongs to $\mathcal{F}$ by the property that $\mathcal{F}$ is a σ-algebra. Hence $\min\{X, Y\}$ is a random variable. $\square$

Borrowing from the earlier examples, we now proceed to show that certain well-designed functions of random variables are also random variables. We make this notion of "well-designedness" precise in the following definition. Akin to how the definition of random variables is coupled tightly to an underlying σ-algebra, the following definition is tightly coupled to the Borel σ-algebra of subsets of $\mathbb{R}$.

Definition 5.2.3. (Borel-measurable functions) Let $f : \mathbb{R} \to \mathbb{R}$ be a real-valued function. We say that f is a *Borel-measurable function* if the following holds:

$$\{x \in \mathbb{R} : f(x) \in B\} \in \mathcal{B}(\mathbb{R}) \quad \text{ for every Borel set } B \in \mathcal{B}(\mathbb{R}).$$

Notice that the above definition is very similar to the definition of a random variable in that if $(\Omega, \mathcal{F})$ is a measurable space and $X : \Omega \to \mathbb{R}$ is an $\mathcal{F}$-measurable random variable, then we know that

$$\{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F} \quad \text{for every Borel set } B \in \mathcal{B}(\mathbb{R}).$$

Definition 5.2.3 is stated by replacing $(\Omega, \mathcal{F})$ with $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Also, just as any random variable X is conceived as a connection between two worlds, one of $(\Omega, \mathcal{F})$ and the other of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so also, a Borel-measurable function $f : \mathbb{R} \to \mathbb{R}$ should be viewed as connecting the world of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

With this picture in mind, we now seek to provide a connection between the world of $(\Omega, \mathcal{F})$ with the world of $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ in two successive stages: first, by establishing the connection $X : \Omega \to \mathbb{R}$, and next by establishing the connection $f : \mathbb{R} \to \mathbb{R}$, with f being Borel-measurable. We show that this procedure of establishing connections successively will indeed be a fruitful one, resulting in a random variable.

Proposition 5.2.4. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable. Suppose $f : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable function. Then, $f(X) : \Omega \to \mathbb{R}$ is also an $\mathcal{F}$-measurable random variable.*

Proof. Since X is a random variable, by definition,

$$X^{-1}(A) \in \mathcal{F} \text{ for every } A \in \mathcal{B}, \tag{5.8}$$

and since f is Borel-measurable,

$$f^{-1}(B) \in \mathcal{B} \text{ for all } B \in \mathcal{B}. \tag{5.9}$$

In order to show that $g = f(X)$ is a random variable, we need to show that

$$g^{-1}(B) \in \mathcal{F} \text{ for every } B \in \mathcal{B}. \tag{5.10}$$

Fix an arbitrary $B \in \mathcal{B}$. Then,

$$\begin{aligned} g^{-1}(B) &= (f(X))^{-1}(B) \\ &= X^{-1}(f^{-1}(B)) \\ &= X^{-1}(A) \\ &\in \mathcal{F}, \end{aligned} \tag{5.11}$$

where $A = f^{-1}(B) \in \mathcal{B}$ from (5.9) since f is a Borel-measurable function, and $X^{-1}(A) \in \mathcal{F}$ from (5.8) since X is a random variable. □

Thus, every Borel-measurable function of any random variable also yields a random variable. Of special interest among all possible Borel-measurable functions are the well-known *continuous functions*. Thus, every continuous function from $\mathbb{R}$ to $\mathbb{R}$ is Borel-measurable (we do not prove this fact. For a proof of this, you may consult any standard textbook on measure theory). As a consequence of this fact, we have the following result (which we do not prove in this course).

Proposition 5.2.5. *Let $(\Omega, \mathcal{F})$ be a measurable space, and let $X_1, \ldots, X_n$ be $\mathcal{F}$-measurable random variables. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Then $f(X_1, \ldots, X_n)$ is also an $\mathcal{F}$-measurable random variable.*

An immediate consequence of proposition 5.2.5 is that if X is a random variable, then so are $|X|$, X^2, $\tan(X)$, $\log(|X|)$, e^X, $\sin(\pi X^2)$, etc. Similarly, if X, Y are two random variables, then so are $X + Y$, $X - Y$, $\min\{X, Y\}$, $\max\{X, Y\}$, XY, $\log(|X + Y|)$, etc.

5.3 Jensen's Inequality

We now present an important inequality concerning expectations of functions of random variables. The specific types of functions for which this result holds are those of *convex* functions and *concave* functions. We define these precisely below. Before we do so, we provide the following definition.

Definition 5.3.1. (Convex Set) A set $S \subseteq \mathbb{R}^n$ is said to be a *convex set* if the following holds: for any two points x_1, x_2 in S and any scalar $\lambda \in [0, 1]$ we have,

$$\lambda x_1 + (1 - \lambda) x_2 \in S.$$

Notice that in the above definition, as λ varies from 0 to 1, all points lying on the line segment joining x_1 and x_2 are covered. Thus, in words, a convex set is a set in which for any two points $x_1 \in S$ and $x_2 \in S$, the line segment connecting x_1 and x_2 lies completely in S.

Examples of convex sets include $\mathbb{R}^n$, $[0, 1]$, a sphere in 3-dimensions, a square in 2-dimensions, a line segment joining any two points in $\mathbb{R}$, etc. Notice that the union of two disjoint convex sets need not in general be a convex set. Also, a kidney-shaped figure is not a convex set.

The following lemma is an equivalent (which means if and only if) statement about convex sets and can be easily proven from the above definition for convex sets. The proof is left as an exercise.

Lemma 5.3.2. *Let $S \subseteq \mathbb{R}^n$ be a set. Then, S is convex if and only if for any given collection of points $x_1, \cdots x_n$ in S, and scalars $\lambda_1, \ldots, \lambda_n$ satisfying $\lambda_i \in [0,1]$, $\sum_{i=1}^{n} \lambda_i = 1$, we have*

$$\sum_{i=1}^{n} \lambda_i x_i \in S.$$

Proof. Exercise. $\square$

We now provide the definition of a convex function.

Definition 5.3.3. Let $S \subseteq \mathbb{R}^n$ be a convex set, and let $f : S \to \mathbb{R}$ be a function defined on S. Then, f is said to be a *convex function* if the following holds: for any $x_1, x_2 \in S$ and any scalar $\lambda \in [0,1]$, we have

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2).$$

For any $\lambda \in [0,1]$, notice that $\lambda x_1 + (1-\lambda)x_2$ denotes a point on the line segment connecting x_1 and x_2, and $\lambda f(x_1) + (1-\lambda) f(x_2)$ denotes a point on the line segment connecting $f(x_1)$ and $f(x_2)$. Thus, for convex functions, the line segment connecting $f(x_1)$ and $f(x_2)$ always lies above the function evaluated at any point on the line segment connecting x_1 and x_2. Thus, in some sense, every convex function looks like a cup (it has a bulge towards the bottom). Also, $f(x) = ax + b$ for any $a, b \in \mathbb{R}$ is convex (see exercise 7(b) below).

If the "$\leq$" in the definition of a convex function is replaced by "$\geq$", we get the definition of what is known as *concave* function. Thus, if f is convex, then $-f$ is concave, and vice-versa.

Examples of convex functions include $f(x) = x^2,\ e^x,\ \log\left(\frac{1}{x}\right)$ for $x > 0$, $|x|$, etc. Notice that the function $f(x) = ax + b$ is both convex and concave.

The following lemma is an equivalent (which means if and only if) statement about convex functions, and can be easily proven from the above definition of convex function. The proof is left as an exercise.

Lemma 5.3.4. *Let $S \subseteq \mathbb{R}^n$ be a convex set, and let $f : S \to \mathbb{R}$ be a function defined on S. Then, f is a convex function on S if and only if for any given collection of points $x_1, \cdots x_n$ in S, and scalars $\lambda_1, \ldots, \lambda_n$ satisfying $\lambda_i \in [0,1]$, $\sum_{i=1}^{n} \lambda_i = 1$, we have*

$$f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \leq \sum_{i=1}^{n} \lambda_i f(x_i).$$

Proof. Exercise. $\square$

We are now ready to state and prove Jensen's inequality formally.

Theorem 5.3.5. *Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function, and let X and $f(X)$*[1] *be such that $\mathbb{E}[f(X)]$ and $\mathbb{E}[X]$ are finite. Then,*

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

Proof. We begin the proof by recollecting some definitions and properties of expectation of random variables.

[1]Note that a convex function is always a continuous. Therefore, $f(X)$ is a well defined random variable.

Definition 5.3.6. (Simple random variable) Let $(\Omega, \mathcal{F})$ be a measurable space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable. Then, X is called *simple* if it takes only finite many values and can be written in the form

$$X = \sum_{i=1}^{n} a_i 1_{A_i}$$

for some real numbers $a_1, \ldots, a_n$ and sets $A_1 \ldots, A_n$ which are subsets of Ω.

In the above definition, without loss of generality, the sets $A_1, \ldots, A_n$ may be assumed to constitute a partition of the sample space Ω. (why?)

Definition 5.3.7. (Expectation of a simple random variable) Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable simple random variable having the representation

$$X = \sum_{i=1}^{n} a_i 1_{A_i}$$

for some real numbers $a_1, \ldots, a_n$ and sets $A_1 \ldots, A_n$ which are a partition of Ω. Then, its expectation $E[X]$ is defined as

$$E[X] = \sum_{i=1}^{n} a_i P(A_i).$$

Notice that in the above definition, since $A_1, \ldots, A_n$ constitute a partition of Ω, we have $\sum\limits_{i=1}^{n} P(A_i) = 1$.

5.3.1 Proof of Jensen's inequality for simple random variables

We now prove Jensen's inequality for simple random variables. With the preceding definitions, we note that $f(X)$ is a simple random variable that can be represented as

$$f(X) = \sum_{i=1}^{n} f(a_i) 1_{A_i}.$$

Furthermore, we have

$$E[f(X)] = \sum_{i=1}^{n} f(a_i) P(A_i).$$

Since $f : \mathbb{R} \to \mathbb{R}$ is a convex function, choosing $x_i = a_i$ and $\lambda_i = P(A_i)$ in Lemma 5.3.4, we get

$$\sum_{i=1}^{n} f(a_i) P(A_i) \geq f\left(\sum_{i=1}^{n} a_i P(A_i) \right) = f(E[X]),$$

which completes the proof for simple random variables.

5.3.2 A general proof of Jensen's inequality

Before we look at a more general proof of Jensen's inequality, we look at a definition for convex functions that is equivalent to definition 5.3.3.

Definition 5.3.8 (First order definition of convex functions)**.** Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable[2] at every point $x \in \mathbb{R}$. Then f is convex if and only if for all $x, y \in \mathbb{R}$, we have

$$f(y) \geq f(x) + f'(x)(y - x).$$

The above definition says that the first order Taylor series approximation of the value of f at x will always be less than the value at y. As pointed out before, definition 5.3.3 can be recovered from definition 5.3.8, and vice-versa; interested readers can refer [1].

We now recall some properties of expectation which will be used in the proof. The reader is invited to check when each of the following properties is used in the proof that follows later.

Properties of expectation

1. $E[aX] = aE[X]$ for any $a \in \mathbb{R}$.
2. $E[X + Y] = E[X] + E[Y]$.
3. If $X(\omega) \geq Y(\omega)$ for all $\omega \in \Omega$, then $E[X] \geq E[Y]$.
4. $E[c] = c$ for a constant $c \in \mathbb{R}$. Thus, for instance $E[E[X]] = E[X]$.

We now prove Jensen's inequality for differentiable convex functions. For each $\omega \in \Omega$, we know that $X(\omega)$ lies in R. Therefore, from the definition of convexity stated above, it follows that for all $\omega \in \Omega$ and for all $x \in \mathbb{R}$, we have

$$\begin{aligned}
&f(X(w)) \geq f(x) + f'(x)(X(w) - x)\\
&\Rightarrow E[f(X)] \geq E[f(x) + f'(x)(X - x)]\\
&\Rightarrow E[f(X)] \geq E[f(x)] + E[f'(x)(X - x)]\\
&\Rightarrow E[f(X)] \geq f(x) + f'(x) \cdot E([X - x]).
\end{aligned}$$

Choosing $x = \mathbb{E}[X]$ in the above inequality, we get

$$\begin{aligned}
&E[f(X)] \geq f(E[X]) + f'(E[X]) \cdot E([X - E[X]])\\
&\Rightarrow E[f(X)] \geq f(E[X]) + f'(E[X]) \cdot (E[X] - E[E[X]])\\
&\Rightarrow E[f(X)] \geq f(E[X]) + f'(E[X])(E[X] - E[X])\\
&\Rightarrow E[f(X)] \geq f(E[X]),
\end{aligned}$$

thereby completing the proof of Jensen's inequality □

5.4 Cauchy-Schwartz Inequality

We now state and prove yet another important inequality in probability theory, namely the Cauchy-Schwartz inequality.

[2] Although differentiability is assumed in the definition we can have similar definition for non-differentiable convex functions, where the gradient at x is replaced with a subgradient.

Theorem 5.4.1. *Let X, Y be two random variables with $E[X^2] < \infty$, $E[Y^2] < \infty$. Then,*

$$(E[XY])^2 \leq E[X^2]E[Y^2].$$

Proof. The proof is exactly the as in [2]. Define

$$Z_x := (xX + Y)^2, \quad x \in \mathbb{R}.$$

Since $Z_x \geq 0$ for all x in $\mathbb{R}$, we have

$$\begin{aligned}
&E[Z_x] \geq 0 \\
&\Rightarrow E[(xX + Y)^2] \geq 0 \\
&\Rightarrow x^2 E[X^2] + x(2E[XY]) + E[Y^2] \geq 0.
\end{aligned}$$

We have thus arrived at a quadratic expression in x which is always greater than or equal to 0. Therefore, it follows that its discriminant $4E[XY]^2 - 4E[X^2]E[Y^2]$ must be non-positive [3], resulting in

$$(E[XY])^2 \leq E[X^2]E[Y^2],$$

thereby completing the proof. □

5.4.1 Equality in Cauchy-Schwartz inequality (next tutorial session)

In this subsection, we provide a necessary and sufficient condition for equality in the Cauchy-Schwartz inequality. Towards this, we have the following fact, the proof of which is left as an exercise.

Exercise: Suppose we have two random variables X, Y such that

$$Y(\omega) = -aX(\omega) \quad \forall \omega \in \Omega$$

holds for some $a \in \mathbb{R}$. Show that the Cauchy-Schwartz inequality holds with equality.

We now wish to claim that $P((Y + aX) = 0) = 1$ for some $a \in \mathbb{R}$ is a necessary condition for the equality to hold in the Cauchy-Schwartz inequality. Towards this, we use the following lemma, a proof of which will be provided in the next tutorial session.

Lemma 5.4.2. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $Z : \Omega \to \mathbb{R}$ be an $\mathcal{F}$-measurable random variable such that $P(Z \geq 0) = 1$. If $E[Z] = 0$, then $P(Z = 0) = 1$.*

We now use this lemma to proceed. Suppose we have equality in the Cauchy-Schwartz inequality. Then, the equation

$$x^2 E[X^2] + x(2E[XY]) + E[Y^2] \geq 0$$

is met with equality, which implies that for some $a \in \mathbb{R}$, we have

$$\mathbb{E}\left[(aX + Y)^2\right] = 0.$$

Since $P((Y + aX)^2 \geq 0) = 1$, using lemma 5.4.2, we get that

$$P((Y + aX) = 0) = 1,$$

which implies that $Y = -aX$ *almost surely* for some $a \in \mathbb{R}$.

[3] Check this fact.

Exercises Assume that $(\Omega, \mathcal{F}, P)$ is a probability space. Assume that all the random variables appearing below are $\mathcal{F}$-measurable and expectation is with respect to $P(\cdot)$.

1. Prove that if X is a random variable, then $-X$ is also a random variable. Use this to argue that $X - Y$ is a random variable.

2. Prove that if X is a random variable, then so are X^2 and aX for any $a \in \mathbb{R}$. Use this and the previous exercise to argue that XY is a random variable if X and Y are two random variables.

3. If X and Y are random variables, show that
$$\{\omega \in \Omega : X(\omega) = Y(\omega)\} \in \mathcal{F}.$$

4. Let $X : \Omega \to \mathbb{R} \cup \{-\infty, +\infty\}$ be a random variable. Then, show that
$$\{\omega \in \Omega : |X(\omega)| = \infty\} \in \mathcal{F}.$$
(In this problem, X is allowed to take the values $-\infty$ and $+\infty$).

5. Prove, by induction, that for any $n \geq 1$, if $X_1, \ldots, X_n$ are random variables, then the following are also $\mathcal{F}$-measurable random variables:
   (a) $S_n := \frac{X_1 + \ldots + X_n}{n}$
   (b) $T_n := \frac{X_1 + \ldots + X_n}{\sqrt{n}}$.

6. Let X be a random variable, and suppose $X_1, X_2, \ldots$ is a sequence of random variables. Then, show that
$$\left\{\omega \in \Omega : |X_n(\omega) - X(\omega)| > \frac{1}{100}\right\} \in \mathcal{F} \text{ for all } n \geq 1.$$

7. (a) Prove that $f(x) = x^2$, $x \in \mathbb{R}$, is a convex function.
   (b) Prove that for any constants a and b, $g(x) = ax + b$, $x \in \mathbb{R}$, is a convex function.
   (c) Prove that $h(x) = e^x$, $x \in \mathbb{R}$, is a convex function.

8. (a) Prove that $E[X^2] \geq (E[X])^2$.
   (b) Prove that $E[e^X] \geq e^{E[X]}$.
   (c) Prove that $\log(E[X]) \geq E[\log(X)]$ for a random variable $X > 0$.

9. Use Cauchy-Schwartz inequality to show that for any two random variables X and Y with finite second moments, we have
$$\mathrm{Cov}(X, Y) \leq \sqrt{\mathrm{Var}(X) \cdot \mathrm{Var(Y)}}.$$

References

1. Bubeck, Sebastien. "Convex optimization: Algorithms and complexity." Foundations and Trends in Machine Learning 8.3-4 (2015): 231-357.

2. Jacod, Jean, and Philip Protter. Probability essentials. Springer Science and Business Media, 2012.