LECTURE 3

Random Variables

3.1 DEFINITION

It is often convenient to represent the outcome of a random experiment by a number. A random variable (r.v.) is such a representation. To be more precise, let \((\Omega, \mathcal{F}, P)\) be a probability space. Then a random variable \(X : \Omega \rightarrow \mathbb{R}\) is a mapping of the outcome.

**Example 3.1.** Let the random variable \(X\) be the number of heads in \(n\) coin flips. The sample space is \(\Omega = \{H, T\}^n\), the possible outcomes of \(n\) coin flips; then

\[
X \in \{0, 1, 2, \ldots, n\}
\]

**Example 3.2.** Consider packet arrival times \(t_1, t_2, \ldots\) in the interval \((0, T]\). The sample space \(\Omega\) consists of the empty string (no packet) and all finite length strings of the form \((t_1, t_2, \ldots, t_n)\) such that \(0 < t_1 \leq t_2 \leq \cdots \leq t_n \leq T\). Define the random variable \(X\) to be the length of the string; then \(X \in \{0, 1, 2, 3, \ldots\}\).

**Example 3.3.** Consider the voltage across a capacitor. The sample space \(\Omega = \mathbb{R}\). Define the random variables

\[
X(\omega) = \omega,
\]

\[
Y(\omega) = \begin{cases} 
+1, & \omega \geq 0, \\
-1, & \text{otherwise}.
\end{cases}
\]
Example 3.4. Let \((\Omega, \mathcal{F}, P)\) be a probability space. For a given event \(A \in \mathcal{F}\), define the indicator random variable

\[
X(\omega) = \begin{cases} 
1, & \omega \in A, \\
0, & \text{otherwise}.
\end{cases}
\]

We use the notation \(1_A(\omega)\) or \(\chi_A(\omega)\) to denote the indicator random variable for \(A\).

Throughout the course, we use uppercase letters, say, \(X, Y, Z, \Phi, \Theta\), to denote random variables, and lowercase letters to denote the values taken by the random variables. Thus, \(X(\omega) = x\) means that the random variable \(X\) takes on the value \(x\) when the outcome is \(\omega\).

As a representation of a random experiment in the probability space \((\Omega, \mathcal{F}, P)\), the random variable \(X\) can be viewed as an outcome of a random experiment on its own. The sample space is \(\mathbb{R}\) and the set of events is the Borel \(\sigma\)-algebra \(\mathcal{B}\). An event \(A \in \mathcal{B}\) occurs if \(X \in A\) and its probability is

\[
P_X(A) = P\{\omega \in \Omega : X(\omega) \in A\},
\]

which is determined by the probability measure \(P\) of the underlying random experiment and the inverse image of \(A\) under the mapping \(X : \Omega \to \mathbb{R}\). Thus, \((\Omega, \mathcal{F}, P)\) induces a probability space \((\mathbb{R}, \mathcal{B}, P_X)\), where

\[
P_X(A) = P\{\omega \in \Omega : X(\omega) \in A\}, \quad A \in \mathcal{B}.
\]

An implicit assumption here is that for every \(A \in \mathcal{B}\), the inverse \(\{\omega \in \Omega : X(\omega) \in A\}\) is an event in \(\mathcal{F}\). A mapping \(X(\omega)\) satisfying this condition is called measurable (with respect to \(\mathcal{F}\)) and we will always assume that a given mapping is measurable.

Since we typically deal with multiple random variables on the same probability space, we will use the notation \(P\{X \in A\}\) instead of the more formal notation \(P_X(A)\) or \(P\{\omega \in \Omega : X(\omega) \in A\}\).

The inverse image of \(A\) under \(X(\omega)\), i.e., \(\{\omega : X(\omega) \in A\}\)
To determine $P\{X \in A\}$ for any Borel set $A$, i.e., any set generated by open intervals via countable unions, intersections, and complements, it suffices to specify $P\{X \in (a, b]\}$ or $P\{X \in (a, b]\}$ for all $-\infty < a < b < \infty$. Then the probability of any other Borel set can be determined by the axioms of probability. Equivalently, it suffices to specify the cumulative distribution function (cdf) of the random variable $X$:

$$F_X(x) = P\{X \leq x\} = P\{X \in (-\infty, x]\}, \quad x \in \mathbb{R}.$$  

The cdf of a random variable satisfies the following properties.

1. $F_X(x)$ is nonnegative, i.e.,

   $$F_X(x) \geq 0, \quad x \in \mathbb{R}.$$ 

2. $F_X(x)$ is monotonically nondecreasing, i.e.,

   $$F_X(a) \leq F_X(b), \quad a < b.$$ 

3. Limits.

   $$\lim_{x \to -\infty} F_X(x) = 0 \quad \text{and} \quad \lim_{x \to \infty} F_X(x) = 1.$$ 

4. $F_X(x)$ is right continuous, i.e.,

   $$F_X(a^+) := \lim_{x \to a^+} F_X(x) = F_X(a).$$ 

5. Probability of a singleton.

   $$P\{X = a\} = F_X(a) - F_X(a^-),$$

   where $F_X(a^-) := \lim_{x \to a^-} F_X(x)$.

Throughout, we use the notation $X \sim F(x)$ means that the random variable $X$ has the cdf $F(x)$.

![Figure 3.2. An illustration of a cumulative distribution function (cdf).](image-url)
3.3 PROBABILITY MASS FUNCTION (PMF)

A random variable $X$ is said to be discrete if $F_X(x)$ consists only of steps over a countable set $\mathcal{X}$ as illustrated in Figure 3.3.

![Figure 3.3. The cdf of a discrete random variable.](image)

A discrete random variable $X$ can be completely specified by its probability mass function (pmf)

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

The set $\mathcal{X}$ is often referred to as the alphabet of $X$. Clearly, $p_X(x) \geq 0$, $\sum_{x \in \mathcal{X}} p_X(x) = 1$, and

$$P(X \in A) = \sum_{x \in A \cap \mathcal{X}} p_X(x).$$  

Throughout, we use the notation $X \sim p(x)$ to mean that $X$ is a discrete random variable $X$ with pmf $p(x)$.

We review a few famous discrete random variables.

**Bernoulli.** $X \sim \text{Bern}(p)$, $p \in [0, 1]$, has the pmf

$$p_X(1) = p \quad \text{and} \quad p_X(0) = 1 - p.$$  

This is the indicator of observing a head from flipping a coin with bias $p$.

**Geometric.** $X \sim \text{Geom}(p)$, $p \in [0, 1]$, has the pmf

$$p_X(k) = p(1 - p)^{k-1}, \quad k = 1, 2, 3, \ldots.$$  

This is the number of independent coin flips of bias $p$ until the first head.

**Binomial.** $X \sim \text{Binom}(n, p)$, $p \in [0, 1]$, $n = 1, 2, \ldots$, has the pmf

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n.$$  

This is the number of heads in $n$ independent coin flips of bias $p$.

**Poisson.** $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$, has the pmf

$$p_X(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, 2, \ldots.$$
3.4 Probability Density Function

A random variable is said to be continuous if its cdf is continuous as illustrated in Figure 3.4.

If $F_X(x)$ is continuous and differentiable (except possibly over a countable set), then $X$ can be completely specified by its probability density function (pdf) $f_X(x)$ such that

$$F_X(x) = \int_{-\infty}^{x} f_X(u) \, du.$$
If $F_X(x)$ is differentiable everywhere, then by the definition of derivative

$$f_X(x) = \frac{dF_X(x)}{dx} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{P(x < X \leq x + \Delta x)}{\Delta x}. \quad (3.1)$$

The pdf of a random variable satisfies the following properties.

1. $f_X(x)$ is nonnegative, i.e., $f_X(x) \geq 0$, $x \in \mathbb{R}$.
2. Normalization.

$$\int_{-\infty}^{\infty} f_X(x) \, dx = 1.$$

3. For any event $A \subset \mathbb{R}$,

$$P\{X \in A\} = \int_{x \in A} f_X(x) \, dx.$$

In particular,

$$P\{a < X \leq b\} = P\{a < X < b\} = P\{a \leq X < b\} = P\{a \leq X \leq b\} = \int_a^b f_X(x) \, dx.$$

Note that $f_X(x)$ should not be interpreted as the probability that $X = x$. In fact, $f_X(x)$ can be greater than 1. In light of (3.1), it is $f_X(x)\Delta x$ that can be interpreted as the approximation of the probability $P\{x < X \leq x + \Delta x\}$ for $\Delta x$ sufficiently small.

Throughout, we use the notation $X \sim f(x)$ to mean that $X$ is a continuous random variable with pdf $f(x)$.

We review a few famous continuous random variables.

**Uniform.** $X \sim \text{Unif}[a, b], a < b$, has the pdf

$$f_X(x) = \begin{cases} \frac{1}{b-a} & x \in [a, b], \\ 0 & \text{otherwise}. \end{cases}$$

This is often used to model quantization noise.

**Exponential.** $X \sim \text{Exp}(\lambda), \lambda > 0$, has the pdf

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & \text{otherwise}. \end{cases}$$

This is often used to model the service time in a queue or the time between two random arrivals. An exponential random variable satisfies the memoryless property

$$P(X > x + t \mid X > t) = \frac{P[X > x + t]}{P[X > t]} = P[X > x], \quad t, x > 0.$$
Example 3.6. Suppose that for every $t > 0$, the number of packet arrivals during time interval $(0, t]$ is a Poisson$(\lambda t)$ random variable, i.e.,

$$p_N(n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \ldots.$$ 

Let $X$ be the time until the first packet arrival. Then the event $\{X > t\}$ is equivalent to the event $\{N = 0\}$ and thus

$$F_X(t) = 1 - P\{X > t\} = 1 - P\{N = 0\} = 1 - e^{-\lambda t}.$$ 

Hence, $f_X(t) = \lambda e^{-\lambda t}$ and $X \sim \text{Exp}(\lambda)$.

**Gaussian.** $X \sim \text{N}(\mu, \sigma^2)$ has the pdf

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$ 

This characterizes many random phenomena such as thermal and shot noise, and is also called a *normal* random variable. The cdf of the *standard normal* random variable $\text{N}(0, 1)$ is

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$$ 

Its complement is

$$Q(x) = 1 - \Phi(x) = P\{X > x\}.$$ 

The numerical values of the $Q$ function is often used to compute probabilities of any Gaussian random variable $Y \sim \text{N}(\mu, \sigma^2)$ as

$$P\{Y > y\} = P\left\{X > \frac{y - \mu}{\sigma}\right\} = Q\left(\frac{y - \mu}{\sigma}\right).$$ (3.2)

### 3.5 FUNCTIONS OF A RANDOM VARIABLE

Let $X$ be a random variable and $g : \mathbb{R} \to \mathbb{R}$ be a given function. Then $Y = g(X)$ is a random variable and its probability distribution can be expressed through that of $X$. For example, if $X$ is discrete, then $Y$ is discrete and

$$p_Y(y) = P\{Y = y\} = P\{g(X) = y\} = \sum_{x : g(x) = y} p_X(x).$$
In general,

\[ F_Y(y) = \mathbb{P}\{Y \leq y\} = \mathbb{P}\{g(X) \leq y\}, \]

which can be further simplified in many cases.

**Example 3.7 (Linear function).** Let \( X \sim F_X(x) \) and \( Y = aX + b, a \neq 0 \). If \( a > 0 \), then

\[ F_Y(y) = \mathbb{P}\{aX + b \leq y\} = \mathbb{P}\left\{ X \leq \frac{y-b}{a} \right\} = F_X\left( \frac{y-b}{a} \right). \]

Taking derivative with respect to \( y \), we have

\[ f_Y(y) = \frac{1}{a} f_X\left( \frac{y-b}{a} \right). \]

We can similarly show that if \( a < 0 \), then

\[ F_Y(y) = 1 - F_X\left( \frac{y-b}{a} \right) \]

and

\[ f_Y(y) = -\frac{1}{a} f_X\left( \frac{y-b}{a} \right). \]
Combining both cases,

\[ f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \]

As a special case, let \( X \sim N(\mu, \sigma^2) \), i.e.,

\[ f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \]

Again setting \( Y = aX + b \), we have

\[
\begin{align*}
  f_Y(y) &= \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \\
         &= \frac{1}{|a|} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2}} \\
         &= \frac{1}{\sqrt{2\pi(a\sigma)^2}} e^{-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}}.
\end{align*}
\]

Therefore, \( Y \sim N(a\mu + b, a^2\sigma^2) \). This result justifies the use of the Q function in (3.2) to compute probabilities for an arbitrary Gaussian random variable.

**Example 3.8 (Quadratic function).** Let \( X \sim F_X(x) \) and \( Y = X^2 \). If \( y < 0 \), then \( F_Y(y) = 0 \). Otherwise,

\[ F_Y(y) = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} = F_X(\sqrt{y}) - F_X\left((-\sqrt{y})^{-}\right) \]

If \( X \) is continuous with pdf \( f_X(x) \), then

\[ f_Y(y) = \frac{1}{2\sqrt{y}} (f_X(-\sqrt{y}) + f_X(\sqrt{y})). \]

![Figure 3.7. A quadratic function.](image-url)
The above two examples can be generalized as follows.

**Proposition 3.1.** Let \( X \sim f_X(x) \) and \( Y = g(X) \) be differentiable. Then

\[
    f_Y(y) = \sum_{i=1}^{n} \frac{f_X(x_i)}{|g'(x_i)|},
\]

where \( x_1, x_2, \ldots \) are the solutions of the equation \( y = g(x) \) and \( g'(x_i) \) is the derivative of \( g \) evaluated at \( x_i \).

The distribution of \( Y \) can be written explicitly even when \( g \) is not differentiable.

**Example 3.9 (Limiter).** Let \( X \) be a r.v. with Laplacian pdf \( f_X(x) = \frac{1}{2}e^{-|x|} \), and let \( Y \) be defined by the function of \( X \) shown in Figure 3.8. Consider the following cases.

- If \( y < -a \), clearly \( F_Y(y) = 0 \).
- If \( y = -a \),

\[
    F_Y(-a) = F_X(-1) = \int_{-\infty}^{-1} \frac{1}{2}e^x \, dx = \frac{1}{2}e^{-1}.
\]

- If \( -a < y < a \),

\[
    F_Y(y) = P\{Y \leq y\} = P\{aX \leq y\} = P\{X \leq \frac{y}{a}\} = F_X\left(\frac{y}{a}\right)
    = \frac{1}{2}e^{-1} + \int_{-1}^{y/a} \frac{1}{2}e^{-|x|} \, dx.
\]

\[\text{Figure 3.8. The limiter function.}\]
∙ If \( y \geq a \), \( F_X(y) = 1 \).

Combining these cases, the cdf of \( Y \) is sketched in Figure 3.9.

![Figure 3.9. The cdf of the random variable \( Y \).](image1)

**3.6 GENERATION OF RANDOM VARIABLES**

Suppose that we are given a uniform random variable \( X \sim \text{Unif}[0, 1] \) and wish to generate a random variable \( Y \) with prescribed cdf \( F(y) \). If \( F(y) \) is continuous and strictly increasing, set

\[
Y = F^{-1}(X).
\]

Then, since \( X \sim \text{Unif}[0, 1] \) and \( 0 \leq F(y) \leq 1 \),

\[
F_Y(y) = P\{Y \leq y\} = P\{F^{-1}(X) \leq y\} = P\{X \leq F(y)\} = F(y).
\]

Thus, \( Y \) has the desired cdf \( F(y) \). For example, to generate \( Y \sim \text{Exp}(\lambda) \) from \( X \sim \text{Unif}[0, 1] \), we set

\[
Y = -\frac{1}{\lambda} \ln(1 - X).
\]
More generally, for an arbitrary cdf $F(y)$, we define

$$F^{-1}(x) := \min\{y: x \leq F(y)\}, \quad x \in (0, 1].$$ \hspace{1cm} (3.4)

Since $F(y)$ is right continuous, the above minimum is well-defined. Furthermore, since $F(y)$ is monotonically nondecreasing, $F^{-1}(x) \leq y$ iff $x \leq F(y)$. We now set $Y = F^{-1}(X)$ as before, but under this new definition of “inverse.” It follows immediately that the equality in (3.3) continues to hold and that $Y \sim F(y)$. For example, to generate $Y \sim \text{Bern}(p)$, we set

$$Y = \begin{cases} 
0 & X \leq 1 - p, \\
1 & \text{otherwise}.
\end{cases}$$

In conclusion, we can generate a random variable with any desired distribution from a $	ext{Unif}[0, 1]$ random variable.

Conversely, a uniform random variable can be generated from any continuous random variable. Let $X$ be a continuous random variable with cdf $F(x)$ and $Y = F(X)$. Since $F(x) \in [0, 1]$, $F_Y(y) = P[Y \leq y] = 0$ for $y < 0$ and $F_Y(y) = 1$ for $y > 1$. For $y \in [0, 1]$, let $F^{-1}(y)$ be defined as in (3.4). Then

$$F_Y(y) = P[Y \leq y]$$
$$= P[F(X) \leq y]$$
$$= P[X \leq F^{-1}(y)]$$
$$= F(F^{-1}(y))$$
$$= y,$$

where the equality in (3.5) follows by the definition of $F^{-1}(y)$. Hence, $Y \sim U[0, 1]$. For example, let $X \sim \text{Exp}(\lambda)$ and

$$Y = \begin{cases} 
1 - \exp(-\lambda X) & X \geq 0, \\
0 & \text{otherwise}.
\end{cases}$$

Then $Y \sim \text{Unif}[0, 1]$. 
Problems

The exact generation of a uniform random variable, which requires an infinite number of bits to describe, is not possible in any digital computer. One can instead use the following approximation. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) Bern$(1/2)$ random variables, and

$$Y = .X_1X_2\ldots X_n$$

be a fraction in base 2 that lies between 0 and 1. Then $Y$ is a discrete random variable uniformly distributed over the set $\{k/2^n : k = 0, 1, \ldots, 2^n - 1\}$ and its cdf $F(y)$ converges to that of a Unif$[0, 1]$ random variable for every $y$ as $n \to \infty$. Thus, by flipping many fair coin flips, one can simulate a uniform random variable.

The fairness of coin flips is not essential to this procedure. Suppose that $Z_1$ and $Z_2$ are i.i.d. Bern$(p)$ random variable. The following procedure due to von Neumann can generate a single Bern$(1/2)$ random variable, even when the bias $p$ is unknown. Let

$$X = \begin{cases} 0 & (Z_1, Z_2) = (0, 1), \\ 1 & (Z_1, Z_2) = (1, 0). \end{cases}$$

If $(Z_1, Z_2) = (0, 0)$ or $(1, 1)$, then the outcome is ignored. Clearly $p_X(0) = p_X(1) = 1/2$. By repeating the same procedure, one can generate a sequence of i.i.d. Bern$(1/2)$ random variables from a sequence of i.i.d. Bern$(p)$ random variables.

PROBLEMS

3.1. Probabilities from a cdf. Let $X$ be a random variable with the cdf shown below.

```
<table>
<thead>
<tr>
<th>F(x)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>30</td>
</tr>
<tr>
<td>50</td>
</tr>
<tr>
<td>70</td>
</tr>
</tbody>
</table>
```

Find the probabilities of the following events.

(a) $\{X = 2\}$.
(b) $\{X < 2\}$.
(c) $\{X = 2\} \cup \{0.5 \leq X \leq 1.5\}$.
(d) $\{X = 2\} \cup \{0.5 \leq X \leq 3\}$.

3.2. Gaussian probabilities. Let $X \sim N(1000, 400)$. Express the following in terms of the Q function.
Random Variables

(a) \(P(0 < X < 1020]\).
(b) \(P(X < 1020 | X > 960]\).

3.3. \textit{Laplacian}. Let \(X \sim f(x) = \frac{1}{2} e^{-|x|}\).
(a) Sketch the cdf of \(X\).
(b) Find \(P(|X| \leq 2 \text{ or } X \geq 0]\).
(c) Find \(P(|X| + |X - 3| \leq 3]\).
(d) Find \(P(X \geq 0 | X \leq 1]\).

3.4. \textit{Distance to the nearest star}. Let the random variable \(N\) be the number of stars in a region of space of volume \(V\). Assume that \(N\) is a Poisson r.v. with pmf
\[
p_N(n) = e^{-\rho V} (\rho V)^n/n!, \quad \text{for } n = 0, 1, 2, \ldots,
\]
where \(\rho\) is the “density” of stars in space. We choose an arbitrary point in space and define the random variable \(X\) to be the distance from the chosen point to the nearest star. Find the pdf of \(X\) (in terms of \(\rho\)).

3.5. \textit{Time until the \(n\)-th arrival}. Let the random variable \(N\) be the number of packets arriving during time \((0, t]\). Suppose that \(N\) is Poisson with pmf
\[
p_N(n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad \text{for } n = 0, 1, 2, \ldots.
\]
Let the random variable \(Y\) be the time to get the \(n\)-th packet. Find the pdf of \(Y\).

3.6. \textit{Uniform arrival}. The arrival time of a professor to his office is uniformly distributed in the interval between 8 and 9 am.
(a) Find the probability that the professor will arrive during the next minute given that he has not arrived by 8:30.
(b) Repeat for 8:50.

3.7. \textit{Lognormal distribution}. Let \(X \sim N(0, \sigma^2)\). Find the pdf of \(Y = e^X\) (known as the lognormal pdf).

3.8. \textit{Random phase signal}. Let \(Y(t) = \sin(\omega t + \Theta)\) be a sinusoidal signal with random phase \(\Theta \sim U[-\pi, \pi]\). Find the pdf of the random variable \(Y(t)\) (assume here that both \(t\) and the radial frequency \(\omega\) are constant). Comment on the dependence of the pdf of \(Y(t)\) on time \(t\).

3.9. \textit{Quantizer}. Let \(X \sim \text{Exp}(\lambda)\), i.e., an exponential random variable with parameter \(\lambda\) and \(Y = \lfloor X \rfloor\), i.e., \(Y = k\) for \(k \leq X < k + 1, k = 0, 1, 2, \ldots\).
(a) Find the pmf of \(Y\).
(b) Find the pdf of the quantization error \(Z = X - Y\).
3.10. *Gambling.* Alice enters a casino with one unit of capital. She looks at her watch to generate a uniform random variable \( U \sim \text{unif}[0, 1] \), then bets the amount \( U \) on a fair coin flip. Her wealth is thus given by the r.v.

\[
X = \begin{cases} 
1 + U, & \text{with probability } 1/2, \\
1 - U, & \text{with probability } 1/2.
\end{cases}
\]

Find the cdf of \( X \).

3.11. *Nonlinear processing.* Let \( X \sim \text{Unif}[-1, 1] \). Define the random variable

\[
Y = \begin{cases} 
X^2 + 1, & \text{if } |X| \geq 0.5 \\
0, & \text{otherwise}.
\end{cases}
\]

Find and sketch the cdf of \( Y \).