

2. Random Variables

Dave Goldsman

H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

1/15/20

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.1 — Introduction

Definition: A **random variable** (RV) is a function from the sample space to the real line. $X : S \rightarrow \mathbb{R}$.

Example: Flip 2 coins. $S = \{HH, HT, TH, TT\}$.

Suppose X is the RV corresponding to the number of H 's.

$$X(TT) = 0, X(HT) = X(TH) = 1, X(HH) = 2.$$

$$P(X = 0) = \frac{1}{4}, P(X = 1) = \frac{1}{2}, P(X = 2) = \frac{1}{4}. \quad \square$$

Notation: Capital letters like X, Y, Z, U, V, W usually represent RVs.

Small letters like x, y, z, u, v, w usually represent particular values of the RVs.

Thus, you can speak of $P(X = x)$.

Example: Let X be the sum of two dice rolls. Then, e.g., $(4, 6)$ is an outcome from the sample space, and of course $X((4, 6)) = 10$.

In addition,

$$P(X = x) = \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \vdots & \\ 6/36 & \text{if } x = 7 \\ \vdots & \\ 1/36 & \text{if } x = 12 \\ 0 & \text{otherwise.} \quad \square \end{cases}$$

Example: Flip a coin.

$$X \equiv \begin{cases} 0 & \text{if } T \\ 1 & \text{if } H. \end{cases}$$

Example: Roll a die.

$$Y \equiv \begin{cases} 0 & \text{if } \{1, 2, 3\} \\ 1 & \text{if } \{4, 5, 6\}. \end{cases}$$

For our purposes, X and Y are the same, since $P(X = 0) = P(Y = 0) = \frac{1}{2}$ and $P(X = 1) = P(Y = 1) = \frac{1}{2}$.

Example: Select a real number at random between 0 and 1.

There are an *infinite* number of “equally likely” outcomes.

Conclusion: $P(\text{we choose the individual point } x) = P(X = x) = 0$, believe it or not!

But $P(X \leq 0.65) = 0.65$ and $P(X \in [0.3, 0.7]) = 0.4$.

If A is any *interval* in $[0,1]$, then $P(X \in A)$ is the length of A .

Definition: If the number of possible values of a RV X is finite or countably infinite, then X is a **discrete** RV. Otherwise,...

A **continuous** RV is one with probability 0 at every point.

Example: Flip a coin — get H or T . Discrete.

Example: Pick a point at random in $[0, 1]$. Continuous.

Example: The amount of time you wait in a line is either 0 (with positive probability) or some positive real number — a *combined* discrete-continuous random variable!

Outline

- 1 Introduction
- 2 Discrete Random Variables**
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.2 — Discrete Random Variables

Definition: If X is a discrete RV, its **probability mass function (pmf)** is $f(x) \equiv P(X = x)$. Note that $0 \leq f(x) \leq 1$, $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Example: Uniform distribution on integers $1, 2, \dots, n$. X can equal $1, 2, \dots, n$, each with prob $1/n$. $f(i) = 1/n, i = 1, 2, \dots, n$.

Example: A discrete RV can have any values. For instance, let X denote the possible profits from an inventory policy, where $f(-5.1) = 0.2$ (lose money), $f(1.3) = 0.5$ (break even), and $f(11) = 0.3$ (big bucks!).

Example/Definition: Let X denote the number of “successes” from n independent trials such that the $P(\text{success})$ at each trial is p ($0 \leq p \leq 1$). Then X has the **Binomial distribution** with parameters n and p .

The trials are referred to as **Bernoulli trials**.

Notation: $X \sim \text{Bin}(n, p)$. “ X has the Binomial distribution.”

Example: Roll a die 3 independent times. Find

$$P(\text{Get exactly two 6's}).$$

“success” (6) and “failure” (1,2,3,4,5).

All 3 trials are indep, and $P(\text{success}) = 1/6$ doesn't change from trial to trial.

Let $X =$ number of 6's. Then $X \sim \text{Bin}(3, \frac{1}{6})$. \square

Theorem: If $X \sim \text{Bin}(n, p)$, then the probability of k successes in n trials is

$$P(X = k) = \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \dots, n,$$

where $q = 1 - p$.

Proof: Consider the particular sequence of successes and failures:

$$\underbrace{SS \cdots S}_{k \text{ successes}} \underbrace{FF \cdots F}_{n - k \text{ failures}} \quad (\text{probability} = p^k q^{n-k}).$$

The number of ways to arrange the sequence is $\binom{n}{k}$. Done. \square

Example (cont'd): Back to the dice example, where $X \sim \text{Bin}(3, \frac{1}{6})$, and we want $P(\text{Get exactly two 6's})$.

$$n = 3, k = 2, p = 1/6, q = 5/6.$$

$$P(X = 2) = \binom{n}{k} p^k q^{n-k} = \binom{3}{2} \left(\frac{1}{6}\right)^2 \left(\frac{5}{6}\right)^1 = \frac{15}{216}.$$

k	0	1	2	3
$P(X = k)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

□

Example: Roll 2 dice and get the sum. Repeat 12 times.

Find $P(\text{Sum will be 7 or 11 exactly 3 times})$.

Let X = the number of times we get 7 or 11.

$$P(7 \text{ or } 11) = P(7) + P(11) = \frac{6}{36} + \frac{2}{36} = \frac{2}{9}.$$

So $X \sim \text{Bin}(12, 2/9)$, and then

$$P(X = 3) = \binom{12}{3} \left(\frac{2}{9}\right)^3 \left(\frac{7}{9}\right)^9. \quad \square$$

Definition: If $P(X = k) = e^{-\lambda} \lambda^k / k!$, $k = 0, 1, 2, \dots$, $\lambda > 0$, we say that X has the **Poisson distribution** with parameter λ .

Notation: $X \sim \text{Pois}(\lambda)$.

Example: Suppose the number of raisins in a cup of cookie dough is $\text{Pois}(10)$. Find the probability that a cup of dough has at least 4 raisins.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X = 0, 1, 2, 3) \\ &= 1 - e^{-10} \left(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!} \right) \\ &= 0.9897. \quad \square \end{aligned}$$

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables**
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.3 — Continuous Random Variables

Example: Pick a point X randomly between 0 and 1, and define the continuous function

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

Intuitively, if $0 \leq a \leq b \leq 1$, then

$$\begin{aligned} P(a < X < b) &= \text{the “area” under } f(x) \text{ from } a \text{ to } b \\ &= b - a. \end{aligned}$$

Definition: Suppose X is a continuous RV. The magic function $f(x)$ is the **probability density function (pdf)** if

- $\int_{\mathbb{R}} f(x) dx = 1$ (area under $f(x)$ is 1).
- $f(x) \geq 0, \forall x$ (always non-negative).
- If $A \subseteq \mathbb{R}$, then $P(X \in A) = \int_A f(x) dx$ (probability that X is in a certain region A).

Remarks: If X is a continuous RV, then

$$P(a < X < b) = \int_a^b f(x) dx.$$

An individual point has probability *zero*, i.e., $P(X = x) = 0$.

Note that $f(x)$ denotes both pmf (**discrete** case) and pdf (**continuous** case) — but they are *different*:

If X is *discrete*, then $f(x) = P(X = x)$ and must have $0 \leq f(x) \leq 1$.

If X is *continuous*, then

- $f(x)$ *isn't a probability* — but it's used to *calculate* probabilities.
- Instead, think of $f(x) dx \approx P(x < X < x + dx)$.
- Must have $f(x) \geq 0$ (and possibly > 1).
- Calculate the probability of an event by integrating,
$$P(X \in A) = \int_A f(x) dx.$$

Example: If X is “equally likely” to be anywhere between a and b , then X has the **uniform distribution** on (a, b) .

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise.} \end{cases}$$

Notation: $X \sim \text{Unif}(a, b)$.

Remark: $\int_{\mathbb{R}} f(x) dx = \int_a^b \frac{1}{b-a} dx = 1$ (as desired).

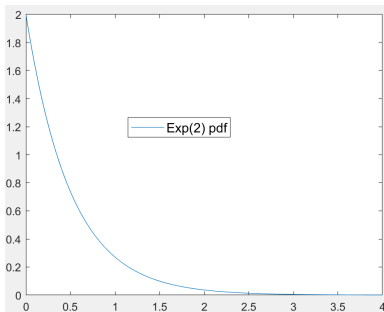
Example: If $X \sim \text{Unif}(-2, 8)$, then

$$P(-1 < X < 6) = \int_{-1}^6 \frac{1}{8 - (-2)} dx = 0.7. \quad \square$$

Example: X has the **exponential distribution** with parameter $\lambda > 0$ if it has pdf $f(x) = \lambda e^{-\lambda x}$, for $x \geq 0$.

Notation: $X \sim \text{Exp}(\lambda)$.

Remark: $\int_{\mathbb{R}} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1$ (as desired).



Example: Suppose $X \sim \text{Exp}(1)$. Then

$$P(X \leq 3) = \int_0^3 e^{-x} dx = 1 - e^{-3}.$$

$$P(X = 3) = \int_3^3 e^{-x} dx = 0.$$

$$P(X \geq 5) = \int_5^{\infty} e^{-x} dx = e^{-5}.$$

$$P(2 \leq X < 4) = P(2 \leq X \leq 4) = \int_2^4 e^{-x} dx = e^{-2} - e^{-4}. \quad \square$$

Example: Suppose X is a continuous RV with pdf

$$f(x) = \begin{cases} cx^2 & \text{if } 0 < x < 2 \\ 0 & \text{otherwise.} \end{cases}$$

First of all, let's find c . Noting that the pdf must integrate to 1, we have

$$1 = \int_{\mathbb{R}} f(x) dx = \int_0^2 cx^2 dx = 8c/3,$$

so that $c = 3/8$.

Now we can calculate any reasonable probabilities, e.g.,

$$P(0 < X < 1) = \int_0^1 \frac{3}{8}x^2 dx = 1/8.$$

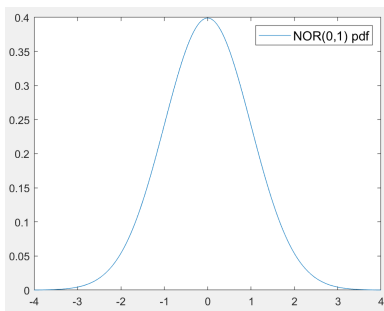
And more-complicated ones, e.g.,

$$\begin{aligned} & P\left(0 < X < 1 \mid \frac{1}{2} < X < \frac{3}{2}\right) \\ &= \frac{P(0 < X < 1 \text{ and } \frac{1}{2} < X < \frac{3}{2})}{P(\frac{1}{2} < X < \frac{3}{2})} \\ &= \frac{P(\frac{1}{2} < X < 1)}{P(\frac{1}{2} < X < \frac{3}{2})} \\ &= \frac{\int_{1/2}^1 \frac{3}{8}x^2 dx}{\int_{1/2}^{3/2} \frac{3}{8}x^2 dx} \\ &= 7/26. \quad \square \end{aligned}$$

Example: X has the **standard normal distribution** if its pdf is

$$\phi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}, \text{ for all } x \in \mathbb{R}.$$

This is the famous “bell curve” distribution.



Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions**
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.4 — Cumulative Distribution Functions

Definition: For any RV X (discrete or continuous), the **cumulative distribution function (cdf)** is defined for all x by

$$F(x) \equiv P(X \leq x).$$

For X discrete,

$$F(x) = \sum_{\{y|y \leq x\}} f(y) = \sum_{\{y|y \leq x\}} P(X = y).$$

For X continuous,

$$F(x) = \int_{-\infty}^x f(y) dy.$$

Discrete cdf's

Example: Flip a coin twice. Let X = number of H 's.

$$X = \begin{cases} 0 \text{ or } 2 & \text{w.p. } 1/4 \\ 1 & \text{w.p. } 1/2. \end{cases}$$

Then the cdf is the following *step function*:

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \leq x < 1 \\ 3/4 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2. \end{cases}$$

Warning! For *discrete* RVs, you have to be careful about “ \leq ” vs. “ $<$ ” at the endpoints of the intervals (where the step function jumps).

Continuous cdf's

Theorem: If X is a *continuous* RV, then $f(x) = F'(x)$ (assuming the derivative exists).

Proof: $F'(x) = \frac{d}{dx} \int_{-\infty}^x f(t) dt = f(x)$, by the Fundamental Theorem of Calculus. \square

Example: $X \sim \text{Unif}(0, 1)$. The pdf and cdf are

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \geq 1. \end{cases} \quad \square$$

Example: $X \sim \text{Exp}(\lambda)$.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

We can use the cdf to find the **median** of X , that is, the point m such that

$$0.5 = F(m) = 1 - e^{-\lambda m}.$$

Solving, we obtain $m = (1/\lambda)\ln(2)$. \square

Properties of all cdf's

$F(x)$ is *non-decreasing* in x , i.e., $a < b$ implies that $F(a) \leq F(b)$.

$\lim_{x \rightarrow \infty} F(x) = 1$ and $\lim_{x \rightarrow -\infty} F(x) = 0$.

$F(x)$ is *right-continuous* at every point x .

Theorem: $P(X > x) = 1 - F(x)$.

Proof: By complements, $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$. \square

Theorem: $a < b \Rightarrow P(a < X \leq b) = F(b) - F(a).$

Proof: Since $a < b$, we have

$$\begin{aligned} P(a < X \leq b) &= P(X > a \cap X \leq b) \\ &= P(X > a) + P(X \leq b) - P(X > a \cup X \leq b) \\ &= 1 - F(a) + F(b) - 1. \quad \square \end{aligned}$$

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations**
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.5 — Great Expectations

The Next Few Lessons:

- Mean (Expected Value)
- Law of the Unconscious Statistician
- Variance
- Probability Inequalities
- Approximations

Definition: The **mean** or **expected value** or **average** of a random variable X is

$$\mu \equiv E[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

The mean gives an indication of a RV's *central tendency*. Think of it as a weighted average of the possible x 's, where the weights are given by $f(x)$.

Example: Suppose X has the **Bernoulli distribution** with parameter p , i.e., $P(X = 1) = p$, $P(X = 0) = q = 1 - p$. Then

$$E[X] = \sum_x x f(x) = (1 \cdot p) + (0 \cdot q) = p.$$

Example: Die toss. $X = 1, 2, \dots, 6$, each w.p. $1/6$. Then

$$E[X] = \sum_x x f(x) = 1 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3.5.$$

Extended Example: Suppose X has the **geometric distribution** with parameter p , i.e., X is the number of $\text{Bern}(p)$ trials until you obtain your first success (e.g., $FFFFS$ gives $X = 5$). Then X has pmf

$$f(x) = (1 - p)^{x-1}p, \quad x = 1, 2, \dots$$

Notation: $X \sim \text{Geom}(p)$.

Here's an application that requires me to admit that I'm not a very good basketball player. 😞 Suppose I take independent foul shots, but the chance of making any particular shot is only 0.4. What's the probability that it'll take me at least 3 tries to make a successful shot?

Answer: The number of tries until my first success is $X \sim \text{Geom}(0.4)$. Thus,

$$\begin{aligned} P(X \geq 3) &= 1 - P(X \leq 2) \\ &= 1 - P(X = 1) - P(X = 2) \\ &= 1 - 0.4 - (0.6)(0.4) = 0.36. \quad \square \end{aligned}$$

Now, let's find the expected value of $X \sim \text{Geom}(p)$.

$$\begin{aligned}
 E[X] &= \sum_x x f(x) = \sum_{x=1}^{\infty} x q^{x-1} p \quad (\text{where } q = 1 - p) \\
 &= p \sum_{x=1}^{\infty} \frac{d}{dq} q^x \\
 &= p \frac{d}{dq} \sum_{x=1}^{\infty} q^x \quad (\text{swap derivative and sum}) \\
 &= p \frac{d}{dq} \frac{q}{1 - q} \quad (\text{geometric sum}) \\
 &= p \left[\frac{(1 - q) - q(-1)}{(1 - q)^2} \right] \quad (\text{ho de hi minus hi dee ho over ho ho}) \\
 &= 1/p. \quad \square
 \end{aligned}$$

So it'll take, on average, $E[X] = 1/p = 1/0.4 = 2.5$ shots. \square

Example: $X \sim \text{Exp}(\lambda)$. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Then

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x f(x) dx \\ &= \int_0^{\infty} x \lambda e^{-\lambda x} dx \\ &= -x e^{-\lambda x} \Big|_0^{\infty} - \int_0^{\infty} (-e^{-\lambda x}) dx \quad (\text{by parts}) \\ &= \int_0^{\infty} e^{-\lambda x} dx \quad (\text{L'Hôpital's rule}) \\ &= 1/\lambda. \quad \square \end{aligned}$$

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance**
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.6 — LOTUS, Moments, and Variance

Law of the Unconscious Statistician (LOTUS)

Theorem: The expected value of a function of X , say $h(X)$, is

$$E[h(X)] \equiv \begin{cases} \sum_x h(x)f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} h(x)f(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

$E[h(X)]$ is a weighted function of $h(x)$, where the weights are the $f(x)$'s.

Remark: It looks like a definition, but it's really a theorem — that's why they call it LOTUS!

Examples: $E[X^2] = \int_{\mathbb{R}} x^2 f(x) dx$, $E[\sin X] = \int_{\mathbb{R}} (\sin x) f(x) dx$.

Just a moment please...

Definition: The k th **moment** of X is

$$E[X^k] = \begin{cases} \sum_x x^k f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x^k f(x) dx & \text{if } X \text{ is continuous.} \end{cases}$$

Example: Suppose $X \sim \text{Bern}(p)$, so that $f(1) = p$ and $f(0) = q$.

$$E[X^k] = \sum_x x^k f(x) = (0^k \cdot q) + (1^k \cdot p) = p \quad \text{for all } k! \quad \square$$

Example: $X \sim \text{Exp}(\lambda)$. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Then

$$\begin{aligned} \mathbb{E}[X^k] &= \int_{\mathbb{R}} x^k f(x) dx \\ &= \int_0^{\infty} x^k \lambda e^{-\lambda x} dx \\ &= \int_0^{\infty} (y/\lambda)^k \lambda e^{-y} (1/\lambda) dy \quad (\text{substitute } y = \lambda x) \\ &= \frac{1}{\lambda^k} \int_0^{\infty} y^{(k+1)-1} e^{-y} dy \\ &= \frac{\Gamma(k+1)}{\lambda^k} \quad (\text{by definition of the gamma function}) \\ &= \frac{k!}{\lambda^k}. \quad \square \end{aligned}$$

Definition: The k th **central moment** of X is

$$\mathbb{E}[(X - \mu)^k] = \begin{cases} \sum_x (x - \mu)^k f(x) & X \text{ is discrete} \\ \int_{\mathbb{R}} (x - \mu)^k f(x) dx & X \text{ is continuous.} \end{cases}$$

Definition: The **variance** of X is the second central moment, i.e., the expected squared deviation of X from its mean,

$$\text{Var}(X) \equiv \mathbb{E}[(X - \mu)^2].$$

Variance is a measure of spread or dispersion.

Notation: $\sigma^2 \equiv \text{Var}(X)$.

Definition: The **standard deviation** of X is $\sigma \equiv +\sqrt{\text{Var}(X)}$.

Example: $X \sim \text{Bern}(p)$, so that $f(1) = p$ and $f(0) = q = 1 - p$. Recall that $\mu = E[X] = p$. Then

$$\begin{aligned}\text{Var}(X) &= E[(X - \mu)^2] \\&= \sum_x (x - p)^2 f(x) \\&= (0 - p)^2 q + (1 - p)^2 p \\&= p^2 q + q^2 p = pq(p + q) \\&= pq. \quad \square\end{aligned}$$

The next results establish the fact that the expected value operator can pass through certain linear functions of X , and then can be used to obtain pleasant expressions for other expected values and variances.

Theorem: For any $h(X)$ and constants a and b — “shift” happens!

$$E[ah(X) + b] = aE[h(X)] + b.$$

Proof (just do continuous case):

$$\begin{aligned} E[ah(X) + b] &= \int_{\mathbb{R}} (ah(x) + b)f(x) dx \\ &= a \int_{\mathbb{R}} h(x)f(x) dx + b \int_{\mathbb{R}} f(x) dx \\ &= aE[h(X)] + b. \quad \square \end{aligned}$$

Corollary: In particular,

$$\mathbb{E}[aX + b] = a\mathbb{E}[X] + b.$$

Similarly, $\mathbb{E}[g(X) + h(X)] = \mathbb{E}[g(X)] + \mathbb{E}[h(X)]$.

Theorem (easier way to calculate variance):

$$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

Proof: By the above results,

$$\begin{aligned}\text{Var}(X) &= \mathbb{E}[(X - \mu)^2] \\ &= \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - \mu^2. \quad \square\end{aligned}$$

Example: Suppose $X \sim \text{Bern}(p)$. Recall that $E[X^k] = p$, for all $k = 1, 2, \dots$. Then

$$\text{Var}(X) = E[X^2] - (E[X])^2 = p - p^2 = pq. \quad \square$$

Example: $X \sim \text{Unif}(a, b)$. $f(x) = \frac{1}{b-a}$, $a < x < b$. Then

$$E[X] = \int_{\mathbb{R}} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

$$E[X^2] = \int_{\mathbb{R}} x^2 f(x) dx = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2 + ab + b^2}{3}$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{(a-b)^2}{12} \quad (\text{algebra}). \quad \square$$

Theorem: Variance doesn't put up with any “shift” b ,

$$\text{Var}(aX + b) = a^2 \text{Var}(X).$$

Proof:

$$\begin{aligned}\text{Var}(aX + b) &= \text{E}[(aX + b)^2] - (\text{E}[aX + b])^2 \\&= \text{E}[a^2X^2 + 2abX + b^2] - (a\text{E}[X] + b)^2 \\&= a^2\text{E}[X^2] + 2ab\text{E}[X] + b^2 \\&\quad - \{a^2(\text{E}[X])^2 + 2ab\text{E}[X] + b^2\} \\&= a^2(\text{E}[X^2] - (\text{E}[X])^2) \\&= a^2\text{Var}(X). \quad \square\end{aligned}$$

Example: $X \sim \text{Bern}(0.3)$. Recall that

$$E[X] = p = 0.3 \quad \text{and} \quad \text{Var}(X) = pq = (0.3)(0.7) = 0.21.$$

Let $Y = h(X) = 4X + 5$. Then

$$E[Y] = E[4X + 5] = 4E[X] + 5 = 6.2$$

and

$$\text{Var}(Y) = \text{Var}(4X + 5) = 16\text{Var}(X) = 3.36. \quad \square$$

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$**
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.7 — Approximations to $E[h(X)]$ and $\text{Var}(h(X))$

Sometimes $Y = h(X)$ is messy, and we may have to approximate $E[h(X)]$ and $\text{Var}(h(X))$ via a Taylor series approach.

Let $\mu = E[X]$ and $\sigma^2 = \text{Var}(X)$, and note that

$$Y = h(\mu) + (X - \mu)h'(\mu) + \frac{(X - \mu)^2}{2}h''(\mu) + R,$$

where R is a remainder term that we'll ignore. Then

$$E[Y] \doteq h(\mu) + E[X - \mu]h'(\mu) + \frac{E[(X - \mu)^2]}{2}h''(\mu) = h(\mu) + \frac{h''(\mu)\sigma^2}{2}$$

and (now an even-cruder approximation)

$$\text{Var}(Y) \doteq \text{Var}(h(\mu) + (X - \mu)h'(\mu)) = [h'(\mu)]^2\sigma^2.$$

Example: Suppose X has pdf $f(x) = 3x^2$, $0 \leq x \leq 1$, and we want to test out our approximations on the “complicated” RV $Y = h(X) = X^{3/4}$.

Well, it's not really that complicated, since we can calculate the *exact moments*:

$$E[Y] = \int_{\mathbb{R}} x^{3/4} f(x) dx = \int_0^1 3x^{11/4} dx = 4/5$$

$$E[Y^2] = \int_{\mathbb{R}} x^{6/4} f(x) dx = \int_0^1 3x^{7/2} dx = 2/3$$

$$\text{Var}(Y) = E[Y^2] - (E[Y])^2 = 2/75 = 0.0267.$$

Before we can do the approximation, note that

$$\mu = E[X] = \int_{\mathbb{R}} x f(x) dx = \int_0^1 3x^3 dx = 3/4$$

$$E[X^2] = \int_{\mathbb{R}} x^2 f(x) dx = \int_0^1 3x^4 dx = 3/5$$

$$\sigma^2 = \text{Var}(X) = E[X^2] - (E[X])^2 = 3/80 = 0.0375.$$

Further,

$$\begin{aligned}h(\mu) &= \mu^{3/4} = (3/4)^{3/4} = 0.8059 \\h'(\mu) &= (3/4)\mu^{-1/4} = (3/4)(3/4)^{-1/4} = 0.8059 \\h''(\mu) &= -(3/16)\mu^{-5/4} = -0.2686.\end{aligned}$$

Thus,

$$E[Y] \doteq h(\mu) + \frac{h''(\mu)\sigma^2}{2} = 0.8059 - \frac{(0.2686)(0.0375)}{2} = 0.8009$$

and

$$\text{Var}(Y) \doteq [h'(\mu)]^2\sigma^2 = (0.8059)^2(0.0375) = 0.0243,$$

both of which are reasonably close to their true values (0.8 and 0.0267, respectively). \square

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions**
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.8 — Moment Generating Functions

Recall that $E[X^k]$ is the k th **moment** of X .

Definition: The **moment generating function** (mgf) of the RV X is

$$M_X(t) \equiv E[e^{tX}].$$

Remark: $M_X(t)$ is a function of t , *not* of X !

Example: $X \sim \text{Bern}(p)$, so that $X = 1$ w.p. p and 0 w.p. q . Then

$$M_X(t) = E[e^{tX}] = \sum_x e^{tx} f(x) = e^{t \cdot 1} p + e^{t \cdot 0} q = pe^t + q. \quad \square$$

Example: If $X \sim \text{Exp}(\lambda)$, then

$$\begin{aligned} M_X(t) &= \mathbb{E}[e^{tX}] \\ &= \int_{\mathbb{R}} e^{tx} f(x) dx \quad (\text{LOTUS}) \\ &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} dx \\ &= \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \quad \square \end{aligned}$$

Big Theorem: Under certain technical conditions (e.g., $M_X(t)$ must exist for all $t \in (-\epsilon, \epsilon)$ for some $\epsilon > 0$), we have

$$E[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0}, \quad k = 1, 2, \dots$$

Thus, you can *generate* the moments of X from the mgf. (Sometimes, it's easier to get moments this way than directly.)

“Proof” (a little non-rigorous):

$$\begin{aligned}
 M_X(t) &= E[e^{tX}] \text{ “=” } E\left[\sum_{k=0}^{\infty} \frac{(tX)^k}{k!}\right] \text{ “=” } \sum_{k=0}^{\infty} E\left[\frac{(tX)^k}{k!}\right] \\
 &= 1 + t E[X] + \frac{t^2 E[X^2]}{2} + \dots
 \end{aligned}$$

This implies

$$\frac{d}{dt} M_X(t) \text{ “=” } E[X] + t E[X^2] + \dots$$

and so

$$\left. \frac{d}{dt} M_X(t) \right|_{t=0} = E[X].$$

Same deal for higher-order moments. \square

Example: $X \sim \text{Bern}(p)$. Then $M_X(t) = pe^t + q$, and

$$E[X] = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = \left. \frac{d}{dt} (pe^t + q) \right|_{t=0} = pe^t \Big|_{t=0} = p.$$

In fact, it's easy to see that $E[X^k] = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0} = p$, for all k . \square

Example: $X \sim \text{Exp}(\lambda)$. Then $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. So

$$E[X] = \left. \frac{d}{dt} M_X(t) \right|_{t=0} = \left. \frac{\lambda}{(\lambda - t)^2} \right|_{t=0} = 1/\lambda$$

$$E[X^2] = \left. \frac{d^2}{dt^2} M_X(t) \right|_{t=0} = \left. \frac{2\lambda}{(\lambda - t)^3} \right|_{t=0} = 2/\lambda^2$$

$$\text{Var}(X) = E[X^2] - (E[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = 1/\lambda^2. \quad \square$$

Other Applications of mgf's

You can do lots of nice things with mgf's. . . .

- Find the mgf of a linear function of X . (Just a second!)
- Identify distributions. (Just a minute!)
- Probability inequality applications. (Next lesson!)
- Find the mgf of the sum of independent random variables. (Later!)
- Convergence of random variables proofs. (Another course!)

Theorem (mgf of a linear function of X): Suppose X has mgf $M_X(t)$ and let $Y = aX + b$. Then

$$M_Y(t) = e^{tb} M_X(at).$$

Proof:

$$M_Y(t) = E[e^{tY}] = E[e^{t(aX+b)}] = e^{tb} E[e^{(at)X}] = e^{tb} M_X(at). \quad \square$$

Example: Let $X \sim \text{Exp}(\lambda)$ and $Y = 3X + 2$. Then

$$M_Y(t) = e^{2t} M_X(3t) = e^{2t} \frac{\lambda}{\lambda - 3t}, \quad \text{if } \lambda > 3t. \quad \square$$

Theorem (identifying distributions): *In this class*, each distribution has a unique mgf.

Proof: Not here!

Example: Suppose that Y has mgf

$$M_Y(t) = e^{2t} \frac{\lambda}{\lambda - 3t}, \quad \text{for } \lambda > 3t.$$

Then by a previous example and the uniqueness of mgf's, it *must* be the case that $Y \sim 3X + 2$, where $X \sim \text{Exp}(\lambda)$. \square

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities**
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.9 — Some Probability Inequalities

Goal: Give results that provide general probability bounds.

Theorem (Markov's Inequality): If X is a nonnegative random variable and $c > 0$, then $P(X \geq c) \leq E[X]/c$. (This is a very crude upper bound.)

Proof: Because X is nonnegative, we have

$$\begin{aligned} E[X] &= \int_{\mathbb{R}} x f(x) dx \\ &= \int_0^{\infty} x f(x) dx \\ &\geq \int_c^{\infty} x f(x) dx \\ &\geq c \int_c^{\infty} f(x) dx \\ &= cP(X \geq c). \quad \square \end{aligned}$$

Theorem (Chebychev's Inequality):¹ Suppose $E[X] = \mu$ and $\text{Var}(X) = \sigma^2$. Then, for any $c > 0$,

$$P(|X - \mu| \geq c) \leq \sigma^2/c^2.$$

Proof: By Markov with $|X - \mu|^2$ in place of X and c^2 in place of c , we have

$$P(|X - \mu| \geq c) = P((X - \mu)^2 \geq c^2) \leq \frac{E[(X - \mu)^2]}{c^2} = \sigma^2/c^2. \quad \square$$

Remarks: Can also write $P(|X - \mu| < c) \geq 1 - \sigma^2/c^2$.

Or, if $c = k\sigma$, then $P(|X - \mu| \geq k\sigma) \leq 1/k^2$.

Chebychev gives a bound on the probability that X deviates from the mean by more than a constant, in terms of the constant and the variance. You can always use Chebychev, but it's crude.

¹There are many, many ways to spell “Chebychev.” See https://en.wikipedia.org/wiki/Talk%3APafnuty_Chebyshev

Example: Suppose $X \sim \text{Unif}(0, 1)$. $f(x) = 1$ for $0 < x < 1$.

Recall that $E[X] = 1/2$, $\text{Var}(X) = 1/12$.

Then Chebychev implies

$$P\left(\left|X - \frac{1}{2}\right| \geq c\right) \leq \frac{1}{12c^2}.$$

In particular, for $c = 1/3$,

$$P\left(\left|X - \frac{1}{2}\right| \geq \frac{1}{3}\right) \leq \frac{3}{4} \quad (\text{upper bound}).$$

Example (cont'd): Let's compare the above bound to the *exact* answer.

$$\begin{aligned}P\left(\left|X - \frac{1}{2}\right| \geq \frac{1}{3}\right) &= 1 - P\left(\left|X - \frac{1}{2}\right| < \frac{1}{3}\right) \\&= 1 - P\left(-\frac{1}{3} < X - \frac{1}{2} < \frac{1}{3}\right) \\&= 1 - P\left(\frac{1}{6} < X < \frac{5}{6}\right) \\&= 1 - \int_{1/6}^{5/6} f(x) dx \\&= 1 - \frac{2}{3} = 1/3.\end{aligned}$$

So Chebychev bound of 3/4 was pretty high by comparison. \square

Bonus Theorem (Chernoff's Inequality): For any c ,

$$P(X \geq c) \leq e^{-ct} M_X(t).$$

Proof: By Markov with e^{tX} in place of X and e^{tc} in place of c , we have

$$P(X \geq c) = P(e^{tX} \geq e^{tc}) = e^{-ct} \mathbb{E}[e^{tX}] = e^{-ct} M_X(t). \quad \square$$

Example: Suppose X has the **standard normal distribution** with pdf $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, for all $x \in \mathbb{R}$, i.e., the famous “bell curve” distribution.

It is easy to show (via a little calculus elbow grease involving completing a square) that the mgf of the standard normal is

$$M_X(t) = \mathbb{E}[e^{tX}] = \int_{\mathbb{R}} e^{tx} \phi(x) dx = e^{t^2/2}.$$

Then using Chernoff with $t = c$ immediately yields the tail probability

$$P(X \geq c) \leq e^{-c^2} M_X(c) = e^{-c^2/2}. \quad \square$$

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable**
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results

Lesson 2.10 — Functions of a Random Variable

The Next Few Lessons:

- Problem Statement with Examples
 - Discrete Case
 - Continuous Case
- Inverse Transform Theorem
- Some Honors Topics

Problem: You have a RV X and you know its pmf/pdf $f(x)$.

Define $Y \equiv h(X)$ (some function of X).

Find $g(y)$, the pmf/pdf of Y .

Remark: Recall that LOTUS gave us results for $E[h(X)]$. But this is much more general than LOTUS, because now we're going to get the *entire distribution* of $h(X)$.

We'll start with the case in which X is a discrete RV, and then we'll go to the continuous X case.

Discrete Case: X discrete implies Y discrete implies

$$g(y) = P(Y = y) = P(h(X) = y) = P(\{x|h(x) = y\}) = \sum_{x|h(x)=y} f(x).$$

Example: X is the number of H 's in 2 coin tosses. We want the pmf for $Y = h(X) = X^3 - X$.

x	0	1	2
$f(x) = P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$y = x^3 - x$	0	0	6

$$g(0) = P(Y = 0) = P(X = 0 \text{ or } 1) = 3/4 \text{ and}$$

$$g(6) = P(Y = 6) = P(X = 2) = 1/4.$$

$$g(y) = \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 6. \end{cases} \quad \square$$

Example: X is discrete with

$$f(x) = \begin{cases} 1/8 & \text{if } x = -1 \\ 3/8 & \text{if } x = 0 \\ 1/3 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2. \end{cases}$$

Let $Y = X^2$ (so Y can only equal 0, 1 or 4).

$$g(y) = \begin{cases} P(Y = 0) = f(0) = 3/8 \\ P(Y = 1) = f(-1) + f(1) = 11/24 \\ P(Y = 4) = f(2) = 1/6. \quad \square \end{cases}$$

Continuous Case: X continuous implies Y can be continuous *or* discrete.

Example: $Y = X^2$ (clearly continuous).

Example: $Y = \begin{cases} 0 & \text{if } X < 0 \\ 1 & \text{if } X \geq 0 \end{cases}$ is *not* continuous.

Method: Compute $G(y)$, the cdf of Y .

$$G(y) = P(Y \leq y) = P(h(X) \leq y) = \int_{\{x|h(x) \leq y\}} f(x) dx.$$

If $G(y)$ is continuous, construct the pdf $g(y)$ by differentiating.

Example: $f(x) = |x|$, $-1 \leq x \leq 1$.

Find the pdf of the RV $Y = h(X) = X^2$.

$$G(y) = P(Y \leq y) = P(X^2 \leq y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1 & \text{if } y \geq 1 \\ (\star) & \text{if } 0 < y < 1, \end{cases}$$

where

$$(\star) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = \int_{-\sqrt{y}}^{\sqrt{y}} |x| dx = y.$$

Thus,

$$G(y) = \begin{cases} 0 & \text{if } y \leq 0 \\ 1 & \text{if } y \geq 1 \\ y & \text{if } 0 < y < 1. \end{cases}$$

This implies

$$g(y) = G'(y) = \begin{cases} 0 & \text{if } y \leq 0 \text{ or } y \geq 1 \\ 1 & \text{if } 0 < y < 1. \end{cases}$$

This means that Y has the Unif(0,1) distribution! \square

Example: Suppose $U \sim \text{Unif}(0, 1)$. Find the pdf of $Y = -\ln(1 - U)$.

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(-\ln(1 - U) \leq y) \\ &= P(1 - U \geq e^{-y}) \\ &= P(U \leq 1 - e^{-y}) \\ &= \int_0^{1-e^{-y}} f(u) du \\ &= 1 - e^{-y} \quad (\text{since } f(u) = 1). \end{aligned}$$

Taking the derivative, we have $g(y) = e^{-y}$, $y > 0$.

Wow! This implies $Y \sim \text{Exp}(\lambda = 1)$. \square

We can generalize this result. . . .

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem**
- 12 Honors Bonus Results

Lesson 2.11 — Inverse Transform Theorem

Here's a terrific result that has lots of applications.

Inverse Transform Theorem:² Suppose X is a continuous random variable having cdf $F(x)$. Then the *random variable* $F(X) \sim \text{Unif}(0, 1)$.

Proof: Let $Y = F(X)$. Then the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \leq y) \\ &= P(F(X) \leq y) \\ &= P(X \leq F^{-1}(y)) \quad (\text{the cdf is mono. increasing}) \\ &= F(F^{-1}(y)) \quad (F(x) \text{ is the cdf of } X) \\ &= y. \quad \text{Uniform!} \quad \square \end{aligned}$$

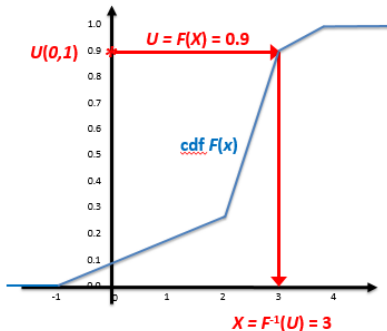
²Also known as the Probability Integral Transform.

Remark: This is a great theorem, since it applies to all continuous RVs X .

Corollary: $X = F^{-1}(U)$, so you can plug a $\text{Unif}(0,1)$ RV into the inverse cdf to generate a realization of a RV having X 's distribution.

Method: Set $F(X) = U$ and solve for $X = F^{-1}(U)$ to generate X .

**Inverse
Transform
Method**
(generate X
from U)

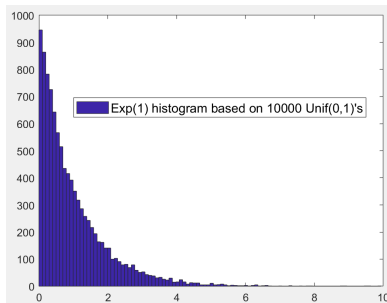


Example: Suppose X is $\text{Exp}(\lambda)$, so that it has cdf $F(x) = 1 - e^{-\lambda x}$. Similar to a previous example, set $F(X) = 1 - e^{-\lambda X} = U$ and generate an $\text{Exp}(\lambda)$ RV by solving for

$$X = F^{-1}(U) = -\frac{1}{\lambda} \ln(1 - U) \sim \text{Exp}(\lambda). \quad \square$$

Remark: If you'd like to generate a nice, beautiful $\text{Exp}(\lambda)$ pdf on a computer, then all you have to do is...

- Generate 10000 $\text{Unif}(0,1)$'s (e.g., use the `rand` function in Excel, or `unifrnd` in Matlab),
- Plug those 10000 into the equation for X above, and
- Plot the histograms of the X 's.



Remark: This trick has tremendous applications in simulation, where we need to generate random variables all the time (e.g., customer arrival times, service times, machine breakdown times, etc.).

Outline

- 1 Introduction
- 2 Discrete Random Variables
- 3 Continuous Random Variables
- 4 Cumulative Distribution Functions
- 5 Great Expectations
- 6 LOTUS, Moments, and Variance
- 7 Approximations to $E[h(X)]$ and $\text{Var}(h(X))$
- 8 Moment Generating Functions
- 9 Some Probability Inequalities
- 10 Functions of a Random Variable
- 11 Inverse Transform Theorem
- 12 Honors Bonus Results**

Lesson 2.12 — Honors Bonus Results

Another Way to Find the pdf of a Function of a Continuous RV

Suppose that $Y = h(X)$ is a monotonic function of a continuous RV X having pdf $f(x)$ and cdf $F(x)$. Let's get the pdf $g(y)$ of Y directly.

$$\begin{aligned}g(y) &= \frac{d}{dy}G(y) = \frac{d}{dy}P(Y \leq y) \\&= \frac{d}{dy}P(h(X) \leq y) \\&= \frac{d}{dy}P(X \leq h^{-1}(y)) \quad (h(x) \text{ is monotonic}) \\&= \frac{d}{dy}F(h^{-1}(y)) \\&= f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| \quad (\text{chain rule}). \quad \square\end{aligned}$$

Example: Suppose that $f(x) = 3x^2$, $0 < x < 1$. Let $Y = h(X) = X^{1/2}$, which is monotone increasing. Then the pdf of Y is

$$\begin{aligned} g(y) &= f(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| \\ &= f(y^2) \left| \frac{d(y^2)}{dy} \right| \\ &= 3y^4(2y) \\ &= 6y^5, \quad 0 < y < 1. \quad \square \end{aligned}$$

Theorem (why LOTUS works): Let's assume that $h(\cdot)$ is monotonically increasing. Then

$$\begin{aligned}
 E[h(X)] &= E[Y] \\
 &= \int_{\mathbb{R}} yg(y) dy \\
 &= \int_{\mathbb{R}} yf(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| dy \\
 &= \int_{\mathbb{R}} h(x)f(x) \left| \frac{dx}{dy} \right| dy \\
 &= \int_{\mathbb{R}} h(x)f(x) dx. \quad \square
 \end{aligned}$$

Next up: Bivariate generalizations of the material in this module!