09: Continuous RVs

Jerry Cain
April 16, 2021
<table>
<thead>
<tr>
<th>Page</th>
<th>Section</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Continuous RVs</td>
<td>09a_continuous_rvs</td>
</tr>
<tr>
<td>15</td>
<td>Uniform RV</td>
<td>09b_uniform</td>
</tr>
<tr>
<td>22</td>
<td>Exponential RV</td>
<td>09c_exponential</td>
</tr>
<tr>
<td>31</td>
<td>Exercises, CDF</td>
<td>LIVE</td>
</tr>
<tr>
<td>55</td>
<td>Extra material</td>
<td>09e_extra</td>
</tr>
</tbody>
</table>
Continuous RVs
Not all values are discrete

```python
import numpy as np
np.random.random()
```
People heights

You are volunteering at the local elementary school.
• To choose a t-shirt for your new buddy Jordan, you need to know their height.

1. What is the probability that your buddy is 54.0923857234 inches tall?
   Essentially 0

2. What is the probability that your buddy is between 52–56 inches tall?
Integrals

Integral = area under a curve

Loving, not scary
Continuous RV definition

A random variable $X$ is **continuous** if there is a **probability density function** $f(x) \geq 0$ such that for $-\infty < x < \infty$:

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

Integrating a PDF must always yield valid probabilities, and therefore the PDF must also satisfy

$$\int_{-\infty}^{\infty} f(x) \, dx = P(-\infty < X < \infty) = 1$$

Also written as: $f_X(x)$
Today’s main takeaway, #1

Integrate $f(x)$ to get probabilities.

PDF Units: probability per units of $X$

$$P(52 \leq X \leq 56) = \int_{52}^{56} f(x) \, dx$$
## PMF vs PDF

<table>
<thead>
<tr>
<th>Discrete random variable $X$</th>
<th>Continuous random variable $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability mass function (PMF): $p(x)$</td>
<td>Probability density function (PDF): $f(x)$</td>
</tr>
<tr>
<td>To get probability: $P(X = x) = p(x)$</td>
<td>To get probability: $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$</td>
</tr>
</tbody>
</table>

Both are measures of how **likely** $X$ is to take on a value.
Computing probability

Let $X$ be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X \geq 1)$?
Computing probability

Let $X$ be a continuous RV with PDF:

$$f(x) = \begin{cases} \frac{1}{2}x & \text{if } 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

What is $P(X \geq 1)$?

**Strategy 1**: Integrate

$$P(1 \leq X < \infty) = \int_{1}^{\infty} f(x)dx = \int_{1}^{2} \frac{1}{2}xdx$$

$$= \left. \frac{1}{2} \left( \frac{1}{2}x^2 \right) \right|_{1}^{2} = \frac{1}{2} \left[ 2 - \frac{1}{2} \right] = \frac{3}{4}$$

**Strategy 2**: Know triangles

$$1 - \frac{1}{2} \left( \frac{1}{2} \right) = \frac{3}{4}$$

Wait...is this even legal?

$$P(0 \leq X < 1) = \int_{0}^{1} f(x)dx \text{ ? ?}$$
Today’s main takeaway, #2

For a continuous random variable $X$ with PDF $f(x)$,

\[
P(X = c) = \int_c^c f(x) \, dx = 0.
\]

Contrast with PMF in discrete case: $P(X = c) = p(c)$
PDF Properties

For a continuous RV $X$ with PDF $f$,

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

True/False:

1. $P(X = c) = 0$

2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$

3. $f(x)$ is a probability

4. In the graphed PDF above, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$
PDF Properties

For a continuous RV $X$ with PDF $f$, 

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$

True/False:

1. $P(X = c) = 0$

2. $P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$

3. $f(x)$ is a probability

4. In the graphed PDF above, $P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$

Compare area under the curve $f$
Uniform RV
Uniform Random Variable

def An **Uniform** random variable $X$ is defined as follows:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

$X \sim \text{Uni}(\alpha, \beta)$

**Support:** $[\alpha, \beta]$ (sometimes defined over $(\alpha, \beta)$)

**PDF**

**Expectation**

$$E[X] = \frac{\alpha + \beta}{2}$$

**Variance**

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$
Quick check

If $X \sim \text{Uni}(\alpha, \beta)$, the PDF of $X$ is:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

What is $\frac{1}{\beta - \alpha}$ if the following graphs are PDFs of Uniform RVs $X$?

1. $\text{pdf}(x)$

2. $\text{pdf}(x)$

3. $\text{pdf}(x)$

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2021
Quick check

If $X \sim \text{Uni}(\alpha, \beta)$, the PDF of $X$ is:

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

What is $\frac{1}{\beta - \alpha}$ if the following graphs are PDFs of Uniform RVs $X$?

1. $f(x)$
   - $f(x) = \frac{1}{25}$
2. $f(x)$
   - $f(x) = \frac{2}{3/2}$
3. $f(x)$
   - $f(x) = \frac{1}{10}$
## Expectation and Variance

<table>
<thead>
<tr>
<th>Discrete RV $X$</th>
<th>Continuous RV $X$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[X] = \sum_x x p(x)$</td>
<td>$E[X] = \int_{-\infty}^{\infty} x f(x) , dx$</td>
</tr>
<tr>
<td>$E[g(X)] = \sum_x g(x) p(x)$</td>
<td>$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) , dx$</td>
</tr>
</tbody>
</table>

### Both continuous and discrete RVs

- $E[aX + b] = aE[X] + b$
- $\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$
- $\text{Var}(aX + b) = a^2 \text{Var}(X)$

---

**TL;DR:** $\sum_{x=a}^{b} \Rightarrow \int_{a}^{b}$

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Spring 2021
Uniform RV expectation

\[
E[X] = \int_{-\infty}^{\infty} x \cdot f(x) \, dx
\]

\[
= \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} \, dx
\]

\[
= \frac{1}{\beta - \alpha} \cdot \frac{1}{2} \left( \beta^2 - \alpha^2 \right)
\]

\[
= \frac{1}{2} \cdot \frac{(\beta + \alpha)(\beta - \alpha)}{\beta - \alpha} = \frac{\alpha + \beta}{2}
\]

Interpretation:
Average the start & end
**Uniform Random Variable**

**def** An **Uniform** random variable $X$ is defined as follows:

$X \sim \text{Uni}(\alpha, \beta)$

PDF

\[
f(x) = \begin{cases} 
\frac{1}{\beta - \alpha} & \text{if } \alpha \leq x \leq \beta \\
0 & \text{otherwise}
\end{cases}
\]

Support: $[\alpha, \beta]$ (sometimes defined over $(\alpha, \beta)$)

Expectation

\[
E[X] = \frac{\alpha + \beta}{2}
\]

Variance

\[
\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}
\]
Exponential RV
Grid of random variables

- **Number of successes**
  - One trial: $\text{Ber}(p)$
  - Several trials: $\text{Bin}(n, p)$
  - Interval of time: $\text{Poi}(\lambda)$

- **Time until success**
  - One success: $\text{Geo}(p)$
  - Several successes: $\text{NegBin}(r, p)$
  - Interval of time to first success: $\text{Exp}(\lambda)$
Consider an experiment that lasts a duration of time until success occurs. An exponential random variable $X$ is the amount of time until success.

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

Support: $[0, \infty)$

Expectation

$$E[X] = \frac{1}{\lambda}$$

Variance

$$\text{Var}(X) = \frac{1}{\lambda^2}$$

Examples:
- Time until next earthquake
- Time for request to reach web server
- Time until end of cell phone contract
Interpreting Exp(\( \lambda \))

**def**  An **Exponential** random variable \( X \) is the amount of time until success.

\[
X \sim \text{Exp}(\lambda) \quad \text{Expectation} \quad E[X] = \frac{1}{\lambda}
\]

Based on the expectation \( E[X] \), what are the units of \( \lambda \)?
Interpreting $\text{Exp}(\lambda)$

**def** An **Exponential** random variable $X$ is the amount of time until success.

$$X \sim \text{Exp}(\lambda)$$

**Expectation**

$$E[X] = \frac{1}{\lambda}$$

Based on the expectation $E[X]$, what are the units of $\lambda$?

- e.g., average # of successes per second

**For both Poisson and Exponential RVs, $\lambda = \# \text{ successes }/\text{time}$.**
Earthquakes

1906 Earthquake
Magnitude 7.8
Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?

We know on average:

\[ \frac{500 \text{ years}}{1 \text{ earthquake}} = 0.002 \frac{\text{earthquakes}}{\text{year}} \]

\[ \frac{1 \text{ earthquake}}{500 \text{ years}} \]

*In California, according to historical data from USGS, 2015.
Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?

Define events/RVs & state goal

- $X$: when next earthquake happens
- $X \sim \text{Exp}(\lambda = 0.002)$
- $\lambda$: year$^{-1} = 1/500$
- Want: $P(X < 30)$

Solve

Recall

$$\int e^{cx} \, dx = \frac{1}{c} e^{cx}$$

*In California, according to historical data from USGS, 2015
Earthquakes

Major earthquakes (magnitude 8.0+) occur once every 500 years.*

1. What is the probability of a major earthquake in the next 30 years?
2. What is the standard deviation of years until the next earthquake?

Define events/ RVs & state goal

\( X \): when next earthquake happens

\( X \sim \text{Exp}(\lambda = 0.002) \)

\( \lambda \): year\(^{-1} \)

Want: \( P(X < 30) \)

Solve

\[ X \sim \text{Exp}(\lambda) \]
\[ E[X] = \frac{1}{\lambda} \]
\[ f(x) = \lambda e^{-\lambda x} \quad \text{if } x \geq 0 \]

*In California, according to historical data form USGS, 2015
09: Continuous RVs

Lisa Yan and Jerry Cain
October 2, 2020
Today’s main takeaway, #1

Integrate \( f(x) \) to get probabilities.

\[
P(a \leq X \leq b) = \int_a^b f(x) \, dx
\]
For a continuous random variable $X$ with PDF $f(x)$,

$$P(X = c) = \int_c^c f(x)dx = 0.$$ 

Implication: $P(a \leq X \leq b) = P(a < X < b)$
Slide 35 has a matching question to go over in Zoom polling. We’ll go over it together afterwards.

Post any clarifications here or in chat!

https://edstem.org/us/courses/5090/discussion/357163

Think by yourself: 1 minute
Determining valid PDFs

Which of the following functions are valid PDFs?

1. \( f(x) \)
   \[ \int_{-\infty}^{\infty} f(x) \, dx = 0.5 \]

2. \( g(x) \)
   \[ \int_{-\infty}^{\infty} g(x) \, dx = 1 \]

3. \( h(x) \)
   \[ \int_{-\infty}^{\infty} h(x) \, dx = 1 \]

4. \( w(x) \)
   \[ \int_{-\infty}^{\infty} w(x) \, dx = 1 \]
Determining valid PDFs

Which of the following functions are valid PDFs?

1. \( f(x) \)
   \[
   \int_{-\infty}^{\infty} f(x) \, dx = 0.5
   \]

2. \( g(x) \)
   \[
   \int_{-\infty}^{\infty} g(x) \, dx = 1
   \]

3. \( h(x) \)
   \[
   \int_{-\infty}^{\infty} h(x) \, dx = 1
   \]

4. \( w(x) \)
   \[
   \int_{-\infty}^{\infty} w(x) \, dx = 1
   \]
Check out the question on the next slide (Slide 38). Post any clarifications here!

https://edstem.org/us/courses/5090/discussion/357163

Breakout rooms: 4 min. Introduce yourself!
Riding the Marguerite Bus

You want to get on the Marguerite bus.

- The bus stops at the Gates building at 15-minute intervals (2:00, 2:15, etc.).
- You arrive at the stop uniformly between 2:00-2:30pm.

\[ P(\text{you wait < 5 minutes for bus})? \]
Riding the Marguerite Bus

You want to get on the Marguerite bus.

• The bus stops at the Gates building at 15-minute intervals (2:00, 2:15, etc.).
• You arrive at the stop uniformly between 2:00-2:30pm.

\[ P(\text{you wait < 5 minutes for bus})? \]

1. Define events/RVs & state goal
   \[ X: \text{time passenger arrives after 2:00} \]
   \[ X \sim \text{Uni}(0,30) \]

2. Solve

\[ f(x) \]

\begin{tabular}{c|c|c}
0 & 15 & 30 \\
\hline
\end{tabular}

wait < 5 min
Cumulative Distribution Function (CDF)

For a random variable $X$, the cumulative distribution function (CDF) is defined as

$$F(a) = F_X(a) = P(X \leq a), \text{ where } -\infty < a < \infty$$

For a discrete RV $X$, the CDF is:

$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$
Cumulative Distribution Function (CDF)

For a random variable $X$, the cumulative distribution function (CDF) is defined as

$$F(a) = F_X(a) = P(X \leq a), \text{ where } -\infty < a < \infty$$

For a discrete RV $X$, the CDF is:

$$F(a) = P(X \leq a) = \sum_{\text{all } x \leq a} p(x)$$

For a continuous RV $X$, the CDF is:

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

CDF is a probability, though PDF is not.

If you learn to use CDFs, you can avoid integrating the PDF.
Think Slide 43 has a matching question to go over by yourself. We’ll go over it together afterwards.

Post any clarifications here or in chat!

https://edstem.org/us/courses/5090/discussion/357163

Think by yourself: 1 min
Using the CDF for continuous RVs

For a **continuous** random variable $X$ with PDF $f(x)$, the CDF of $X$ is

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Matching (choices are used 0/1/2 times)

1. $P(X < a)$  
   A. $F(a)$
2. $P(X > a)$  
   B. $1 - F(a)$
3. $P(X \geq a)$  
   C. $F(a) - F(b)$
4. $P(a \leq X \leq b)$  
   D. $F(b) - F(a)$
Using the CDF for continuous RVs

For a **continuous** random variable $X$ with PDF $f(x)$, the CDF of $X$ is

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

Matching (choices are used 0/1/2 times)

1. $P(X < a)$ ——— A. $F(a)$
2. $P(X > a)$ ——— B. $1 - F(a)$
3. $P(X \geq a)$ ——— C. $F(a) - F(b)$
4. $P(a \leq X \leq b)$ ——— D. $F(b) - F(a)$ (next slide)
Using the CDF for continuous RVs

For a continuous random variable $X$ with PDF $f(x)$, the CDF of $X$ is

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

4. $P(a \leq X \leq b) = F(b) - F(a)$

Proof:

$$F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx$$

$$= \left(\int_{-\infty}^{a} f(x)dx + \int_{a}^{b} f(x)dx\right) - \int_{-\infty}^{a} f(x)dx$$

$$= \int_{a}^{b} f(x)dx$$
Addendum to today’s main takeaway, #1

Integrate $f(x)$ to get probabilities.*

*If you have $F(a)$, you already have probabilities, since $F(a) = \int_{-\infty}^{a} f(x) \, dx$

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$
CDF of an Exponential RV

\( X \sim \text{Exp}(\lambda) \quad F(x) = 1 - e^{-\lambda x} \quad \text{if } x \geq 0 \)

Proof:

\[
F(x) = P(X \leq x) = \int_{y=-\infty}^{x} f(y)dy = \int_{y=0}^{x} \lambda e^{-\lambda y}dy
\]

\[
= \left. \frac{1}{-\lambda} e^{-\lambda y} \right|_{0}^{x}
\]

\[
= -1(e^{-\lambda x} - e^{-\lambda 0})
\]

\[
= 1 - e^{-\lambda x}
\]
PDF/CDF \( X \sim \text{Exp}(\lambda = 1) \)

\( X \sim \text{Exp}(\lambda) \)
\( f(x) = \lambda e^{-\lambda x} \)
\( F(x) = 1 - e^{-\lambda x} \)

\( P(X \leq 2) \)
\( P(X > 2) \)

\( \int_0^2 \lambda e^{-\lambda x} \, dx \approx 0.86 \)

\( \int_2^\infty \lambda e^{-\lambda x} \, dx \approx 0.14 \)

\( 1 - e^{-2\lambda} \approx 0.86 \)

\( 1 - F(2) = e^{-2\lambda} \approx 0.14 \)
Check out the question on the next slide (Slide 50). Post any clarifications here!

https://edstem.org/us/courses/5090/discussion/357163

Breakout rooms: 3 min.
Earthquakes

Major earthquakes (magnitude 8.0+) occur independently on average once every 500 years.*

What is the probability of zero major earthquakes next year?

*In California, according to historical data from USGS, 2015
Earthquakes

Major earthquakes (magnitude 8.0+) occur independently on average once every 500 years.*

What is the probability of zero major earthquakes next year?

Strategy 1: Exponential RV

Define events/RVs & state goal

\( T \): when first earthquake happens

\( T \sim \text{Exp}(\lambda = 0.002) \)

Want: \( P(T > 1) = 1 - F(1) \)

Solve

\[
P(T > 1) = 1 - (1 - e^{-\lambda \cdot 1}) = e^{-\lambda}
\]

*In California, according to historical data from USGS, 2015
Earthquakes

Major earthquakes (magnitude 8.0+) occur independently on average once every 500 years.*

What is the probability of zero major earthquakes next year?

**Strategy 1:** Exponential RV

Define events/RVs & state goal

- \( T \): when first earthquake happens
- \( T \sim \text{Exp}(\lambda = 0.002) \)

Want: \( P(T > 1) = 1 - F(1) \)

Solve

\[
P(T > 1) = 1 - (1 - e^{-\lambda \cdot 1}) = e^{-\lambda}
\]

*In California, according to historical data from USGS, 2015*

**Strategy 2:** Poisson RV

Define events/RVs & state goal

- \( N \): number of earthquakes next year
- \( N \sim \text{Poi}(\lambda = 0.002) \)

Want: \( P(N = 0) \)

Solve

\[
P(N = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda} \approx 0.998
\]

Read more in Ross! (section 9.1)
Happy Friday
Extra
Expectation of the Exponential

\[ X \sim \text{Exp}(\lambda) \]

Expectation

\[ E[X] = \frac{1}{\lambda} \]

Proof:

\[
E[X] = \int_{-\infty}^{\infty} x f(x) \, dx = \int_{0}^{\infty} x \lambda e^{-\lambda x} \, dx
\]

\[
= -xe^{-\lambda x}\bigg|_{0}^{\infty} + \int_{0}^{\infty} e^{-\lambda x} \, dx
\]

\[
= -xe^{-\lambda x}\bigg|_{0}^{\infty} - \frac{1}{\lambda} e^{-\lambda x}\bigg|_{0}^{\infty} + \int_{0}^{\infty} dx
\]

\[
= [0 - 0] + [0 - \left(\frac{-1}{\lambda}\right)]
\]

\[
= \frac{1}{\lambda}
\]

Integration by parts

\[
\int x\lambda e^{-\lambda x} \, dx = \int u \cdot dv
\]

\[
u = x \quad dv = \lambda e^{-\lambda x} \, dx
\]

\[
du = dx \quad v = -e^{-\lambda x}
\]

\[
\int u \cdot dv = u \cdot v - \int v \cdot du
\]

\[
-xe^{-\lambda x} - \int -e^{-\lambda x} \, dx
\]
Website visits

Suppose a visitor to your website leaves after $X$ minutes.

- On average, visitors leave the site after 5 minutes.
- The length of stay, $X$, is exponentially distributed.

1. $P(X > 10)$?

2. $P(10 < X < 20)$?
Website visits

Suppose a visitor to your website leaves after $X$ minutes.

- On average, visitors leave the site after 5 minutes.
- The length of stay, $X$, is exponentially distributed.

1. $P(X > 10)$?

   Define
   
   $X$: when visitor leaves
   $X \sim \text{Exp}(\lambda = 1/5 = 0.2)$

   Solve
   
   $P(X > 10) = 1 - F(10)$
   
   $= 1 - (1 - e^{-10/5}) = e^{-2} \approx 0.1353$

2. $P(10 < X < 20)$?

   Solve
   
   $P(10 < X < 20) = F(20) - F(10)$
   
   $= (1 - e^{-4}) - (1 - e^{-2}) \approx 0.1170$
Replacing your laptop

Let $X = \# \text{ hours of use until your laptop dies}$. 
- $X$ is distributed as an Exponential RV, where 
- On average, laptops die after 5000 hours of use. 
- You use your laptop 5 hours a day.

What is $P(\text{your laptop lasts 4 years})$?
Replacing your laptop

Let $X = \#$ hours of use until your laptop dies.
- $X$ is distributed as an Exponential RV, where
- On average, laptops die after 5000 hours of use.
- You use your laptop 5 hours a day.

What is $P(\text{your laptop lasts 4 years})$?

Define

$X$: # hours until laptop death  
$X \sim \text{Exp}(\lambda = 1/5000)$

Want: $P(X > 5 \cdot 365 \cdot 4)$

Solve

$P(X > 7300) = 1 - F(7300)$

$= 1 - (1 - e^{-7300/5000}) = e^{-1.46} \approx 0.2322$

Better plan ahead if you’re co-terming!

- 5-year plan:  
  $P(X > 9125) = e^{-1.825} \approx 0.1612$

- 6-year plan:  
  $P(X > 10950) = e^{-2.19} \approx 0.1119$