09: Continuous RVs

Lisa Yan
April 24, 2020
# Quick slide reference

<table>
<thead>
<tr>
<th>Page</th>
<th>Topic</th>
<th>File Name</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?  

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

**Discrete** random variable $X$

Probability mass function (PMF):

$$p(x)$$

To get probability:

$$P(X = x) = p(x)$$

**Continuous** random variable $X$

Probability density function (PDF):

$$f(x)$$

To get probability:

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

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**

$$P(1 \leq X < \infty) = \int_1^\infty f(x)dx = \int_1^2 \frac{1}{2}x\,dx$$

$$= \frac{1}{2} \left( \frac{1}{2}x^2 \right) \bigg|_1^2 = \frac{1}{2} \left[ 2 - \frac{1}{2} \right] = \frac{3}{4}$$

**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,
   $$P(x_1 \leq X \leq x_2) > P(x_2 \leq X \leq x_3)$$
Uniform RV
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

$$\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$$
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. $f(x)$
   ![Graph 1](image)

2. $f(x)$
   ![Graph 2](image)

3. $f(x)$
   ![Graph 3](image)
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. $f(x)$
2. $f(x)$
3. $f(x)$

Lisa Yan, CS109, 2020
Expectation and Variance

Discrete RV $X$

\[
E[X] = \sum_x x \cdot p(x)
\]
\[
E[g(X)] = \sum_x g(x) \cdot p(x)
\]

Continuous RV $X$

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

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: $\sum_{x=a}^{b} \Rightarrow \int_{a}^{b}$
Uniform RV expectation

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

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

\[ = \frac{1}{\beta - \alpha} \cdot \frac{1}{2} x^2 \bigg|_{\alpha}^{\beta} \]

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

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

Interpretation: Average the start & end
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**

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

Number of successes

One trial

Ber($p$)

$Bin(n, p)$

$Poi(\lambda)$

Several trials

$Geo(p)$

$NegBin(r, p)$

Interval of time

Time until success

$n = 1$

$r = 1$

$Exp(\lambda)$

Interval of time to first success

One success

Several successes
Consider an experiment that lasts a duration of time until success occurs. An Exponential random variable $X$ is the amount of time until success.

**PDF**

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

**Support:** $[0, \infty)$

**Expectation**

$$E[X] = \frac{1}{\lambda}$$ (in extra slides)

**Variance**

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

**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.

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

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

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.

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

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:

\[
\begin{align*}
500 & \quad \text{years} \\
\text{earthquake} & \\
0.002 & \quad \text{earthquakes} \\
\text{year} & \\
1 & \quad \text{earthquakes} \\
\text{500 years} &
\end{align*}
\]

*In California, according to historical data form 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
\( X \sim \text{Exp}(\lambda = 0.002) \)

\( \lambda: \text{year}^{-1} = 1/500 \)

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

Recall

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

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

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

\[ f(x) = \lambda e^{-\lambda x} \quad \text{if} \quad x \geq 0 \]

*In California, according to historical data form 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

\[ X: \text{ when next earthquake happens} \]

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

\[ \lambda: \text{year}^{-1} \]

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 from USGS, 2015
09: Continuous RVs

Slides by Lisa Yan
July 10, 2020
Today’s main takeaway, #1

Integrate $f(x)$ to get probabilities.

$$P(a \leq X \leq b) = \int_{a}^{b} f(x) \, dx$$
Today’s main takeaway, #2

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

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

Implication: $P(a \leq X \leq b) = P(a < X < b)$
Think

The next slide has a matching question to go over by yourself. We’ll go over it together afterwards.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/88677

Think by yourself: 2 min
Determining valid PDFs

Which of the following functions are valid PDFs?

1. $f(x)$
   - $\int_{-\infty}^{\infty} f(x) dx = 0.5$
   - $P(a \leq X \leq b) = \int_{a}^{b} f(x) dx$

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

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

4. $w(x)$
   - $\int_{-\infty}^{\infty} w(x) dx = 1$

(by yourself)
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. Post any clarifications here!

https://us.edstem.org/courses/667/discussion/88677

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.

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

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

2. Solve

Want: wait < 5 min
Interlude for fun/announcements
Announcements

Midterm Quiz

Time frame: Mon.-Tues. July 20-21 5pm-5pm PT
Covers: Up to and including Lecture 11


Note: If you have an emergency situation during the quiz, please contact Oishi and Cooper. We will try our best to accommodate.
Interesting probability news

NYC subway math

Distribution of time until the next subway arrival
Probably Beta RV (Week 8)
Ethics in Probability: Disclosure Avoidance

The 2020 US Census is using “differential privacy” for disclosure avoidance. Differential privacy uses the Laplace Distribution (a two-tailed Exponential distribution) to guarantee probabilistic privacy protection.

https://www.census.gov/about/policies/privacy/statistical_safeguards/disclosure-avoidance-2020-census.html

https://hdsr.mitpress.mit.edu/pub/dgg03vo6/release/1
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

The next slide has a matching question to go over by yourself. We’ll go over it together afterwards.

Post any clarifications here!

https://us.edstem.org/courses/667/discussion/88677

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

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

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

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

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

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

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

Proof:

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

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

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

Stanford University
CDF of an Exponential RV

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

Proof:

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

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

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

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

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

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

\[ P(X \leq 2) \]

\[ 1 - e^{-2\lambda} \approx 0.86 \]

\[ P(X > 2) \]

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

https://us.edstem.org/courses/667/discussion/88677

Breakout rooms: 4 min. Introduce yourself!
Earthquakes

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

What is the probability of zero major earthquakes next year?

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

Major earthquakes (magnitude 8.0+) occur once every 500 years.*
What is the probability of zero major earthquakes next year?

Strategy 1: Exponential RV

Define events/RVs & state goal

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

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

Solve

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

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

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

What is the probability of zero major earthquakes next year?

**Strategy 1: Exponential RV**

Define events/RVs & state goal

\[ X: \text{when first earthquake happens} \]
\[ X \sim \text{Exp}(\lambda = 0.002) \]

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

Solve

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

**Strategy 2: Poisson RV**

Define events/RVs & state goal

\[ X: \text{# earthquakes next year} \]
\[ X \sim \text{Poi}(\lambda = 0.002) \]

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

Solve

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

*In California, according to historical data from USGS, 2015*
Next up: Normal RVs
Extra slides

Expectation of the Exponential

Extra problems
Extra
**Expectation of the Exponential**

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

Expectation

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

Proof:

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

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

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

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

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

Integration by parts

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

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

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

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

\[
= -xe^{-\lambda x} - \int -e^{-\lambda x} dx
\]

\[
= -xe^{-\lambda x} + \frac{1}{\lambda} e^{-\lambda x} + C
\]
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. $P(10 < X < 20)$?
   
   Define
   
   $X$: when visitor leaves
   
   $X \sim \text{Exp}(\lambda = 1/5 = 0.2)$
   
   Solve
   
   $P(10 < X < 20) = F(20) - F(10)$
   
   $= (1 - e^{-4}) - (1 - e^{-2}) \approx 0.1170$
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 $P$(your laptop lasts 4 years)?

<table>
<thead>
<tr>
<th>Define</th>
<th>Solve</th>
</tr>
</thead>
</table>
| $X$: # hours until laptop death  
$X \sim\text{Exp}(\lambda = 1/5000)$ | $P(X > 7300) = 1 - F(7300)$

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

Better plan ahead if you’re co-terming!

• 5-year plan: $P(X > 9125) = e^{-1.825} \approx 0.1612$
• 6-year plan: $P(X > 10950) = e^{-2.19} \approx 0.1119$