

Chapter 31-35: Some Important Continuous RVs

Meike Niederhausen and Nicky Wakim

2024-11-13

Learning Objectives

1. Distinguish between Uniform, Exponential, Gamma, and Normal distributions when reading a story.
2. Identify the variable and the parameters in a story, and state what the variable and parameters mean.
3. Use the formulas for the pdf/CDF, expected value, and variance to answer questions and find probabilities.

Where are we?

Basics of probability

- Outcomes and events
- Sample space
- Probability axioms
- Probability properties
- Counting
- Independence
- Conditional probability
- Bayes' Theorem
- Random Variables

Probability for discrete random variables

- Functions: pmfs/CDFs
- Important distributions
- Joint distributions
- Expected values and variance

Probability for continuous random variables

- Calculus ✓
 - Functions: pdfs/CDFs ✓
 - Important distributions ✓
 - Joint distributions → 2 RVs
 - Expected values and variance ✓
- single RVs*

Advanced probability

- Central limit theorem
- Functions: moment generating functions

Chapter 31: Uniform RVs

Properties of continuous uniform RVs

- **Scenario:** Events are equally likely to happen anywhere or anytime in an interval of values
- Shorthand: $X \sim U[a, b]$

pdf

$$f_X(x) = \frac{1}{b-a}, \text{ for } \underline{a \leq x \leq b}$$

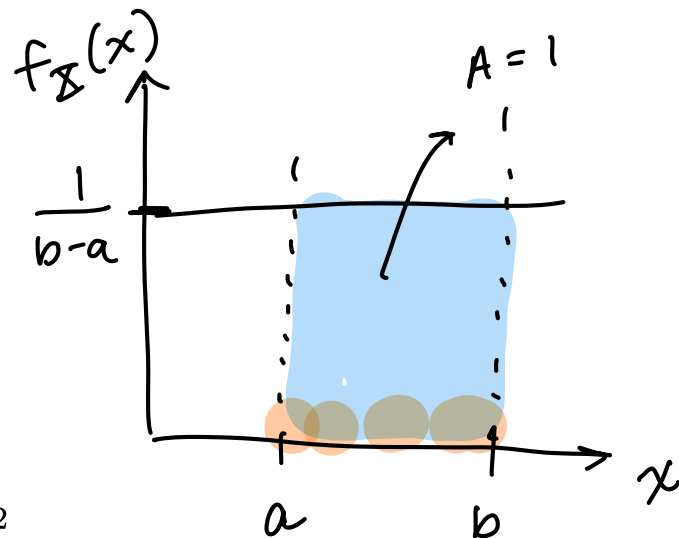
CDF

$$F_X(x) = \int_{-\infty}^x f_X(s) ds$$

$$F_X(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

$$\underline{E(X)} = \frac{a+b}{2}, \quad \underline{\text{Var}(X)} = \frac{(b-a)^2}{12}$$

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$



Identifying continuous uniform RV from word problems

- Look for some indication that all events are **equally likely**
 - Could also say “uniformly distributed”
- Look for an interval
 - **Time example:** Customer in your store will approach the cash register in next 30 minutes. Approaching the register throughout the 30 minutes is equally likely.
 - **Length example:** You have a 12 inch string that you need to cut. You are equally likely to cut anywhere on the string.
- Different than the discrete uniform
 - Discrete usually includes a *countable* number of events that are equally likely
 - Continuous is not countable
 - Exact time and length can be measured with infinite decimal places

Helpful R code

Let's say we're looking at equally likely arrival times between 10 am and 11 am.

- If we want to know the probability that someone arrives at 10:30am or earlier:

```
1 punif(q = 30, min = 0, max = 60)
[1] 0.5
```

$$P(X \leq 30)$$

dunif → output
 $f_X(x) \neq \text{prob}$

- If we want to know the time, say t , where the probability of arriving at t or earlier is 0.35:

```
1 qunif(p = 0.35, min = 0, max = 60)
[1] 21
```

$$P(X \leq t) = 0.35$$

0 to 60
min

- If we want to know the probability that someone arrives between 10:14 and 10:16 am:

```
1 punif(q = 16, min = 0, max = 60) - punif(q = 14, min = 0, max = 60)
[1] 0.03333333
```

$$P(14 < X < 16)$$

- If we want to sample 20 arrival times from the distribution:

```
1 runif(n = 20, min = 0, max = 60)
[1] 14.678120 51.978743 20.053322 30.961489 18.560075 23.306970 6.055852
[8] 58.759085 28.970863 42.766885 37.308040 47.688816 4.226385 46.083334
[15] 6.912673 31.533889 26.919587 35.281129 2.206351 46.045267
```

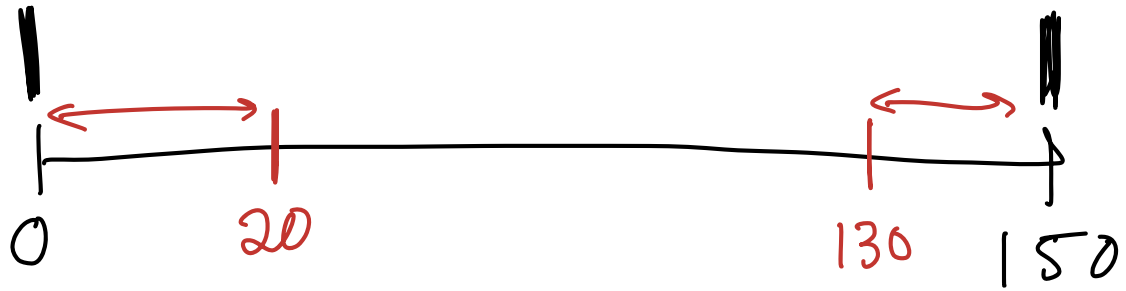
$$= P(X < 16) - P(X < 14) \\ = F_X(16) - F_X(14)$$

Bird on a wire (TB 31.5)

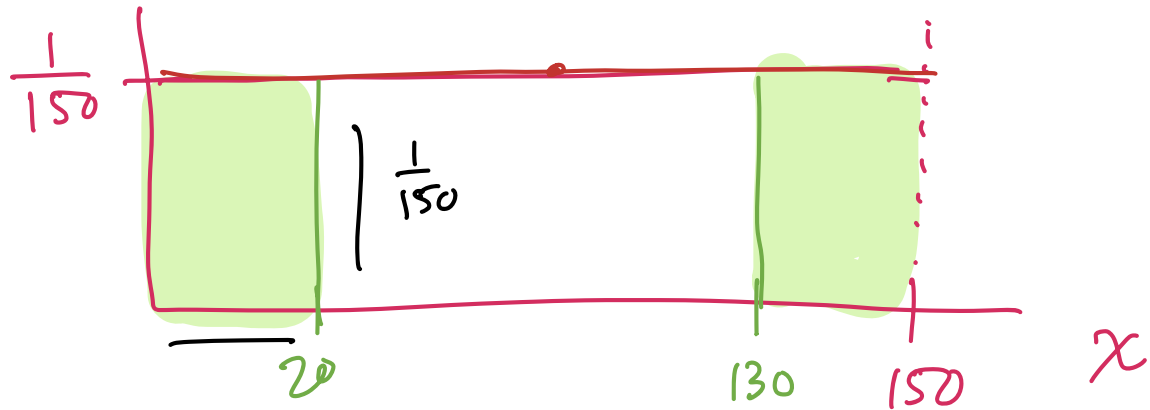
Example 1

A bird lands at a location that is Uniformly distributed along an electrical wire of length 150 feet. The wire is stretched tightly between two poles.

What is the probability that the bird is 20 feet or less from one or the other of the poles?



$$P(X < 20) + P(X > 130)$$



$$2 P(X < 20) = 2 \cdot \frac{1}{150} \cdot 20 = 0.2667$$

Chapter 32: Exponential RVs

Properties of exponential RVs

- **Scenario:** Modeling the time until the next (first) event
- Continuous analog to the geometric distribution!
- Shorthand: $X \sim \text{Exp}(\lambda)$

$$f_X(x) = \lambda e^{-\lambda x} \text{ for } x > 0, \underline{\lambda > 0}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda}$$

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

$\lambda =$ avg rate of
successes

OR

$\lambda =$ avg # of
arrivals/
successes
per time
period

Memoryless Property

If $b > 0$,

$$P(X > a + b | X > a) = P(X > b)$$

- This can be interpreted as:

- If you have waited a seconds (or any other measure of time) without a success
- Then the probability that you have to wait b more seconds is the same as the probability of waiting b seconds initially.

Identifying exponential RV from word problems

- Look for time between events/successes
- Look for a rate of the events over time period
- How does it differ from the geometric distribution?
 - Geometric is *number of trials* until first success
 - Exponential is *time* until first success
- Relation to the Poisson distribution?
 - When the time between arrivals is exponential, the number of arrivals in a fixed time interval is Poisson with the mean λ

Helpful R code

Let's say we're sitting at the bus stop, measuring the time until our bus arrives. We know the bus comes every 10 minutes on average.

- If we want to know the probability that the bus arrives in the next 5 minutes: $P(X \leq 5)$

```
1 pexp(q = 5, rate = 1/10)
```

```
[1] 0.3934693
```

\rightarrow 1 bus / 10 min

- If we want to know the time, say t , where the probability of the bus arriving at t or earlier is 0.35:

```
1 qexp(p = 0.35, rate = 1/10)
```

```
[1] 4.307829
```

$$P(X \leq t) = 0.35$$

- If we want to know the probability that the bus arrives between 3 and 5 minutes: $P(3 \leq X \leq 5)$

```
1 pexp(q = 5, rate = 1/10) - pexp(q = 3, rate = 1/10)
```

```
[1] 0.1342876
```

- If we want to sample 20 bus arrival times from the distribution:

```
1 rexp(n = 20, rate = 1/10)
```

```
[1] 2.6536492 9.9241230 9.8590037 16.7313808 1.7358231 9.5964185  
[7] 1.3491135 2.2182230 9.6290820 0.6212311 53.4122497 4.6816007  
[13] 14.0487481 7.0381356 7.5556238 6.3182355 3.5183107 27.6916935  
[19] 0.4412962 28.0393721
```

Transformation of independent exponential RVs

Revisit after joint notes:

Example 1

Let $X_i \sim \text{Exp}(\lambda_i)$ be independent RVs, for $i = 1 \dots n$. Find the pdf for the first of the arrival times.

Chapter 33: Gamma RVs

Properties of gamma RVs

- **Scenario:** Modeling the time until the r^{th} event.
- Continuous analog to the Negative Binomial distribution
- Shorthand: $X \sim \text{Gamma}(r, \lambda)$

$$f_X(x) = \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} \text{ for } x > 0, \lambda > 0, \Gamma(r) = (r-1)!$$

Handwritten notes: A green circle around $\Gamma(r)$ is crossed out with a green 'X'. Below it, $(r-1)!$ is written in green. A red arrow points from the word "Gamma" in pink to the green circle around $\Gamma(r)$.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} \sum_{j=0}^{r-1} \frac{(\lambda x)^j}{j!} & x \geq 0 \end{cases}$$

$$E(X) = \frac{r}{\lambda}, \text{ Var}(X) = \frac{r}{\lambda^2}$$

Common to see $\alpha = r$ and $\beta = \lambda$

$$\begin{aligned} X &\sim \text{Gamma}(r, \lambda) \\ X &\sim \text{Gamma}(\alpha, \beta) \end{aligned}$$

Handwritten notes: A pink bracket connects the r in the first equation to the α in the second equation.

Identifying gamma RV from word problems

- Gamma distribution with $r = 1$ is same as exponential
 - Just like Negative Binomial with $r = 1$ is same as the geometric distribution
- Similar to exponential
 - Look for time between or until events/successes
 - BUT now we are measuring time until more than 1 success
 - Look for a rate of the events over time period

Helpful R code

Let's say we're sitting at the bus stop, measuring the time until 4 buses arrive. We know the bus comes every 10 minutes on average. $\lambda = 1 \text{ bus}/10 \text{ min}$

- If we want to know the probability that the 4 buses arrive in the next 50 minutes: $P(X \leq 50)$

```
1 pgamma(q = 50, rate = 1/10, shape = 4)
```

[1] 0.7349741

λ

```
1 pgamma(q = 50, scale = 10, shape = 4)
```

[1] 0.7349741

α β

- If we want to know the time, say t , where the probability of the 4 buses arriving at t or earlier is 0.35:

```
1 qgamma(p = 0.35, rate = 1/10, shape = 4)
```

[1] 29.87645

$$P(X \leq t) = 0.35$$

- If we want to know the probability that the 4 buses arrives between 30 and 50 minutes:

```
1 pgamma(q = 50, scale = 10, shape = 4) - pgamma(q = 30, scale = 10, shape = 4)
```

[1] 0.382206

$$P(30 \leq X \leq 50)$$

- If we want to sample 20 arrival times for the 4 buses:

```
1 rgamma(n = 20, scale = 10, shape = 4)
```

[1]	43.334506	11.669883	107.085600	68.256312	6.594443	40.773304
[7]	20.604468	39.343215	22.625283	44.203254	30.295759	42.049533
[13]	28.457668	47.731174	46.818598	60.548348	45.123801	36.140067
[19]	37.573537	31.463989				

Remarks

- The parameter r in a $\text{Gamma}(r, \lambda)$ distribution does NOT need to be a positive integer
 - r is usually a positive integer
- When r is a positive integer, the distribution is sometimes called an $\text{Erlang}(r, \lambda)$ distribution
- When r is any positive real number, we have a general gamma distribution that is usually instead parameterized by $\alpha > 0$ and $\beta > 0$, where:
 - $\alpha = \text{shape parameter}$: same as k , the total number of events we must witness
 - In R code example: 4 buses to wait for
 - $\beta = \text{scale parameter}$: same as λ , the rate parameter
 - In R code example: 1 bus per 10 minutes (1/10)

Sending money orders

Example 1

On average, someone sends a money order once per 15 minutes. What is the probability someone sends 10 money orders in less than 3 hours?

$$\lambda = \frac{1}{15}$$

$$\underline{X} \sim \text{Gamma}(r=10, \underbrace{\lambda = \frac{1}{15}}_{\substack{\downarrow \\ \text{in minutes}}})$$

\downarrow
time

\underline{X} time in hours

$$\underline{X} \sim \text{Gamma}(10, 4)$$

$$\lambda = \frac{1 \text{ order}}{15 \text{ min}} \cdot \frac{60 \text{ min}}{\text{hr}} = \frac{60}{15} = 4$$

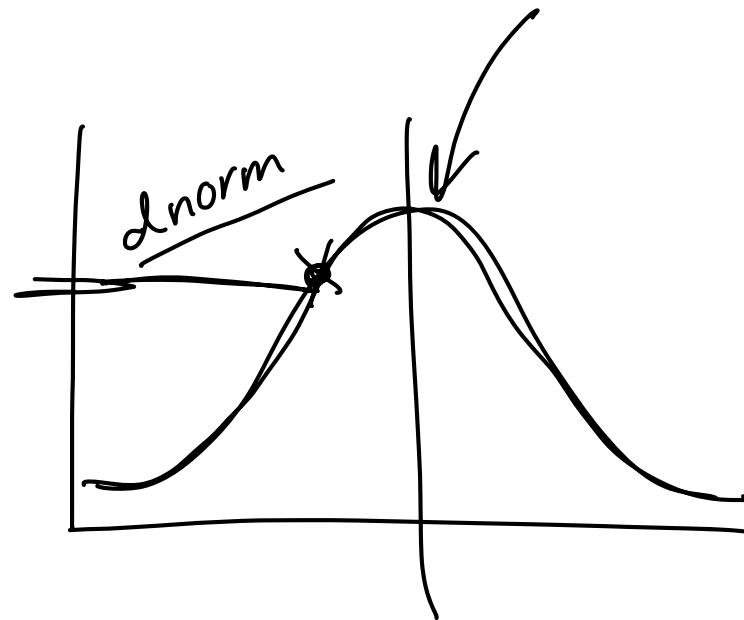
$$P(\underline{X} \leq 3) = 0.7576$$

$$\text{pgamma}(q=3, \text{shape}=10, \text{rate}=4)$$

Additional Resource

- Another helpful site with R code: <https://rpubs.com/mpfoley73/459051>

Chapter 35: Normal RVs



Properties of Normal RVs

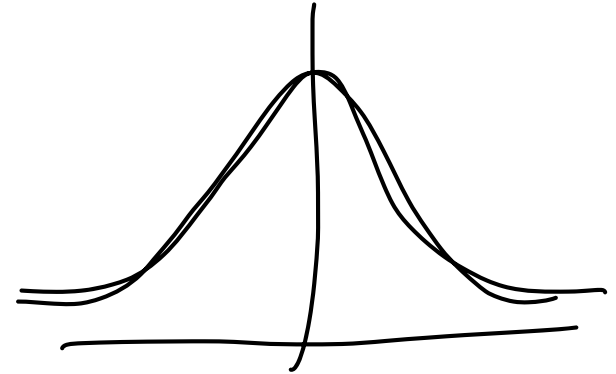
- No scenario description here because the Normal distribution is so universal
 - Central Limit Theorem (~~next class~~) makes it applicable to many types of events

• Shorthand: $X \sim \text{Normal}(\mu, \sigma^2)$ $\underline{X} \sim \text{Norm}(\mu, \sigma)$

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/(2\sigma^2)}, \text{ for } -\underline{\text{inf}} < x < \underline{\text{inf}}$$

$$E(X) = \mu$$

$$\text{Var}(X) = \sigma^2$$



Helpful R code

Let's say we're measuring the high temperature today. The average high temperature on this day across many, many years is 50 degrees with a standard deviation of 4 degrees.

- If we want to know the probability that the high temperature is below 45 degrees: $P(X < 45)$

```
1 pnorm(q = 45, mean = 50, sd = 4)
[1] 0.1056498
```

- If we want to know the temperature, say t , where the probability of that the temperature is at t or lower is 0.35:

$$P(X \leq t) = 0.35$$

```
1 qnorm(p = 0.35, mean = 50, sd = 4)
[1] 48.45872
```

- If we want to know the probability that the temperature is between 45 and 50 degrees: $P(45 \leq X \leq 50)$

```
1 pnorm(q = 50, mean = 50, sd = 4) - pnorm(q = 45, mean = 50, sd = 4)
[1] 0.3943502
```

- If we want to sample 20 days' temperature (over the years) from the distribution:

```
1 rnorm(n = 20, mean = 50, sd = 4)
[1] 58.19334 48.69453 48.01503 46.60855 42.96407 50.10835 49.07470 50.74432
[9] 43.80155 47.44748 50.30352 53.64828 47.59478 50.79763 52.01850 52.46709
[17] 51.06833 46.64400 60.99100 55.08259
```

$$= \int_{45}^{50} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Movie night while studying

Example 1

Children's movies run an average of 98 minutes with a standard deviation of 10 minutes. You check out a random movie from the library to entertain your kids so you can study for your test. Assume that your kids will be occupied for the entire length of the movie.

- What is the probability that your kids will be occupied for at least the 2 hours you would like to study?
- What is range for the bottom quartile (lowest 25%) of time they will be occupied?

$$\bar{X} \sim N(\mu = 98, \sigma = 10)$$

$\sigma^2 = 100$

$$a) P(\bar{X} \geq 2 \cdot 60) = P(\bar{X} > 120)$$

hrs $\frac{\text{min}}{\text{hr}}$

$$\text{pnorm}(q = 120, \text{mean} = 98, \text{sd} = 10, \text{lower.tail} = F)$$

$$= 0.0139$$

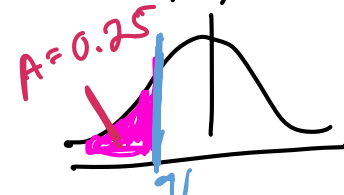
→ $P(X > ?)$

$$b) P(\bar{X} \leq t) = 0.25$$

T: $P(X < ?)$

$$\text{qnorm}(p = 0.25, \text{mean} = 98, \text{sd} = 10)$$

$$= 91.255$$



Standard Normal Distribution

$$Z \sim \text{Normal}(\mu = 0, \sigma^2 = 1)$$

- Used to be more helpful when computing was not as advanced
 - Use tables of the standard normal
 - You can convert any normal distribution to a standard normal through transformation

$$Z = \frac{X - \mu_X}{\sigma_X}$$

$$X \sim N(\mu_X, \sigma_X) \rightarrow Z = \frac{X - \mu_X}{\sigma_X} \sim$$

- Comes from $X = \sigma_X Z + \mu_X$

- Since σ_X and μ_X are constants, then $E(X) = \mu_X$ and $SD(X) = \sigma_X$

$$N(0, 1)$$

