Chapter 4. Continuous Random Variables

4.2: Zoo of Continuous RVs

Slides (Google Drive) Alex Tsun Video (YouTube)

Now that we've learned about the properties of continuous random variables, we'll discover some frequently used RVs just like we did for discrete RVs! In this section, we'll learn the continuous Uniform distribution, the Exponential distribution, and Gamma distribution. In the next section, we'll finally learn about the Normal/Gaussian (bell-shaped) distribution which you all may have heard of before!

4.2.1 The (Continuous) Uniform RV

The continuous uniform random variable models a situation where there is no preference for any particular value over a bounded interval. This is very similar to the discrete uniform random variable (e.g., roll of a fair die), except extended to include decimal values. The probability of equalling any particular value is again 0 since we are dealing with a continuous RV.

Definition 4.2.1: Uniform (Continuous) RV

$X \sim \text{Unif}(a, b)$ (continuous) where $a < b$ are real numbers, if and only if X has the following pdf:

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

X is equally likely to be take on any value in $[a, b]$. Note the similarities and differences it has with the discrete uniform! The value of the density function is constant at $\frac{1}{b-a}$, for any input $x \in [a, b]$, and makes it a rectangle whose area integrates to 1.

$$\mathbb{E}[X] = \frac{a+b}{2}, \quad \mathsf{Var}(X) = \frac{(b-a)^2}{12}$$

The cdf is

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

Proof of Expectation and Variance of Uniform. I'm setting up the integrals but omitting the steps that are not relevant to your understanding of probability theory (computing integrals):

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x)dx = \int_a^b x \cdot \frac{1}{b-a}dx = \frac{a+b}{2}$$

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x)dx = \int_a^b x^2 \cdot \frac{1}{b-a}dx = \frac{a^2+ab+b^2}{3}$$

$$\mathsf{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{a^2+ab+b^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}$$

□

Example(s)

Suppose we think that a Hollywood movie's overall rating is equally likely to be any decimal value in the interval $[1, 5]$ (this may not be realistic). You may be able to do these questions "in your head", but I encourage you to formalize the questions and solutions to practice the notation and concepts we've learned. (You probably wouldn't be able to do them "in your head" if the movie rating wasn't uniformly distributed!)

1. A movie is considered average if it is overall rating is between 1.5 and 4.5. What is the probability that is average?
2. A movie is considered a huge success if its overall rating is at least 4.5. What is the probability that it is a huge success?
3. A movie is considered legendary if its overall rating is at least 4.95. *Given that* a movie is a huge success, what is the probability it is legendary?

Solution Before starting, we can write that the overall rating of a movie is $X \sim \text{Unif}(1, 5)$. Hence, its density function is $f_X(x) = \frac{1}{5-1} = \frac{1}{4}$ for $x \in [1, 5]$ (and 0 otherwise).

1. We know the probability of being in the range $[1.5, 4.5]$ is the area under the density function from 1.5 to 4.5, so
$$\mathbb{P}(1.5 \le X \le 4.5) = \int_{1.5}^{4.5} f_X(x)dx = \int_{1.5}^{4.5} \frac{1}{4}dx = \frac{3}{4}$$
You could have also drawn a picture of this density function (which is flat at $1/4$), and exploited geometry to figure that the base of the rectangle is 3 and the height is $1/4$.
2. Similarly,
$$\mathbb{P}(X \ge 4.5) = \int_{4.5}^{\infty} f_X(x)dx = \int_{4.5}^{5} \frac{1}{4}dx = \frac{1}{8}$$
Note that the density function for values $x \ge 5$ is zero, so that's why the integral changed its upper bound from ∞ to 5 when replacing the density!
3. We'll use Bayes Theorem:
$$\mathbb{P}(X \ge 4.95 \mid X \ge 4.5) = \frac{\mathbb{P}(X \ge 4.5 \mid X \ge 4.95)\,\mathbb{P}(X \ge 4.95)}{\mathbb{P}(X \ge 4.5)}$$
Note that $\mathbb{P}(X \ge 4.5 \mid X \ge 4.95) = 1$ (why?) and $\mathbb{P}(X \ge 4.95) = \frac{1}{80}$ (do a similar integral again or use geometry), so plugging in these numbers gives
$$= \frac{1 \cdot \frac{1}{80}}{\frac{1}{8}} = \frac{1}{10}$$
Think about why this also might make sense intuitively!

□

4.2.2 The Exponential RV

Now we'll learn a distribution which is typically used to model waiting time until an event, like a server failure or the bus arriving. This is a *continuous* RV since the time taken has decimal places, like 3.5341109 minutes or 9.9324 seconds. This is like the continuous extension of the Geometric (discrete) RV which is the number of trials until a success occurs.

Recall the Poisson Process with parameter $\lambda > 0$ has events happening at average rate of λ per unit time forever. The exponential RV measures the *time* (e.g., 4.33212 seconds, 9.382 hours, etc.) until the first occurrence of an event, so is a continuous RV with range $[0, \infty)$ (unlike the Poisson RV, which counts the *number of occurrences* in a unit of tine, with range $\{0, 1, 2, ...\}$ and is a discrete RV).

Let $Y \sim \text{Exp}(\lambda)$ be the time until the first event. We'll first compute its CDF $F_Y(t)$ and then differentiate it to find its PDF $f_Y(t)$.

Let $X(t) \sim \text{Poi}(\lambda t)$ be the number of events in the first t units of time, for $t \geq 0$ (if average is is λ per unit of time, then it is λt per t units of time). Then, $Y > t$ (wait longer than t units of time until the first event) *if and only if* $X(t) = 0$ (no events happened in the first t units of time). This allows us to relate the Exponential CDF to the Poisson PMF.

$$\mathbb{P}(Y > t) = \mathbb{P}(\text{no events in the first t units}) = \mathbb{P}(X(t) = 0) = e^{-\lambda t}\frac{(\lambda t)^0}{0!} = e^{-\lambda t}$$

Note that we plugged in the $\text{Poi}(\lambda t)$ PMF at 0 in the second last equality. Now, the CDF is just the complement of the probability we computed:

$$F_Y(t) = \mathbb{P}(Y \leq t) = 1 - \mathbb{P}(Y > t) = 1 - e^{-\lambda t}$$

Remember since the CDF was the integral of the PDF, the PDF is the derivative of the CDF by the fundamental theorem of calculus:

$$f_Y(t) = \frac{d}{dt}F_Y(t) = \lambda e^{-\lambda t}$$

Definition 4.2.2: The Exponential RV

$X \sim \text{Exp}(\lambda)$, if and only if X has the following pdf (and range $\Omega_X = [0, \infty)$):

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

X is the waiting time until the first occurrence of an event in a Poisson Process with parameter λ.

$$\mathbb{E}[X] = \frac{1}{\lambda}, \qquad \mathsf{Var}(X) = \frac{1}{\lambda^2}$$

The cdf is

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

Proof of Expectation and Variance of Exponential. You can use integration by parts if you want to solve these integrals, or you can use WolframAlpha. Again, I'm omitting the steps that are not relevant to your understanding of probability theory (computing integrals):

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$\mathbb{E}\left[X^2\right] = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^{\infty} x^2 \cdot \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

$$\mathsf{Var}(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

□

If you usually skip examples, please don't skip the next two. The first example here highlights the relationship between the Poisson and Exponential RVs, and the second highlights the memoryless property!

Example(s)

Suppose that, on average, 13 car crashes occur each *day* on Highway 101. What is the probability that no car crashes occur in the next *hour*? Be careful of units of time!

Solution We will solve this problem with three equivalent approaches! Take the time to understand why each of them work.

1. Then, on average there are $\frac{13}{24}$ car crashes per *hour*. So the number of crashes in the next hour is $X \sim \text{Poi}\left(\lambda = \frac{13}{24}\right)$.

$$\mathbb{P}(X = 0) = e^{-13/24} \frac{(13/24)^0}{0!} = e^{-13/24}$$

2. Similar to above, the time (in *hours*) until the first car crash is $Y \sim \text{Exp}\left(\lambda = \frac{13}{24}\right)$, since on average 13/24 car crashes happen per *hour*. Then, the probability no car crashes happen in the next hour is

$$\mathbb{P}(Y > 1 \text{ (hour)}) = 1 - \mathbb{P}(Y \leq 1) = 1 - F_Y(1) = 1 - (1 - e^{-13/24 \cdot 1}) = e^{-13/24}$$

3. If we don't want to change the units, then we can say the waiting time until the next car crash (in *days*) is $Z \sim \text{Exp}(\lambda = 13)$. since on average 13 car crashes happen per *day*. Then, the probability no car crashes occur in the next hour (1/24 of a day) is the probability that we wait longer than 1/24 day:

$$\mathbb{P}(Z > 1/24) = 1 - \mathbb{P}(Z \leq 1/24) = 1 - F_Z(1/24) = 1 - (1 - e^{-13 \cdot 1/24}) = e^{-13/24}$$

Hopefully the first and second solutions show you the relationship between the Poisson and Exponential RVs (they both come from the Poisson process), and the second and third solution show you how to be careful with units and that you'll get the same answer as long as you are consistent. □

Example(s)

Suppose the average battery life of an AAA battery is approximately 50 hours.

1. What is the probability the battery lasts more than 60 hours?
2. What is the probability the battery lasts more than 40 hours?
3. What is the probability the battery lasts more than 100 hours, *given* that the battery has

already lasted 60 hours? That is, what is the probability it can last 40 additional hours? Relate this to your answer from the previous part!

Solution Since we want to model battery life, we should use an Exponential distribution. Since we know the average battery life is 50 hours, and that the expected value of an exponential RV is $1/\lambda$ (see above), we should say that the battery life is $X \sim \text{Exp}\left(\lambda = \frac{1}{50} = 0.02\right)$.

1. If we want the probability the battery lasts more than 60 hours, then we want

$$\mathbb{P}(X \geq 60) = \int_{60}^{\infty} f_X(t)dt = \int_{60}^{\infty} 0.02e^{-0.02t}dt = e^{-1.2}$$

   But continuous distributions have a CDF which we can and should take advantage of! We can look up the CDF above as well:

$$\mathbb{P}(X \geq 60) = 1 - \mathbb{P}(X < 60) = 1 - F_X(60) = 1 - (1 - e^{-0.02 \cdot 60}) = e^{-1.2}$$

   We made a step above that said $\mathbb{P}(X < 60) = F_X(60)$, but $F_X(60) = \mathbb{P}(X \leq 60)$. It turns out they are the same for continuous RVs, since the probability $X = 60$ exactly is zero!

2. Similarly,

$$\mathbb{P}(X \geq 40) = 1 - \mathbb{P}(X < 40) = 1 - F_X(40) = 1 - (1 - e^{-0.02 \cdot 40}) = e^{-0.8}$$

3. By Bayes Theorem,

$$\mathbb{P}(X \geq 100 \mid X \geq 60) = \frac{\mathbb{P}(X \geq 60 \mid X \geq 100)\,\mathbb{P}(X \geq 100)}{\mathbb{P}(X \geq 60)}$$

   Note that $\mathbb{P}(X \geq 60 \mid X \geq 100) = 1$ (why?) and $\mathbb{P}(X \geq 100) = e^{-0.02 \cdot 100} = e^{-2}$ (same process as above), so plugging in these numbers gives

$$= \frac{1 \cdot e^{-2}}{e^{-1.2}} = e^{-0.8}$$

   Note that this is exactly the same as $\mathbb{P}(X \geq 40)$ above, the probability we the battery lasted at least 40 hours. This says that the previous 60 hours don't matter - $\mathbb{P}(X \geq 40 + 60 \mid X \geq 60) = \mathbb{P}(X \geq 40)$. This property is called memorylessness, since the battery essentially forgets that it was alive for 60 hours! We'll discuss this more formally below and prove it.

□

4.2.3 Memorylessness

Definition 4.2.3: Memorylessness

A random variable X is **memoryless** is for all $s, t \geq 0$,

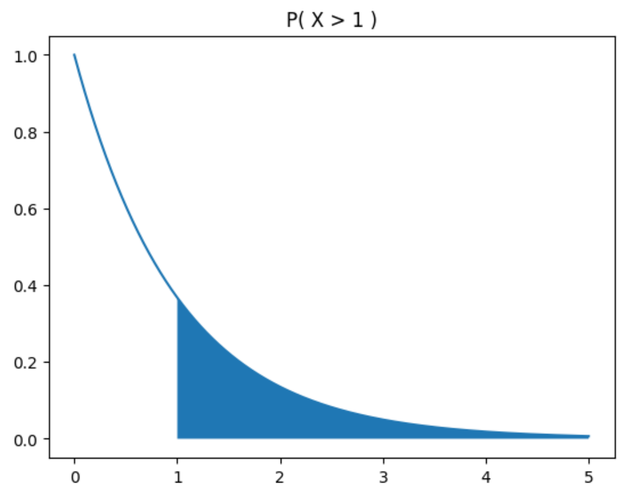
$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

We just saw a concrete example above, but let's see another. Let $s = 7, t = 2$. So $\mathbb{P}(X > 9 \mid X > 7) = \mathbb{P}(X > 2)$.

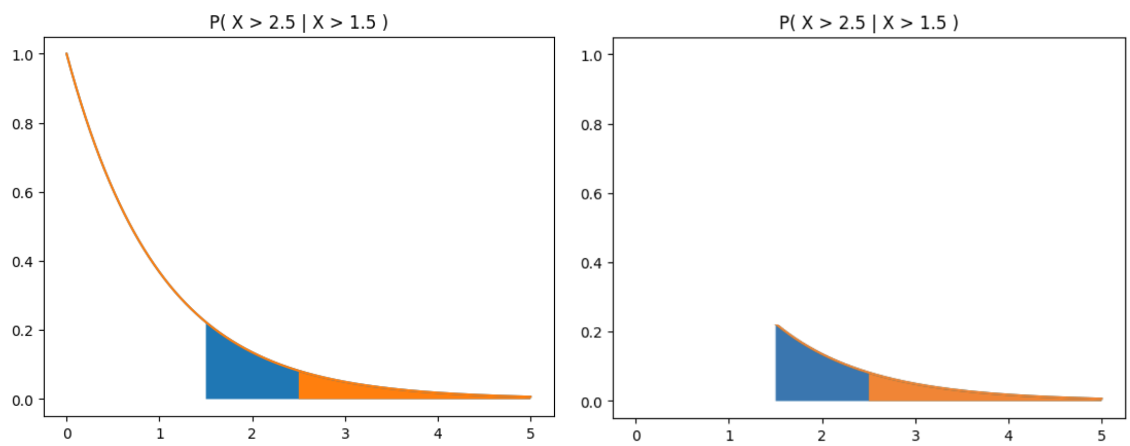
This memoryless property says that, given we've waited (at least) 7 minutes, the probability we wait (at least) 2 more minutes, is the same as the probability we waited (at least 2) more from the beginning. That is, the random variable "forgot" how long we've already been waiting.

The only memoryless RVs are the **Geometric** (discrete) and **Exponential** (Continuous)! This is because events happen *independently* over time/trials, and so the past doesn't matter.

We've seen it algebraically and intuitively, but let's see it pictorially as well. Here is a picture of the probability is greater than 1 for an exponential RV. It is the area to the right of 1 of the density function $\lambda e^{-\lambda x}$ for $x \geq 0$ (shaded in blue).



Below is a picture of the probability $X > 2.5$ given $X > 1.5$ (shaded in orange and blue). If you hide the area to the left of 1.5, you can see the ratio of the orange area (right of 2.5) to the entire shaded region (right of 1.5) is the same as $\mathbb{P}(X > 1)$ above. So this exponential density function has memorylessness built in!



Theorem 4.2.1: Memorylessness of Exponential

If $X \sim \text{Exp}(\lambda)$, then X has the memoryless property.

Proof of Memorylessness of Exponential.

If $X \sim \text{Exp}(\lambda)$ and $x \geq 0$, then recall

$$\mathbb{P}(X > x) = 1 - F_X(x) = 1 - (1 - e^{-\lambda x}) = e^{-\lambda x}$$

$$\begin{aligned}
\mathbb{P}(X > s + t \mid X > s) &= \frac{\mathbb{P}(X > s \mid X > s + t)\,\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} && \text{[Bayes Theorem]} \\
&= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} && [\mathbb{P}(X > s \mid X > s + t) = 1] \\
&= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} && \text{[plug in formula above]} \\
&= e^{-\lambda t} \\
&= \mathbb{P}(X > t)
\end{aligned}$$

□

Theorem 4.2.2: Memorylessness of Geometric

If $X \sim \text{Geo}(p)$, then X has the memoryless property.

Proof of Memorylessness of Geometric.

If $X \sim \text{Geo}(p)$, then for $k \in \Omega_X = \{1, 2, \dots\}$, then by independence of the trials,

$$\mathbb{P}(X > k) = \mathbb{P}(\text{no successes in first } k \text{ trials}) = (1 - p)^k$$

Then, I'll leave it to you to do the same computation as above (using Bayes Theorem). You'll see it work out almost exactly the same way! □

4.2.4 The Gamma RV

Just like the Exponential RV is the continuous extension of the Geometric RV (from discrete trials to continuous time), we have a Gamma RV which models the time until the r-th event. This should remind you of the Negative Binomial RV, which modelled the number of trials until the r-th success, and so was the sum of r independent and identically distributed (iid) $Geo(p)$ RVs.

Definition 4.2.4: Gamma RV

$X \sim \text{Gamma}(r, \lambda)$ if and only if X has the following pdf:

$$f_X(x) = \begin{cases} \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

X is the sum of r independent $\text{Exp}(\lambda)$ random variables.
Gamma is to Exponential as Negative Binomial to Geometric. It is the waiting time until the r-th event, rather than just the first event. So you can write it as a sum of r independent exponential random variables.

$$\mathbb{E}[X] = \frac{r}{\lambda}, \qquad \mathsf{Var}(X) = \frac{r}{\lambda^2}$$

X is the waiting time until the r^{th} occurrence of an event in a Poisson Process with parameter λ. Notice that $\text{Gamma}(1, \lambda) \equiv \text{Exp}(\lambda)$. By definition, if X, Y are independent with $X \sim \text{Gamma}(r, \lambda)$ and $Y \sim \text{Gamma}(s, \lambda)$, then $X + Y \sim \text{Gamma}(r + s, \lambda)$.

Proof of Expectation and Variance of Gamma. The PDF of the Gamma looks very ugly and hard to deal with, so let's use our favorite trick: Linearity of Expectation! As mentioned earlier, if $X \sim \text{Gamma}(r, \lambda)$, then $X = \sum_{i=1}^{r} X_i$ where each $X_i \sim \text{Exp}(\lambda)$ is independent with $\mathbb{E}[X_i] = \frac{1}{\lambda}$ and $\mathsf{Var}(X_i) = \frac{1}{\lambda^2}$. So by LoE,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{r} X_i\right] = \sum_{i=1}^{r} \mathbb{E}[X_i] = \sum_{i=1}^{r} \frac{1}{\lambda} = \frac{r}{\lambda}$$

Now, we can use the fact that the variance of a sum of *independent* rvs is the sum of the variances (we have yet to prove this fact).

$$\mathsf{Var}(X) = \mathsf{Var}\left(\sum_{i=1}^{r} X_i\right) = \sum_{i=1}^{r} \mathsf{Var}(X_i) = \sum_{i=1}^{r} \frac{1}{\lambda^2} = \frac{r}{\lambda^2}$$

□

4.2.5 Exercises

1. Suppose that on average, 40 babies are born every hour in Seattle.
   (a) What is the probability that no babies are born in the next minute? Try solving this in two different but equivalent ways - using a Poisson and Exponential RV.
   (b) What is the probabilities that it takes more than 20 minutes for the first 10 babies to be born? Again, try solving this in two different but equivalent ways - using a Poisson and Gamma RV.
   (c) What is the expected time until the 5th baby is born?

   Solution:

   (a) The number of babies born in the next minute is $X \sim \text{Poi}(40/60)$, so $\mathbb{P}(X = 0) = e^{-40/60} \approx 0.5134$. Alternatively, the time in minutes until the next baby is born is $Y \sim \text{Exp}(40/60)$, and we want the probability that no babies are born in the next minute; i.e., it takes at least one minute for the first baby to be born. Hence,

   $$\mathbb{P}(Y > 1) = 1 - F_Y(1) = 1 - (1 - e^{-2/3 \cdot 1}) = e^{-2/3}$$

   We got the same answer with two different approaches!

   (b) The number of babies born in the next 20 minutes is $W \sim \text{Poi}(40/3)$, so

   $$\mathbb{P}(W \leq 10) = \sum_{k=0}^{10} e^{-40/3} \frac{(40/3)^k}{k!}$$

   Alternatively, the time in minutes until the tenth baby is born is $Z \sim \text{Gamma}(10, 40/60)$, and we are asking what's the probability this is over 20 minutes, is

   $$\mathbb{P}(Z > 20) = 1 - F_Z(20) = 1 - \int_0^{20} \frac{(40/60)^{10}}{(10-1)!} x^{10-1} e^{-(40/60)x} dx$$

Unfortunately, there isn't a nice closed form for the Gamma CDF, but this would evaluate to the same result!

(c) The time in minutes until the 5th baby is born is $V \sim \text{Gamma}(5, 40/60)$, so $\mathbb{E}[V] = \frac{r}{\lambda} = \frac{5}{40/60} =$ 7.5 minutes.

2. You are waiting for a bus to take you home from CSE. You can either take the E-line, U-line, or Cline. The distribution of the waiting time in minutes for each is the following:

- E-Line: $E \sim \text{Exp}(\lambda = 0.1)$.
- U-Line: $U \sim \text{Unif}(0, 20)$ (continuous).
- C-Line: Has range $(1, \infty)$ and PDF $f_C(x) = 1/x^2$.

Assume the three bus arrival times are independent. You take the first bus that arrives

(a) Find the CDFs of E, U, and C, $F_E(t)$, $F_U(t)$ and $F_C(t)$. Hint: The first two can be looked up in our distributions handout!

(b) What is the probability you wait more than 5 minutes for a bus?

(c) What is the probability you wait more than 30 minutes for a bus?

Solution:

(a) The CDF of E for $t > 0$ is $F_E(t) = 1 - e^{-0.1t}$ (see above).
The CDF of U for $0 < t < 20$ is $F_U(t) = \frac{t}{20}$.
The CDF of C for $t > 1$ is $F_C(t) = \int_1^t f_C(x)dx = 1 - \frac{1}{t}$.

(b) Let $B = \min\{E, U, C\}$ be the time until the first bus. Then, the probability we wait more than 5 minutes is the probability that all of them take longer than 5 minutes to arrive. We can then multiply the individual probabilities due to independence.

$$\mathbb{P}(B > 5) = \mathbb{P}(E > 5, U > 5, C > 5) = \mathbb{P}(E > 5)\,\mathbb{P}(U > 5)\,\mathbb{P}(C > 5)$$

Then, writing in terms of the CDF and plugging in:

$$= (1 - F_E(5))(1 - F_U(5))(1 - F_C(5)) = e^{-0.5} \cdot \frac{15}{20} \cdot \frac{1}{5} = \frac{3}{20}e^{-0.5}$$

(c) The same exact logic applies here! But be careful of the range of U when plugging in the CDF. It is true that

$$\mathbb{P}(B > 30) = \mathbb{P}(E > 30)\,\mathbb{P}(U > 30)\,\mathbb{P}(C > 30)$$

But when plugging in $\mathbb{P}(U > 30) = 1 - F_U(30)$, we have to remember that $F_U(30) = 1$ because U must be in $[0, 20]$. That's why it is so important to define the piecewise function! This probability is indeed 0 since bus U will always come within 20 minutes.