Lecture 7 Worksheet

Chrysafis Vogiatzis

Every worksheet will work as follows.

1. You will be asked to form a group with other students in the class: you can make this as big or as small as you'd like, but groups of 4-5 work best.
2. Read through the worksheet, discussing any questions with the other members of your group.
   - You can call me at any time for help!
   - I will also be interrupting you for general guidance and announcements at random points during the class time.
3. Answer each question (preferably in the order provided) to the best of your knowledge.
4. While collaboration between students is highly encouraged and expected, each student has to submit their own version.
5. You will have 24 hours (see gradescope) to submit your work.

Activity 1: Basic continuous probability distribution properties

Let X be a continuous random variable measuring the current (in milliamperes, mA) in a wire with probability density function (pdf) given by $f(x) = 0.05$, for $0 \leq x \leq \alpha$. Answer the following questions.

Problem 1: Valid pdf?

What is α in order for $f(x)$ to be a valid pdf? [1]

Answer to Problem 1.

[1] Recall that this means that $f(x) \geq 0$ (which is clearly true here) **and** that $\int f(x)dx = 1$ over *all values* that random variable X is allowed to take...

Problem 2: Constructing cumulative distribution functions

What is the cumulative distribution function? [2]

Answer to Problem 2.

Problem 3: Calculating probabilities

The wire is said to be overheating if the current is more than 10mA. It is also said to be working within the normal range of operations if it is between 5 and 10mA. Answer the following questions:

- What is the probability the wire is overheating?
- What is the probability the wire is within its normal range?

Answer to Problem 3.

[2] How is a cdf defined for *continuous* random variables?

Activity 2: The exponential distribution

The time until the next customer arrives is exponentially distributed with a rate of 1 customer every 10 minutes. Answer the following questions.

Problem 4: Calculating probabilities

Let T be the time until the next customer arrives. What is the probability the next customer shows up in the next:

a) 1 minute? b) 5 minutes? c) 10 minutes? d) 20 minutes?

Answer to Problem 4.

a) 1 minute: $P(T \leq 1) =$

b) 5 minutes: $P(T \leq 5) =$

c) 10 minutes: $P(T \leq 10) =$

d) 20 minutes: $P(T \leq 20) =$

This is the big difference between using a rate for a Poisson or for an exponential distribution. For the Poisson distribution, we first convert the rate into the time units of the question. For the exponential distribution, we simply multiply the rate by the necessary time interval length!

For example, for a rate $\lambda = 3$ per minute, the probability we see 5 customers in 3 minutes would be written as

$$P(X = x) = e^{-\lambda} \cdot \frac{\lambda^x}{x!} \implies P(X = 10) = e^{-9} \cdot \frac{9^5}{5!},$$

because λ would be set equal to 9 (customers per 3 minutes).

On the other hand, for the same rate, the probability the next customer shows up within the first 3 minutes would be written as

$$P(T \leq t) = 1 - e^{-\lambda t} \implies P(T \leq 3) = 1 - e^{-3 \cdot 3} = 1 - e^{-9}.$$

Problem 5: Inverting the question

The employee of the store wants to take a break, but they do not want to miss the next customer arrival. How long should the break be for in order to have a 50% chance of not missing the next customer? [3]

[3] In essence, we want the time $\tilde{t}$ such that the probability $P(T \leq \tilde{t}) = 0.5$.

Answer to Problem 5.

Problem 6: Taking a break

In the same store, assume that the employee takes a 2 minute break during which no customer arrived. What is the probability the next customer does not arrive in the next 5 minutes now? Contrast it to your answer for $P(T \leq 5)$ in Problem 4.

Answer to Problem 6.

Based on your answer in Problem 6, we must be deducing that the exponential distribution is indeed **memoryless**. In the next activity, we prove that property for all exponentially distributed random variables.

As a reminder, we say that a distribution is memoryless if for random variable X following that distribution, we have that

$$P(X > s + t | X > s) = P(X > t).$$

For example, in the case of the time of arrival of the next customer (random variable T) in more than 5 minutes from now, we would write $P(T > 5)$. After we take a break for 2 minutes, then that same probability would be $P(T > 7 | T > 2)$.

Activity 3: Memorylessness

Let's check the proof!

Problem 7: Memorylessness proof

In the first step of the proof, we will calculate the right hand side of what we want to show: $P(X > s+t|X > s)$. Using conditional probabilities [4], what is this equal to?

[4] Remember that $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

Answer to Problem 7.

$P(X > s+t|X > s) =$

Problem 8: Memorylessness proof (cont'd)

In the second step of the proof, we need to make one observation. If $A \subseteq B$ then $A \cap B = A$. Is this the case here for the denominator of your answer in Problem 7? Specifically what can you write about $P(X > s+t \cap X > s)$? [5]

[5] Recall what you can say for $P(A \cap B)$ if $B \subseteq A$...

Answer to Problem 8.

$P(X > s+t \cap X > s) =$

Problem 9: Memorylessness proof (cont'd)

In the third step of the proof, consider the original identity we are trying to show: $P(X > s+t|X > s) = P(X > t)$. Assume that X is exponentially distributed with rate λ: that is, $P(X > t) = e^{-\lambda t}$. Fill in the blanks below to get the result. Hints are given along the way (look at the final equation!).

Answer to Problem 9.

$$P(X > s+t|X > s) = \frac{P(X > s+t \cap X > s)}{P(X > s)} = \qquad \text{(from Prob. 7)}$$

$$= \frac{P(X > s+t)}{P(X > s)} = \qquad \text{(from Prob. 8)}$$

$$= \frac{\qquad}{\qquad\qquad} =$$

$$= \underline{\qquad\qquad} = P(X > t).$$

The next worksheet problem comes pre-filled. However, going through this will help with understanding why the exponential distribution is unique.

Problem 10: Only the exponential distribution!

So, is the exponential distribution one of many distributions to be memoryless? The answer is a resounding no: it is the **only** continuous distribution to be memoryless. [6] To show that, we need to use all of the tools in our toolbox.

[6] Again: the **only continuous** distribution. We saw a discrete distribution that is memoryless not too long ago!

Answer to Problem 10.

Like we did in Problem 9:

$$P(X > s+t | X > s) = \frac{P(X > s+t \ \cap \ X > s)}{P(X > s)} = \qquad \text{(from Prob. 7)}$$
$$= \frac{P(X > s+t)}{P(X > s)}. \qquad \text{(from Prob. 8)}$$

We now note that if X is memoryless, then $P(X > s+t|X > s) = P(X > t)$:

$$P(X > s+t|X > s) = P(X > t) \implies \frac{P(X > s+t)}{P(X > s)} = P(X > t)$$
$$\implies P(X > s+t) = P(X > s) \cdot P(X > t).$$

For simplicity, define $\overline{F}(x) = P(X > x)$ and hence our previous equality becomes:

$$\overline{F}(s+t) = \overline{F}(s) \cdot \overline{F}(t).$$

Now, take the logarithms on both sides!

$$\ln \overline{F}(s+t) = \ln \overline{F}(s) \cdot \overline{F}(t) = \ln \overline{F}(s) + \ln \overline{F}(t).$$

Once again, for simplicity, define $g(x) = \ln \overline{F}(x)$, rendering our final equality as:

$$g(s+t) = g(s) + g(t).$$

The **only continuous function** that satisfies $g(s+t) = g(s) + g(t)$ is the linear one (see the citation for the proof in the end), hence we deduce that $g(x) = \beta \cdot x$, for some β.
Combine everything:

1. We need $P(X > s+t) = P(X > s) \cdot P(X > t)$ for memorylessness to hold.
2. Equivalently, after some definitions, we saw this is equivalent to $g(s+t) = g(s) + g(t)$, where $g(x) = \ln \overline{F}(x)$.
3. We saw though that this is only true if $g(s)$ is linear, that is $g(x) = \beta \cdot x$.

We replace backwards: $g(s) = \beta \cdot s \implies \ln \overline{F}(s) = \beta \cdot s \implies \overline{F}(s) = e^{\beta \cdot s} \implies F(s) = 1 - e^{-\beta \cdot s}$. Recall that the exponential distribution has $F(x) = 1 - e^{-\lambda \cdot x}$. Pretty close! Letting $\beta = -\lambda$ finishes the proof!

Problem 11: Just checking

Circle the correct choice in each of the three questions.

Answer to Problem 11.

The exponential distribution is the only *continuous* probability distribution that is memoryless.

a) True b) False

The exponential distribution is the only probability distribution that is memoryless.

a) True b) False

Assume that the time until the next event is exponentially distributed. At 1pm, we calculated the probability that the next event happens before or on 1.30pm is 30%. It is 1.30pm now, and no event has happened yet. What is the probability that the next event happens before or on 2pm? Pick the correct answer in the multiple choice question.

a) It is higher than 30%: it was 30% at 1pm and it never happened, so now it should happen with higher certainty.

b) It is lower than 30%: it was 30% at 1pm and it never happened, so now we should consider that maybe our probability calculation was off to begin with.

c) It is equal to 30%: it was 30% at 1pm and while it never happened, memorylessness states that the probability is still 30%.

Activity 4: Exponential and Poisson

We will play with this relationship (and introduce Gamma/Erlang distributed random variables, too) even more in the next lecture. That said, let us practice the fact that **the exponential distribution and the Poisson distribution go hand-in-hand**. [7]

[7] See the notes for more details!

If the time between events is **exponentially distributed** with rate λ then the number of events in some time t is **Poisson distributed** with rate $\lambda \cdot t$, and vice versa.