Math 3081 (Probability and Statistics)

Lecture #13 of 27 ∼ July 27th, 2021

Applications of Poisson and Exponential Distributions

- Poisson Distributions and the Poisson Limit Theorem
- Memoryless Processes and Exponential Distributions
- Modeling With Normal, Poisson, and Exponential Distributions

This material represents §2.3.3-2.3.4 from the course notes, and problems 12-20 from WeBWorK 4. It is also the last of the material this week for Midterm 2.

Erstwhile....

In the last two lectures, we discussed the normal distribution, motivated its common appearance (namely, via the central limit theorem), and discussed some of its applications to answering various statistical questions.

The goal today is to discuss two other important random variable models that arise frequently: the Poisson distribution and the exponential distribution, and to explain the reasons for their utility as models.

Recall

We already mentioned the Poisson distribution at the end of last lecture:

Definition

The Poisson distribution with parameter $\lambda > 0$ is the discrete random variable X that takes the nonnegative integer value n with probability $P(X = n) = \dfrac{\lambda^n e^{-\lambda}}{n!}$.

Some other facts about the Poisson distribution:

- The peak occurs when n is the greatest integer less than or equal to λ (and when λ is exactly an integer, the peak is shared between $\lambda - 1$ and λ).
- The expected value is λ and the variance is also λ.

The Poisson Limit Theorem, I

The Poisson distribution arises in the analysis of systems having a large number of independent events each of which occurs rarely.

- More specifically, suppose we would like to model the probability distribution of how often a rare event will occur in a fixed time interval.
- We hypothesize that on average the event will occur λ times in the interval and we also assume that occurrences are independent (meaning that the occurrence of one event does not affect the probability that a second will occur).

The Poisson Limit Theorem, II

- We can approximate this situation by dividing the time interval into W possible "small windows" in which a rare event (occurring with probability $p = \lambda/W$) can either occur or not occur: we wish to find the probability distribution for the number of events that do occur.
- With this description, the probability distribution of this approximation will be the binomial distribution with W independent events and event probability $p = \lambda/W$, meaning that the probability of observing exactly n events is equal to $\binom{W}{n} p^n (1-p)^{W-n} = \binom{W}{n} (\lambda/W)^n \cdot (1 - \lambda/W)^{W-n}$.
- However, this is an only an approximation to the original problem: to find the answer to the original question, we need to take the limit as $W \to \infty$.

The Poisson Limit Theorem, III

Our main result is that taking the limit yields a Poisson distribution:

Theorem (Poisson Limit Theorem)

Suppose $\lambda > 0$ *is a fixed constant and* $p = \lambda / W$. *Then* $\lim_{W \to \infty} \binom{W}{n} p^n (1-p)^{W-n} = \frac{\lambda^n e^{-\lambda}}{n!}$.
Therefore, the probability distribution of the number of rare independent events occurring in a fixed interval, under the assumption that the average number of events per interval is λ, *is Poisson with parameter* λ.

This theorem is also often called the law of rare events since it describes the distribution of rare events.

The Poisson Limit Theorem, IV

Proof:

- We first rewrite the binomial probability:
$\binom{W}{n} p^n (1-p)^{W-n} = \frac{W(W-1)\cdots(W-n+1)}{n!} \cdot (\frac{\lambda}{W})^n (1-\frac{\lambda}{W})^{W-n}$
$= \frac{W(W-1)\cdots(W-n+1)}{W \cdot W \cdot W \cdot \cdots \cdot W} \cdot \frac{\lambda^n}{n!} \cdot (1-\frac{\lambda}{W})^W \cdot (1-\frac{\lambda}{W})^{-n}$.
- The first term $\frac{W(W-1)(W-2)\cdots(W-n+1)}{W \cdot W \cdot W \cdot \cdots \cdot W} = \frac{W}{W} \cdot \frac{W-1}{W} \cdots \frac{W-n+1}{W}$ has limit 1 as $W \to \infty$.
- The second term $\lambda^n/n!$ is a constant, so its limit is itself.
- The third term has limit $e^{-\lambda}$ by a standard application of L'Hôpital's rule.
- The last term has limit 1, since $(1-\lambda/W) \to 1$.
- Thus, the product has limit $1 \cdot \frac{\lambda^n}{n!} \cdot e^{-\lambda} \cdot 1 = \frac{\lambda^n e^{-\lambda}}{n!}$, as claimed.

The Poisson Limit Theorem, V

The Poisson limit theorem serves as a sort of complement to the central limit theorem for binomial distributions.

- The central limit theorem says that as $n \to \infty$, the binomial distribution tends to a normal distribution when np and $n(1-p)$ are moderately large.
- The Poisson limit theorem can be reinterpreted as saying that the binomial distribution tends to a Poisson distribution when $\lambda = np$ is small. (The equality $\lambda = np$ follows because both count the expected number of successes.)

The practical outcome of the Poisson limit theorem is that the Poisson distribution can be used to model the occurrences of independent rare events.

The Poisson Limit Theorem, VI

Examples (of quantities with a Poisson model):

- The number of soldiers killed by horse-kicks each year in the Prussian cavalry. (One of the first historical applications.)
- The number of telephone calls received by a customer service center.
- The number of mutations created on a DNA strand during replication.
- The number of customers arriving at a restaurant or shop.
- The number of insurance claims during a given month.
- The number of earthquakes during a given month.
- The number of goals scored by a hockey or soccer team during a game.
- The number of decay events observed in a radioactive sample with a long half-life.

Poisson Models, I

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

1. In the next hour, there are no calls.
2. In the next hour, there is exactly one call.
3. In the next two hours, there are no calls.
4. In the next two hours, there are at least 3 calls.
5. In the next 30 minutes, there is at least one call.

Poisson Models, I

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

1. In the next hour, there are no calls.
2. In the next hour, there is exactly one call.
3. In the next two hours, there are no calls.
4. In the next two hours, there are at least 3 calls.
5. In the next 30 minutes, there is at least one call.

- A Poisson model is reasonable for this problem, because calls are fairly rare (based on the average of 1.2 per hour) and they should be essentially independent of one another.

Poisson Models, II

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

1. In the next hour, there are no calls.

Poisson Models, II

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

1. In the next hour, there are no calls.
   - The given information says that the number of calls X in a one-hour window will have a Poisson distribution with parameter $\lambda = 1.2$.
   - Thus, the probability of having no calls is $P(X = 0) = \frac{1.2^0 e^{-1.2}}{0!} \approx 0.3012$.
2. In the next hour, there is exactly one call.

Poisson Models, II

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

1. In the next hour, there are no calls.
   - The given information says that the number of calls X in a one-hour window will have a Poisson distribution with parameter $\lambda = 1.2$.
   - Thus, the probability of having no calls is $P(X = 0) = \dfrac{1.2^0 e^{-1.2}}{0!} \approx 0.3012$.
2. In the next hour, there is exactly one call.
   - The probability of exactly one call is $P(X = 1) = \dfrac{1.2^1 e^{-1.2}}{1!} \approx 0.3614$.

Poisson Models, III

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

3. In the next two hours, there are no calls.

Poisson Models, III

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

3. In the next two hours, there are no calls.

- The number of calls Y in a two-hour window will also have a Poisson distribution, by the exact same logic.
- Since the average number of calls in 2 hours is $2 \cdot 1.2 = 2.4$, the corresponding parameter is $\lambda = 2.4$.
- Then $P(Y = 0) = \frac{2.4^0 e^{-2.4}}{0!} \approx 0.0907$.
- Remark: This event is the intersection of [no calls in hour 1] and [no calls in hour 2], which are independent and both have probability 0.3012. Indeed, $0.3012^2 = 0.0907$.

4. In the next two hours, there are at least 3 calls.

Poisson Models, III

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

3. In the next two hours, there are no calls.

- The number of calls Y in a two-hour window will also have a Poisson distribution, by the exact same logic.
- Since the average number of calls in 2 hours is $2 \cdot 1.2 = 2.4$, the corresponding parameter is $\lambda = 2.4$.
- Then $P(Y = 0) = \frac{2.4^0 e^{-2.4}}{0!} \approx 0.0907$.
- Remark: This event is the intersection of [no calls in hour 1] and [no calls in hour 2], which are independent and both have probability 0.3012. Indeed, $0.3012^2 = 0.0907$.

4. In the next two hours, there are at least 3 calls.

- The probability of at least 3 calls is $P(Y \geq 3) = 1 - P(Y \leq 2) = 1 - \frac{2.4^0 e^{-2.4}}{0!} - \frac{2.4^1 e^{-2.4}}{1!} - \frac{2.4^2 e^{-2.4}}{2!} \approx 0.4303$.

Poisson Models, IV

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

5. In the next 30 minutes, there is at least one call.

Poisson Models, IV

Example: At a call center, customer service calls come in at a rate of 1.2 per hour. Find the probabilities of the following events:

5. In the next 30 minutes, there is at least one call.

- The distribution of the number of calls Z in a 30-minute window will also have a Poisson distribution, but now with parameter $\lambda = 0.5 \cdot 1.2 = 0.6$.
- The probability of having no calls is therefore $P(Z = 0) = \dfrac{0.6^0 e^{-0.6}}{0!} \approx 0.5488$.
- So the probability of having at least one call is $1 - 0.5488 = 0.4512$.

Poisson Models, V

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

1. The probability of 0 patients between 2am and 3am today.
2. The probability of 10+ patients between 2am and 3am today.
3. The probability that a total of exactly 35 patients arrive between 2am and 3am this week.
4. The distribution of the total number of patients arriving between 2am and 3am over the full 366-day year.
5. The approximate probability of 2000+ total patients this year.
6. The approximate probability that at least 15 times this year, the hospital will see 10+ patients between 2am and 3am.
7. The approximate probability that at least twice this year, the hospital will see 0 patients between 2am and 3am.

Poisson Models, VI

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

1. The probability of 0 patients between 2am and 3am today.

Poisson Models, VI

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

1. The probability of 0 patients between 2am and 3am today.

- A Poisson model is appropriate here: the arrival of patients is fairly rare based on the given average of 5.3 per hour, and it is also reasonable to assume that the arrivals of patients at the emergency room are essentially independent of one another.
- The given information says that the number of patients for one day will have a Poisson distribution with parameter $\lambda = 5.3$.
- Thus, the probability of having no patients today is $e^{-5.3} \approx 0.00499$, which is about 0.5%.

Poisson Models, VII

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

2. The probability of 10+ patients between 2am and 3am today.

Poisson Models, VII

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

2. The probability of 10+ patients between 2am and 3am today.

- The number of patients has a Poisson distribution with parameter $\lambda = 5.3$.
- The easiest way to evaluate the tail sum of the Poisson distribution is to compute the probability of the complement.
- We see that $P(X \geq 10) = 1 - P(X < 10)$, and $P(X < 10) = P(X = 0) + P(X = 1) + \cdots + P(X = 9)$ $= \sum_{n=0}^{9} \frac{\lambda^n e^{-\lambda}}{n!} \approx 0.9559$.
- Thus, the probability $P(X \geq 10)$ of having 10+ patients today is $1 - 0.9559 = 0.0441$, about 4.41%.

Poisson Models, VIII

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

3. The probability that a total of exactly 35 patients arrive between 2am and 3am this week.

Poisson Models, VIII

<u>Example</u>: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

3. The probability that a total of exactly 35 patients arrive between 2am and 3am this week.

- The average number of patients for one week is $7 \cdot 5.3 = 37.1$.
- The distribution here will also be Poisson, with parameter $\lambda = 37.1$.
- The desired probability is then $P(P_{37.1} = 35) = \dfrac{37.1^{35} e^{-37.1}}{35!} \approx 0.0633$, about 6.33%.

Poisson Models, IX

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

4. The distribution of the total number of patients arriving between 2am and 3am over the full 366-day year.

Poisson Models, IX

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

4. The distribution of the total number of patients arriving between 2am and 3am over the full 366-day year.

- In principle, the total number of patients arriving over the full year will also be Poisson-distributed (with parameter $\lambda = 366 \cdot 5.3$), under the same logic as before.
- However, since we are taking such a large sample, we would also expect, by the central limit theorem, that the distribution should be approximately normal with mean $\mu = 366 \cdot 5.3$ and standard deviation $\sigma = \sqrt{366} \cdot \sqrt{5.3}$.
- Indeed, these two models are consistent, since the predicted means and standard deviations are the same.

Poisson Models, X

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

5. The approximate probability of 2000+ total patients this year.

Poisson Models, X

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

5. The approximate probability of 2000+ total patients this year.

- We can use either the Poisson model ($\lambda = 366 \cdot 5.3 = 1939.8$) or the normal model ($\mu = 366 \cdot 5.3 = 1939.8$, $\sigma = \sqrt{366} \cdot \sqrt{5.3} = 44.04$).
- The normal model is much easier to calculate with: using a continuity correction since the number of patients is discrete yields $P(\# \geq 2000) = P(N_{\mu,\sigma} > 1999.5) = P(N_{0,1} > 1.3555) = 0.0876$, about 8.76%.
- Using a computer to evaluate the Poisson sum yields $P(\# \geq 2000) = P(P_\lambda \geq 2000) = 0.0881$, about 8.81%.

Poisson Models, XI

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

6. The approximate probability that at least 15 times this year, the hospital will see 10+ patients between 2am and 3am.

Poisson Models, XI

<u>Example</u>: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

6. The approximate probability that at least 15 times this year, the hospital will see 10+ patients between 2am and 3am.

- We previously found that the probability of having 10+ patients on any given day is approximately 0.0441.
- Thus, the number of times that the hospital has 10+ patients during the year will be binomially distributed with $n = 366$ and $p = 0.0441$.
- In this case, since $np = 16.10$ and $5p/(1-p) = 108.38$, the normal approximation to the binomial should be fairly good.
- The mean is $\mu = np = 16.10$ with standard deviation $\sigma = \sqrt{np(1-p} = 3.9279$, so the approximation is $P(N_{\mu,\sigma} > 14.5) = P(N_{0,1} > -0.4177) = 0.6619$.

Poisson Models, XI

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

7. The approximate probability that at least twice this year, the hospital will see 0 patients between 2am and 3am.

- We previously found that the probability of having 0 patients on any given day is approximately 0.00499.

Poisson Models, XI

Example: Based on past history, a hospital determines that the average number of patients arriving at the emergency room between 2am and 3am is 5.3 per day. Find the following:

7. The approximate probability that at least twice this year, the hospital will see 0 patients between 2am and 3am.

- We previously found that the probability of having 0 patients on any given day is approximately 0.00499.
- Thus, the number of times that the hospital has 0 patients during the year will be binomially distributed with $n = 366$ and $p = 0.00499$.
- In this case, since $np = 1.8269$, the normal distribution is not a good approximation, but a Poisson distribution will be.
- The parameter is $\lambda = np = 1.8269$, so the approximation is $P(P_\lambda \geq 2) = 1 - P(P_\lambda = 0) - P(P_\lambda = 1) = 0.5450$.

Exponential Models, I

We now discuss processes modeled by exponential distributions. Recall the definition:

Definition

The exponential distribution with parameter $\lambda > 0$ is the continuous random variable with probability density function $p(x) = \lambda e^{-\lambda x}$ for $x \geq 0$, and is 0 for negative x.

- We found that the cdf is $c(x) = 1 - e^{-\lambda x}$ for $x \geq 0$, and that the expected value and standard deviation are both $1/\lambda$.

Exponential Models, II

The exponential distribution is used to model "memoryless" processes, as follows:

Definition

Suppose X is a continuous random variable measuring the waiting time for an event. We say that X is memoryless if X has the property that the subsequent waiting time is independent of the amount of time already waited.

Some examples of waiting times include the time until the failure of a piece of equipment, the time before the next customer arrives at a shop, and the time until a radioactive isotope decays.

Exponential Models, III

We can rephrase the memoryless condition in terms of the density function of the random variable:

- If a represents the total time already waited, and b represents the additional time before the event occurs, then the memoryless condition says that $P(X > a + b | X > a) = P(X > b)$ for every a and b.
- Equivalently, by the conditional probability formula, this means $P(X > a + b) = P(X > a) \cdot P(X > b)$, since the event $X > a + b$ includes the event $X > a$.
- But now, if X has an exponential distribution, then $P(X > a+b) = e^{-\lambda(a+b)} = e^{-\lambda a}e^{-\lambda b} = P(X > a) \cdot P(X > b)$.

Thus, exponentially-distributed random variables are memoryless.

Exponential Models, IV

In fact, the exponential distributions are the only memoryless continuous probability distributions:

Proposition (Memoryless Distributions)

If X is a memoryless continuous random variable, then in fact X has an exponential distribution.

Exponential Models, V

Proof:

- The condition $P(X > a + b) = P(X > a) \cdot P(X > b)$ implies that $P(X > 2) = P(X > 1)^2$, $P(X > 3) = P(X > 1)^3$, $P(X > 4) = P(X > 1)^4$, and so forth.
- By the same logic, we have $P(X > 1/n) = P(X > 1)^{1/n}$ for every integer n, so combining this reasoning with the argument above shows that $P(X > a) = P(X > 1)^a$ for every rational number $a > 0$.
- But since $P(X > a)$ is a nondecreasing function of a, this means in fact $P(X > a) = P(X > 1)^a$ for every real $a > 0$.
- Now writing $\lambda = -\ln[P(X > 1)]$ yields $P(X > a) = e^{-\lambda a}$, and so the cumulative distribution function agrees with that of the exponential distribution with parameter λ.
- This means X must be exponentially distributed with parameter λ, as claimed.

Exponential Models, VI

We can also analyze memoryless discrete random variables.

- In fact, by modifying the proof slightly, we can show without too much effort that the only memoryless discrete random variables are the geometric distributions, for which $P(G_p = n) = p^n(1-p)$ for integers $n = 0, 1, 2, \dots$, where p is a fixed parameter with $0 \le p < 1$.
- The geometric distributions are the discrete analogue of the exponential distributions.
- The geometric distribution given above arises as follows: if we repeatedly flip an unfair coin with probability p of landing tails until we obtain heads for the first time, then G_p counts the total number of tails obtained.
- Think for yourself why this is a memoryless waiting problem.

Exponential Models, VII

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution. Customer surveys indicate that 20% of the refrigerators must be repaired within their first year of operation.

1. Find the parameter λ for the distribution.
2. Find the percentage of refrigerators that will last at least 5 years without needing to be repaired.
3. Find the probability that a refrigerator will fail within 4 years, given that it did not fail in its first 3 years of operation.

Exponential Models, VIII

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution. Customer surveys indicate that 20% of the refrigerators must be repaired within their first year of operation.

1. Find the parameter λ for the distribution.

Exponential Models, VIII

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution. Customer surveys indicate that 20% of the refrigerators must be repaired within their first year of operation.

1. Find the parameter λ for the distribution.

- If X is the waiting time for repair, the given information says that $P(X < 1) = 0.20$.
- If the parameter is λ then since $P(X < 1) = 1 - e^{-\lambda}$ we see $1 - e^{-\lambda} = 0.20$.
- This yields $\lambda = -\ln(0.80) \approx 0.2231$.

Exponential Models, IX

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution with $\lambda = -\ln(0.8) = 0.2231$.

2. Find the percentage of refrigerators that will last at least 5 years without needing to be repaired.

Exponential Models, IX

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution with $\lambda = -\ln(0.8) = 0.2231$.

2. Find the percentage of refrigerators that will last at least 5 years without needing to be repaired.

- The proportion of refrigerators that will last at least 5 years is $P(X \geq 5) = e^{-5\lambda} = (0.80)^5 \approx 0.3277$, which is about 33%.
- Alternatively, we could use the memoryless property to calculate this probability.
- The given information says that 80% of refrigerators last one year without being repaired: thus, of these, 80% will last another year, while 80% of those will last a third year, and so forth.
- The overall proportion that will last 5 years is then $(0.80)^5$, exactly as above.

Exponential Models, X

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution with $\lambda = -\ln(0.8) = 0.2231$.

3. Find the probability that a refrigerator will fail within 4 years, given that it did not fail in its first 3 years of operation.

Exponential Models, X

Example: The usage time before a certain refrigerator model needs to be repaired is modeled as an exponential distribution with $\lambda = -\ln(0.8) = 0.2231$.

3. Find the probability that a refrigerator will fail within 4 years, given that it did not fail in its first 3 years of operation.

- This probability is $P(X \leq 4 | X > 3)$, which equals $P(X \leq 4 \cap X > 3)/P(X > 3) = P(3 < X \leq 4)/P(X > 3)$.
- We compute $P(3 < X \leq 4) = e^{-3\lambda} - e^{-4\lambda} = 0.1024$ and $P(X > 3) = e^{-3\lambda} = 0.512$.
- Thus, the desired probability is $0.1024/0.512 = 0.2$.
- Alternatively, using the memoryless property: if we "forget" the functional first three years of operation, we are now asking that the refrigerator functions for one more year, which occurs 20% of the time.

Exponential Models, XI

Example: You call an unreliable ride-share service to take you to the airport. You expect to wait 45 minutes on average, but feel that the total amount of time you have waited so far has no relationship to the amount of additional time you will have to wait.

1. Describe the distribution of the random variable measuring your waiting time.
2. Find the probability that you will have to wait longer than your expected average of 45 minutes.
3. Find the probability that the car actually does arrive within the next 5 minutes.
4. If you use this service 40 times a year, describe the distribution of the total number of times the car shows up within 5 minutes.
5. Find the Poisson and normal estimates to the probability that the car shows up within 5 minutes at least 6 times out of 40 uses. Which estimate is better?

Exponential Models, XII

Example: An unreliable taxi service is supposed to take you to the airport. You expect to wait 45 minutes on average, but feel that the total amount of time you have waited has no effect on the amount of additional time you will have to wait.

1. Describe the distribution of the random variable measuring your waiting time.

Exponential Models, XII

Example: An unreliable taxi service is supposed to take you to the airport. You expect to wait 45 minutes on average, but feel that the total amount of time you have waited has no effect on the amount of additional time you will have to wait.

1. Describe the distribution of the random variable measuring your waiting time.

- The given information is describing a memoryless waiting time. Thus, by our results, the waiting time will be exponentially distributed.
- Since the expected value is $1/\lambda$, we must have $\lambda = 1/45$.

2. Find the probability that you will have to wait longer than your expected average of 45 minutes.

Exponential Models, XII

Example: An unreliable taxi service is supposed to take you to the airport. You expect to wait 45 minutes on average, but feel that the total amount of time you have waited has no effect on the amount of additional time you will have to wait.

1. Describe the distribution of the random variable measuring your waiting time.

- The given information is describing a memoryless waiting time. Thus, by our results, the waiting time will be exponentially distributed.
- Since the expected value is $1/\lambda$, we must have $\lambda = 1/45$.

2. Find the probability that you will have to wait longer than your expected average of 45 minutes.

- This is $P(X > 45) = e^{-45 \cdot 1/45} = e^{-1} = 0.3679$.

Exponential Models, XIII

Example: An unreliable taxi service is supposed to take you to the airport. Your wait time is exponentially distributed with $\lambda = 1/45$.

3. Find the probability that the car actually does arrive within the next 5 minutes.

Exponential Models, XIII

Example: An unreliable taxi service is supposed to take you to the airport. Your wait time is exponentially distributed with $\lambda = 1/45$.

3. Find the probability that the car actually does arrive within the next 5 minutes.

- This is $P(X < 5) = 1 - e^{-5 \cdot 1/45} = 1 - e^{-1/9} = 0.1052$.

4. If you use this service 40 times a year, describe the distribution of the total number of times the car shows up within 5 minutes.

Exponential Models, XIII

Example: An unreliable taxi service is supposed to take you to the airport. Your wait time is exponentially distributed with $\lambda = 1/45$.

3. Find the probability that the car actually does arrive within the next 5 minutes.

- This is $P(X < 5) = 1 - e^{-5 \cdot 1/45} = 1 - e^{-1/9} = 0.1052$.

4. If you use this service 40 times a year, describe the distribution of the total number of times the car shows up within 5 minutes.

- As we just calculated, the probability that the car shows up within 5 minutes is $1 - e^{-1/9} = 0.1052$.
- If the service is used 40 times, assuming that the individual uses are independent, then the exact distribution will be binomial with parameters $n = 40$ and $p = 0.1052$.

Exponential Models, XIV

Example: An unreliable taxi service is supposed to take you to the airport. Your wait time is exponentially distributed with $\lambda = 1/45$.

5. Find the Poisson and normal estimates to the probability that the car shows up within 5 minutes at least 6 times out of 40 uses. Which estimate is better?

- Since $n = 40$ and $p = 0.1052$

Exponential Models, XIV

Example: An unreliable taxi service is supposed to take you to the airport. Your wait time is exponentially distributed with $\lambda = 1/45$.

5. Find the Poisson and normal estimates to the probability that the car shows up within 5 minutes at least 6 times out of 40 uses. Which estimate is better?

- Since $n = 40$ and $p = 0.1052$, for the Poisson estimate we have $\lambda = np = 4.2064$. We then compute $P(P_\lambda \geq 6) = 1 - P(P_\lambda < 6) = 0.2479$.
- For the normal estimate we have $\mu = np = 4.2064$ and $\sigma = \sqrt{np(1-p)} = 1.9404$. We then compute $P(N_{\mu,\sigma} > 5.5) = P(N_{0,1} > 0.6658) = 0.2528$.
- The actual value is 0.2406.
- Since $np = 4.2064$ and n is moderately large, we are in the range where we would expect the Poisson approximation to be better (which it is, though not by a whole lot).

Exponential and Poisson Randomness, I

There is a connection between the Poisson distribution and the exponential distribution, arising from our interpretations of the processes they model.

- The Poisson distribution models the number of occurrences of independently-occurring rare events in a particular interval of time, while the exponential distribution models the waiting time for a memoryless process.
- Now suppose we have a Poisson-distributed phenomenon, and we ask: how long do we have to wait between two occurrences of the phenomenon?
- Because the Poisson events are independent and rare, the occurrence of one does not affect the waiting time for the next one. Since this precisely describes a memoryless process, the distribution of waiting times between Poisson events will have an exponential distribution.

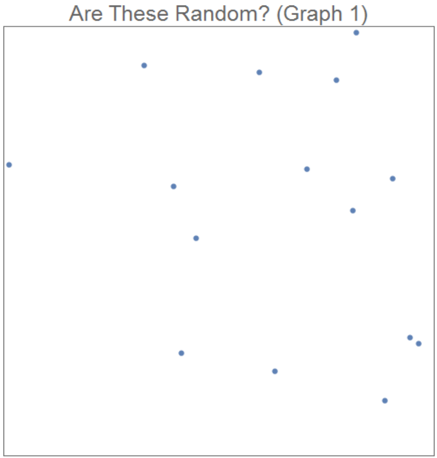
Exponential and Poisson Randomness, II

The fact that waiting times between Poisson events have an exponential distribution leads to some unintuitive results.

- For example, the exponential distribution decreases rapidly, starting from 0.
- This fact tells us that the distances between Poisson events are more likely to be "small" rather than "big": if the average distance is D (so the exponential parameter is $1/D$), the probability of obtaining a distance less than the average is $1 - e^{-1/D \cdot D} = 1 - e^{-1} \approx 0.6321$.
- Thus, despite the fact that the Poisson events will be uniformly distributed inside the time interval (since they are, after all, independent and occur randomly), it is nonetheless likely that we will observe "clusters" of occurrences.
- This merely reflects the general fact that randomly-occurring events will still tend to appear in clusters.

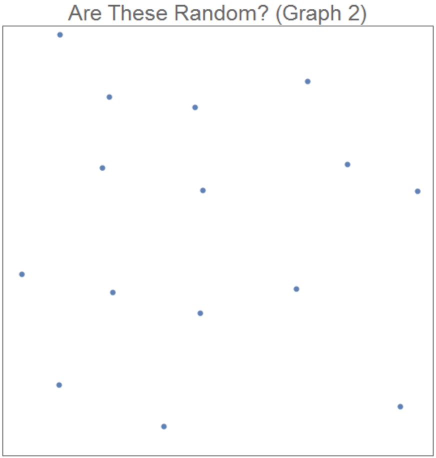
Exponential and Poisson Randomness, III

As an illustration, each of these plots has 15 points. Which one(s) were randomly generated and which ones were not?



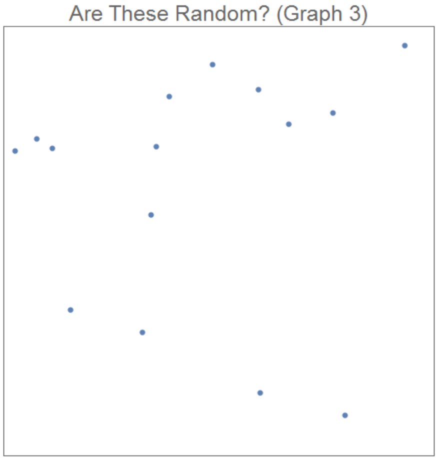
Exponential and Poisson Randomness, III

As an illustration, each of these plots has 15 points. Which one(s) were randomly generated and which ones were not?



Exponential and Poisson Randomness, III

As an illustration, each of these plots has 15 points. Which one(s) were randomly generated and which ones were not?



More Models, I

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

1. Describe the distribution of the random variable X measuring the total number of typos in one week.
2. What is the probability there are exactly 5 typos this week?
3. What is the probability there are no typos in today's lecture?
4. Describe the distribution of the random variable Y measuring the total amount of time before the next typo is made.
5. After one typo, what is the probability that at least 4 full lectures pass before another typo is made?
6. Estimate the probability of obtaining more than 180 typos if the course runs for 52 weeks.

More Models, II

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

1. Describe the distribution of the random variable X measuring the total number of typos in one week.

More Models, II

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

1. Describe the distribution of the random variable X measuring the total number of typos in one week.

- From the description, typos are a relatively rare event, and it is reasonable to assume that they are independent.
- So, the distribution of typos will be Poisson, and the parameter will be the average number of typos per week, which is $\lambda = 3$.

2. What is the probability there are exactly 5 typos this week?

More Models, II

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

1. Describe the distribution of the random variable X measuring the total number of typos in one week.

- From the description, typos are a relatively rare event, and it is reasonable to assume that they are independent.
- So, the distribution of typos will be Poisson, and the parameter will be the average number of typos per week, which is $\lambda = 3$.

2. What is the probability there are exactly 5 typos this week?

- This is $P(P_\lambda = 5) = \dfrac{3^5 e^{-3}}{5!} \approx 0.1008$.

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

3. What is the probability there are no typos in today's lecture?

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

3. What is the probability there are no typos in today's lecture?

- The average number of typos in one lecture is $3/4$, so the number of typos in today's lecture will be Poisson-distributed with parameter $\lambda = 3/4$.
- Then the probability of no typos is $e^{-3/4} \approx 0.4724$.[1]

4. Describe the distribution of the random variable Y measuring the total time, in weeks, before the next typo is made.

[1] Since their arr sevral typoes in thiss footnorte, the actua; probability is 0.

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

3. What is the probability there are no typos in today's lecture?

- The average number of typos in one lecture is 3/4, so the number of typos in today's lecture will be Poisson-distributed with parameter $\lambda = 3/4$.
- Then the probability of no typos is $e^{-3/4} \approx 0.4724$.[1]

4. Describe the distribution of the random variable Y measuring the total time, in weeks, before the next typo is made.

- The waiting time will be exponential because the probability of obtaining a typo is independent of the amount of time since the last typo.
- The average time between typos is 3/4 of a week, so $\lambda = 3/4$.

[1] Since their arr sevral typoes in thiss footnorte, the actua; probability is 0.

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

5. After one typo, what is the probability that at least 4 full lectures pass before another typo is made?

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

5. After one typo, what is the probability that at least 4 full lectures pass before another typo is made?

- Since the waiting time is exponential, this is $P(E_\lambda \geq 4) = e^{-4 \cdot 3/4} = e^{-3} \approx 0.0498$.

6. Estimate the probability of obtaining more than 180 typos if the course runs for 52 weeks.

More Models, III

On average, a certain Math 3081 instructor makes 3 typos per week (there are 4 lectures per week).

5. After one typo, what is the probability that at least 4 full lectures pass before another typo is made?

- Since the waiting time is exponential, this is $P(E_\lambda \geq 4) = e^{-4 \cdot 3/4} = e^{-3} \approx 0.0498$.

6. Estimate the probability of obtaining more than 180 typos if the course runs for 52 weeks.

- The exact distribution is Poisson with $\lambda = 52 \cdot 3 = 156$, but it is very cumbersome to evaluate the exact probability this way.
- Instead, by the central limit theorem, we can observe that the distribution is approximately normal with mean $\mu = 52\lambda = 156$ and standard deviation $\sqrt{52\lambda} \approx 12.4900$.
- Including a continuity correction, the approximate probability is $P(N_{156,12.4900} \geq 179.5) = P(N_{0,1} \geq 1.8815) = 0.0300$.

Summary

We proved the Poisson limit theorem and discussed examples of phenomena that have Poisson models.

We discussed how exponential distributions model memoryless processes.

We did some additional examples of modeling using normal, Poisson, and exponential models.

Next lecture: Maximum likelihood estimates