Solutions to CS 70 Challenge Problems: Distributions

Other worksheets and solutions at https://alextseng.net/teaching/cs70/

Alex Tseng

1 Distributions to Use

Find the distribution (discrete or continuous) that is most appropriate for each of these scenarios. Then, answer the question using the specified distribution.

Possible distributions are uniform, binomial, Poisson, geometric, exponential, and Gaussian

(a) Every day, 1000 people play the lottery. Each person buys a ticket with randomized numbers. The chance of winning is $\frac{1}{10000}$. What is the probability that fewer than 3 people win?
Binomial, $(\frac{9999}{10000})^{1000} + 1000(\frac{1}{10000})(\frac{9999}{10000})^{999} + \binom{1000}{2}(\frac{1}{10000})^2(\frac{9999}{10000})^{998}$
There are discrete trials (of playing the lottery), each with an independent and equivalent chance of succeeding ($\frac{1}{10000}$). We are interested in number of successes. This is clearly binomial. The probability that fewer than 3 people win is the probability that 0, 1, or 2 people win: $P(X=0)+P(X=1)+P(X=2) = (\frac{9999}{10000})^{1000} + 1000(\frac{1}{10000})(\frac{9999}{10000})^{999} + \binom{1000}{2}(\frac{1}{10000})^2(\frac{9999}{10000})^{998} \approx 0.99985$

(b) Due to human impact on climate and habitats, fewer and fewer Roan antelope can be seen in South Africa's Kruger National Park. If there are an average of 2 sightings of Roan antelope per day, what is the probability that there are 7 sightings in a week?
Poisson, $2e^{-2}$
We are given the average rate of how many times an event happens over a continuous time scale, and we are interested in the number of times the event happens, so this distribution is best described by a Poisson distribution. Here, $\lambda = 2$ per day. 7 sightings in a week is equivalent to 1 sighting per day. For a Poisson distribution, $P(X=1) = \frac{\lambda^1 e^{-\lambda}}{1!} = 2e^{-2} \approx 0.27067$

(c) A card is drawn at random from a 52-card deck. What is the probability that the card drawn is a face card (i.e. jack, queen, or king)?
Uniform, $\frac{3}{13}$
There is equivalent chance of the card being drawn as anything, so the distribution is completely uniform. We are interested in the probability that our card is one of 3 outcomes out of the possible 13. Thus, this probability is $\frac{3}{13}$.

(d) On average, there are 2 car crashes per hour in the county. How many minutes are expected until the next car crash?
Exponential, 30 minutes
An average is given, and we wish to know the time (which is a continuous domain) until the next event. Recall that the exponential and Poisson distributions are closely related. They both deal with rare events over a continuous domain. We are given the expectation of the Poisson event (a car crash), which has parameter $\lambda = 2$. However, we are asked about the time until the first occurrence of such a Poisson event (rather than the number of events in a certain time period), so we are looking for the expectation of the corresponding exponential distribution. The expectation of an exponential distribution is $\frac{1}{\lambda} = \frac{1}{2}$, or half an hour (30 minutes).

(e) The average score on a recent exam was 70%. Only 10 students scored above 90%. How many students are expected to have scored below 50%?
Gaussian, 10
Exam scores (at least ideally) are distributed Normally. Gaussian distributions are symmetric about the mean, which in this case is 70. We know the number of students who scored 20 points above the mean, so the number of students who scored 20 points below the mean are expected to be the same.

(f) Steven the basketball player is trying to make 3-point shots. The probability he can make a 3-point shot is always $\frac{1}{10}$. If each shot takes 20 seconds total, how long do we expect to wait until he finally

successfully shoots one into the basket?
Geometric, 200 seconds
There are discrete trials, where there is an identical and independent chance of a success (making a basket). We are not interested in the number of baskets made, but the number of trials until the first basket is made. Thus, we turn to the geometric distribution. Here, the only parameter needed is the probability of success in each trial: $\frac{1}{10}$. We know the expectation of the geometric distribution is the reciprocal of this probability, so we expect to wait 10 trials, or 200 seconds.

2 Combining Distributions

(a) A company's server uses 2 hard drives—one for normal use and the other for backup. Both hard drives are identical, and each is expected to fail in 4 years. As soon as the first hard drive fails, the company will start using the second one (which had not been used until now). What is the expected time until both hard drives fail? What is the variance?
Expectation: 8 years, variance: 32 years
Let D_1 and D_2 be the time until the first drive and the backup drive fail, respectively. D_1 and D_2 are both exponential random variables, since we are interested in the time until an event happens (an event whose average we know). Here, the event has an average rate of $\lambda = \frac{1}{4}$ per year (equivalent to 1 per 4 years).
The time it takes until both drives fail is simply $D_1 + D_2$, since we use one after the other.
$\text{E}[D_1 + D_2] = \text{E}[D_1] + \text{E}[D_2] = \frac{1}{\lambda} + \frac{1}{\lambda} = 4 + 4 = 8$ years.
To find $\text{Var}[D_1 + D_2]$, we use the fact that D_1 and D_2 are independent—the drives fail independently. Thus, $\text{Var}[D_1 + D_2] = \text{Var}[D_1] + \text{Var}[D_2] = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = 16 + 16 = 32$ years.

(b) *Challenge* Given two independent Poisson random variables X and Y with means λ_x and λ_y, respectively, show that $X + Y$ is also a Poisson random variable. What is its mean?
Let $Z = X + Y$. In order to show that Z has a Poisson distribution, we just need to show that $P(Z = k) = \frac{\lambda_z^k e^{-\lambda_z}}{k!}$ for some λ_z.
For Z to be equal to k, we need the values of X and Y to sum to k. It is important to note that X and Y, since they have a Poisson distribution, are non-negative.
So we start with the equality of $P(Z = k) = \sum_{i=0}^{k} P(X = i, Y = k - i)$ (the probability that Z is k is the sum of probabilities of all the ways that X and Y sum to k).
$P(Z = k) = \sum_{i=0}^{k} P(X = i, Y = k - i) = \sum_{i=0}^{k} P(X = i)P(Y = k - i)$ by the independence of X and Y
$= \sum_{i=0}^{k} \frac{\lambda_x^i e^{-\lambda_x}}{i!} \frac{\lambda_y^{k-i} e^{-\lambda_y}}{(k-i)!}$ by the definition of the Poisson distribution
$= \sum_{i=0}^{k} \frac{1}{i!(k-i)!} \lambda_x^i \lambda_y^{k-i} e^{-(\lambda_x + \lambda_y)}$ through some rearranging
$= e^{-(\lambda_x + \lambda_y)} \sum_{i=0}^{k} \frac{1}{i!(k-i)!} \lambda_x^i \lambda_y^{k-i}$ after pulling out terms that don't depend on i
$= \frac{e^{-(\lambda_x + \lambda_y)}}{k!} \sum_{i=0}^{k} \frac{k!}{i!(k-i)!} \lambda_x^i \lambda_y^{k-i}$ multiply by $\frac{k!}{k!}$
$= \frac{e^{-(\lambda_x + \lambda_y)}}{k!} \sum_{i=0}^{k} \binom{k}{i} \lambda_x^i \lambda_y^{k-i}$ using the definition of the binomial coefficient
$= \frac{e^{-(\lambda_x + \lambda_y)}}{k!} (\lambda_x + \lambda_y)^k$ by the binomial theorem (think polynomial multiplication)
Here, we see $P(Z = k)$ of the desired form, proving that Z has a Poisson distribution, where the mean $\lambda_z = \lambda_x + \lambda_y$.

3 Central Limit Theorem

z-score table:

z	$P(X \leq z)$ where $X \sim N(0,1)$
2.5	0.9938
3.5	0.9999

(a) You buy 100 identical lightbulbs from the hardware store. The manufacturer of these lightbulbs promises that on average, they have a lifespan of 2 years. Assume that these 100 lightbulbs are a random sample from the factory. You start using these 100 lightbulbs and record how long it takes each one to fail. Let T be the average time it took to fail (over all 100 lightbulbs). What is the probability that $T \leq 1.5$? Use the provided z-score table.

0.0062

The lightbulbs follow an exponential distribution. The expected time until they die is 2 years. Here, $\lambda = \frac{1}{2}$ (recall, the expectation of an exponential distribution is the reciprocal of the rate λ). That makes the variance of the lightbulbs $\frac{1}{\lambda^2} = 4$. Let the mean of each individual lightbulb be $\mu = 2$, and let the variance be $\sigma^2 = 4$.

By the central limit theorem, T is normally distributed with mean that is equal to μ and variance equal to $\frac{\sigma^2}{n}$, where $n = 100$. Thus, the standard deviation of T is $\frac{\sigma}{10} = \frac{2}{10} = 0.2$

In order to query probabilities from a normal distribution, we use the z-score. For $T \leq 1.6$, we have a z-score of $\frac{1.5-2}{0.2} = -2.5$. Our z-score table only has positive z-scores, but since Guassian distributions are symmetric about the mean, we can simply flip the sign of the z-score and subtract our probability from 1. Thus, we get $P(T \leq 1.5) = 1 - 0.9938 = 0.0062$

(b) Hoping for better performance, you want the probability that $T \leq 1.5$ to be smaller. How many lightbulbs would you need to buy in order for T, the average lifespan, to be no more than 0.0001?

196 lightbulbs

Now, we go backwards. We want to ensure that $P(T \leq 1.5) \leq 0.0001$. From the table above, we see that to do this, we need a z-score that is at least as negative as -3.5. Again, draw out the distribution and shade in the area if the table is confusing.

Here, we wish to solve for a sufficient value of n: $\frac{1.5-2}{\frac{\sigma}{\sqrt{n}}} \leq -3.5$ where $\sigma = 2$

$\frac{1.5-2}{\frac{2}{\sqrt{n}}} \leq -3.5 \Rightarrow 0.5 \geq 3.5\frac{2}{\sqrt{n}} \Rightarrow \sqrt{n} \geq 14 \Rightarrow n \geq 196$