CS 70 Discrete Mathematics and Probability Theory
Spring 2015 Vazirani HW 13

Reminder: Deadline for HW12 self-grade is Thursday, April 23 at noon.

Reading: Note 19, 20.

Due Monday April 27

1. **Distributions**

   What distribution would best model each of the following scenarios? Choose from Poisson, geometric, exponential, uniform, or normal distribution. Provide a brief justification of your answer. (It might help to recall the interpretations of the distributions above as repeated coin tossing processes.)

   (a) Number of errors while transmitting a message of n characters.

   (b) Amount of time you need to wait until a fly enters through your window.

   (c) Number of shooting stars seen on a given night.

   (d) Average height of students in CS70.

   (e) Angle between two needles when we spin them at random about a common point.

   (f) Number of trees in a plot of 100 square meters in the forest.

   (g) Number of times you have to refresh a webpage until it loads.

   (h) Average grade in a CS70 exam.

   (i) Number of times a web server is accessed per minute.

   (j) Amount of time until the next time your telephone rings.

   Answer:

   (a) Either Poisson or the normal distribution. It is reasonable to assume that errors happen independently (this would be false for example if we lived in a world where messages tend to get corrupted in runs). Then for each character we are flipping a biased coin to decide whether it gets corrupted or not. Therefore the number of errors can be modeled as a binomial distribution (the number of heads in n coin flips). When n gets large enough, depending on the probability of error for each character, this distribution gets close to either the Poisson distribution or the normal distribution.

   (b) Exponential distribution. A reasonable argument is that the amount of time until a fly enters through the window is a memoryless variable. If we wait 5 minutes and see no fly, the distribution on the amount of time we still have to wait should not change. Therefore the exponential distribution is the correct answer here.

   (c) Poisson distribution. Moments in time where we see a shooting star form what is called a Poisson process. Very reasonably one can argue that whether or not we see a shooting star at any sufficiently small interval of time (like 1 second) is akin to a coin toss with very small probability

of success. So if we divide the night into many many small intervals, the number of shooting stars becomes like a binomial distribution with a large number of trials and a small probability of success for each trial. And we know that these tend to the Poisson process as the mean is kept constant. The mean here is just the expected number of shooting stars, which does not depend on the number of intervals we break the night into.

(d) Normal distribution. The average height of students is an average of many many (turn your head around the class if you do not believe this) random variables. If the height of any particular student comes from a distribution $\mathscr{D}$, then we are taking the average of many independent and identically distributed random variables. The central limit theorem states that this average becomes more and more like a normal random variable as the number of students grows. It is noteworthy that even the height of a single student $\mathscr{D}$ very closely follows a normal distribution likely because of the fact that many independent factors affect the height (genetics, nutrition, etc.).

(e) Uniform distribution. Let us assume that the first needle stops parallel to the x-axis. Then the second needle can stop at any angle with equal probability. Therefore the angle it makes with the x-axis would follow a uniform distribution. This argument would not change if the first needle was stopped at any other angle, and the uniform distribution would remain the same. Therefore the angle between the two needles follows the uniform distribution.

(f) Poisson distribution. If we divide the land into very small regions of equal surface area, then whether or not in each of those regions lies a tree becomes like a coin flip (since two trees are very unlikely to grow in the same tiny region). The average number of trees is fixed while the regions become smaller and smaller. Therefore we have binomial distributions with fixed means, and they tend towards the Poisson distribution.

(g) Geometric distribution. Each time we refresh, whether or not the webpage loads can be thought of as a coin flip. The number of times we flip the coin until we see heads (a page load) is the definition of the geometric distribution.

(h) Normal distribution. It's fair to assume that the grade of each student follows some distribution $\mathscr{D}$ fairly independently of the other students. Therefore when we take the average of many of these grades, according to the central limit theorem, we get closer and closer to the normal distribution.

(i) Poisson distribution. We can divide the time into minuscule intervals and see whether the web server was accessed in each interval. If the intervals become short enough, the probability of more than one access in one interval becomes negligible. Now in each interval whether or not the server was accessed can be thought of as a coin flip. Therefore we have a sum of coin flips (i.e. the binomial distribution) whose mean is fixed. As the intervals become shorter and shorter, the binomial distribution becomes more and more like a Poisson distribution.

(j) Exponential distribution. The amount of time until the telephone rings is a memoryless property. Knowing that no one has called yet, should not change the distribution on the amount of time we still have to wait. We can also reasonably model this using the following method: divide the time horizon into tiny intervals and in each interval flip a biased coin to see whether the phone rings. As the intervals become shorter, the probability of each coin flip has to become smaller too, to maintain the average rate of telephone calls per unit of time. This is exactly how the exponential distribution is defined.

2. **Poisson**

(a) It is fairly reasonable to model the number of customers entering a shop during a particular hour as a Poisson random variable. Assume that this Poisson random variable X has mean λ. Suppose that whenever a customer enters the shop they leave the shop without buying anything with probability p. Assume that customers act independently, i.e. you can assume that they each simply flip a biased coin to decide whether to buy anything at all. Let us denote the number of customers that buy something as Y and the number of them that do not buy anything as Z (so $X = Y + Z$). What is the probability that $Y = k$ for a given k? How about $\Pr[Z = k]$? Prove that Y and Z are Poisson random variables themselves.
Hint: you can use the identity $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$.

(b) Assume that you were given two independent Poisson random variables X_1, X_2. Assume that the first has mean λ_1 and the second has mean λ_2. Prove that $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$.

Answer:

(a) We consider all possible ways that the event $Y = k$ might happen: namely, $k + j$ people enter the shop ($X = k + j$) and then exactly k of them choose to buy something. That is,

$$\begin{aligned}
\Pr[Y = k] &= \sum_{j=0}^{\infty} \Pr[X = k + j] \cdot \Pr[Y = k \mid X = k + j] \\
&= \sum_{j=0}^{\infty} \left(\frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \right) \cdot \left(\binom{k+j}{k} p^j (1-p)^k \right) \\
&= \sum_{j=0}^{\infty} \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \cdot \frac{(k+j)!}{k!\, j!} p^j (1-p)^k \\
&= \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot \sum_{j=0}^{\infty} \frac{(\lambda p)^j}{j!} \\
&= \frac{(\lambda(1-p))^k e^{-\lambda}}{k!} \cdot e^{\lambda p} \\
&= \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!}
\end{aligned}$$

Hence, Y follows the Poisson distribution with parameter $\lambda(1-p)$. The case for Z is completely analogous: $\Pr[Z = k] = \frac{(\lambda p)^k e^{-\lambda p}}{k!}$ and Z follows the Poisson distribution with parameter λp.

(b) Let X, Y, and Z be defined as in part (a), with $\lambda = \lambda_1 + \lambda_2$ and $p = \frac{\lambda_1}{\lambda_1 + \lambda_2}$. Then Y and X_1 follow the Poisson distribution with parameter λ_1, and Z and X_2 follow the Poisson distribution with parameter λ_2. Moreover, Y and Z are independent because

$$\begin{aligned}
\Pr[Y = k, Z = j] &= \Pr[X = k + j] \cdot \Pr[Y = k \mid X = k + j] \\
&= \frac{\lambda^{k+j}}{(k+j)!} e^{-\lambda} \cdot \frac{(k+j)!}{k!\, j!} p^j (1-p)^k \\
&= \frac{(\lambda(1-p))^k e^{-\lambda(1-p)}}{k!} \cdot \frac{(\lambda p)^j e^{-\lambda p}}{j!} \\
&= \Pr[Y = k] \cdot \Pr[Z = j].
\end{aligned}$$

Hence, $Y+Z$ must have the same distribution as X_1+X_2. We complete the proof by observing that $X = Y+Z$ is a Poisson random variable with mean $\lambda = \lambda_1+\lambda_2$.

3. **Proof or Counterexample**

For each of the following statements determine whether it is true or false and justify your answer (either by a proof or a counter-example).

(a) Given a non-negative continuous random variable X the following holds:

$$\mathrm{E}[X] = \int_0^\infty \Pr[X \geq t]\, dt.$$

Note: this is just the continuous analogue of a similar statement about discrete random variables. One approach is to try the same proof as in the discrete case with summations replaced by integrals.

(b) There is a finite number that bounds *all* density functions $f(x)$ of continuous random variables, i.e., there exists a constant $0 < C < \infty$ such that for any probability density function f, $f(x) \leq C$ for all x.

Answer:

(a) True. Let $f(x)$ denote the density function for X, and let us mimic the proof of Theorem 19.1.

$$\begin{aligned}
\mathrm{E}[X] &= \int_0^\infty f(x)x\, dx \\
&= \int_0^\infty f(x) \int_0^x 1\, dt\, dx \\
&= \int_0^\infty \int_0^x f(x)\, dt\, dx \\
&= \int_0^\infty \int_0^\infty \mathbf{1}[t \leq x] f(x)\, dt\, dx \\
&= \int_0^\infty \int_0^\infty \mathbf{1}[t \leq x] f(x)\, dx\, dt \\
&= \int_0^\infty \int_t^\infty f(x)\, dx\, dt \\
&= \int_0^\infty \Pr[X \geq t]\, dt
\end{aligned}$$

(Note: $\mathbf{1}[t \leq x]$ denotes the function that is 1 if $t \leq x$ and 0 otherwise.)

(b) False. Suppose there exists such $C > 0$. If $f(x)$ is the exponential density function with parameter $2C$, $f(0) = 2C > C$. Contradiction.

4. **Exponential Distribution is Memoryless**

Let random variable X have exponential distribution with parameter λ. Show that, for any positive s,t we have $\Pr[X > s+t \mid X > t] = \Pr[X > s]$.

Answer: By definition of exponential distribution,

$$\Pr[X > a] = 1 - \Pr[X \leq a] = 1 - \int_0^a \lambda e^{-\lambda x} dx = 1 - \left[(-e^{-\lambda x})\right]_0^a = e^{-\lambda a}.$$

Hence,

$$\Pr[X > s+t \mid X > t] = \frac{\Pr[X > s+t \cap X > t]}{\Pr[X > t]} = \frac{\Pr[X > s+t]}{\Pr[X > t]} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} = e^{-\lambda s} = \Pr[X > s].$$

5. **Exponential**

There are two rare events that have the potential to cancel classes. One of them is a power outage and the other one is a fire on campus. Let's model the length of the time from now until the next power outage as a random variable X and the length of the time until the next fire on campus as Y (they are both measured in the number of days, but can be fractional, i.e. it might take 0.3 days until the next power outage). For the sake of simplicity assume that X and Y are independent (in real life, they really are not independent). It is reasonable to model X and Y as exponential random variables. From historic measurements you conclude that the mean of X is 100, and the mean of Y is 50.

(a) For any t what is $\Pr[X \geq t]$ and what is $\Pr[Y \geq t]$?

(b) You really only care about the next power outage or fire (either one will cancel classes). The time from now until the next earliest such event is $\min(X,Y)$. Let us call this random variable Z. Prove that Z is itself an exponential random variable. What is the expected value of Z?
Hint: What is $\Pr[Z \geq t]$? You can simplify this expression using the binomial theorem.

Answer:

(a) If A is exponentially distributed with mean $\frac{1}{\lambda}$,

$$\Pr[A \geq t] = \int_t^\infty \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x}\right]_t^\infty = e^{-\lambda t}$$

Hence, $\Pr[X \geq t] = e^{-t/100}$ and $\Pr[Y \geq t] = e^{-t/50}$.

(b) Note that $Z \geq t$ if and only if $X \geq t$ and $Y \geq t$. Since X and Y are independent,

$$\Pr[Z \geq t] = \Pr[X \geq t \cap Y \geq t] = \Pr[X \geq t] \cdot \Pr[Y \geq t] = e^{-t/100} \cdot e^{-t/50} = e^{-t \cdot (3/100)}$$

Hence, Z is an exponential random variable with mean $100/3$.

6. **Sum of Independent Normal Distributions**

One of the important properties of the normal distribution is that the sum of two independent normal random variables is also normal. In the note we have seen a geometric proof of this fact. Here we present a different argument based on the central limit theorem.

(a) Let X be a binomial random variable with parameter n and p. Recall that X can be written as the sum of n independent coin flips with bias p. Use central limit theorem to argue that the distribution of $\hat{X} = (X - np)/\sqrt{np(1-p)}$ converges to the standard normal distribution as $n \to \infty$.

(b) Let X and Y be independent binomial random variables with the same parameter n and p, and let $W = X + Y$. Note that W can be written as the sum of $2n$ independent coin flips with bias p. Let $\hat{W} = (W - 2np)/\sqrt{2np(1-p)}$, $\hat{X} = (X - np)/\sqrt{np(1-p)}$, and $\hat{Y} = (Y - np)/\sqrt{np(1-p)}$. Write $\hat{W}$ in terms of $\hat{X}$ and $\hat{Y}$, and use part (a) to argue that the sum of two independent standard normal distribution is also normal. What is the resulting mean and variance?

Answer:

(a) The expectation and variance of a single coin flip with bias p is p and $p(1-p)$ respectively. Then the central limit theorem states that $(X-np)/\sqrt{np(1-p)}$ converges to the standard normal distribution as $n \to \infty$, as desired.

(b) With $\hat{X}$, $\hat{Y}$, and $\hat{Z}$ defined as in the question,

$$\hat{X}+\hat{Y} = \frac{X-np}{\sqrt{np(1-p)}} + \frac{Y-np}{\sqrt{np(1-p)}} = \frac{X+Y-2np}{\sqrt{np(1-p)}} = \frac{W-2np}{\sqrt{np(1-p)}} = \sqrt{2}\hat{W}.$$

By part (a), each of $\hat{X}$, $\hat{Y}$, and $\hat{W}$ converges to the standard normal distribution as $n \to \infty$. By Lemma 20.1, this means that $\sqrt{2}\hat{W}$ converges to the normal distribution with mean 0 and variance 2. Hence, the above equality shows that the sum of two independent standard normal variables is also normal, with mean 0 and variance 2.

7. **Extra Credit: Baseball Cards**

Suppose you are collecting baseball cards with n different types of cards in all, using the same process as in the coupon collector problem, i.e., you collect one card at random sequentially until you get all n cards. What is the probability that when you are finally done collecting all the cards, you have exactly one card for Barry Bonds?

Hint: Think about the event that you got the Bonds card after collecting j other players' cards.

Answer: Suppose that we are collecting cards, and we keep a set S which contains the names of the baseball players we have so far collected. Now let A_j be the event that the j-th player whose name enters S is Bonds.

First of all let us prove that $\Pr[A_j] = \frac{1}{n}$. For a fixed j, consider the following events: the j-th player whose name enters S is x, for x being one of the n different players. These are n events which partition the space. But by symmetry they all have the same probability. Therefore the probability of each one is $\frac{1}{n}$ (since the probabilities have to add up to 1).

Now if we name the event we are interested in (having only one Bonds card after we finish) as B, then to compute the probability of B we can use the following formula:

$$\Pr[B] = \sum_{j=1}^{n} \Pr[B \mid A_j] \Pr[A_j] = \frac{1}{n} \sum_{j=1}^{n} \Pr[B \mid A_j].$$

So we just need to compute the conditional probabilities $\Pr[B \mid A_j]$. Assume that we have j players' names (including Bonds which was the last card we collected before reaching j). Now we have exactly one card for Bonds, one or more cards for $j-1$ other players and 0 cards for $n-j$ players. We can safely ignore the cards we collect for the $j-1$ other players because they are neither Bonds cards nor do they affect how soon we finish. So let us just look at the cards we collect for the $n-j+1$ players from now on, one of whom is Bonds.

Now assume that instead of stopping the process after we have a card for everyone, we stop the process when we have one card for everyone and at least two cards for Bonds. Now the event B becomes the event that the last card we collect before stopping is Bonds. It is easy to see that if the last card we collect before stopping is Bonds, it means that before that we only had one card for Bonds and the event B was satisfied. But otherwise the event B was not satisfied, because we had more than one card for Bonds before we collected the last card, and since we kept continuing the process we had no card

for at least one player other than Bonds; this means that in the original process as well, we would have more than one card for Bonds before stopping.

So now conditioning on A_j, we have collected the first card for Bonds, and no card for $n-j$ players. In the modified process we need two cards for Bonds and one for the $n-j$ players. So we can ignore the first card we collected for Bonds, and run the coupon collector process for $n-j+1$ players and just compute the probability that the last card collected is Bonds. But note that the last card collected can be any of the $n-j+1$ players, and by symmetry the probability of it being any of the players is the same. So the probability that it is a particular player (Bonds) is simply $\frac{1}{n-j+1}$. This means that $\Pr[B \mid A_j] = \frac{1}{n-j+1}$. Therefore we have

$$\Pr[B] = \frac{1}{n}\sum_{j=1}^{n}\frac{1}{n-j+1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{i}.$$

Side note: The quantity $\sum_{i=1}^{n}\frac{1}{i}$ is called H_n, the n-th Harmonic number, and it is well-approximated by $\ln(n)$. So the probability $\Pr[B]$ is well-approximated by $\ln(n)/n$.