CS 70 Spring 2022	Discrete Mathematics and Probability Theory Satish Rao and Koushik Sen	HW 13

1 Estimating π

In this problem, we discuss some interesting ways that you could probabilistically estimate π, and see how good these techniques are at estimating π.

Technique 1: Buffon's needle is a method that can be used to estimate the value of π. There is a table with infinitely many parallel lines spaced a distance 1 apart, and a needle of length 1. It turns out that if the needle is dropped uniformly at random onto the table, the probability of the needle intersecting a line is $\frac{2}{\pi}$. We have seen a proof of this in the notes.

Technique 2: Consider a square dartboard, and a circular target drawn inscribed in the square dartboard. A dart is thrown uniformly at random in the square. The probability the dart lies in the circle is $\frac{\pi}{4}$.

Technique 3: Pick two integers x and y independently and uniformly at random from 1 to M, inclusive. Let p_M be the probability that x and y are relatively prime. Then

$$\lim_{M\to\infty} p_M = \frac{6}{\pi^2}.$$

Let $p_1 = \frac{2}{\pi}, p_2 = \frac{\pi}{4}$, and $p_3 = \frac{6}{\pi^2}$ be the probabilities of the desired events of **Technique 1**, **Technique 2**, and **Technique 3**, respectively. For each technique, we apply each technique N times, then compute the proportion of the times each technique occurred, getting estimates $\hat{p}_1, \hat{p}_2$, and $\hat{p}_3$,, respectively.

(a) For each $\hat{p}_i$, compute an expression X_i in terms of $\hat{p}_i$ that would be an estimate of π.

(b) Using Chebyshev's Inequality, compute the minimum value of N such that X_2 is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ.

For X_1 and X_3, computing the minimum value of N will be more tricky, as the expressions for X_1 and X_3 are not as nice as X_2.

(c) For $i = 1$ and 3, compute a constant c_i such that

$$|X_i - \pi| < \varepsilon \iff |\hat{p}_i - p_i| < c_i\varepsilon + o(\varepsilon^2),$$

where the $o(\varepsilon^2)$ represents terms containing powers of ε that are 2 or higher (i.e. $\varepsilon^2, \varepsilon^3$, etc.).

(Hint: You may find the following Taylor series helpful: For x close to 0,

$$\frac{1}{a-x} = \frac{1}{a} + \frac{x}{a^2} + o(x^2)$$
$$\frac{1}{(a-x)^2} = \frac{1}{a^2} + \frac{2x}{a^3} + o(x^2).$$

The $o(x^2)$ represents terms that have x^2 powers or higher.)

In this problem, we assume ε is close enough to 0 such that $o(\varepsilon^2)$ is 0. In other words,

$$\mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon + o(\varepsilon^2)\right] = \mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon\right].$$

Combining with part (c) then gives

$$\mathbb{P}\left[|X_i - \pi| < \varepsilon\right] = \mathbb{P}\left[|\hat{p}_i - p_i| < c_i\varepsilon\right].$$

(d) For $i = 1$ and 3, use Chebyshev's Inequality and the above work to compute the minimum value of N such that X_i is within ε of π with $1-\delta$ confidence. Your answer should be in terms of ε and δ.

(e) Which technique required the lowest value for N? Which technique required the highest?

Solution:

(a) $\hat{p}_1$ is an estimate of $\frac{2}{\pi}$, so $X_1 = \frac{2}{\hat{p}_1}$ would be an estimate of π. Similarly, $X_2 = 4\hat{p}_2$ and $X_3 = \sqrt{\frac{6}{\hat{p}_3}}$ are estimates of π.

(b) We have

$$\mathbb{P}\left[|X_2 - \pi| \geq \varepsilon\right] = \mathbb{P}\left[\left|\hat{p}_2 - \frac{\pi}{4}\right| \geq \frac{1}{4}\varepsilon\right] \\ \geq \frac{\mathrm{Var}(\hat{p}_2)}{\left(\frac{1}{4}\varepsilon\right)^2}$$

by Chebyshev's Inequality and using the fact that $X_2 = 4\hat{p}_2$. We want our estimate to have confidence $1-\delta$, so we want $\frac{\mathrm{Var}(\hat{p}_2)}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta$. Since $\hat{p}_2$ is a Binomial(N, p_2) variable, it has variance $\frac{p_2(1-p_2)}{N}$. Combining everything gives

$$\frac{\frac{p_2(1-p_2)}{N}}{\left(\frac{1}{4}\varepsilon\right)^2} < \delta \implies N > \frac{16p_2(1-p_2)}{\delta\varepsilon^2} = \frac{\pi(4-\pi)}{\delta\varepsilon^2}.$$

(c) For $i = 1$, we have

$$\begin{aligned}
|X_1 - \pi| < \varepsilon &\iff \left|\frac{2}{\hat{p}_1} - \pi\right| < \varepsilon \\
&\iff \pi - \varepsilon < \frac{2}{\hat{p}_1} < \pi + \varepsilon \\
&\iff \frac{2}{\pi + \varepsilon} < \hat{p}_1 < \frac{2}{\pi - \varepsilon} \\
&\iff \frac{2}{\pi + \varepsilon} - \frac{2}{\pi} < \hat{p}_1 - \frac{2}{\pi} < \frac{2}{\pi - \varepsilon} - \frac{2}{\pi} \\
&\iff -\frac{2\varepsilon}{\pi(\pi + \varepsilon)} < \hat{p}_1 - \frac{2}{\pi} < \frac{2\varepsilon}{\pi(\pi - \varepsilon)}.
\end{aligned}$$

We apply the Taylor series expansion $\frac{1}{a-x} = \frac{1}{a} + o(x)$ to get

$$\frac{2\varepsilon}{\pi(\pi - \varepsilon)} = \frac{2\varepsilon}{\pi}\left(\frac{1}{\pi} + o(\varepsilon)\right) = \frac{2\varepsilon}{\pi^2} + o(\varepsilon^2)$$

and

$$-\frac{2\varepsilon}{\pi(\pi + \varepsilon)} = -\frac{2\varepsilon}{\pi}\left(\frac{1}{\pi} + o(\varepsilon)\right) = -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2).$$

We conclude that

$$|X_1 - \pi| < \varepsilon \implies -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2) < \hat{p}_1 - \frac{2}{\pi} < -\frac{2\varepsilon}{\pi^2} + o(\varepsilon^2) \implies |\hat{p}_1 - \frac{2}{\pi}| < \frac{2\varepsilon}{\pi^2} + o(\varepsilon^2),$$

so $c_1 = \frac{2}{\pi^2}$.

Similarly, for $i = 3$, we have

$$\begin{aligned}
|X_3 - \pi| < \varepsilon &\iff \left|\sqrt{\frac{6}{\hat{p}_3}} - \pi\right| < \varepsilon \\
&\iff \pi - \varepsilon < \sqrt{\frac{6}{\hat{p}_3}} < \pi + \varepsilon \\
&\iff \frac{6}{(\pi + \varepsilon)^2} < \hat{p}_3 < \frac{6}{(\pi - \varepsilon)^2} \\
&\iff \frac{6}{(\pi + \varepsilon)^2} - \frac{6}{\pi^2} < \hat{p}_3 - \frac{6}{\pi^2} < \frac{6}{(\pi - \varepsilon)^2} - \frac{6}{\pi^2} \\
&\iff -\frac{12\pi\varepsilon + 6\varepsilon^2}{\pi^2(\pi + \varepsilon)^2} < \hat{p}_3 - \frac{6}{\pi^2} < \frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2(\pi - \varepsilon)^2}.
\end{aligned}$$

We apply the Taylor series expansion $\frac{1}{(a-x)^2} = \frac{1}{a^2} + o(x)$ to get

$$\frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2(\pi - \varepsilon)^2} = \frac{12\pi\varepsilon - 6\varepsilon^2}{\pi^2}\left(\frac{1}{\pi^2} + o(\varepsilon)\right) = \frac{12\pi\varepsilon}{\pi^4} + o(\varepsilon^2) = \frac{12\varepsilon}{\pi^3} + o(\varepsilon^2)$$

and

$$-\frac{12\pi\varepsilon+6\varepsilon^2}{\pi^2(\pi+\varepsilon)^2} = -\frac{12\pi\varepsilon+6\varepsilon^2}{\pi^2}\left(\frac{1}{\pi^2}+o(\varepsilon)\right) = -\frac{12\pi\varepsilon}{\pi^4}+o(\varepsilon^2) = -\frac{12\varepsilon}{\pi^3}+o(\varepsilon^2).$$

We conclude that

$$|X_3-\pi|<\varepsilon \implies -\frac{12\varepsilon}{\pi^3}+o(\varepsilon^2) < \hat{p}_3-\frac{6}{\pi^2} < \frac{12\varepsilon}{\pi^3}+o(\varepsilon^2) \implies \left|\hat{p}_3-\frac{6}{\pi^2}\right| < \frac{12\varepsilon}{\pi^3}+o(\varepsilon^2),$$

so $c_3 = \frac{12}{\pi^3}$.

(d) By Chebyshev's Inequality,

$$\begin{aligned}\mathbb{P}\left[|X_i-\pi|\geq\varepsilon\right] &= \mathbb{P}\left[|\hat{p}_i-p_i|\geq c_i\varepsilon\right] \\ &\geq \frac{\operatorname{Var}(\hat{p}_i)}{(c_i\varepsilon)^2}.\end{aligned}$$

We want our estimate to have confidence $1-\delta$, so we want $\frac{\operatorname{Var}(\hat{p}_i)}{(c_i\varepsilon)^2}<\delta$. Since $\hat{p}_i$ is a Binomial(N,p_i) variable, it has variance $\frac{p_i(1-p_i)}{N}$. Combining everything gives

$$\frac{\frac{p_i(1-p_i)}{N}}{(c_i\varepsilon)^2}<\delta \implies N > \frac{p_i(1-p_i)}{c_i^2\varepsilon^2\delta}.$$

Plugging in $i=1$ gives

$$N > \frac{\frac{2}{\pi}\left(1-\frac{2}{\pi}\right)}{\left(\frac{2}{\pi^2}\right)^2\varepsilon^2\delta} = \frac{\pi^2(\pi-2)}{2\varepsilon^2\delta}.$$

and plugging in $i=3$ gives

$$N > \frac{\frac{6}{\pi^2}\left(1-\frac{6}{\pi^2}\right)}{\left(\frac{12}{\pi^3}\right)^2\varepsilon^2\delta} = \frac{\pi^2(\pi^2-6)}{24\varepsilon^2\delta}.$$

(e) Looking at our values from parts (b) and (d),

$$\frac{\pi^2(\pi^2-6)}{24\varepsilon^2\delta} < \frac{\pi(4-\pi)}{\delta\varepsilon^2} < \frac{\pi^2(\pi-2)}{2\varepsilon^2\delta},$$

so technique 3 required the lowest value for N, while technique 1 required the highest value for N.

2 Random Cuckoo Hashing

Cuckoo birds are parasitic beasts. They are known for hijacking the nests of other bird species and evicting the eggs already inside. Cuckoo hashing is inspired by this behavior. In cuckoo hashing, when we get a collision, the element that was already there gets evicted and rehashed.

We study a simple (but ineffective, as we'll see) version of cuckoo hashing, where all hashes are random. Let's say we want to hash n pieces of data $d_1, d_2, \ldots, d_n$ into n possible hash buckets labeled $1, \ldots, n$. We hash the $d_1, \ldots, d_n$ in that order. When hashing d_i, we assign it a random bucket chosen uniformly from $1, \ldots, n$. If there is no collision, then we place d_i into that bucket. If there is a collision with some other d_j, we evict d_j and assign it another random bucket uniformly from $1, \ldots, n$. (It is possible that d_j gets assigned back to the bucket it was just evicted from!) We again perform the eviction step if we get another collision. We keep doing this until there is no more collision, and we then introduce the next piece of data, d_{i+1} to the hash table.

(a) What is the probability that there are no collisions over the entire process of hashing $d_1, \ldots, d_n$ to buckets $1, \ldots, n$? What value does the probability tend towards as n grows very large?

(b) Assume we have already hashed $d_1, \ldots, d_{n-1}$, and they each occupy their own bucket. We now introduce d_n into our hash table. What is the expected number of collisions that we'll see while hashing d_n? (*Hint*: What happens when we hash d_n and get a collision, so we evict some other d_i and have to hash d_i? Are we at a situation that we've seen before?)

(c) Generalize the previous part: Assume we have already hashed $d_1, \ldots, d_{k-1}$ successfully, where $1 \le k \le n$. Let C_k be the number of collisions that we'll see while hashing d_k. What is $\mathbb{E}[C_k]$?

(d) Let C be the total number of collisions over the entire process of hashing $d_1, \ldots, d_n$. What is $\mathbb{E}[C]$? You may leave your answer as a summation.

Solution:

(a) When hashing d_i, there are $(n-i+1)$ empty buckets, as $(i-1)$ of them are already occupied by $d_1, \ldots, d_{i-1}$. If we want no collisions over this entire hashing process, we must choose an empty bucket on the first go for each d_i. This gives:

$$\mathbb{P}[\text{no collisions}] = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \cdots \cdot \frac{1}{n} = \frac{n!}{n^n}$$

To understand what happens as n grows very large, we can upper bound the probability as follows:

$$\mathbb{P}[\text{no collisions}] = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \cdots \cdot \frac{1}{n} \le 1 \cdot \cdots \cdot 1 \cdot \frac{1}{n} = \frac{1}{n}$$

We are upper bounding each term in the product above by 1, except the very last term, which we leave as $\frac{1}{n}$. When n is large, this upper bound goes to 0, so $\mathbb{P}[\text{no collisions}]$ will also tend to 0.

Another way to obtain the $\frac{n!}{n^n}$ probability is to see that considering the first bucket to which each datum gets hashed is a uniform sample space with size n^n. The number of sample points in our event (no collisions) is the number of ways of assigning each datum a unique bucket to be placed in, i.e. the number of ways to permute the datum within the buckets, or $n!$.

(b) Let C_n be the number of collisions experienced when hashing a single datum into a table with $(n-1)$ buckets already populated. (Note that we don't specify that we hash d_n in particular when defining C.)

First, it is possible that we end with 0 collisions. This happens with probability $\frac{1}{n}$. Otherwise, we get a collision, and we have to evict some other datum d_i. Now, we are back in the original situation; the number of collisions experienced after re-hashing d_i is also C because we are again in the situation of introducing a single datum into a table with $(n-1)$ buckets already populated. However, we do need to count the fact that we already had one collision–the one that evicted d_i. This gives us:

$$\mathbb{E}[C_n] = 0 \cdot \frac{1}{n} + (\mathbb{E}[C_n]+1) \cdot \frac{n-1}{n}$$

Solving for $\mathbb{E}[C_n]$ above, we get an expected $(n-1)$ collisions.

Remark: It is also perfectly valid to use an infinite sum based solution.

(c) We take a similar approach to the previous part. Let C_k be the number of collisions experienced when hashing a single datum into a table with $(k-1)$.

When we hash d_k we have probability $\frac{n-(k-1)}{n}$ of not getting a collision and finishing the process with 0 collisions. Otherwise, we evict some other datum and are left with the same situation. This gives us:

$$\mathbb{E}[C_k] = 0 \cdot \frac{n-k+1}{n} + (\mathbb{E}[C_k]+1) \cdot \frac{k-1}{n}$$

Solving for $\mathbb{E}[C_k]$ above, we get an expected $\frac{k-1}{n-k+1}$ collisions.

(d) Let C_k be the random variable denoting number of collisions which occur while hashing the kth datum, d_k. Let C be the total number of collisions which occur over the entire process. That is, $C = C_1 + C_2 + \cdots + C_n$. Then we have:

$$\mathbb{E}[C] = \mathbb{E}\left[\sum_{k=1}^{n} C_k\right] = \sum_{k=1}^{n} \mathbb{E}[C_k] = \sum_{k=1}^{n} \frac{k-1}{n-k+1} = \sum_{k=0}^{n-1} \frac{k}{n-k}$$

The second step uses linearity of expectation, and the third step makes use of the result from the previous part.

3 Coupon Collector Variance

It's that time of the year again—Safeway is offering its Monopoly Card promotion. Each time you visit Safeway, you are given one of n different Monopoly Cards with equal probability. You need to collect them all to redeem the grand prize.

(a) Let X be the number of visits you have to make before you can redeem the grand prize. Show that $\mathrm{Var}(X) = n^2 \left(\sum_{i=1}^{n} i^{-2}\right) - \mathbb{E}[X]$.

(b) The series $\sum_{i=1}^{\infty} i^{-2}$ converges to the constant value $\pi^2/6$. Using this fact and Chebyshev's Inequality, find a lower bound on β for which the probability you need to make more than $\mathbb{E}[X] + \beta n$ visits is less than $1/100$, for large n. *[Hint: Use the approximation $\sum_{i=1}^{n} i^{-1} \approx \ln n$ as n grows large.]*

Solution:

(a) Note that this is the coupon collector's problem, but now we have to find the variance. Let X_i be the number of visits we need to make before we have collected the ith unique Monopoly card actually obtained, given that we have already collected $i-1$ unique Monopoly cards. Then $X = \sum_{i=1}^{n} X_i$ and each X_i is geometrically distributed with $p = (n-i+1)/n$. Moreover, the indicators themselves are independent, since each time you collect a new card, you are starting from a clean slate.

$$\begin{aligned}
\mathrm{Var}(X) &= \sum_{i=1}^{n} \mathrm{Var}(X_i) && \text{(as the } X_i \text{ are independent)} \\
&= \sum_{i=1}^{n} \frac{1-(n-i+1)/n}{[(n-i+1)/n]^2} && \text{(variance of a geometric r.v. is } (1-p)/p^2\text{)} \\
&= \sum_{j=1}^{n} \frac{1-j/n}{(j/n)^2} && \text{(by noticing that } n-i+1 \text{ takes on all values from 1 to } n\text{)} \\
&= \sum_{j=1}^{n} \frac{n(n-j)}{j^2} \\
&= \sum_{j=1}^{n} \frac{n^2}{j^2} - \sum_{j=1}^{n} \frac{n}{j} \\
&= n^2 \left(\sum_{j=1}^{n} \frac{1}{j^2}\right) - \mathbb{E}[X] && \text{(using the coupon collector problem expected value).}
\end{aligned}$$

(b) We are looking for the smallest value of β for which we can say that $\mathbb{P}[X \geq \mathbb{E}[X] + \beta n] < 1/100$ for all values of n.

We have:

$$\begin{aligned}
\mathbb{P}[X \geq \mathbb{E}(X) + \beta n] &= \mathbb{P}[X - \mathbb{E}[X] \geq \beta n] \\
&\leq \mathbb{P}[|X - \mathbb{E}[X]| \geq \beta n] \\
&\leq \frac{\mathrm{Var}\,(X)}{(\beta n)^2} && \text{(by Chebyshev's inequality)} \\
&= \frac{n^2 \sum_{i=1}^n i^{-2} - \mathbb{E}[X]}{(\beta n)^2} \\
&\approx \frac{n^2 \sum_{i=1}^n i^{-2} - n \ln n}{(\beta n)^2} \\
&= \frac{\sum_{i=1}^n i^{-2}}{\beta^2} - \frac{\ln n}{n \beta^2}
\end{aligned}$$

Therefore, we desire a lower bound on β such that the following is satisfied:

$$\frac{\sum_{i=1}^n i^{-2}}{\beta^2} - \frac{\ln n}{n\beta^2} < \frac{1}{100}.$$

But as $n \to \infty$, the second term approaches zero, since n grows faster than $\ln n$, and the first term approaches $\frac{\pi^2}{6}$. Therefore, we simply need to satisfy

$$\frac{\pi^2}{6\beta^2} < \frac{1}{100}.$$

This requires $\beta > 10\pi/\sqrt{6} \approx 12.825$.

4 Short Answer

(a) Let X be uniform on the interval $[0,2]$, and define $Y = 2X + 1$. Find the PDF, CDF, expectation, and variance of Y.

(b) Let X and Y have joint distribution

$$f(x,y) = \begin{cases} cxy + \frac{1}{4} & x \in [1,2] \text{ and } y \in [0,2] \\ 0 & \text{otherwise.} \end{cases}$$

Find the constant c. Are X and Y independent?

(c) Let $X \sim \mathrm{Exp}(3)$.

(i) Find probability that $X \in [0,\, 1]$.

(ii) Let $Y = \lfloor X \rfloor$. For each $k \in \mathbb{N}$, what is the probability that $Y = k$? Write the distribution of Y in terms of one of the famous distributions; provide that distribution's name and parameters.

(d) Let $X_i \sim \text{Exp}(\lambda_i)$ for $i = 1, \ldots, n$ be mutually independent. It is a (very nice) fact that $\min(X_1, \ldots, X_n) \sim \text{Exp}(\mu)$. Find μ.

Solution:

(a) Let's begin with the CDF. It will first be useful to recall that

$$F_X(t) = \mathbb{P}[X \leq t] = \begin{cases} 0 & t \leq 0 \\ \frac{t}{2} & t \in [0,2] \\ 1 & t \geq 2 \end{cases}.$$

Since Y is defined in terms of X, we can compute that

$$\begin{aligned} F_Y(t) = \mathbb{P}[Y \leq t] &= \mathbb{P}[2X+1 \leq t] \\ &= \mathbb{P}\left[X \leq \frac{t-1}{2}\right] \\ &= F_X\left(\frac{t-1}{2}\right) \\ &= \begin{cases} 0 & t \leq 1 \\ \frac{t-1}{4} & t \in [1,5] \\ 1 & t \geq 5 \end{cases} \end{aligned}$$

where in the third line we have used the PDF for X. We know that the PDF can be found by taking the derivative of the CDF, so

$$f_Y(t) = \frac{\mathrm{d}F_Y(t)}{\mathrm{d}t} = \begin{cases} \frac{1}{4} & t \in [1,5] \\ 0 & \text{else} \end{cases}.$$

By linearity of expectation $\mathbb{E}[Y] = \mathbb{E}[2X+1] = 2\,\mathbb{E}[X]+1 = 3$, and similarly

$$\text{Var}(Y) = \text{Var}(2X+1) = 4\,\text{Var}(X) = 4 \cdot \frac{4}{12} = \frac{4}{3}.$$

(b) To find the correct constant, we use the fact that a PDF must integrate to one. In particular,

$$1 = \int_1^2 \int_0^2 (cxy + 1/4)\,\mathrm{d}y\,\mathrm{d}x = 3c + \frac{1}{2},$$

so $c = 1/6$. In order to check independence, we need to first find the marginal distributions of X and Y:

$$\begin{aligned} f_X(x) &= \int_0^2 f(x,y)\,\mathrm{d}y = 1/2 + x/3 \\ f_Y(y) &= \int_1^2 f(x,y)\,\mathrm{d}x = 1/4 + y/4. \end{aligned}$$

Since

$$f_X(x) f_Y(y) = \frac{1}{8} + \frac{y}{8} + \frac{x}{12} + \frac{xy}{12} \neq \frac{1}{4} + \frac{xy}{6} = f(x,y),$$

the random variables are not independent.

(c) (i) Since $X \sim \text{Exp}(3)$, the CDF of X is $F(x) = 1 - e^{-3x}$. Thus we have

$$\mathbb{P}[X \in [0,1]] = \int_0^1 f(x)\,\mathrm{d}x = F(1) - F(0) = (1 - e^{-3}) - (1 - e^0) = 1 - e^{-3}.$$

(ii) Similarly, if $Y = \lfloor X \rfloor$, then $Y = k$ exactly when $X \in [k, k+1)$, so

$$\begin{aligned}
\mathbb{P}[Y = k] &= \mathbb{P}[X \in [k, k+1)] \\
&= \int_k^{k+1} f(x)\,\mathrm{d}x \\
&= F(k+1) - F(k) \\
&= (1 - e^{-3(k+1)}) - (1 - e^{-3k}) \\
&= e^{-3k} - e^{-3(k+1)} \\
&= e^{-3k}\left(1 - e^{-3}\right) = (e^{-3})^k \left(1 - e^{-3}\right).
\end{aligned}$$

In other words, $Y = W - 1$ for $W \sim \text{Geometric}(1 - e^{-3})$.

(d) Since the X_i are independent,

$$\begin{aligned}
\mathbb{P}[\min(X_1, \dots, X_n) \le t] &= 1 - \mathbb{P}[X_1 > t, X_2 > t, \cdots\cdot X_n > t] \\
&= 1 - \mathbb{P}[X_1 > t] \cdot \mathbb{P}[X_2 > t] \cdot \cdots \cdot \mathbb{P}[X_n > t] \qquad \text{(by independence)} \\
&= 1 - e^{-\lambda_1 t} e^{-\lambda_2 t} \cdot \cdots \cdot e^{-\lambda_n t} \\
&= 1 - e^{-(\lambda_1 + \lambda_2 + \cdots + \lambda_n)t}.
\end{aligned}$$

This is exactly the CDF of an $\text{Exp}(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ random variable, so $\mu = \lambda_1 + \cdots + \lambda_n$.

5 Useful Uniforms

Let X be a continuous random variable whose image is all of $\mathbb{R}$; that is, $\mathbb{P}[X \in (a,b)] > 0$ for all $a, b \in \mathbb{R}$ and $a \neq b$.

(a) Give an example of a distribution that X could have, and one that it could not.

(b) Show that the CDF F of X is strictly increasing. That is, $F(x + \varepsilon) > F(x)$ for any $\varepsilon > 0$. Argue why this implies that $F : \mathbb{R} \to (0,1)$ must be invertible.

(c) Let U be a uniform random variable on $(0,1)$. What is the distribution of $F^{-1}(U)$?

(d) Your work in part (c) shows that in order to sample X, it is enough to be able to sample U. If X was a discrete random variable instead, taking finitely many values, can we still use U to sample X?

Solution:

(a) Any random variable with density $f(x) > 0$ for all x works as a positive example; e.g. $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ (corresponding to the normal distribution) or

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if } |x| < 1, \\ \frac{1}{4|x|^2}, & \text{if } |x| \geq 1 \end{cases}.$$

Any distribution of density f such that $f(x) = 0$ for all $x \in (a,b)$ for some $a, b \in \mathbb{R}$, $a \neq b$ works as a negative example; e.g. (corresponding to an exponential random variable)

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

or (corresponding to a uniform variable on $[0,1]$),

$$f(x) = \begin{cases} 1, & \text{if } x \in [0,1], \\ 0, & \text{otherwise} \end{cases}.$$

(b) We have

$$\begin{aligned} F(x+\varepsilon) &= \mathbb{P}[X \leq x+\varepsilon] \\ &= \mathbb{P}[X \leq x] + \mathbb{P}[X \in (x, x+\varepsilon]] \\ &\geq F(x) + \mathbb{P}[X \in (x, x+\varepsilon)] \\ &> F(x) \end{aligned}$$

where in the very last inequality we used the fact that $\mathbb{P}[X \in (a,b)] > 0$ with $a = x$ and $b = x+\varepsilon$.

Next, to show invertibility, we need to show (i) injectivity and (ii) surjectivity.

Injectivity: If $x \neq y$, then either $x < y$ or $y < x$ and so either $F(x) < F(y)$ or $F(y) < F(x)$. In either case, $F(x) \neq F(y)$, and so F must be injective.

Invertibility: F is continuous (in fact, differentiable with derivative f), approaching 1 as $x \to \infty$, and approaching 0 as $x \to -\infty$. Therefore, it must assume all values between 0 and 1, and hence is surjective.

(c) $\mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x)$, where $\{F^{-1}(U) \leq x\} = \{U \leq F(x)\}$ since F is strictly increasing. Thus $F^{-1}(U)$ and X have the very same CDF, which means that $F^{-1}(U)$ and X share the same distribution.

(d) Yes, we can! Assume X took values in a discrete set $\mathscr{A} = \{a_1, a_2, \ldots, a_n\} \subset \mathbb{R}$ with probabilities $\mathbb{P}[X = a_k] = p_k$. Then mimicking the argument from part (c), we can define $G : [0,1] \to \mathscr{A}$ as

$$G(x) = \begin{cases} a_1, & \text{if } x \leq p_1, \\ a_2, & \text{if } x \in (p_1, p_1+p_2], \\ a_3, & \text{if } x \in (p_1+p_2, p_1+p_2+p_3], \\ \vdots & \vdots \\ a_{n-1}, & \text{if } x \in (\sum_{k=1}^{n-2} p_k, \sum_{k=1}^{n-1} p_k], \\ a_n, & \text{if } x \in (\sum_{k=1}^{n-1} p_k, 1] \end{cases}$$

(draw a picture of G's graph!), for which we have $\mathbb{P}[G(U)=a_k] = \sum_{j=1}^{k} p_j - \sum_{j=1}^{k-1} p_j = p_k = \mathbb{P}[X=a_k]$. That is, $G(U)$ and X have the same distribution as desired.

6 It's Raining Fish

A hurricane just blew across the coast and flung a school of fish onto the road nearby the beach. The road starts at your house and is infinitely long. We will label a point on the road by its distance from your house (in miles). For each $n \in \mathbb{N}$, the number of fish that land on the segment of the road $[n, n+1]$ is independently $\text{Poisson}(\lambda)$ and each fish that is flung into that segment of the road lands uniformly at random within the segment. Keep in mind that you can cite any result from lecture or discussion without proof.

(a) What is the distribution of the number of fish arriving in segment $[0, n]$ of the road, for some $n \in \mathbb{N}$?

(b) Let $[a, b]$ be an interval in $[0, 1]$. What is the distribution of the number of fish that lands in the segment $[a, b]$ of the road?

(c) Let $[a, b]$ be any interval such that $a \geq 0$. What is the distribution of the number of fish that land in $[a, b]$?

(d) Suppose you take a stroll down the road. What is the distribution of the distance you walk (in miles) until you encounter the first fish?

(e) Suppose you encounter a fish at distance x. What is the distribution of the distance you walk until you encounter the next fish?

Solution:

(a) From lecture, we learned that if X and Y are independent, and $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, then $X+Y \sim \text{Poisson}(\lambda+\mu)$. We know the number of fish to land in the segment $[0, n]$ is the sum of the number of fish to land in $[i, i+1]$ for each $i \in [0, 1, \ldots, n-1]$. Thus the number of fish is in $[0, n]$ is $\text{Poisson}(n\lambda)$.

(b) The probability that a particular fish lands in the interval $[a, b]$ is $b-a$ since its location is uniformly distributed within $[0, 1]$. Thus, the distribution is $\text{Poisson}((b-a)\lambda)$.

(c) The answer is still $\text{Poisson}((b-a)\lambda)$. Clearly, this is true if $[a, b]$ is contained within some interval $[n, n+1]$. If it's not, then let i be the smallest integer such that $i \geq a$ and let j be the largest integer such that such that $j \leq b$. Then the distribution is

$$\text{Poisson}((i-a)\lambda) + \text{Poisson}((j-i)\lambda) + \text{Poisson}((b-j)\lambda) = \text{Poisson}((b-a)\lambda).$$

(d) The distance is $\text{Exp}(\lambda)$. To prove this, it suffices to show that the cdf matches the exponential cdf. Let X be the distance of the first fish from the house. Note that $\mathbb{P}[X \geq t] =$

$\mathbb{P}$[no fish in $[0,t]$]. By the previous parts, we know that the number of fish in $[0,t]$ is Poisson(λt), which is equal to 0 with probability

$$\frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}.$$

Thus, we have that $\mathbb{P}[X < t] = 1 - e^{-\lambda t}$ which is exactly the exponential cdf.

(e) Still Exp(λ). Using the same logic as the previous part, we have that $\mathbb{P}[X \geq t]$ is equal to the probability that no fish lands in the segment $[x, x+1]$. This in turn is equal to $e^{-\lambda t}$, because the number of fish in that segment is distributed as Poisson$(-\lambda t)$.

7 Waiting For the Bus

Edward and Jerry are waiting at the bus stop outside of Soda Hall.

Like many bus systems, buses arrive in periodic intervals. However, the Berkeley bus system is unreliable, so the length of these intervals are random, and follow Exponential distributions.

Edward is waiting for the 51B, which arrives according to an Exponential distribution with parameter λ. That is, if we let the random variable X_i correspond to the difference between the arrival time ith and $(i-1)$st bus (also known as the inter-arrival time) of the 51B, $X_i \sim \text{Expo}(\lambda)$.

Jerry is waiting for the 79, whose inter-arrival times also follows Exponential distributions with parameter μ. That is, if we let Y_i denote the inter-arrival time of the 79, $Y_i \sim \text{Expo}(\mu)$. Assume that all inter-arrival times are independent.

(a) What is the probability that Jerry's bus arrives before Edward's bus?

(b) After 20 minutes, the 79 arrives, and Jerry rides the bus. However, the 51B still hasn't arrived yet. Let D be the additional amount of time Edward needs to wait for the 51B to arrive. What is the distribution of D?

(c) Lavanya isn't picky, so she will wait until either the 51B or the 79 bus arrives. Find the distribution of Z, the amount of time Lavanya will wait before catching her bus.

(d) Khalil doesn't feel like riding the bus with Edward. He decides that he will wait for the second arrival of the 51B to ride the bus. Find the distribution of $T = X_1 + X_2$, the amount of time that Khalil will wait to ride the bus.

Solution:

(a) Let f_{Y_i} be the pdf of Y_i. By total probability,

$$\begin{aligned}
\mathbb{P}[X_i > Y_i] &= \int_{t=0}^{\infty} f_{Y_i}(t) \cdot \mathbb{P}[X_i > Y_i \mid Y_i = t] \, \mathrm{d}t \\
&= \int_{t=0}^{\infty} f_{Y_i}(t) \cdot \mathbb{P}[X_i > t] \, \mathrm{d}t \\
&= \int_{t=0}^{\infty} f_{Y_i}(t) \cdot (1 - F_{X_i}(t)) \, \mathrm{d}t \\
&= \int_{t=0}^{\infty} \mu e^{-\mu t} (e^{-\lambda t}) \, \mathrm{d}t \\
&= \mu \int_{t=0}^{\infty} e^{-(\lambda+\mu)t} \, \mathrm{d}t \\
&= \frac{\mu}{\lambda+\mu} \int_{t=0}^{\infty} (\lambda+\mu) e^{-(\lambda+\mu)t} \, \mathrm{d}t \\
&= \frac{\mu}{\lambda+\mu},
\end{aligned}$$

where the integral in the second-to-last line evaluates to 1, since it is the total integral of the Exponential$(\lambda+\mu)$ density.

(b) We observe that $\mathbb{P}[D > d] = \mathbb{P}[X > 20 + d \mid X \geq 20]$. Then, we apply Bayes Rule:

$$\begin{aligned}
\mathbb{P}[X > 20 + d \mid X \geq 20] &= \frac{\mathbb{P}[X > 20 + d]}{\mathbb{P}[X \geq 20]} \\
&= \frac{1 - F_X(20+d)}{1 - F_X(20)} \\
&= \frac{e^{-\lambda(20+d)}}{e^{-20\lambda}} \\
&= e^{-\lambda d}
\end{aligned}$$

Thus, the CDF of D is given by $\mathbb{P}[D \leq d] = 1 - \mathbb{P}[D > d] = 1 - e^{-\lambda d}$. This is the CDF of an exponential, so D is exponentially distributed with parameter λ.

One can also directly apply the memoryless property of the exponential distribution to arrive at this answer.

(c) Lavanya's waiting time is the minimum of the time it takes for the 51B and the time it takes for the 79 to arrive. Thus, $Z = \min(X, Y)$.

$$\begin{aligned}
\mathbb{P}[Z > t] &= \mathbb{P}[X > t \cap Y > t] \\
&= \mathbb{P}[X > t] \cdot \mathbb{P}[Y > t] \\
&= (1 - F_X(t))(1 - F_Y(t)) \\
&= (1 - (1 - e^{-\mu t}))(1 - (1 - e^{-\lambda t})) \\
&= e^{-\mu t} e^{-\lambda t} \\
&= e^{-(\mu+\lambda)t}
\end{aligned}$$

It follows that the CDF is Z, $\mathbb{P}[Z \le t] = 1 - e^{-(\mu+\lambda)t}$. Thus, Z is exponentially distributed with parameter $\mu + \lambda$.

(d) Let $t > 0$. By total probability,

$$
\begin{aligned}
\mathbb{P}[T \le t] &= \mathbb{P}[X_1 + X_2 \le t] \\
&= \int_0^\infty \mathbb{P}[X_1 + X_2 \le t \mid X_1 \in \mathrm{d}x] \cdot \mathbb{P}[X_1 \in \mathrm{d}x] \\
&= \int_0^t \mathbb{P}[X_1 + X_2 \le t \mid X_1 \in \mathrm{d}x] \cdot \mathbb{P}[X_1 \in \mathrm{d}x] + \int_t^\infty 0 \cdot \mathbb{P}[X_1 \in \mathrm{d}x] \\
&= \int_0^t \mathbb{P}[X_2 \le t - X_1 \mid X_1 \in \mathrm{d}x] \cdot \mathbb{P}[X_1 \in \mathrm{d}x] + 0 \\
&= \int_0^t \mathbb{P}[X_2 \le t - x] \cdot \mathbb{P}[X_1 \in \mathrm{d}x] \\
&= \int_0^t F_{X_2}(t-x) \cdot f_{X_1}(x)\,\mathrm{d}x \\
&= \int_0^t \left(1 - e^{-\lambda(t-x)}\right) \cdot \lambda e^{-\lambda x}\,\mathrm{d}x \\
&= \int_0^t \lambda e^{-\lambda x} - \lambda e^{-\lambda t}\,\mathrm{d}x \\
&= \int_0^t \lambda e^{-\lambda x} - \lambda e^{-\lambda t} \int_0^t \mathrm{d}x \\
&= F_{X_1}(t) - \lambda e^{-\lambda t} \cdot t \\
&= 1 - e^{-\lambda t} - \lambda t e^{-\lambda t}
\end{aligned}
$$

Upon differentiating the CDF, we have

$$
\begin{aligned}
f_T(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{P}[T \le t] &= \lambda e^{-\lambda t} - \lambda e^{-\lambda t} + \lambda^2 t e^{-\lambda t} \\
&= \lambda^2 t e^{-\lambda t}, \quad \text{for } t > 0.
\end{aligned}
$$