HW5

Nikhil Unni

1. In order to estimate the probability of a head in a coin flip, p, you flip a coin n times and count the number of heads, S_n. You use the estimator $\hat{p} = \frac{S_n}{n}$.

   (a) You choose the sample size n to have a guarantee

   $$P(|\hat{p} - p| \geq \epsilon) \leq \delta$$

   Using Chebyshev inequality, determine n with the following parameters:

   i. Compare the value of n when $\epsilon = 0.05$, $\delta = 0.1$ to when $\epsilon = 0.1, \delta = 0.1$.

   First, we find the variance of $\hat{p}$. The variance of just S_n is clearly $np(1-p)$, since it's binomial. Then, the variance of $\hat{p}$ is just dividing that by n^2, giving us $\frac{p(1-p)}{n}$.
   Next, Chebyshev's Inequality gives us:

   $$P(|\hat{p} - p| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} = \delta = \frac{p(1-p)}{\epsilon^2 n}$$

   So plugging in the values we get:

   $$\frac{p(1-p)}{0.05^2 * n} = 0.1$$

   Then, $n = \frac{p(1-p)}{0.05^2 * 0.1} = 4000p(1-p)$.

   With the other numbers, we get: $n = \frac{p(1-p)}{0.1^2 * 0.1} = 1000p(1-p)$.

   So the second set of numbers, we need a fourth as much as the first set.

   ii. Compare the value of n when $\epsilon = 0.1$, $\delta = 0.05$ to when $\epsilon = 0.1, \delta = 0.1$.

   With the first set:

   $$n = \frac{p(1-p)}{0.1^2 * 0.05} = 2000p(1-p)$$

   And we already know the second set is $1000p(1-p)$, which is half the first set.

   (b) Now, we change the scenario slightly. You know that $p \in (0.4, 0.6)$ and would now like to determine the smallest n such that:

   $$P(\frac{|\hat{p} - p|}{p} \leq 0.05) \geq 0.95$$

   We already have the equation $\delta = \frac{p(1-p)}{\epsilon^2 n}$. Plugging in 0.95, and 0.05, we get:

   $$n = \frac{p(1-p)}{0.05^2 * 0.95} = 421.052631579p(1-p)$$

   So, the smallest possible number is $\lceil 421.052631579p(1-p) \rceil$.

2. Let $X_i, 1 \leq i \leq n$ be a sequence of i.i.d random variables distributed uniformly in $[-1, 1]$. Show that the following sequences converge in probability to some limit.

   (a) $Y_n = (X_n)^n$

   I assert that the sequence converges in probability to 0. This would mean:

   $$\lim_{n \to \infty} P(|Y_n - 0| > \epsilon) = 0$$

   Or:

   $$\lim_{n \to \infty} P(Y_n > \epsilon) = 0$$

   From Chebyshev:

   $$P(|Y_n - E[Y_n]| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

   But we know that $\lim_{n \to \infty} a = 0$, for some $0 \geq a < 1$. So regardless of what X_n is, because it is boudned by 0 and 1, $(X_n)^n$ approaches 0 w.p. 1. So $E[Y_n]$ is clearly 0, and the variance is just $0^2 - 0^2 = 0$. So w.p. 1, we get: $\lim_{n \to \infty} P(Y_n > \epsilon) = 0$.

   (b) $Y_n = \Pi_{i=1}^n X_i$

   I assert that the sequence converges in probability to 0. This would mean:

   $$\lim_{n \to \infty} P(|Y_n - 0| > \epsilon) = 0$$

   Or:

   $$\lim_{n \to \infty} P(Y_n > \epsilon) = 0$$

   From Chebyshev:

   $$P(|Y_n - E[Y_n]| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

   Because all of the X_i are independent, $E[Y_n] = \Pi_{i=1}^n E[X_i] = 0$. To calculate $\text{Var}(X_1 \cdots X_n)$, we see:

   $$= E[(X_1 \cdots X_n)^2] - (E[X_1 \cdots X_n])^2$$
   $$= E[(X_1^2 \cdots X_n^2)] - 0^2 = E[X_1^2 \cdots X_n^2]$$
   $$= \Pi_{i=1}^n \text{Var}(X_i) + E[X_i]^2$$
   $$= \Pi_{i=1}^n \text{Var}(X_i) = \Pi_{i=1}^n \frac{1}{12}(2)^2$$
   $$= (\frac{1}{3})^n$$

   So finally, we have:

   $$P(|Y_n - 0| > \epsilon) \leq \frac{1}{3^n \epsilon^2}$$

   And $\lim_{n \to \infty} \frac{1}{3^n \epsilon^2} = 0$. So we have:

   $$\lim_{n \to \infty} P(Y_n > \epsilon) = 0$$

(c) $Y_n = max\{X_1, X_2, \cdots, X_n\}$

Using the same strategy as part (b), let's try to find $E[Y_n]$ and $\sigma^2_{Y_n}$ first.

$$f_{Y_n}(y) = nf_X(y)[F_X(y)]^{n-1} = n(\frac{1}{2})(\frac{y+1}{2})^{n-1}$$

So then:

$$\begin{aligned} E[Y_n] &= \int_{y=-1}^{1} ny(\frac{1}{2})(\frac{y+1}{2})^{n-1}dy \\ &= \frac{1}{2}n\int_{y=-1}^{1} y(\frac{y+1}{2})^{n-1}dy \\ &= \frac{1}{2}n(\frac{2(n-1)}{n(n+1)}) \\ &= \frac{n-1}{n+1} \end{aligned}$$

Similarly:

$$\begin{aligned} E[Y_n^2] &= \int_{y=-1}^{1} ny^2(\frac{1}{2})(\frac{y+1}{2})^{n-1}dy \\ &= \frac{1}{2}n\int_{y=-1}^{1} y^2(\frac{y+1}{2})^{n-1}dy \\ &= \frac{1}{2}n(\frac{2[n(n-1)+2]}{n(n+1)(n+2)}) \\ &= \frac{n^2-n+2}{(n+1)(n+2)} \end{aligned}$$

So we have $E[Y_n] = \frac{n-1}{n+1}$ and $\text{Var}(Y_n) = \frac{n^2-n+2}{(n+1)(n+2)} - \frac{(n-1)^2}{(n+1)^2}$. Evaluating these limits as $n \to \infty$, with L'Hoppital's Rule, we get $\lim_{n\to\infty} E[Y_n] = \frac{1}{1} = 1$, and $\lim_{n\to\infty} \sigma^2_{Y_n} = \frac{1}{1} - \frac{1}{1} = 0$.

So, starting from the Chebyshev Inequality:

$$P(|Y_n - E[Y_n]| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

So, evaluating at the limit we get:

$$\lim_{n\to\infty} P(|Y_n - E[Y_n]| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

$$\lim_{n\to\infty} P(|Y_n - 1| > \epsilon) \leq \frac{0}{\epsilon^2}$$
$$\lim_{n\to\infty} P(|Y_n - 1| > \epsilon) = 0$$

So Y_n converges in probability to 1.

3. I break a stick n times in the following manner : the ith time I break the stick, I keep a fraction X_i of the remaining stick where X_i is uniform on the interval $[0,1]$ and $X_1, X_2, \cdots, X_n$ are i.i.d. Let $P_n = \Pi_{i=1}^n X_i$ be the fraction of the original stick that I end up with. Find $\lim_{n\to\infty} P_n^{1/n}$ and $E[P_n]^{1/n}$.

For $\lim_{n\to\infty} P_n^{1/n}$:

$$\begin{aligned}\lim_{n\to\infty} P_n^{1/n} &= \lim_{n\to\infty}(X_1\cdots X_n)^{1/n}\\ &= \lim_{n\to\infty}(X_1)^{1/n}\cdots(X_n)^{1/n}\end{aligned}$$

All of the exponents will approach 0, and so every $(X_i)^0$ will be 1, regardless of the value of (X_i). So we just get:

$$\lim_{n\to\infty} P_n^{1/n} = \lim_{n\to\infty}\Pi_{i=1}^n 1 = 1$$

For $E[P_n]^{1/n}$:

$$E[P_n]^{1/n} = E[\Pi_{i=1}^n X_i]^{1/n}$$

Since all of the X_i are independent, this is just:

$$(\Pi_{i=1}^n E[X_i])^{1/n}$$

And all of the $E[X_i]$ are clearly $\frac{1}{2}$, so we have:

$$E[P_n]^{1/n} = (\Pi_{i=1}^n \frac{1}{2})^{1/n}) = (\frac{1}{2})^{n^{1/n}} = \frac{1}{2}$$

4. Consider the Markov chain of Figure 1 (not pictured), where $a, b \in (0,1)$.

   (a) Find the invariant distribution.

   First, we note the transition matrix:

$$P = \begin{pmatrix}1-a & a & 0\\ 1-b & 0 & b\\ 0 & 1 & 0\end{pmatrix}$$

   The invariant distribution is some π s.t. $\pi P = \pi$. So we have:

$$\begin{pmatrix}x & y & z\end{pmatrix}\begin{pmatrix}1-a & a & 0\\ 1-b & 0 & b\\ 0 & 1 & 0\end{pmatrix} = \begin{pmatrix}x & y & z\end{pmatrix}$$

   This gives us : $x(1-a)+y(1-b) = x$, $ax+z = y$, and $by = z$. Notice that the rank is 2, but we can also throw in the fact that $x+y+z = 1$. Throwing this all together, with some painstaking algebra we get:

$$\pi = \begin{pmatrix}x & y & z\end{pmatrix} = \begin{pmatrix}\frac{1-b}{1+a-b+ab} & \frac{a}{1+a-b+ab} & \frac{ab}{1+a-b+ab}\end{pmatrix}$$

   (b) Calculate $P(X(1) = 1, X(2) = 0, X(3) = 0, X(4) = 1 | X(0) = 0)$.

$$\begin{aligned}P(X(1) = 1|X(0) = 0) &= a\\ P(X(2) = 0|X(1) = 1) &= 1-b\\ P(X(3) = 0|X(2) = 0) &= 1-a\\ P(X(4) = 1|X(3) = 0) &= a\end{aligned}$$

   So the probability of all of those things happening is the probability that, at each time step, we made the "right" decision. This is just all of the above probabilities multiplied by each other. In other words.

$$\begin{aligned}&P(X(1) = 1, X(2) = 0, X(3) = 0, X(4) = 1|X(0) = 0)\\ &= P(X(1) = 1|X(0) = 0)P(X(2) = 0|X(1) = 1)P(X(3) = 0|X(2) = 0)P(X(4) = 1|X(3) = 0)\\ &= a^2(1-a)(1-b)\end{aligned}$$

(c) Show that the Markov chain is aperiodic.

Since $a, b \in (0.4, 0.6)$, then $0 < 1-a, 1-b < 1$. So all of the transitions happen with nonzero probability, meaning that the Markov Chain is irreducible. Since node 0 has a self loop, and the Markov Chain is irreducible, the period of nodes 0, and therefore all other nodes, is 1, making the Markov Chain aperiodic.

5. Let $\{X_n, n \geq 0\}$ be a Markov chain with two states, -1 and 1, and transition probabilities $P(-1, 1) = P(1, -1) = a$ for $a \in (0, 1)$. Define,
$$Y_n = X_0 + \cdots + X_n$$
Is $\{Y_n, n \geq 0\}$ a Markov chain? Prove or disprove.

Y is a Markov chain if $P(Y_n = j | Y_0, \cdots, Y_{n-2}, Y_{n-1} = i) = P(i, j)$. But notice that $Y_0 = X_0$, and $Y_1 = X_0 + X_1$, so we can recover X_1 from that information. And so on and so forth. So if we're given $Y_0, \cdots, Y_{n-1}$, then we know $X_0, \cdots X_{n-1}$. So then:
$$P(Y_n = j | Y_0, \cdots, Y_{n-2}, Y_{n-1} = i) = P(Y_n = j | X_0, \cdots X_{n-1}, Y_{n-1} = i) = P(X_n = j - i | X_{n-1})$$
And clearly, unless $|j - i| = 1$, the probability is 0. But $P(X_n = j - i | X_{n-1})$ is just some constant probability. It's either a or $1 - a$, depending on the choice of j and i (and the recovered X_{n-1} information).

However, $P(Y_n = j | Y_{n-1} = i)$ tells us less information. Although we can recover the number of X_i states that were 1 and the number of states that were -1 from just Y_{n-1} alone, (number of 1 states is $\frac{i+n}{2}$), the actual placements of the states is unknown, and thus, X_{n-1} could be either 1 or -1, whereas in the previous case we knew explicitly what X_{n-1} was from the given information. Explicitly we know:
$$P(Y_n = j | Y_{n-1} = i) = P(X_n = j - i | Y_{n-1} = i)$$
Since there are more "paths" to get to $j - i$, the probability is distinct from the case where we know all the prior information.

Thus, Y is **not** a Markov chain.

6. You have a database of an infinite number of movies. Each movie has a rating that is uniformly distributed in $\{0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5\}$ independent of all other movies. You want to find two movies such that the sum of their ratings is greater than 7.5 (7.5 is not included).

(a) A Stanford student chooses two movies each time and calculates the sum of their ratings. If it is less than or equal to 7.5, the student throws away these two movies and chooses two other movies. The student stops when he/she finds two movies such that the sum of their ratings is greater than 7.5. What is the expected number of movies that this student needs to choose from the database?

Notice that this R.V., let's call it Y, is almost Geometrically distributed. Let's call the sum of the movies $Z = X_1 + X_2$, and we'll reuse it for multiple movie attempts. Then the pdf of Y is just:
$$P(Y = y) = \Sigma_{i=0}^{\infty} (P(X_1 + X_2) \leq 7.5)^i (P(X_1 + X_2) > 7.5)$$
We can easily calculate $P(X_1 + X_2 > 7.5)$ just by counting. Out of all $11 * 11 = 121$ possible outcomes of $X_1 + X_2$, 15 of them result in a score ¿ 7.5. So $P(X_1 + X_2 > 7.5) > 7.5 = \frac{15}{121}$, and therefore $P(X_1 + X_2 \leq 7.5) = \frac{106}{121}$. So now we have a Geometric distribution Y w.p. $p = \frac{15}{121}$. And the expected value of a Geometric R.V. is $1/p = \frac{121}{15} \approx 8.06$. So if we try $\frac{121}{15}$ times, then we see $\frac{121}{15} * 2 \approx 16.1$ movies.

(b) A Berkeley student chooses movies from the database one by one and keeps the movie with the highest rating. The student stops when he/she finds that the sum of the ratings of the last movie that he/she has chosen and the movie with the highest rating among all the previous movies is greater than 7.5. What is the expected number of movies that the student will have to choose?

Let's call the max of everything we've seen so far $Y_n = max\{X_1, \cdots, X_{n-1}\}$. Then, the number of attempts before we reach 7.5 is given by some R.V. called Z. The pdf of Z is:

$$P(Z = z) = \Sigma_{i=0}^{\infty}[\Pi_{n=2}^{i}(P(Y_n + X_n) \leq 7.5)](P(Y_i + X_i) > 7.5)$$

Now we need to count (more tediously...). For any given time i, let's call our R.V.s Y and X. If $Y = 5$, there are 5 choices of X that will get a score of over 7.5. Similarly, if $Y = 4.5$, there are 4 choices for X, and so on until $Y = 3$ with one choice for X (which is 5, obviously). So then we have:

$$P(Y_i + X_i > 7.5) = f_Y(5)(\frac{5}{11}) + f_Y(4.5)(\frac{4}{11}) + \cdots + f_Y(3)(\frac{1}{11})$$

Now consider the ratings of movies can be real numbers and assume that the ratings are i.i.d uniformly distributed in $[0, 5]$.

(c) What is the expected number of movies that the Stanford student will have to choose in order to find two movies such that the sum of their ratings is greater than 7.5?

Again, we have the same distribution of Y, and we need to find $P(X_1+X_2 > 7.5)$ and $P(X_1+X_2 \leq 7.5)$. So then $P(X_1 + X_2 > 7.5) = \int_{x_1=2.5}^{5} \int_{x_2=7.5-x_1}^{5} \frac{1}{25}\, dx_2\, dx_1 = 0.125$. So now we have a geometric distribution with $p = 0.125$, so then expected value is $1/p = 8$, or an expected 16 movies.

(d) (Optional) What is the expected number of movies that the Berkeley student will have to choose in order to find two movies such that the sum of their ratings is greater than 7.5?