Probability 1
HW 3

Daniel Yu

September 27, 2024

1. Let E_n be the number of empty bins. Show that

$$\lim_{n\to\infty} \frac{1}{n} E[E_n].$$

exists and compute as function of c.

Proof. We know from the previous homework that the probability of k empty bins is

$$P[E_n = k] = \frac{(n-k)^r}{n^r}.$$

Alternatively, consider $E_n = B_1 + B_2 + \ldots + B_n$ where B_i is a random variable representing if the ith bin is empty. Since expected value is linear,

$$\begin{aligned} E[E_n] &= E[B_1] + E[B_2] + \ldots + E[B_n] \\ &= \sum_{i=1}^{n} E[B_i] \\ &= \sum_{i=1}^{n} [0 + 1 \cdot P[B_i] = 1] \\ &= n\frac{(n-1)^r}{n^r}. \end{aligned}$$

As n, r go to infinity as $\frac{r}{n} \to c \Leftrightarrow r \to cn$.

$$E[E_n] = n\frac{(n-1)^{cn}}{n^{cn}}.$$

Then,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} E[E_n] &= \lim_{n\to\infty} \frac{1}{n} n \frac{(n-1)^{cn}}{n^{cn}} \\ &= \lim_{n\to\infty} \frac{(n-1)^{cn}}{n^{cn}} \\ &= \lim_{n\to\infty} (\frac{n-1}{n})^{cn} \\ &= \lim_{n\to\infty} \exp(n \ln(1 - \frac{1}{n}))^c \\ &= \lim_{n\to\infty} \exp(n \cdot -\frac{1}{n})^c \\ &= \lim_{n\to\infty} \exp(-1)^c \\ &= e^{-c}. \end{aligned}$$

□

2. Let N be the number of rolls required for a fair six sided die to have the same number show up twice in a row. Find expected value of N

Proof. Denote G_i as the random variable representing the number of rolls until the first ith value shows up. Since $G_i \sim Geo(\frac{1}{6})$, so $E[G_i] = \frac{1}{6}$. By the law of iterated expectations, we know that

$$E[N] = E_{G_i}[E_N[N|G_i]].$$

Now, consider X_{G+1} the result of the roll after G_i. Condition $E[N|G_i]$ on X_{G+1}.

$E[N|G_i, X_{G+1} = i] = G + 1$
$E[N|G_i, X_{G+1} \neq 1] = G + 1 + E[N]$,we reset the experiment with X_{G+1} as the new X_1

.

This gives,

$$\begin{aligned}
E[N|G_i] &= E[N|G_i, X_{G+1} = i] \cdot P[X_{G+1} = i] + E[N|G_i, X_{G+1} \neq i] \cdot P[X_{G+1} \neq i] \\
&= (G+1) \cdot \frac{1}{6} + (G+1+E[N]) \cdot \frac{5}{6} \\
&= G + 1 + \frac{5}{6}E[N].
\end{aligned}$$

Backsubsituting for $E[N]$:

$$\begin{aligned}
E[N] = E_{G_i}[E_N[N|G_i]] &= E_{G_i}[G + 1 + \frac{5}{6}E[N]] \\
&= 1 + 1 + \frac{5}{6}E[N] \\
&= 2 + \frac{5}{6}E[N] \\
\frac{1}{6}E[N] &= 2 \\
&= 12.
\end{aligned}$$

The expected value of N is 12. □

3. For a variant of the Monty Hall problem, should the contestant switch screens?

   Proof. Consider the following setup. There are 5 screens, A, B, C, D, E, F and without loss of generality assume A, B have prizes and D, E, F have goats. If the contestant picks $x \in \{A, B\}$ then Monty Hall has to open the one of $\{A, B\}$ which the contestant did not choose. If the contestant switches, then he gets a goat. If the contestant stays, then he gets the prize.

   Alternatively, if the contestant picks $x \in \{D, E, F\}$, then Monty Hall has to open $y \in \{X, Y\}$. If the contestant stays, he loses, but if the contestant switches then there is a $\frac{1}{3}$ chance of getting a prize.

   Now, let's got back to the setup. Assuming the contestant uniformly chooses one of $\{A, B, C, D, E, F\}$ uniformly then if he doesn't switch, $P[\text{win}|\text{No switch}, x \in \{A, B\}] = 1$, $P[\text{win}|\text{switch}, x \in \{A, B\}] = 0$ and $P[x \in \{A, B\}] = \frac{2}{5}$ while $P[\text{win}|\text{No switch}, x \in \{D, E, F\}] = 0$ and $P[\text{win}|\text{Switch}, x \in \{D, E, F\}] = \frac{1}{3}$, $P[x \in \{D, E, F\}]$. Then

$$\begin{aligned}
P[\text{win}|\text{switch}] &= P[\text{win}|\text{switch}, x \in \{A, B\}] + P[\text{win}|\text{switch}, x \in \{D, E, F\}] \\
&= 0 \cdot \frac{2}{5} + \frac{1}{3} \cdot \frac{3}{5} \\
&= \frac{1}{5}.
\end{aligned}$$

   and,

$$\begin{aligned}
P[\text{win}|\text{no switch}] &= P[\text{win}|\text{no switch}, x \in \{A, B\}] + P[\text{win}|\text{no switch}, x \in \{D, E, F\}] \\
&= 1 \cdot \frac{2}{5} + 0 \cdot \frac{3}{5} \\
&= \frac{2}{5}.
\end{aligned}$$

   Thus, the contestant should not switch as $P[\text{win}|\text{no switch}] = \frac{2}{5} > P[\text{win}|\text{switch}] = \frac{1}{5}$ □

4. Let p, q be two real numbers in $(0, 1)$. Let $V \sim Geo(p)$ be number of people that visit a store in a given day. Let q be the probability each customer buys a bar. Each customer is independent.

   (a) What is the expected number of chocolate bars sold in a day?

Proof. Let C be the number of chocolate bars sold in a day. Represent each customer as an independent identically distributed random variable X_i representing whether or not they bought a chocolate bar. Then, $P[x_i = 1] = q$ and $P[X_i = 0] = 1 - q$. Then, $C = \sum_{i=1}^{V} X_i$ ad $P[C = c|V] = \binom{V}{c} q^c (1-q)^{V-c}$.

$$\begin{aligned}
E[C] = E_v[E_c[C|V]] &= E_V[\sum_{k=1}^{v} k \cdot P[C = k|V = v]] \\
&= E_V[\sum_{k=1}^{v} \binom{v}{k} q^k (1-q)^{v-k}] \\
&= \sum_{v=1}^{\infty}[\sum_{k=1}^{v} \binom{v}{k} q^k (1-q)^{v-k}] \cdot P[V = v] \\
&= \sum_{k=1}^{v} q^k \sum_{v=1}^{\infty} \binom{v}{k} (1-q)^{v-k} \cdot P[V = v]
\end{aligned}$$

.

$$\begin{aligned}
E[C] = E_v[E_C[C|V]] &= E_v[E_C[bin(V, q)]] \\
&= E_v[V \cdot q] \\
&= q \cdot E_v[V] \\
&= q \cdot \frac{1}{p} \\
&= \frac{q}{p}.
\end{aligned}$$

□

(b) What is the probability that the number of chocolate bars sold is equal to the number of customers that visited the store on a particular day?

Proof. This is asking $P[C = V|V] = \binom{V}{V} \cdot q^V = q^V$. So we are considering, $P[C = V]$.

$$\begin{aligned}
P[C = V] &= \sum_{k}^{\infty} P[C = k \cap V = k] \\
&= \sum_{k}^{\infty} P[C = k|V = k] \cdot P[V = k] \\
&= \sum_{k}^{\infty} q^k \cdot P[V = k] \\
&= \sum_{k}^{\infty} q^k (1-p)^{k-1} p \\
&= qp \sum_{k}^{\infty} q^{k-1} (1-p)^{k-1} \\
&= qp \sum_{k}^{\infty} (q(1-p))^{k-1} \\
&= \frac{qp}{1 - q(1-p)}.
\end{aligned}$$

As $k \to \infty$, $P[C = V] \to 0$ because $q, p < 0$ □

5. Let X, Y be two independent exponential random variables with $\lambda = 1$. Conditional on X, Y, let Z be a uniform random variable on $[-X, Y]$ what is the mean and variance of Z?

Proof. Define the distribution of Z as:

$$P[Z=z]=\begin{cases}\frac{1}{Y+X}, \text{-X} \leq z \leq Y \\ 0,\ \text{otherwise}\end{cases}.$$

Substituting e^{-t} for X, Y:

$$P[Z=z]=\begin{cases}\frac{1}{2e^{-t}}, z \in [-e^{-t}, e^{t}], t \geq 0 \\ 0,\ \text{otherwise}\end{cases}.$$

The cdf of Z is:

$$F_Z(z)=\begin{cases}\frac{z+e^{-t}}{2e^{-t}}, z \in [-e^{-t}, e^{t}], t \geq 0 \\ 0,\ \text{otherwise}\end{cases}.$$

The pdf of Z is:

$$f_Z(z)=\frac{d}{dz}F_Z(z)=\frac{d}{dz}[\frac{z}{2e^{-t}}+\frac{e^{-t}}{2e^{-t}}]=\frac{1}{2e^{-t}}.$$

In this case, as expected the expected value is the center of the interval, 0.

$$\begin{aligned}E[Z]&=\int_{-\infty}^{\infty} z\cdot\frac{1}{2e^{-t}}dz\\&=\int_{-e^{-t}}^{e^{-t}} z\cdot\frac{1}{2e^{-t}}dz\\&=\frac{1}{2}[\frac{z^2}{2e^{-t}}]_{-e^{-t}}^{e^{-t}}\\&=\frac{1}{2}[\frac{e^{-2t}}{2e^{-t}}-\frac{e^{-2t}}{2e^{-t}}]\\&=0.\end{aligned}$$

The variance of Z:

$$\begin{aligned}Var(Z)&=E[Z^2]-E[Z]^2\\&=\int_{-\infty}^{\infty} z^2\cdot\frac{1}{2e^{-t}}dz-0\\&=\int_{-e^{-t}}^{e^{-t}} z^2\cdot\frac{1}{2e^{-t}}dz\\&=\frac{1}{3}[\frac{z^3}{2e^{-t}}]_{-e^{-t}}^{e^{-t}}\\&=\frac{1}{3}[\frac{e^{-3t}}{2e^{-t}}-\frac{-e^{-3t}}{2e^{-t}}]\\&=\frac{1}{3}\cdot\frac{e^{-3t}}{e^{-t}}\\&=\frac{1}{3}e^{-2t}.\end{aligned}$$

□