STA2001 Probability and statistical Inference I
Tutorial 4

1. 2.3-17. Let X equal the number of flips of a fair coin that are required to observe the first head-to-tail transition on consecutive fips. (Similar to 2.3-16)

   (a) Find the pmf of X. Hint: Draw a tree diagram.

   (b) Show that the mgf of X is $M(t) = e^{2t}/(e^t - 2)^2$.

   (c) Use the mgf to find the values of (i) the mean and (ii) the variance of X.

   (d) Find the values of (i) $P(X \leq 3)$, (ii) $P(X \geq 5)$, and (iii) $P(X = 3)$.

   Solution:

   (a) The description can be transformed in the following way: Once we observe "a head is followed by a tail" we stop the experiment and let X be number of all flips. The tree diagram can be drawn inversely (from the x^{th} flip to the 1^{st} flip) Note that in order to terminate at the first heads-tails flips with x flips in total, we have $x-1$ possible outcomes (i.e., the number of possible sequence to get the desired result), and each outcome has the same probability which is $P = (\frac{1}{2})^x$. Hence the pmf for X is

$$
\begin{aligned}
f(x) &= (\frac{1}{2})^x (x-1) \\
&= \frac{x-1}{2^x}
\end{aligned}
$$

   where $x \in S_X = \{2, 3, 4, \cdots\}$.

   (b) Recall the definition of moment generating function, we have

$$
\begin{aligned}
M(t) &= E(e^{tX}) \\
&= \sum_{x \in S_X} e^{tx} \frac{(x-1)}{2^x} \\
&= \sum_{x \in S_X} \left(\frac{e^t}{2}\right)^x (x-1) \\
&= \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 \times 2 + \left(\frac{e^t}{2}\right)^4 \times 3 + \cdots
\end{aligned}
$$

   Times $e^t/2$ for both sides,

$$
\left(\frac{e^t}{2}\right) M(t) = \left(\frac{e^t}{2}\right)^3 + \left(\frac{e^t}{2}\right)^4 \times 2 + \cdots
$$

Take the difference of $M(t)$ and $e^t/2 \cdot M(t)$,

$$\left(1 - \frac{e^t}{2}\right) M(t) = \left(\frac{e^t}{2}\right)^2 + \left(\frac{e^t}{2}\right)^3 + \left(\frac{e^t}{2}\right)^4 + \cdots = \sum_{k=2}^{\infty} (\frac{e^t}{2})^k$$

It's obvious that above series is a geometric series excluding the first two terms, so if $\frac{e^t}{2} < 1$ then the series converges, and we have

$$M(t) = \left(\frac{e^t}{2}\right)^2 \frac{1}{1 - \frac{e^t}{2}} \left(1 - \frac{e^t}{2}\right)^{-1} = \frac{e^{2t}}{(e^t - 2)^2}$$

(c) Recall that the n-th derivative of mgf evaluated at $t = 0$ equals to the n-th moment of the random variable, given that the quantities are well defined.

Take the first and second derivatives of $M(t)$,

$$\begin{aligned} M'(t) &= \frac{8e^{2t} - 4e^{3t}}{(e^t - 2)^4} \\ M''(t) &= \frac{(16e^{2t} - 12e^{3t})(e^t - 2)^4 - (8e^{2t} - 4e^{3t}) \cdot 4 \cdot (e^t - 2)^3 e^t}{\left[(e^t - 2)^4\right]^2} \end{aligned}$$

Set $t = 0$, then we have

$$\begin{aligned} M'(0) &= 4 = \mu \\ M''(0) &= 20 = E(X^2) \\ \sigma^2 &= E(X^2) - \mu^2 = 4 \end{aligned}$$

Note that we usually let μ denotes expectation and σ^2 denotes variance.

(d) From (a), we could easily compute

$$\begin{aligned} P(X \leq 3) &= \sum_{x=2}^{3} \frac{x-1}{2^x} = \frac{1}{4} + \frac{2}{8} = \frac{1}{2} \\ P(X \geq 5) &= 1 - \sum_{x=2}^{4} \frac{x-1}{2^x} = \frac{5}{16} \\ P(X = 3) &= \frac{3-1}{2^3} = \frac{1}{4} \end{aligned}$$

2. 2.4-15. A hospital obtains 40% of its flu vaccine from Company A, 50% from Company B, and 10% from Company C. From past experience, it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective. The hospital tests five vials from each shipment (Each shipment comes from a single company). If at least one of the five is ineffective, find the conditional probability of that shipment's having come from C. (Please recall the fair/unfair coin problem" in the tutorial of Bayes' Theorem)

Solution:

Let A be the events that "the shipment is from A"

Let B be the events that "the shipment is from B"

Let C be the events that "the shipment is from C"

Let I be the events that "at least one of the five is ineffective"

Therefore,

$$P(A) = 40\% \qquad P(B) = 50\% \qquad P(C) = 10\%$$

Let X_c be the R.V. of the numbers of ineffective vials if that shipment is from C. Thus, X_c follows a binomial distribution, $X_c \sim b(5, 5\%)$, and we have

$$\begin{aligned} P(I|C) = P(X_C \geq 1) = 1 - P(X_C = 0) &= 1 - \binom{5}{0}(5\%)^0(95\%)^5 \\ &= 1 - 0.95^5 \end{aligned}$$

Similarly, we could define X_A and X_B, and find that

$$P(I|B) = 1 - 0.98^5 \qquad P(I|A) = 1 - 0.97^5$$

Since A, B, C are mutually exclusive and exhaustive,

$$P(C|I) = \frac{P(I|C)P(C)}{P(I|A)P(A) + P(I|B)P(B) + P(I|C)P(C)} = 0.178$$

3. 2.5-9. (Coupon collector's problem) One of four different prizes was randomly put into each box of a cereal. If a family decided to buy this cereal until it obtained at least one of each of the four different prizes, what is the expected number of boxes of cereal that must be purchased? (Similar to 2.5-10)

   Solution:

   Let X be number of boxes we purchased to obtain all 4 different prizes.

   Let X_1 be the number of boxes we purchased to obtain the 1st prize.

   Let X_2 be the number of boxes we purchased to obtain the 2nd prize provided that we have got the 1st prize.

   Let X_3 be the number of boxes we purchased to obtain the 3rd prize provided that we have got 2 different prizes.

   Let X_4 be the number of boxes we purchased to obtain the 4th prize provided that we have got 32 different prizes.

   Apparently, we have

$$X = X_1 + X_2 + X_3 + X_4$$

   Since it always takes only one box to obtain the 1st prize, so

$$X_1 = 1 \text{ with probability } 1$$

   To obtain the 2nd prize, we must have obtained the 1st prize first. To obtain the 3rd prize, we must have obtained the 2nd prize first, and similar argument holds for the 4th prize.

   Therefore, X_2, X_3 and X_4 follow geometric distributions:

$$X_2 \sim \text{geometric } (\frac{3}{4})$$
$$X_3 \sim \text{geometric } (\frac{2}{4})$$
$$X_4 \sim \text{geometric } (\frac{1}{4})$$

   Thus, the expectation of X is given by

$$\begin{aligned} E(X) &= E(X_1 + X_2 + X_3 + X_4) \\ &= E(X_1) + E(X_2) + E(X_3) + E(X_4) \\ &= 1 + \frac{1}{\frac{3}{4}} + \frac{1}{\frac{2}{4}} + \frac{1}{\frac{1}{4}} \\ &= \frac{25}{3} \end{aligned}$$

4. 2.6-9. A store selling newspapers orders only $n = 4$ of a certain newspape because the manager does not get many calls for that publication. If the number of requests per day follows a Poisson distribution with mean 3,

   (a.) What is the expected value of the number sold?

   (b.) What is the minimum number that the manager should order so that the chance of having more requests than available newspapers is less than 0.05?

   Solution:

   (a.) Let X be the number of requests per day, says $X \sim$ Poisson (3) with pmf $f(x)$, and $x \in S_X = \{0, 1, 2, 3, 4, \cdots\}$.
   Let $u(X)$ be the number of newspaper sold, then

$$u(x) = \begin{cases} x & x = 0, 1, 2, 3 \\ 4 & x \geq 4 \end{cases}$$

   Therefore the expectation is computed as

$$\begin{aligned} E(u(X)) &= \sum_{x \in S_X} u(x) f(x) \\ &= \sum_{k=0}^{3} k \cdot \frac{3^k e^{-3}}{k!} + \sum_{k=4}^{\infty} 4 \cdot \frac{3^k e^{-3}}{k!} \\ &\approx 2.681 \end{aligned}$$

   (b.) Let y be the minimum number that the manager should order (it should be an non-negative integer), and let $F(x)$ be the cdf of the Poisson random variable X we mentioned in (a).
   Then,

$$\begin{aligned} & P(X > y) < 0.05 \\ \Leftrightarrow & 1 - P(X \leq y) < 0.05 \\ \Leftrightarrow & P(X \leq y) > 0.95 \\ \Leftrightarrow & F(y) > 0.95 \\ \Rightarrow & y = 6 \end{aligned}$$

   The final result is due to $F(6) = 0.966$ and $F(5) = 0.916$, where we got the figures by checking the table of Poisson distribution.