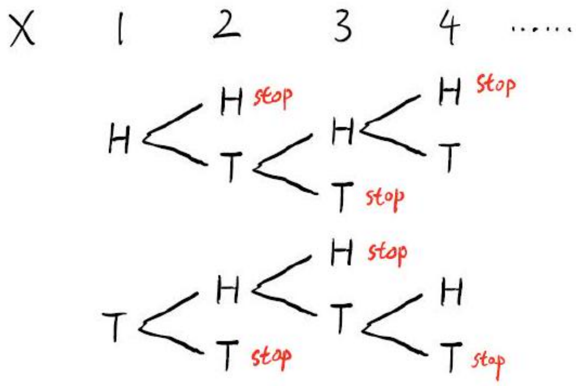
STA2001 Assignment 4

Deadline: 23:59 pm, Mar 1st

1. (2.3-16). Let X equal the number of flips of a fair coin that are required to observe the same face on consecutive flips.

   (a) Find the pmf of X. Hint: Draw a tree diagram.
   (b) Find the moment-generating function of X.
   (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) $P(X = 3)$.

   (a) It's obvious that X only takes values not less than 2, i.e. $X = 2, 3, 4, \cdots$. Now we draw graph to show the this fair coin flipping experiment and mark those scenarios we stop.



Now we can see that the pmf $f(x)$ is given by

$$P(X = x) = f(x) = \left(\frac{1}{2}\right)^x \cdot 2 = \left(\frac{1}{2}\right)^{x-1}, \quad \forall x = 2, 3, 4, \ldots$$

(b) Using the pmf and according to the definition of a mgf,

$$\begin{aligned} M(t) &= E\left[e^{tx}\right] \\ &= \sum_{x=2}^{\infty} e^{tx} \left(\frac{1}{2}\right)^{x-1} \\ &= 2\sum_{x=2}^{\infty} \left(\frac{e^t}{2}\right)^x \\ &= 2\sum_{x=0}^{\infty} \left(\frac{e^t}{2}\right)^x - 2\left(1 + \frac{e^t}{2}\right) \\ &= 2 \times \frac{1}{e^t/2} - 2 \times \frac{2 + e^t}{2} \\ &= \frac{e^{2t}}{2 - e^t}, \quad t < \ln 2 \end{aligned}$$

where we have used the result of a geometric series in the 5th equality (and thus we derive the interval of convergence $t < \ln 2$).

(c) We use the property that $E\left(X^k\right) = M^{(k)}(0)$ for positive integers k to find the mean and variance of random variable X. First we compute the derivatives of mgf,

$$M'(t) = \frac{d}{dt}M(t) = \frac{4e^{2t} - e^{3t}}{(2-e^t)^2}$$

$$M''(t) = \frac{d}{dt}M'(t) = \frac{(2-e^t)^2 \left(8e^{2t} - 3e^{3t}\right) - \left(4e^{2t} - e^{3t}\right) 2 \times (2-e^t)(-e^t)}{(2-e^t)^4}$$

Then we have

$$E(X) = M'(0) = 3$$
$$E\left(X^2\right) = M''(0) = 11$$
$$\sigma^2 = E\left(X^2\right) - [E(X)]^2 = 11 - 3^2 = 2$$

(d) Using the pmf we obtained from question (a), we have

(i). $P(X \leq 3) = P(X=2) + P(X=3) = \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 = \frac{3}{4}$

(ii). $P(X \geq 5) = 1 - P(X \leq 4) = 1 - \left(\frac{1}{2} + \frac{1}{4} + \frac{1}{8}\right) = \frac{1}{8}$

(iii). $P(X = 3) = \left(\frac{1}{2}\right)^2 = \frac{1}{4}$

2. Given a random permutation of the integers in the set $\{1,2,3,4,5\}$, let X equal the number of integers that are in their natural position. The moment-generating function of X is

$$M(t) = \frac{44}{120} + \frac{45}{120}e^t + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}$$

(a) Find the mean and variance of X.
(b) Find the probability that at least one integer is in its natural position.
(c) Draw a graph of the probability histogram of the pmf of X.

(a) We will use a property of moment generating function to compute the mean and variance.

First, we compute the first and second derivative of mgf,

$$\begin{aligned} M'(t) &= \frac{d}{dt}M(t) \\ &= \frac{45}{120}e^t + \frac{20}{120} \cdot 2 \cdot e^{2t} + \frac{10}{120} \cdot 3 \cdot e^{3t} + \frac{1}{120} \cdot 5 \cdot e^{5t} \\ &= \frac{3}{8}e^t + \frac{1}{3}e^{2t} + \frac{1}{4}e^{3t} + \frac{1}{24}e^{5t} \\ M''(t) &= \frac{d}{dt}M'(t) = \frac{3}{8}e^t + \frac{2}{3}e^{2t} + \frac{3}{4}e^{3t} + \frac{5}{24}e^{5t} \end{aligned}$$

Then use the property that $E\left(X^k\right) = M^{(k)}(0)$, we have

$$E(X) = M'(0) = \frac{3}{8} + \frac{1}{3} + \frac{1}{4} + \frac{1}{24} = 1$$
$$E\left(X^2\right) = M''(0) = \frac{3}{8} + \frac{2}{3} + \frac{3}{4} + \frac{5}{24} = 2$$

Finally, the variance is given by

$$\sigma^2 = E\left(X^2\right) - [E(X)]^2 = 2 - 1^2 = 1$$

(b) Let $f(x), x = 0, 1, \cdots, 5$, be the pmf of X.

Note that a mgf is unique if it exists, so by definition of the mgf, we have

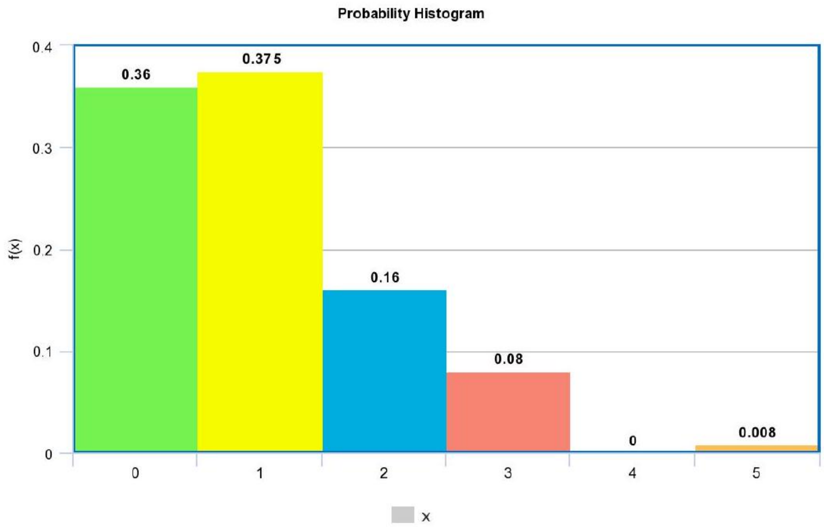
$$M(t) = \sum_{x=0}^{5} f(x)e^{tx} = \frac{44}{120} + \frac{45}{120}e^t + \frac{20}{120}e^{2t} + \frac{10}{120}e^{3t} + \frac{1}{120}e^{5t}$$

By matching the coefficients, we have the entire probability mass function $f(x)$,

$$f(0) = \frac{44}{120}, \quad f(1) = \frac{45}{120}, \quad f(2) = \frac{20}{120}, \quad f(3) = \frac{10}{120}, \quad f(4) = 0, \quad f(5) = \frac{1}{120}$$

Therefore, $P(X \geq 1) = 1 - P(X = 0) = \frac{76}{120} = \frac{19}{30}$.

(c) The graph is given below (I have converted the fractions to decimals).



3. (2.4-5). In a lab experiment involving inorganic syntheses of molecular precursors to organometallic ceramics, the final step of a five-step reaction involves the formation of a metalCmetal bond. The probability of such a bond forming is $p = 0.20$. Let X equal the number of successful reactions out of $n = 25$ such experiments.

(a) Find the probability that X is at most 4.
(b) Find the probability that X is at least 5.
(c) Find the probability that X is equal to 6.
(d) Give the mean, variance, and standard deviation of X.

X follows a binomial distribution with parameter $p = 0.2$, that is, $X \sim B(25, 0.2)$. Therefore we know its pdf is given by

$$P(X = x) = \binom{25}{x} 0.2^x 0.8^{25-x}$$

(a) $P(X \leq 4) = \sum_{x=0}^{4} \binom{25}{x} 0.2^x 0.8^{25-x} = 0.42$

(b) $P(X \geq 5) = 1 - P(X \leq 4) = 0.58$

(c) $P(X = 6) = \binom{25}{6} (0.2)^6 (0.8)^{19} = 0.16$

(d) Using the property of a binomial distribution, we have

$$E(x) = np = 25 \times 0.2 = 5$$
$$\sigma^2 = np(1 - p) = 25 \times 0.2 \times 0.8 = 4$$
$$\sigma = \sqrt{4} = 2$$

4. (2.4-7). Suppose that 2000 points are selected independently and at random from the unit square $\{(x, y) : 0 \leq x < 1, 0 \leq y < 1\}$. Let W equal the number of points that fall into $A = \{(x, y) : x^2 + y^2 < 1\}$.
(a) How is W distributed?
(b) Give the mean, variance, and standard deviation of W.
(c) What is the expected value of $W/500$?

(a) Let $S = \{(x, y) : 0 \leq x < 1, 0 \leq y < 1\}$. We know that the region $A \subset S$, and both regions A and S are defined on the first quadrant. So probability that one of the 2000 points fall in the region A, is $\pi/4$, as the region A is essentially a quarter-circle (i.e. 1/4 of a unit circle). As each point is generated independently, so the random variable W follows a binomial distribution, that is,

$$W \sim b(2000, \pi/4)$$

(b) Using the properties of a binomial distribution, we have

$$E(W) = np = 2000 \times \frac{\pi}{4} = 500\pi$$
$$\text{Var}(W) = np(1-p) = 2000 \times \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right) \approx 337$$
$$\sigma = \sqrt{337} = 18.36$$

(c) Note that expectation is a linear operator, we have

$$E\left(\frac{W}{500}\right) = \frac{1}{500}E(W) = \pi$$

5. (2.4-20). (i) Give the name of the distribution of X (if it has a name), (ii) find the values of μ and σ^2, and (iii) calculate $P(1 \leq X \leq 2)$ when the moment-generating function of X is given by
(a) $M(t) = (0.3 + 0.7e^t)^5$
(b) $M(t) = \frac{0.3e^t}{1-0.7e^t} \quad , \quad t < -\ln(0.7)$
(c) $M(t) = 0.45 + 0.55e^t$
(d) $M(t) = 0.3e^t + 0.4e^{2t} + 0.2e^{3t} + 0.1e^{4t}$
(e) $M(t) = \sum_{x=1}^{10}(0.1)e^{tx}$

In the following, we identity the distribution if it's possible and then use the corresponding properties to compute the mean and variance. However, you should be able to derive these quantities by using the mgf only, or using the definition only (i.e. using more fundamental approach).

(a) It's a binomial distribution, that is, $X \sim b(5, 0.7)$

$$\mu = np = 5 \times 0.7 = 3.5$$
$$\sigma^2 = np(1-p) = 3.5 \times 0.3 = 1.05$$
$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = \binom{5}{1} \times 0.7 \times 0.3^4 + \binom{5}{2} \times 0.7^2 \times 0.3^3 = 0.1607$$

(b) It's a geometric distribution with parameter $p = 0.3$.

$$\mu = \frac{1}{p} = \frac{10}{3}$$
$$\sigma^2 = \frac{1-p}{p^2} = \frac{0.7}{0.3^2} = \frac{70}{9}$$
$$P(1 \leq X \leq 2) = P(X = 1) + P(X = 2) = 0.3 + 0.3 \times 0.7 = 0.51$$

(c) It's a Bernoulli distribution with parameter $p = 0.55$.

$$\mu = p = 0.55$$
$$\sigma^2 = p(1-p) = 0.55 \times 0.45 = 0.2475$$
$$P(1 \le X \le 2) = P(X = 1) = 0.55$$

(d) This discrete random variable doesn't have a name. The pmf of this random variable can be obtained by matching coefficients of the given moment generating function (compare the coefficients of the closed form and its definition), as it uniquely determines a distribution. Once we obtain the pmf $f(x)$, we have

$$\mu = \sum_{x=1}^{4} x \cdot f(x) = 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.2 + 4 \times 0.1 = 2.1$$
$$\sigma^2 = E\left[(X-\mu)^2\right] = 1.1^2 \times 0.3 + 0.1^2 \times 0.4 + 0.9^2 \times 0.2 + 1.9^2 \times 0.1 = 0.89$$
$$P(1 \le X \le 2) = P(X = 1) + P(X = 2) = 0.3 + 0.4 = 0.7$$

(e) It's a discrete uniform distribution on the set $\{1, 2, \ldots, 10\}$. Similar with question (d), we compare the coefficients and obtain the pmf. Then we could compute

$$\mu = \sum_{x=1}^{10} x \cdot 0.1 = 5.5$$
$$\sigma^2 = E\left[(X-\mu)^2\right] = 0.1 \times \sum_{x=1}^{10} (x - 5.5)^2 = 8.25$$
$$P(1 \le X \le 2) = 0.1 + 0.1 = 0.2$$

6. (2.5-10). In 2012, Red Rose tea randomly began placing 1 of 12 English porcelain miniature figurines in a 100 -bag box of the tea, selecting from 12 nautical figurines.

(a) On the average, how many boxes of tea must be purchased by a customer to obtain a complete collection consisting of the 12 nautical figurines?
(b) If the customer uses one tea bag per day, how long can a customer expect to take, on the average, to obtain a complete collection?

(a) Let X_k be the number of box of tea we need to purchase in order to have a nautical figurine that is different from the k nautical figurines we already got.

Now, X_k can be viewed as a geometric random variable with parameter $p_k = \frac{12-k}{12}$, for $k = 0, 1, \cdots, 11$. Note that in this case $X_0 = 1$ with probability 1 , so it's also a degenerate random variable.

Let X be the total number of box we need to purchase to get the whole collection, so the mean is given by

$$E(X) = E\left(\sum_{k=0}^{11} X_k\right) = \sum_{k=0}^{11} E\left(X_k\right) = \sum_{k=0}^{11} \frac{1}{p_k} = \sum_{k=0}^{11} \frac{12}{12-k} = \frac{86021}{2310} = 37.2385$$

(b) As each box contains 100 bags of tea, so we have

$$100 \times E(X) \approx 3724 \text{ days}$$

which is approximately equal to 10.2 years.

7. The number of times that an individual contracts a cold in a given year is a Poisson random variable with mean $\theta = 6$. Suppose a new wonder drug (based on large quantities of vitamin C) has just been marketed that reduces the Poisson mean to $\theta = 4$ for 60 percent of the population. For the other 40 percent of the population the drug has no appreciable effect on colds. If an individual tries the drug for a year and has 3 colds in that time, how likely is it that the drug is beneficial for him/her?

Solution:

Let X be the no. of colds the individual contracts within a year. Then $X \sim \mathrm{Po}(4)$ if the drug is effective and $X \sim \mathrm{Po}(6)$ if the drug is not effective. Denote E as the event that the drug is effective, then

$$\mathrm{P}(X = 3 \mid E) = \frac{e^{-4}4^3}{3!} = \frac{32e^{-4}}{3}, \quad \mathrm{P}(E) = 0.6$$
$$\mathrm{P}\left(X = 3 \mid E^c\right) = \frac{e^{-6}6^3}{3!} = 36e^{-6}, \quad \mathrm{P}\left(E^c\right) = 0.4$$

Hence using law of total Probability and Bayes' theorem,

$$\mathrm{P}(X = 3) = 0.6 \times \frac{32e^{-4}}{3} + 0.4 \times 36e^{-6} = 6.4e^{-4} + 14.4e^{-6}$$
$$\mathrm{P}(E \mid X = 3) = \frac{\mathrm{P}(X = 3 \mid E)\,\mathrm{P}(E)}{\mathrm{P}(X = 3)} = \frac{6.4e^{-4}}{6.4e^{-4} + 14.4e^{-6}} = 0.7666$$

8. (2.6-5). Flaws in a certain type of drapery material appear on the average of one in 150 square feet. If we assume a Poisson distribution, find the probability of at most one flaw appearing in 225 square feet. The parameter λ is the average number of flaws appearing 225 square feet, so it's calculated by

$$\lambda = \frac{1}{150} \times 225 = 1.5$$

So $X \sim \text{Poisson}(1.5)$, and the probability is given by

$$P(X \leq 1) = \sum_{x=0}^{1} \frac{1.5^x \cdot e^{-1.5}}{x!} = 0.558$$

9. (2.6-8). Suppose that the probability of suffering a side effect from a certain flu vaccine is 0.005. If 1000 persons are inoculated, find the approximate probability that

(a) At most 1 person suffers.
(b) 4, 5, or 6 persons suffer.

Let X be the number of persons that have side effect among 100 persons. It's obvious that X follows a binomial distribution with probability of success $p = 0.005$. However, as $n = 1000$ is a pretty large number, this makes the probability calculation difficult.

To get the result, we use a Poisson distribution to approximate the original binomial distribution, that is,

$$\binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k}{k!} e^{-np}, \quad k = 0, 1, 2, \cdots$$

which means the parameter for this Poisson distribution is given by $\lambda \approx np$, and this approximation is good provided that n is large and p is small (the proof of this approximation can be found on the textbook or on the internet).

In this question, we know that $\lambda \approx np = 1000 \times 0.005 = 5$. (a) $P(X \leq 1) \approx \sum_{k=0}^{1} \frac{5^k \cdot e^{-5}}{k!} = 0.04$ (b) $P(X = 4 \text{ or } X = 5 \text{ or } X = 6) = P(X = 4) + P(X = 5) + P(X = 6) \approx \sum_{k=4}^{6} \frac{5^k \cdot e^{-5}}{k!} = 0.497$

10. (2.6-11). An airline always overbooks if possible. A particular plane has 95 seats on a flight in which a ticket sells for \$300. The airline sells 100 such tickets for this flight.

(a) If the probability of an individual not showing up is 0.05, assuming independence, what is the probability that the airline can accommodate all the passengers who do show up?
(b) If the airline must return the \$300 price plus a penalty of \$400 to each passenger that cannot get on the flight, what is the expected payout (penalty plus ticket refund) that the airline will pay?

(a) Let X be the number of passengers not showing up, so $X \sim b(100, 0.05)$. Therefore,

$$
\begin{aligned}
& P(\text{ the airline can accommodate all the passengers who do show up }) \\
=& P(X \geq 5) \\
=& 1 - P(X \leq 4) \\
=& 1 - \sum_{k=0}^{4} \binom{100}{k} 0.05^k 0.95^{100-k} \\
=& 0.564
\end{aligned}
$$

(b) Note that the company must pay penalty and refund if $X = 0, 1, 2, 4$, that is, there are some extra passengers showing up.

By the definition of expectation, we could compute

$$
\begin{aligned}
E(\text{ payout }) &= 700 \times [1 \times P(X = 4) + 2 \times P(X = 3) + 3 \times P(X = 2) + 4 \times P(X = 1) + 5 \times P(X = 0)] \\
&= 598.56
\end{aligned}
$$

Note: it may easier to compute the results of (a) and (b) if we use Poisson distribution to approximate this binomial distribution, and the approximated results are actually quite close to the true results, that is, 0.56 for (a) and 613.9 for (b).