Probability and Statistics for JMC
Solutions 3 — Discrete Random Variables

1. An experiment involves tossing two fair coins

   (a) What is the sample space for this experiment?

   $S = \{HH, HT, TH, TT\}$

   (b) What is the probability mass function (pmf) of the random variable X, which takes value 2 if two heads show, 1 if one head shows, and 0 if no heads show?

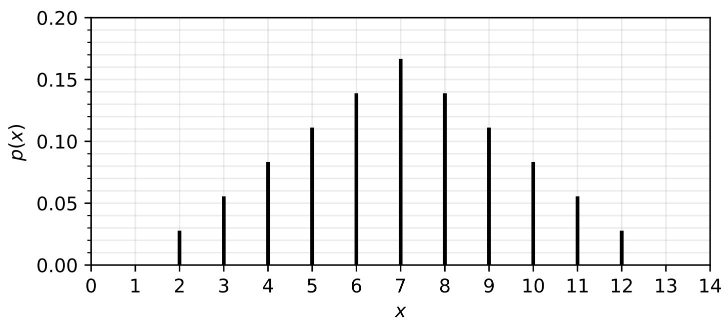
   $p_X(0) = \mathrm{P}(\{TT\}) = \frac{1}{4}, p_X(1) = \mathrm{P}(\{HT, TH\}) = \frac{1}{2}, p_X(2) = \mathrm{P}(\{HH\}) = \frac{1}{4}$, and $p_X(x) = 0$ if $x \notin \{0, 1, 2\}$.

   (c) What is the probability mass function of the random variable Y, which takes the value 3 if at least one head shows and 1 if no heads show?

   $p_Y(3) = \mathrm{P}(\{HH, HT, TH\}) = \frac{3}{4}, \quad p_Y(1) = \mathrm{P}(\{TT\}) = \frac{1}{4}$

2. Suppose that two fair dice are thrown and define a random variable X as the total number of spots showing. Make a table showing the probability mass function, $p(x)$ of X and plot a graph of $p(x)$.

x	2	3	4	5	6	7	8	9	10	11	12
$p(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$



3. In tossing a fair coin four times, what is the probability that one will obtain

   (a) four heads? $p(4) = \left(\frac{1}{2}\right)^4 = \frac{1}{16}$

   (b) three heads? $p(3) = \binom{4}{3}\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^1 = \frac{1}{4}$

   (c) at least two heads? $p(2) + p(3) + p(4) = \binom{4}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^2 + p(3) + p(4) = \frac{11}{16}$

   (d) not more than one head? $1 - \mathrm{P}(\text{at least two heads}) = 1 - \frac{11}{16} = \frac{5}{16}$

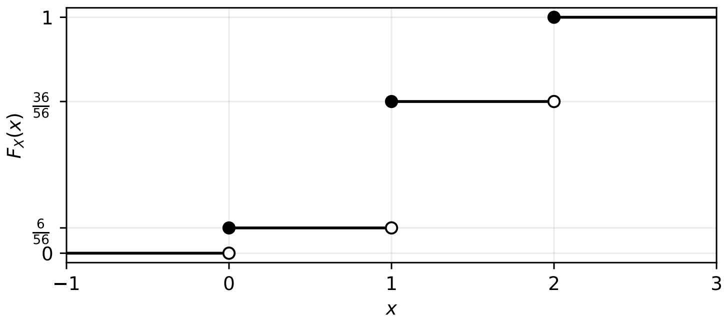
4. An urn holds 5 white and 3 black marbles.

   (a) If two marbles are drawn at random without replacement and X denotes the number of white marbles

      i. find the probability mass function of X

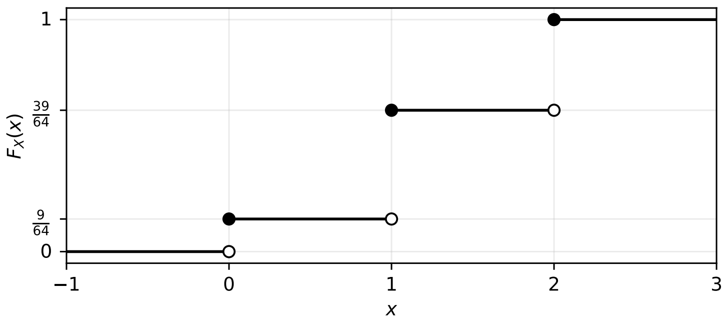
         Range of X is $\{0, 1, 2\}$.
         $p(0) = \mathrm{P}(B_1 B_2) = \mathrm{P}(B_1 B_2 | B_1)\mathrm{P}(B_1) = \frac{2}{7} \cdot \frac{3}{8} = \frac{6}{56}$
         $p(1) = \mathrm{P}(\{B_1 W_2, W_1 B_2\}) = \mathrm{P}(B_1 W_2) + \mathrm{P}(W_1 B_2) = \frac{5}{7} \cdot \frac{3}{8} + \frac{3}{7} \cdot \frac{5}{8} = \frac{30}{56}$
         $p(2) = \mathrm{P}(W_1 W_2) = \frac{4}{7} \cdot \frac{5}{8} = \frac{20}{56}$

      ii. plot the cumulative distribution function of X.



   (b) Repeat 4a if the marbles are drawn with replacement.

      $p(0) = \mathrm{P}(B_1 B_2) = \mathrm{P}(B_1 B_2 | B_1)\mathrm{P}(B_1) = \frac{3}{8} \cdot \frac{3}{8} = \frac{9}{64}$
      $p(1) = \mathrm{P}(\{B_1 W_2, W_1 B_2\}) = \mathrm{P}(B_1 W_2) + \mathrm{P}(W_1 B_2) = \frac{5}{8} \cdot \frac{3}{8} + \frac{3}{8} \cdot \frac{5}{8} = \frac{30}{64}$
      $p(2) = \mathrm{P}(W_1 W_2) = \frac{5}{8} \cdot \frac{5}{8} = \frac{25}{64}$



5. The probability that a student will pass a particular course is 0.4. Find the probability that, out of 5 students

   (a) none pass $= \binom{5}{0} 0.4^0\, 0.6^5 = 0.078$

   (b) one passes$= \binom{5}{1} 0.4^1\, 0.6^4 = 0.26$

   (c) at least one passes $= 1 - p(0) = 0.922$

6. (a) If each student in a class of 110 has the same probability, 0.8, of passing an examination, what is

      i. the expected number of passes?

         Let X = number that pass. Then $X \sim \text{Binomial}(n = 110, p = 0.8)$.
         $E(X) = np = 110 \cdot 0.8 = 88.0$.

      ii. the standard deviation of the number of passes?

         $\text{sd}(X) = \sqrt{\text{Var}(X)} = \sqrt{np(1-p)} = 4.195$

   (b) If each student in a college of 11000 has the same probability, 0.8, of graduating, what is

   i. the expected number of graduates? Same as above but $n = 11000$. $E(X) = 8800$
   ii. the standard deviation of the number of graduates? $\text{sd}(X) = 41.95$

7. Banach match problem. The pipe-smoking mathematician Banach has a matchbox in each of his two pockets. Whenever he needs a match he reaches into one of his pockets at random. Each matchbox starts out with N matches. At the moment Banach discovers one of the boxes is empty what is the probability that the other box contains exactly k matches (for $k = 0, 1, 2, \ldots, N$)? (Discovering a box to be empty means reaching for a box that contains zero matches — it does not mean taking the last match from a box.)

   Write an outcome as a sequence of L's and R's (i.e. LLRL... means go for left pocket, then left again, then right, then left, ...). Let's partition outcomes by whether the discovery of the empty box is in the left or right pocket. Then,

   $$\begin{aligned} p(k) &= p(k \text{ and last pick is L}) + p(k \text{ and last pick is R}) \\ &= 2p(k \text{ and last pick is L}) \qquad \text{(by symmetry of L and R)}. \end{aligned}$$

   So we reduce to dealing with the case where we end on L and there are k matches remaining in the R box. Therefore, outcomes contributing to $p(k$ and last pick is L) are sequences containing $N + 1$ L's, $N - k$ R's, and where the last letter is L. Fixing the final L, there are $\binom{2N-k}{N-k}$ ways of arranging the first $2N - k$ letters and each sequence has a probability of $\left(\frac{1}{2}\right)^{2N-k+1}$.
   Therefore, $p(k) = 2\binom{2N-k}{N-k}\left(\frac{1}{2}\right)^{2N-k+1} = \binom{2N-k}{N-k}\left(\frac{1}{2}\right)^{2N-k}$.

8. Compute the mean, sd, and the skewness for the following binomial distributions, and comment on the trends as n and p change:

   For Binomial(n, p), mean$= np$, sd$=\sqrt{np(1-p)}$, skewness$=(1-2p)/\sqrt{np(1-p)}$.

   (a) Binomial(100, 0.9) mean=90, sd=3.0, skew=-0.267
   (b) Binomial(100, 0.7) mean=70, sd=4.6, skew=-0.087
   (c) Binomial(100, 0.5) mean=50, sd=5.0, skew=0
   (d) Binomial(1000, 0.9) mean=900, sd=9.5, skew=-0.084
   (e) Binomial(1000, 0.7) mean=700, sd=14.5, skew=-0.028
   (f) Binomial(1000, 0.5) mean=500, sd=15.8, skew=0

   Skewness decreases as n increases and as p gets closer to $\frac{1}{2}$.

9. In a class of 20 students taking an examination,
   2 have probability 0.4 of passing;
   4 have probability 0.6 of passing;
   5 have probability 0.7 of passing;
   7 have probability 0.8 of passing;
   2 have probability 0.9 of passing.

   Each student either passes or doesn't. This is a binary outcome that can be modeled as a Bernoulli RV, $X_i \sim \text{Bernoulli}(p_i)$, where p_i is the probability that student i passes.

   (a) What is the expected number of passes?

   $$E(X_1 + X_2 + \ldots + X_{20}) = E(X_1) + \ldots E(X_{20}) = p_1 + \ldots p_{20} = 14.1.$$

   (b) What is the standard deviation of the number of passes?

   $\text{Var}(X_1 + X_2 + \ldots + X_{20}) = \text{Var}(X_1) + \ldots \text{Var}(X_{20}) = p_1(1-p_1) + \ldots p_{20}(1-p_{20}) = 3.79$. And sd $= \sqrt{\text{Var}} = 1.95$

10. A computer class has a limited number of terminals available for use. You notice that, on average, there is a 0.4 chance that there will be a free terminal each time you try to use a machine.

   (a) What is the average number of times you will have to try use a machine until you are successful?

   The number of tries until success is a geometric random variable with trial success probability $p = 0.4$. The mean is $1/p = 2.5$.

   (b) What is your chance of being successful the first time you try? 0.4

   (c) What is your probability of being successful the first time on each of three different occasions? $0.4^3 = 0.064$

11. (a) What is the mean and variance of a sum of n independent Bernoulli random variables, each with parameter p?

   mean= np, var=$np(1-p)$, see answer to question 9a.

   (b) What if they have different parameters, $(p_1, p_2, \ldots, p_n)$?

   mean= $\sum_{i=1}^{n} p_i$, var=$\sum_{i=1}^{n} p_i(1-p_i)$, see answer to question 9a.

   (c) What can you say if I now tell you that they are not independent?

   The mean is unaffected but we need further information to know if the variance will be affected.

12. Let $X \sim \text{Poisson}(\lambda)$, for some $\lambda > 0$.

   (a) Verify that the Poisson distribution has a valid probability mass function (non-negative and normalized).

   $p(k) = e^{-\lambda}\lambda^k/k!$ for $k = 0, 1, 2, \ldots$ and zero otherwise
   (1) $p(k) \geq 0$ since each factor in the pmf is > 0.
   (2) $\sum_{k=0}^{\infty} p(k) = e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k/k! = e^{-\lambda}e^{\lambda} = 1$, using the power series for $\exp(\lambda)$.

   (b) Find $E(X^2)$.

   $E(X^2) = \sum_{k=0}^{\infty} k^2 p(k) = e^{-\lambda} \sum_{k=0}^{\infty} k^2 \lambda^k/k! = e^{-\lambda} \sum_{k=0}^{\infty} \lambda \frac{d}{d\lambda} \lambda \frac{d}{d\lambda} \lambda^k/k! = e^{-\lambda} \lambda \frac{d}{d\lambda} \lambda \frac{d}{d\lambda} \sum_{k=0}^{\infty} \lambda^k/k! =$
   $e^{-\lambda}\lambda\frac{d}{d\lambda}\lambda\frac{d}{d\lambda}e^{\lambda} = e^{-\lambda}\lambda\frac{d}{d\lambda}(\lambda e^{\lambda}) = e^{-\lambda}\lambda(e^{\lambda} + \lambda e^{\lambda}) = \lambda(1+\lambda)$.

   Alternatively, looking up the variance for a Poisson RV we have $\text{Var}(X) = \lambda^2$. Use the fact that $\text{Var}(X) = E(X^2) - E(X)$, along with $E(X) = \lambda$ to solve for $E(X^2)$.

   (c) Suppose Z is another discrete random variable with $P(Z = 16) = 0.2$, and $P(Z = z) \propto P(X = z)$ whenever $z = \mathbb{R}\backslash\{16\}$.
   Find the probability mass function of Z.

   When $z \neq 16$ we have $P(Z = z) = ce^{-\lambda}\lambda^z/z!$, for some constant c. Use normalization condition to get constant: (add and subtract the poisson pmf term for $z = 16$)
   $1 = P(Z=0) + P(Z=1) + \ldots + P(Z=16) + \ldots = c(e^{-\lambda}\lambda^0/0! + e^{-\lambda}\lambda^1/1! + \ldots + e^{-\lambda}\lambda^{16}/16! + \ldots) + 0.2 - ce^{-\lambda}\lambda^{16}/16! = c + 0.2 - ce^{-\lambda}\lambda^{16}/16!$. Solving for c gives $c = 0.8/(1 - e^{-\lambda}\lambda^{16}/16!)$.

13. If X is a geometric random variable, show that

$$P(X = n + k \mid X > n) = P(X = k)$$

and give a verbal explanation of why this makes sense, given the situation that a geometric random variable describes.

A geometric RV corresponds to the experiment where we keep performing independent trials until we get a success. "Given $X > n$" means that we condition on the fact that we have already made it to the n^{th} trial without success. The probability of going an additional k trials to get a success is just the same as starting from scratch and needing to go k trials to get a success.

$$\begin{aligned} P(X = n + k \mid X > n) = \frac{P(X = n + k, X > n)}{P(X > n)} &= \frac{P(X = n + k)}{P(X > n)} \\ &= \frac{(1-p)^{n+k-1}p}{(1-p)^n} \\ &= (1-p)^{k-1}p = P(X = k). \end{aligned}$$

The second equality is because the event "$X = n + k$ AND $X > n$" is just the event "$X = n + k$" since we must have $X > n$ in order for $X = n + k$. And $P(X > n)$ is the probability of getting a failure in the first n trials, so it is equal to $(1-p)^n$.

14. There are n fish in a pond. You catch a fish at random, tag it and then throw it back. You keep doing this until you catch a fish that has already been tagged. What is the pmf for X, the number of fish you catch?

When $X = k$ it means that the first $k - 1$ fish are different and then next is one of the $k - 1$ that has already been tagged. Let E_i be the event that the i^{th} fish you catch is new. The probability of catching a new fish $k - 1$ times in a row is,

$$\begin{aligned} P(\overline{E_k}, E_{k-1}, \ldots E_1) &= P(\overline{E_k}, E_{k-1}, \ldots E_2 | E_1)P(E_1) \\ &= P(\overline{E_k}, E_{k-1}, \ldots E_3 | E_2, E_1)P(E_2|E_1)P(E_1) \\ &\vdots \\ &= P(\overline{E_k}|E_{k-1}, \ldots E_1)P(E_{k-1}|E_{k-2}, \ldots E_1) \cdots P(E_2|E_1)P(E_1). \end{aligned}$$

It's probably easier to think about it in words from right to left. There's a probability of 1 that the first catch is a new fish, probability $\frac{n-1}{n}$ that the second catch is new, ..., probability $\frac{n-k+2}{n}$ that the $k-1$ catch is new, and finally a probability $\frac{k-1}{n}$ that the k^{th} catch is one of the first $k - 1$ fish.

So the answer is $p(k) = \dfrac{k-1}{n}\dfrac{n}{n}\dfrac{n-1}{n} \cdots \dfrac{n-k+2}{n} = \dfrac{k-1}{n^k}\dfrac{n!}{(n-k+1)!}$.

15. We set up a new server that is to run for 1 year. From our experience with similar machines we estimate that crashes occur about once every 130,000 minutes (i.e. we monitored the other machines over a time T, observed k crashes and found that $k/T = 130{,}000$ mins.

Let X be the number of times our new server will crash.

(a) What kind of distribution should we use to model the random variable X? Explain your reasoning.

A Poisson RV is a good choice. If we imagine dividing the year up into small increments of time (e.g. minutes) it seems that there is a very small

chance of a crash during any given interval. Further, it seems reasonable to model that crashes in different time intervals are independent events. The "average rate" of crashes is 1 per 130,000 minutes so the expected number of crashes over a year is (1 year)$\times$1/(130,000 min)$\approx$ 4.0.
A reasonable model would therefore be $X \sim$ Poisson($\lambda = 4.0$).

(b) What is the probability that our server will not crash even once over the course of the year?

We need the probability of zero crashes. The Poisson pmf gives $p(0) = e^{-\lambda}\lambda^0/0! = e^{-4.0} = 0.019$, i.e. only about a 2% chance of not having a crash during the year.