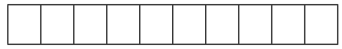
STAT 230 Book Problem Set

Zhilin (Catherine) Zhou

February 28, 2020

1. Suppose the letters of the word STATISTICS are arranged at random. Find the probability of the event G that the arragement begins and ends with S

Solution. We construct the arragements by filling ten boxes corresponding to the positions in the arragement



We can choose the position for the three S's in $\binom{10}{3}$ ways. For each of these choices, we can choose the positions for the three T's in $\binom{7}{3}$ ways, then we can place the two I's in $\binom{4}{2}$ ways, then the C in $\binom{2}{1}$ ways and finally the A in $\binom{1}{1}$ ways. The number of equally probable outcomes in the sample space is

$$\binom{10}{3}\binom{7}{3}\binom{4}{2}\binom{2}{1}\binom{1}{1} = \frac{10!7!4!2!1!}{3!7! \cdot 3!4! \cdot 2!2! \cdot 1!1! \cdot 1!0!} = \frac{10!}{3!3!2!1!1!}$$

■

2. The probability of a randomly selected male is colour-blind is 0.05, whereas the probability a female is colour-blind is only 0.0025. If the population is 50% male, what is the fraction that is colour-blind?

   Solution. We denote that
   C – the person selected is colour-blind
   M – the person selected is male
   F = $\overline{M}$ – the person selected is female.
   We need to find P(C). Given that

$$\begin{aligned} P(C|M) &= 0.05 \\ P(C|F) &= 0.0025 \\ P(M) &= 0.5 = P(F) \end{aligned}$$

   We need to know what is P(C). From Definition 4.1, we know that

$$\begin{aligned} P(M) &= \frac{P(C \cap M)}{P(C|M)} \\ P(F) &= \frac{P(C \cap F)}{P(C|F)} \\ C &= (C \cap M) \cup (C \cap \overline{M}) \end{aligned}$$

   Therefore, we have

$$\begin{aligned} P(C \cap M) &= 0.05 \times 0.5 = 0.025 \\ P(C \cap F) &= 0.0025 \times 0.5 = 0.00125 \\ P(C) &= 0.025 + 0.00125 = 0.02625 \end{aligned}$$

■

3. In an insurance portfolio 10% of the policy holders are in class A_1, 40% are in Class A_2, 10% are in Class A_3. The probability there is a claim on a Class A_1 policy in a given year is 0.10; similar probabilities for Classes A_2 and A_3 are 0.05 and 0.02. Find the probability that if a claim is made, it is made on a Class A_1 policy.

   Solution. We denote
   B – policy has a claim
   A_i – policy is of Class A_i, i = 1, 2, 3

$$
\begin{aligned}
P(A_1) &= 0.1 & \qquad P(B|A_1) &= 0.10 \\
P(A_2) &= 0.4 & \qquad P(B|A_2) &= 0.05 \\
P(A_3) &= 0.5 & \qquad P(B|A_3) &= 0.02
\end{aligned}
$$

   We need to find P(A_1|B). We know that

$$
\begin{aligned}
P(A_1|B) &= \frac{P(A_1 \cap B)}{P(B)} \\
P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B) \\
&= P(B|A_1) \cdot P(A_1) + P(B|A_2) \cdot P(A_2) + P(B|A_3) \cdot P(A_3) \\
&= 0.10 \times 0.1 + 0.05 \times 0.4 + 0.02 \times 0.5 = 0.01 + 0.02 + 0.01 = 0.04 \\
P(A_1 \cap B) &= P(B|A_1) \cdot P(A_1) = 0.10 \times 0.1 = 0.01 \\
\implies P(A_1|B) &= \frac{0.01}{0.04} = 0.25
\end{aligned}
$$

■

4. **Testing for HIV**

Tests used to diagnose medical conditions are often imperfect, and give false positive or false negative results, as described in Problem 2.6 of Chapter 2. A fairly cheap blood test for the Human Immunodeficiency Virus (HIV) that causes AIDS (Acquired Immune Deficiency Syndrome) has the following characteristics: the false negative rate is 2% and the false positive rate is 0.5%. It is assumed that around 0.04% of Canadian males are infected with HIV. Find the probability that if a male tests positive for HIV, he actually has HIV.

Solution. We can define

$$\begin{aligned} A &= \text{male has HIV} \\ B &= \text{blood test is positive} \end{aligned}$$

We need to find the value of $P(A|B)$ (when we know that the blood test is positive, it is a male). We know that

$$\begin{aligned} P(A) &= 0.0004 & P(\overline{A}) &= 0.9996 \\ P(B|A) &= 0.98 & P(B|\overline{A}) &= 0.05 \end{aligned}$$

Since by Theorem 4.2, $P(AB) = P(A)P(B|A)$, we can conclude that

$$\begin{aligned} P(AB) &= 0.0004 \times 0.98 \approx 0.000392 \\ P(\overline{A}B) &= P(\overline{A})P(B|\overline{A}) = 0.9996 \times 0.05 \approx 0.004998 \\ P(B) &= P(AB) + P(\overline{A}B) \approx 0.00539 \\ P(A|B) &= \frac{P(AB)}{P(B)} \approx 0.0727 \end{aligned}$$

■

5. Find

$$\sum_{x=0}^{\infty} x(x-1)\binom{a}{x}\binom{b}{n-x}$$

Solution. Since for $x = 0$ or $x = 1$ the term becomes 0, we can start the sumation with $x = 2$.

$$\begin{aligned}
\sum_{x=0}^{\infty} x(x-1)\binom{a}{x}\binom{b}{n-x} &= \sum_{x=2}^{\infty} x(x-1)\frac{a!}{x(x-1)(x-2)!(a-x)!}\binom{b}{n-x} \\
&= \sum_{x=2}^{\infty} \frac{a!}{(x-2)!(a-x)!}\binom{b}{n-x} \\
&= \sum_{x=2}^{\infty} \frac{a(a-1)(a-2)!}{(x-2)![(a-2)-(x-2)]!}\binom{b}{n-x} \\
&= a(a-1)\sum_{x=2}^{\infty}\binom{a-2}{x-2}\binom{b}{(n-2)-(x-2)} \\
&= a(a-1)\binom{a+b-2}{n-2} \qquad \text{(by Hypergeometric Identity)}
\end{aligned}$$

■

6. In Lotto $6/49$ a player selects a set of six numbers (with no repeats) from the set $\{1, 2, \ldots, 40\}$. In the lottery draw six numbers are selected at random. Find the probability function for X, the number from your set which are drawn.

Solution. Think of your numbers as S (success) objects and the remainder as F (failure) objects. Then X has a Hypergeometric distribution with $N = 49, r = 6, n = 6$, so we have

$$f(x) = P(X = x) = \frac{\binom{6}{x}\binom{43}{6-x}}{\binom{49}{6}} \text{ for } x = 0, 1, \ldots, 6$$

■

7. Suppose we have 15 cans of soup with no labels, but 6 are tomato and 9 are pea soup. We randomly pick 8 cans and open them. Find the probability three of them are tomato.

   Solution. The correct solution uses the Hypergeometric distribution. and is (with X = number of tomato soup cans picked)

$$P(X=3) = \frac{\binom{6}{3}\binom{9}{5}}{\binom{15}{8}} \approx 0.3916$$

   If we incorrectly used the Binomial distribution, we would obtain

$$\binom{8}{3}(\frac{6}{15})^3(\frac{9}{15})^5 \approx 0.2787$$

   However, if we had 1500 cans: 600 tomato and 900 pea, we are not likely to get the same can again even if we did replace each of the 8 cans after opening it. (Put another way, to probability we get a tomato soup on each pick is very close to 0.4, regardless of what the other picks give). The Hypergeometric probability gives

$$\frac{\binom{600}{3}\binom{900}{5}}{\binom{1500}{8}} \approx 0.2974$$

   The Binomial probability,

$$\binom{8}{3}(\frac{600}{1500})^3(\frac{900}{1500})^5 \approx 0.2787$$

   which is a very good approximation. ■

8. A specific blood type T is 0.08 (8%). For blood donation purposes it is necessary to find 5 people with type T blood. If randomly selected individuals form the population are tested one after another, then

   (a) What is the probability y persons have to be teseted to get 5 type T persons?

   Solution. Let a type T person as a success (S) and a non-type T as an F. Let Y = number of persons who have to be tested and let X = number of non-type T persons in order to get 5 S's. Then X has a Negative Binomial distribution with $k = 5$ and $p = 0.8$ and

$$P(X = x) = f(x) = \binom{x+4}{x}(0.08)^5(0.92)^x \text{ for } x = 0, 1, 2, \ldots$$

   We are actually asked here about $Y = X + 5$. Thus,

$$\begin{aligned} P(Y = y) &= P(X = y - 5) \\ &= f(y - 5) \\ &= \binom{y-1}{y-5}(0.08)^5(0.92)^{y-5} \text{ for } y = 5, 6, 7, \ldots \end{aligned}$$

■

   (b) What is the probability that over 80 people have to be tested?

   Solution.

$$\begin{aligned} P(Y > 80) &= P(X > 75) \\ &= 1 - P(X \le 75) \\ &= 1 - \sum_{x=0}^{75} f(x) \\ &= 0.2235 \end{aligned}$$

■

9. There are 200 people at party. What is the probability that 2 of them were born on January 1?

Solution. Assuming all days of the year are equally likely for a birthday (and ignore February 29) and that the birthdays are independent (e.g. no twins!) We can use Bionmial distribution with $n = 200$ and $p = \frac{1}{365}$ for X = number born on January 1, giving

$$f(2) = \binom{200}{2}\left(\frac{1}{365}\right)^2\left(1 - \frac{1}{365}\right)^{198} = 0.0876767$$

Since n is large and p is close to 0, we can use the Poisson distribution to approximate this Binomial probability, with

$$\mu = np = \frac{200}{365}$$

giving

$$f(2) = \frac{(\frac{200}{365})^2 e^{-\frac{200}{365}}}{2!} = 0.086791$$

■

10. At a nuclear power station an average of 8 leaks of heavy water are reported per year. Find the probability of 2 or more leaks in 1 month, if leaks follow a Poisson process.

Solution. A month is $\frac{1}{12}$ of a year. Let X be the number of leaks in one month. Then X has the Poisson distribution with $\lambda = 8$ and $t = \frac{1}{12}$, so $\mu = \lambda t = \frac{8}{12}$. Thus,

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) \\ &= 1 - [f(0) + f(1)] \\ &= 1 - [\frac{(\frac{8}{12})^0 e^{-\frac{8}{12}}}{0!} + \frac{(\frac{8}{12})^1 e^{-\frac{8}{12}}}{1!}] \\ &= 0.1443 \end{aligned}$$

■

11. A very large (essentially infinite) number of ladybugs is released in a large orchard. They scatter randomly so that on average a tree has 6 ladybugs on it.Trees are all the same size:

(a) Find the probability a tree has > 3 ladybugs on it.

Since the ladybugs are **randomly scattered**, we can use the Poisson distirbution to solve this. In this case, $\lambda = 6$ and $v = 1$ (i.e. any tree has a "volume" of one unit), so $\mu = 6$ and

$$
\begin{aligned}
P(X > 3) = 1 - P(X \leq 3) &= 1 - [f(0) + f(1) + f(2) + f(3)] \\
&= [\frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!} + \frac{6^3 e^{-6}}{3!}] \\
&= 0.8488
\end{aligned}
$$

(b) When 10 trees are picked at random, what is the probability 8 of these trees have > 3 ladybugs on them?

We can use the Binomial Distribution where "success" means to have > 3 ladybugs on a tree. We have $n = 10$, and

$$f(8) = \binom{10}{8}(0.8488)^8(1 - 0.8488)^{10-8} = 0.2772$$

(c) Trees are checked until 5 wtih > 3 ladybugs are found. Let X be the total number of trees checked. Find the probability function, $f(x)$

We can use the Negative Binomial Distribution. We need the number of success, k, to be 5, and the number of failures to be $(x - 5)$. Then,

$$f(x) = \binom{x-1}{4}(0.8488)^5(1 - 0.8488)^{x-5} \text{ for } x = 5, 6, 7, \ldots$$

(d) Find the probability a tree with > 3 ladybugs on it has exactly 6.

This is a conditional probability. Let A = {6 ladybugs} and B = {a tree with > 3 ladybugs}. Then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\frac{6^6 e^{-6}}{6!}}{0.8488} = 0.1892$$

(e) On 2 trees there are a total of t ladybugs. Find the probability that x of these are on the first of these 2 trees.

This is also a conditional probability. Let A ={x on the first tree}, B ={$t - x$ on second tree}, and C ={2 trees has a total of t ladybugs}. Then we have

$$
\begin{aligned}
P(X = x) = P(A|C) &= \frac{P(AC)}{P(C)} = \frac{P(AB)}{P(C)} \\
&= \frac{P(A)P(B)}{P(C)}
\end{aligned}
$$

We can use Poisson Distribution at this point to calculate each, with $\mu = 6 \times 2 = 12$in the denominator since there are 2 trees.

$$
\begin{aligned}
P(A|C) &= \frac{\left(\frac{6^x e^{-6}}{x!}\right)\left(\frac{6^{t-x} e^{-6}}{(t-x)!}\right)}{\frac{12^t e^{-12}}{t!}} \\
&= \frac{t!}{x!(t-x)!}\left(\frac{6}{12}\right)^x \left(\frac{6}{12}\right)^{t-x} \\
&= \binom{t}{x}\left(\frac{1}{2}\right)^x \left(1-\frac{1}{2}\right)^{t-x} \text{ for } x = 0, 1, \ldots,\ t
\end{aligned}
$$

12. **Expected Winnings in a Lottery**
A small lottery sells 1000 tickets numbered 000, 001,. . . ,999; the tickets cost \$10 each. When all the tickets have been sold, the draw takes place. This consists of a single ticket from 000 to 999 being chosen at random. For ticket holders the prize structure as follows:

- Your ticket is drawn – win \$5000
- Your ticket has the same first two numbers as the winning ticket – win \$100
- Your ticket has the same first number as the winning ticket – win \$10
- All other cases – win nothing.

Let the random variable X be the winnings from a given ticket. Find $E(X)$

Solution. The possible values for X are 0, 10, 100, 500 (dollars). First, we need to find the probability function for X. We find (make sure you can do this) that $f(x) = P(X = x)$ has values

$$f(0) = 0.9, \qquad f(10) = 0.09, \qquad f(100) = 0.009, \qquad f(5000) = 0.001$$

The expected winnings are thus the expected value of X, or

$$E(X) = \sum_{\text{all } x} xf(x) = 10 \times 0.09 + 100 \times 0.009 + 5000 \times 0.001 = \$6.80$$

Thus, the gross expected winnings per ticket are \$6.80. However, since a ticket costs \$10 your expected net winings are negative, -\$3.20 (that is, an expected loss of \$3.20) ■

13. **Diagnostic Medical Tests**

Often there are cheaper, less accurate tests for diagnosing the presence of some conditions in a person, along with more expensive, accurate tests. Suppose we have two cheap tests and one expensive test, with the following characteristics. All three tests are positive if a person has the condition (there are no "false negatives"), but the cheap tests give "false positives". Let a person be chosen at random, and let D = {person has the condition}. For the three tests the probability of a false positive and cost are:

Test	P(positive test|$\bar{D}$)	Cost (in dollars)
1	0.05	5
2	0.03	8
3	0	40

We want to check a large number of people for the condition, and have to choose among three testing strategies:

(a) Use Test 1, followed by Test 3 if Test 1 is positive

(b) Use Test 2, followed by Test 3 if Test 2 is positive

(c) Use Test 3

Determine the expeceted cost per person under each of strategies 1, 2, and 3. We will then choose the strategy with the lowest expecetd cost. It is known that about 0.001 of the population have the condition

$$P(D) = 0.001, \qquad P(\bar{D}) = 0.999$$

Assume that given D or $\bar{D}$, tests are independent of one another.

Solution. For a person tested chosen at random and tested, define the random variable X as follows:

$$X = 1 \qquad \text{if the initial test is negative}$$
$$X = 2 \qquad \text{if the initial test is positive}$$

Let $g(x)$ be the total cost of testing the person, the expected cost per person is then

$$E[g(X)] = \sum_{x=1}^{2} g(x)f(x)$$

The probability function $f(x)$ for X and the function $g(x)$ differ for strategies 1, 2, 3. Consider for example strategy 1. Then

$$\begin{aligned} P(X=2) &= P(\text{initial test positive}) \\ &= P(D) + P(\text{positive}|\bar{D})P(\bar{D}) \\ &= 0.001 + (0.005)(0.999) \\ &= 0.0510 \end{aligned}$$

The rest of the probabilities, associated with the values of $g(X)$ and $E[g(X)]$ are obtained below.

(a) Strategy 1

$$\begin{aligned} f(2) &= 0.0510 \text{ obtained above} \\ f(1) &= P(X=1) = 1 - f(2) = 1 - 0.0510 = 0.949 \\ g(1) &= 5 \qquad g(2) = 45 \\ E[g(X)] &= 5(0.949) + 45(0.0510) = \$7.04 \end{aligned}$$

(b) Strategy 2

$$\begin{aligned} f(2) &= 0.001 + (0.03)(0.999) = 0.03097 \\ f(1) &= 1 - f(2) = 0.96903 \\ g(1) &= 8 \qquad g(2) = 48 \\ E[g(X)] &= 8(0.96903) + 48(0.03097) = \$9.2388 \end{aligned}$$

(c) Strategy 3

$$\begin{aligned} f(2) &= 0.001, \qquad f(1) = 0.999 \\ g(2) &= g(1) = 40 \\ E[g(X)] &= \$40.00 \end{aligned}$$

Therefore, the cheapest strategy is strategy 1. ■

14. Let $X \sim Binomial(n, p)$. Find $E(X)$.

Solution.

$$\begin{aligned}\mu = E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}\end{aligned}$$

When $x = 0$ the value of the expression is 0. We can therefore begin our sum at $x = 1$. Provided $x \neq 0$, we can expand $x!$ as $x(x-1)!$. Therefore

$$\begin{aligned}\mu &= \sum_{x=1}^{n} \frac{n(n-1)!}{(x-1)![(n-1)-(x-1)]!} p p^{x-1} (1-p)^{(n-1)-(x-1)} \\ &= np(1-p)^{n-1} \sum_{x=1}^{n} \binom{n-1}{x-1} \left(\frac{p}{1-p}\right)^{x-1}\end{aligned}$$

Let $y = x - 1$ in the sum to get

$$\begin{aligned}\mu &= np(1-p)^{n-1} \sum_{y=0}^{n-1} \binom{n-1}{y} \left(\frac{p}{1-p}\right)^{y} \\ &= np(1-p)^{n-1} \left(1 + \frac{p}{1-p}\right)^{n-1} \quad \text{by the Binomial Theorem} \\ &= np(1-p)^{n-1} \frac{(1-p+p)^{n-1}}{(1-p)^{n-1}} \\ &= np\end{aligned}$$

■

15. Let X have a Poisson distribution where λ is the average rate of occurance and the time interval is of length t. Find $\mu = E(X)$

Solution. Since the probability function of X is

$$f(x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!} \text{ for } x = 0, 1, \ldots$$

then

$$\mu = E(X) = \sum_{x=0}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

As in the Binomial example, we can eliminate the term when $x = 0$ and expand $x!$ to $x(x-1)!$ for $x = 1, 2, \ldots$ to obtain

$$\begin{aligned}
\mu &= \sum_{x=1}^{\infty} x \frac{(\lambda t)^x e^{-\lambda t}}{x(x-1)!} \\
&= \sum_{x=1}^{\infty} (\lambda t) e^{-\lambda t} \frac{(\lambda t)^{x-1}}{(x-1)!} \\
&= (\lambda t) e^{-\lambda t} \sum_{x=1}^{\infty} \frac{(\lambda t)^{x-1}}{(x-1)!} \\
&= (\lambda t) e^{-\lambda t} \sum_{y=0}^{\infty} \frac{(\lambda t)^{y}}{y!} \text{ letting } y = x - 1 \text{ in the sum} \\
&= (\lambda t) e^{-\lambda t} e^{\lambda t} \text{ since } e^x = \sum_{y=0}^{\infty} \frac{x^y}{y!} \\
&= \lambda t
\end{aligned}$$

Note that we used the symbol $\mu = \lambda t$ earlier in connection with the Poisson model: this was because we knew (but couldn't show until now) that $E(X) = \mu$. ■

16. Suppose X is a random variable with probability function given by

x	1	2	3	4	5	6	7	8	9	Total
$f(x)$	0.07	0.10	0.12	0.13	0.16	0.13	0.12	0.10	0.07	1

Find $E(X)$ and $\text{Var}(X)$.

Solution.

$$\begin{aligned}\mu &= E(X)\\ &= 1(0.07)+2(0.10)+3(0.12)+4(0.13)+5(0.16)\\ &+6(0.13)+7(0.12)+8(0.10)+9(0.07)\\ &= 5\end{aligned}$$

$E(X) = 5$ should be obvious by looking at the histogram. If a probability historgram is symmetric about the line $x = \mu$ then $E(X) = \mu$ with any calculation.

Without doing any calculations we also know that $\text{Var}(X) = \sigma^2 \leq 16$. This is because the possible values of X are {1,2,...,9} and so the maximum possible value for $(X-\mu)^2$ is $(9-5)^2$ or $(1-5)^2 = 16$. Therefore,

$$\begin{aligned}\text{Var}(X) = E[(X-5)^2] &= \sum_{x=1}^{9}(x-5)^2 P(X=x)\\ &\leq \sum_{x=1}^{9}(9-5)^2 P(X=x) = 16\sum_{x=1}^{9} P(X=x) = 16(1) = 16\end{aligned}$$

An expected value of a funciton, say $E[g(X)]$ is always somewhere between the minimum and the maximum value of the function $g(x)$ so in this case $0 \leq \text{Var}(X) \leq 16$. Since

$$\begin{aligned}E(X^2) &= (1)^2(0.07)+(2)^2(0.10)+(3)^2(0.12)+(4)^2(0.13)+(5)^2(0.16)\\ &+(6)^2(0.13)+(7)^2(0.12)+(8)^2(0.10)+(9)^2(0.07)\\ &= 30.26\end{aligned}$$

Therefore,

$$\begin{aligned}\sigma^2 = \text{Var}(X) &= E(X^2)-\mu^2\\ &= 30.26-(5)^2\\ &= 5.26\end{aligned}$$

and

$$\begin{aligned}\sigma &= \sqrt{\text{Var}(X)}\\ &= \sqrt{5.26}\\ &= 2.2935\end{aligned}$$

■

17. **Variance of Binomial random variable**
Let $X \sim Binomial(n, p)$. Find $\text{Var}(X)$.

Solution. The probability function for X is

$$f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \text{ for } x = 0, 1, \ldots, n$$

so we'll use formula (2) above,

$$E[X(X-1)] = \sum_{x=0}^{n} x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

If $x = 0$ or $x = 1$ the value of the term is 0, so we can begin summing at $x = 2$. For $x \neq 0$ or 1, we can expand the $x!$ as $x(x-1)(x-2)!$ Therefore,

$$E[X(X-1)] = \sum_{x=2}^{n} \frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x}$$

Now re-group to fit the Binomial Theorem, since that was the summation technique used to show $\sum f(x) = 1$ and to derive $\mu = np$

$$\begin{aligned}
E[X(X-1)] &= \sum_{x=2}^{n} \frac{n(n-1)(n-2)!}{(x-2)![(n-2)-(x-2)]!}p^2p^{x-2}(1-p)^{(n-2)-(x-2)} \\
&= n(n-1)p^2(1-p)^{n-2}\sum_{x=2}^{n}\binom{n-2}{x-2}\left(\frac{p}{1-p}\right)^{x-2}
\end{aligned}$$

Let $y = x - 2$ in the sum, giving

$$\begin{aligned}
E[X(X-1)] &= n(n-1)p^2(1-p)^{n-2}\sum_{y=0}^{n-2}\binom{n-2}{y}\left(\frac{p}{1-p}\right)^{y} \\
&= n(n-1)p^2(1-p)^{n-2}\left(1+\frac{p}{1-p}\right)^{n-2} \\
&= n(n-1)p^2(1-p)^{n-2}\frac{(1-p+p)^{n-2}}{(1-p)^{n-2}} \\
&= n(n-1)p^2
\end{aligned}$$

Then

$$\begin{aligned}
\sigma^2 &= E[X(X-1)] + \mu - \mu^2 \\
&= n(n-1)p^2 + np - (np)^2 \\
&= n^2p^2 - np^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}$$

■

18. **Variance of Poisson random variable**
Suppose X has a Poisson(μ) distribution. Find $\mathrm{Var}(X)$.

Solution. The probability function for X is

$$f(x) = \frac{\mu^x e^{-\mu}}{x!} \text{ for } x = 0, 1, 2, \ldots$$

from which we obtain

$$\begin{aligned} E[X(X-1)] &= \sum_{x=0}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x!} \\ &= \sum_{x=2}^{\infty} x(x-1)\frac{\mu^x e^{-\mu}}{x(x-1)(x-2)!} \text{ setting the lower limit to 2 and expanding } x! \\ &= \mu^2 e^{-\mu} \sum_{x=2}^{\infty} \frac{\mu^{x-2}}{(x-2)!} \end{aligned}$$

Let $y = x - 2$ we get

$$\begin{aligned} E[X(X-1)] &= \mu^2 e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} = \mu^2 e^{-\mu} e^{\mu} = \mu^2 \\ \sigma^2 &= E[X(X-1)] + \mu - \mu^2 \\ &= \mu^2 + \mu - \mu^2 = \mu \end{aligned}$$

■

19. Let X be a continuous random variable with probability density function

$$f(x) = \begin{cases} kx^2 & 0 < x \leq 1 \\ k(2-x) & 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find:

(a) the constant k

(b) the cumulative distribution function $F(x) = P(X \leq x)$

(c) $P(0.5 < X < 1.5)$

Solution. (a) When finding teh area of region bounded by different functions we split the integral into pieces.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x)dx \\ &= \int_{-\infty}^{0} 0dx + \int_{0}^{1} kx^2dx + \int_{1}^{2} k(2-x)dx + \int_{2}^{\infty} 0dx \\ &= 0 + k\int_{0}^{1} x^2dx + k\int_{1}^{2}(2-x)dx + 0 \\ &= k\frac{x^3}{3}|_0^1 + k\left(2x - \frac{x^2}{2}|_1^2\right) \\ &= \frac{5k}{6} \text{ and therefore } k = \frac{6}{5} \end{aligned}$$

(b) let us start with the easy pieces (which are unfortunately often left out) first:

$F(x) = P(X \leq x) = 0$ if $x \leq 0$

$F(x) = P(X \leq x) = 1$ if $x \geq 2$ since the probability density function equals 0 for all $x \geq 2$

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)dz = 0 + \int_{-\infty}^{x} \frac{6}{5}z^2dz = \frac{6}{5}\frac{x^3}{3}|_0^x = \frac{2x^3}{5}$ if $0 < x < 1$

$F(x) = P(X \leq x) = 0 + \int_{0}^{1} \frac{6}{5}z^2dz + \int_{1}^{x} \frac{6}{5}(2-z)dz = \frac{6}{5}\frac{x^3}{3}|_0^1 + \frac{6}{5}(2x - \frac{x^2}{2})|_1^x$

$= \frac{12x - 3x^2 - 7}{5}$ if $1 < x < 2$

Therefore,

$$F(x) = P(X \leq x) = \begin{cases} 0 & x \leq 0 \\ \frac{2x^3}{5} & 0 < x \leq 1 \\ \frac{12x-3x^2-7}{5} & 1 < x < 2 \\ 1 & x \geq 2 \end{cases}$$

As a rough check, $F(x)$ should have the same value as we approach each boudary point from above and from below.

For example,

$$\begin{aligned}
&\text{as } x \to 0^+,\ \frac{2x^3}{5} \to 0\\
&\text{as } x \to 1^-,\ \frac{2x^3}{5} \to \frac{2}{5}\\
&\text{as } x \to 1^+,\ \frac{12x - 3x^2 - 7}{5} \to \frac{2}{5}\\
&\text{as } x \to 2^-,\ \frac{12x - 3x^2 - 7}{5} \to 1
\end{aligned}$$

This quick check won't prove your answer is right, but will detect many careless errors.

(c)

$$\begin{aligned}
P(0.5 < X < 1.5) &= \int_{0.5}^{1.5} f(x)dx = F(1.5) - F(0.5)\\
&= \frac{12(1.5) - 3(1.5)^2 - 7}{5} - \frac{2(0.5^3)}{5} = 0.8
\end{aligned}$$

■

20. For the problem above, find

    (a) the 0.4 quantile (40th percentile) of the distribution

    (b) the median of the distribution

Solution. (a) Since $F(1) = 0.4$, the 0.4 quantile is equal to 1.

(b) The median is the solution to

$$F(x) = \frac{12x - 3x^2 - 7}{5} = 0.5$$

or

$$24x - 6x^2 - 19 = 0$$

which has two solutions. Since $F(1) = 0.4$ we know that the mexian lie betwene 1 and 2 and we choose the solution $x \approx 1.087$. The median is approximately equal to 1.087.

■

21. We have

$$f(x) = \begin{cases} \frac{1}{4} & 0 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

and

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \frac{x}{4} & 0 < x < 4 \\ 1 & x \geq 4 \end{cases}$$

Find the probability density function of $Y = X^{-1}$

Solution. Step 1 from above becomes:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^{-1} \leq y) \\ &= P(X \geq y^{-1}) = 1 - P(X < y^{-1}) \\ &= 1 - F_X(y^{-1}) \end{aligned}$$

For step (2), we can substitute $\frac{1}{y}$ in place of x in $F_X(x))$ giving:

$$F_Y(y) = 1 - \frac{y^{-1}}{4} = 1 - \frac{1}{4y}$$

and then differentiate to obtain the probability density function

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{4y^2} \text{ for } y \geq \frac{1}{4}$$

(Note that as x goes from 0 to 4, $y = \frac{1}{x}$ goes between ∞ and $\frac{1}{4}$).

Alternatively, and a little more generally, we can use the chain rule:

$$\begin{aligned} f_Y(y) = \frac{d}{dy} F_Y(y) &= \frac{d}{dy}[1 - F_X(y^{-1})] \\ &= -f_X(y^{-1}) \frac{d}{dy}(y^{-1}) \text{ since } \frac{d}{dx} F_X(x) = f_X(x) \\ &= -f_X(y^{-1})(-y^{-2}) = \frac{1}{4}(-y^{-2}) \\ &= \frac{1}{4y^2} \text{ for } y \geq \frac{1}{4} \end{aligned}$$

■

22. For the earlier spinner exmaple,

$$f(x) = \begin{cases} \frac{1}{4} & 0 < x \leq 4 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value and variance.

Solution.

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx = 0 + \int_{0}^{4} x \frac{1}{4} dx + 0 = \frac{1}{4} \left(\frac{x^2}{2} \right) |_0^4 = 2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = 0 + \int_{0}^{4} x^2 \frac{1}{4} dx + 0 = \frac{1}{4} \left(\frac{x^3}{3} \right) |_0^4 = \frac{16}{3}$$

$$\sigma^2 = \text{Var}(X) = E(X^2) - \mu^2 = \frac{16}{3} - (2)^2 = \frac{4}{3}$$

■

23. Let X have probability density funciton

$$f(x) = \begin{cases} \frac{6x^2}{5} & 0 < x \leq 1 \\ \frac{6}{5}(2-x) & 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value and variance.

Solution.

$$\begin{aligned}
\mu = E(X) &= \int_{-\infty}^{\infty} xf(x)dx = 0 + \int_0^1 x\frac{6}{5}x^2 dx + \int_1^2 x\frac{6}{5}(2-x)dx + 0 \\
&= \frac{6}{5}\left[\frac{x^4}{4}|_0^1 + \left(x^2 - \frac{x^3}{3}|_1^2\right)\right] = \frac{11}{10} = 1.1 \\
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)dx = 0 + \int_0^1 x^2\frac{6}{5}x^2 dx + \int_1^2 x^2\frac{6}{5}(2-x)dx + 0 \\
&= \frac{6}{5}\left[\frac{x^5}{5}|_0^1 + 2\left(\frac{x^3}{3}\right)|_1^2 - \frac{x^4}{4}|_1^2\right] = \frac{67}{50} \\
\sigma^2 = \mathrm{Var}(X) &= E(X^2) - \mu^2 = \frac{67}{50} - \left(\frac{11}{10}\right)^2 \\
&= \frac{13}{100} \\
&= 0.13
\end{aligned}$$

■