1. (10) Show that the variance of a random variable can be calculated by using the formula

$$\operatorname{Var}(X) = \mathrm{E}\left(X^2\right) - \left(\mathrm{E}(X)\right)^2$$

$$\begin{aligned}\operatorname{Var}(X) &= \mathrm{E}\left[\left(X - \mathrm{E}(X)\right)^2\right] = \mathrm{E}\left[X^2 - 2X\mathrm{E}(X) + \left(\mathrm{E}(X)\right)^2\right] \\ &= \mathrm{E}\left(X^2\right) - 2\mathrm{E}(X)\,\mathrm{E}(X) + \left(\mathrm{E}(X)\right)^2 \\ &= \mathrm{E}\left(X^2\right) - \left(\mathrm{E}(X)\right)^2\end{aligned}$$

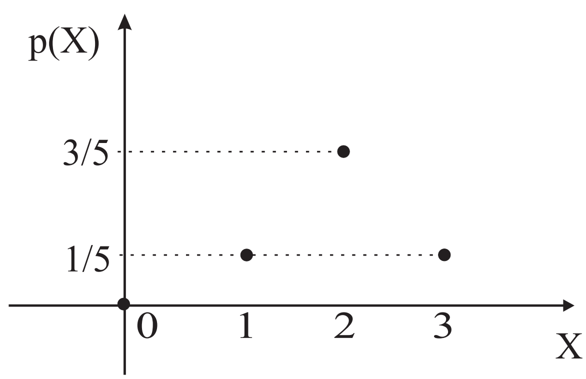
2. (20) An urn contains 4 red marbles and 2 black marbles. Three marbles are picked without replacement. Let X be the random variable which equals the number of red marbles picked.

   (a) Find and plot the values of the probability mass (density) function, p, of the random variable X.

$$p(0) = P\{X = 0\} = 0 \qquad p(1) = P\{X = 1\} = \frac{\binom{4}{1}\binom{2}{2}}{\binom{6}{3}} = \frac{1}{5}$$

$$p(2) = \frac{\binom{4}{2}\binom{2}{1}}{\binom{6}{3}} = \frac{3}{5} \qquad p(3) = P\{X = 3\} = \frac{\binom{4}{3}\binom{2}{0}}{\binom{6}{3}} = \frac{1}{5}$$

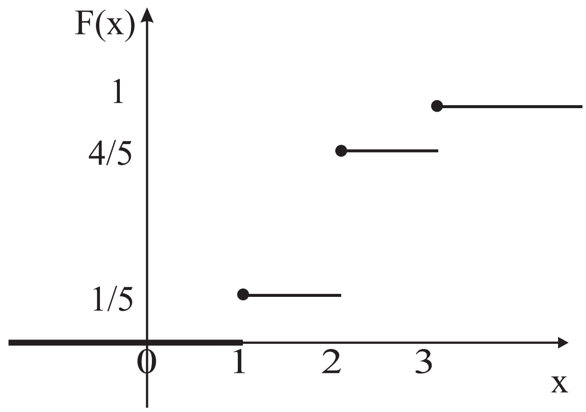
   The plot of this mass density function is shown below.



(b) Find and plot the values of the cumulative distribution function, F, of the random variable X.

$$F(x) = P\{X \leq x\} \text{ so for the values of } X \text{ we have}$$
$$F(0) = 0, \qquad F(1) = \frac{1}{5}, \qquad F(2) = \frac{4}{5}, \qquad F(3) = 1$$

A plot of $F(x)$ is shown below



(c) Compute the expected value and variance of X.

$$\begin{aligned} \mathrm{E}(X) &= \sum_{i=0}^{3} x_i p(i) = 0(0) + 1\left(\frac{1}{5}\right) + 2\left(\frac{3}{5}\right) + 3\left(\frac{1}{5}\right) \\ &= \frac{1+6+3}{5} = 2 \end{aligned}$$

$$\begin{aligned} \mathrm{Var}(X) &= \mathrm{E}\left(X^2\right) - (\mathrm{E}(X))^2 = \sum_{i=0}^{3} x_i^2 p(i) - 4 \\ &= 1\left(\frac{1}{5}\right) + 4\left(\frac{3}{5}\right) + 9\left(\frac{1}{5}\right) - 4 = \frac{1+12+9}{5} - 4 = \frac{2}{5} \end{aligned}$$

(d) Compute the expected value of $\ln(1+X)$.

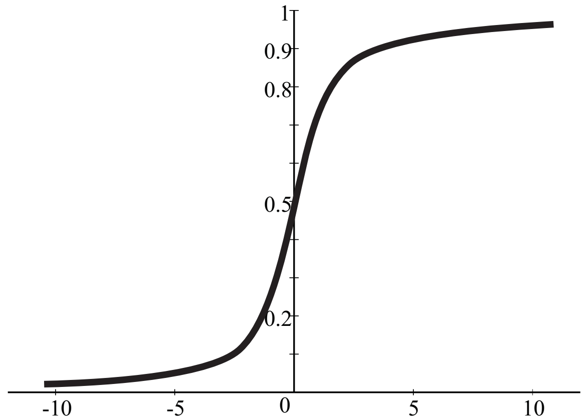
$$\begin{aligned} \mathrm{E}(\ln(1+X)) &= \sum_{i=0}^{3} \ln(1+x_i)\, p(i) = \ln(2)\frac{1}{5} + \ln(3)\frac{3}{5} + \ln(4)\frac{1}{5} \\ &= \frac{\ln 2 + \ln 27 + \ln 4}{5} = \frac{\ln(216)}{5} \approx 1.075 \end{aligned}$$

3. (10) The average number of automobile accidents in the Bryan College Station area is 5 per day. What is the probability that in any given day the number of such accidents will exceed 4? Be sure to explain your reasoning.

Think of this as a large number of independent Bernoulli trials. Each trial being a trip by an automobile. Assume there are n such trips, then the probability of any trip generating an accident is $\frac{5}{n}$. Thus, the expected number of accidents is $n\frac{5}{n} = 5$. Now approximate the random variable which equals the number of accidents per day with a random variable, N, which has a Poisson distribution with parameter 5. Thus, the probability that the number of accidents will exceed 4 is approximately

$$
\begin{aligned}
P\{N > 4\} &= 1 - [P\{N = 0\} + P\{N = 1\} + P\{N = 2\} \\
&\qquad + P\{N = 3\} + P\{N = 4\}] \\
&= 1 - \left[e^{-5}\left(1 + \frac{5}{1} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!}\right)\right] \\
&\approx 1 - [0.4405] \\
&= 0.5595
\end{aligned}
$$

4. (10) The plot below is that of the cumulative distribution function of a random variable X.



(a) Can it be inferred from the plot as to whether or not X is a discrete or a continuous random variable? An explanation is required.

Yes, this is the cdf of a continuous random variable. The graph is smooth so the function $F(x)$ is differentiable. Call its derivative $f(x)$. Then

$$\begin{aligned} \mathrm{P}\{X \le x\} &= F(x) = F(x) - \lim_{b \to -\infty} F(b) \\ &= \lim_{b \to -\infty} \int_b^x f(t)\, dt \\ &= \int_{-\infty}^{x} f(t)\, dt \end{aligned}$$

From this one easily sees that $\mathrm{P}\{a < X \le b\} = \int_a^b f(t)\, dt$

(b) Estimate $P\left\{X \le \frac{1}{2}\right\}$.

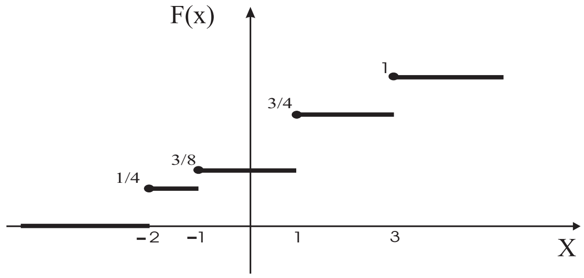
The graph in the plot above crosses the y axis at $y = 0.5$. Thus,

$$\mathrm{P}\left\{X \le \frac{1}{2}\right\} = F(1/2) \approx 0.55$$

(c) Estimate $P\{X > 5\}$.

$$\begin{aligned} P\{X > 5\} &= 1 - P\{X \le 5\} \\ &= 1 - F(5) \approx 1 - 0.94 \\ &= 0.06 \end{aligned}$$

5. (15) The plot given below is that of the cumulative distribution function for the discrete random variable X.



(a) How many distinct values does X take on, and what are they?
X can take on any of four different values, and they are

$$-2, \ -1, \ 1, \text{ and } 3$$

(b) What does $P\{X \leq 2\}$ equal?

$$P\{X \leq 2\} = \frac{3}{4}$$

(c) What does $P\{X = 1\}$ equal?

$$P\{X = 1\} = \frac{3}{8}$$

(d) What does $P\{X = -2\}$ equal?

$$P\{X = -2\} = \frac{1}{4}$$

6. (10) If $f(t)$ and $F(t)$ are the density and distribution function respectively of a random variable X whose value is the lifetime of some event, the hazard rate function $\lambda(t)$ is defined as

$$\lambda(t) = \frac{f(t)}{1 - F(t)}$$

Explain what $\lambda(t)$ measures, and justify your answer.

The number $\lambda(t)$ is a measure of how likely the lifetime of the event will end in the interval $[t, t+\delta]$, given that the event has made it to time t. To see this we compute the probability that $X \in [t, t+\delta]$ given that $X \geq t$.

$$\begin{aligned}
P\{X \in [t, t+\delta] \mid X \geq t\} &= \frac{P\{X \in [t, t+\delta], X \geq t\}}{P\{X \geq t\}} \\
&= \frac{P\{X \in [t, t+\delta]\}}{P\{X \geq t\}} \\
&= \frac{\int_t^{t+\delta} f(\xi)\, d\xi}{1 - F(t)} \approx \frac{f(t)\,\delta}{1 - F(t)} \\
&= \lambda(t)\,\delta
\end{aligned}$$

7. (10) Let Z be a random variable which has the standard normal density function. Show first that

$$P\{Z \leq -x\} = P\{Z \geq x\},$$

and then use this to infer that $P\{Z \leq -x\} = 1 - P\{Z \leq x\}$.

The density function of the random variable Z, equals $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. The thing to observe is that this function is an even function. Thus,

$$\begin{aligned} P\{Z \leq -x\} &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-t^2/2}\, dt \qquad (\text{set } t = -\tau) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\infty}^{x} e^{-\tau^2/2}\, (-d\tau) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-\tau^2/2}\, d\tau \\ &= P\{Z \geq x\} \end{aligned}$$

To get the last equation we have

$$\begin{aligned} P\{Z \leq -x\} &= P\{Z \geq x\} = 1 - P\{Z < x\} \\ &= 1 - P\{Z \leq x\} \end{aligned}$$

8. (5) Let X be a random variable which is normally distributed with parameters $\mu = -2$ and $\sigma^2 = 9$. Express $P\{-5 \leq X \leq 0\}$ in terms of a standard normally distributed random variable Z.

$$\begin{aligned} P\{-5 \leq X \leq 0\} &= P\{-3 \leq X - (-2) \leq 2\} \\ &= P\left\{\frac{-3}{3} \leq \frac{X+2}{3} \leq \frac{2}{3}\right\} \\ &= P\left\{-1 \leq Z \leq \frac{2}{3}\right\} \end{aligned}$$

9. (10) A satellite orbiting Jupiter sends back a stream of characters, either 0's or 1's. The probability that a particular character is incorrectly received is 0.01. Assuming a particular transmission contains 75,000 characters, and that the correct reception of any particular character is independent of whether or not any other character is received correctly, what is the probability that at least 80% of the characters are correctly received?

The probability that a character is correctly received is 0.99. Think of this problem as a sequence of Bernoulli trials. The number of trials being $75,000$. If X is the random variable which counts the number of successes in these trials, then X counts the number of characters correctly received. Thus,

$$\begin{aligned} P\{X = k\} &= \binom{75,000}{k} (0.99)^k (0.01)^{75000-k} \\ E(X) &= 75,000\,(0.99) = 74,250 \\ Var(X) &= 74250\,(0.1) = 742.5 \\ \sigma &= \sqrt{742.5} \approx 27.249 \end{aligned}$$

Given the large number of trials the random variable X can be quite closely approximated by a random variable, N, which is normally distributed with parameters $\mu = 74,250$ and $\sigma^2 = 7,425$. Thus,

$$\begin{aligned} P\{X \geq (0.8)\,75,000\} &\approx P\{N \geq 60,000\} \\ &= P\left\{\frac{N - 74,250}{27.249} \geq \frac{60,000 - 74,250}{27.249}\right\} \\ &= P\{Z \geq -522.95\} \\ &= 1 \end{aligned}$$

There is no need to even consult any table of values of the distribution function of the random variable Z, as -522.95 is quite a few standard deviations away from the mean of Z.