Notes:
1. For Problem 1b and 1c, assume there is only one contaminated bag in the 100 bags.
2. For Problem 2. Typo: 'earthquakes' should be 'hurricanes'. Skip 2b).
3.For problem 3. The 11 days in a row start from the first day. The student should not need to wait at the 12^{th} day.
4. For Problem 5. You may need to use the table normal_cdf.pdf

1) ***15 points, 5 each.*** **A quality control study in the Trader Joe's distribution center showed that 1.2% of all frozen bags of broccoli were contaminated. An immediate recall is performed on many thousands of broccoli bags that have been sold at the Davis location, and they are collected and examined for contamination before they are returned to customers. Define (for the 3 questions bellows) the random variable of interest, name its distribution and give the appropriate pmf (in equation form), as well as answering the questions.**

(a) In a sample of 100 recalled bags, what is the probability of getting at least 2 contaminated items?

Random Variable: The number of defective items out of 100 sampled items

Distribution: Binomial

pmf:
$$p(x) = \begin{cases} C_x^{100}(0.012)^x(0.988)^{100-x} & x = 0 \quad 1 \quad 2\ldots100 \\ 0 & otherwise \end{cases}$$

(If you already substituted in x = 0 or 1 in the pmf, that is OK)

Want $P(x \geq 2)$ so, calculate $1 - P(x<2) = 1 - p(0) - p(1) = 1-.299-.3632 = 0.3378$

(b) Assume in 100 bags, there is only one bag contaminated. What is the probability that every single bag from the 100 item sample will have to be examined for 1 contaminated item to be found?

Random Variable: The number of items sampled until the first defective item is found

Distribution: Geometric

pmf:
$$p(x) = \begin{cases} (0.988)^{x-1}(0.012) & x = 1 \quad 2 \quad 3\ldots \\ 0 & otherwise \end{cases}$$

$$p(100) = (0.988)^{100-1}(0.012) = 0.00363$$

(c) Same assumption and the way of examination as in 1b). What is the expected number of bags that will have to be examined before a contaminated item is found?

E[X] for a Geometrically Distributed RV is: $1/p = 1/0.012 = 83.3$

2) 10 ***points, 5 each*****. The West Coast is, on average, hit by 5 hurricanes each year.**

(a) What is the probability that fewer than four hurricanes will hit the region in a given year?

Use the Poisson distribution. X = the number of earthquakes in a time span, T. T = 1 year, $\lambda = 5$/year.

$$p(x) = \frac{e^{-5T}(5T)^x}{x!} = e^{-5}\frac{(5)^x}{x!},\ x = 0,1,2\ldots;\ 0 \quad elswhere$$

$$P(X < 4) = p(0) + p(1) + p(2) + p(3) = \frac{e^{-5}5^0}{0!} + \frac{e^{-5}5^1}{1!} + \frac{e^{-5}5^2}{2!} + \frac{e^{-5}5^3}{3!} = 0.265$$

(b) Find the expected value and variance for the time *between* hurricanes.

(Give full credits for 2b, since it is beyond the scope of the lectures)

For this problem use the exponential distribution. Y = time between earthquakes, and $\lambda = 5$/year.

$E[Y] = 1/\lambda = 1/5 = 0.2$ years

$\text{Variance}[Y] = 1/\lambda^2 = 1/25 = 0.04\ \text{years}^2$

3) *15 points, 5 each*. A student from the ECI 114 class walks to campus every morning and has to stop at the Russell and La Rue intersection. Given his/her interest in statistics and probability, he/she recorded the status of the walk signal (pass/no pass) when he/she arrives at the intersection. After many and many days of walking up to the intersection, the student realized that the walk signal will be in the pass mode 16% of the time. Assume every walk is an independent trial and we are starting from tomorrow. Define the random variable, the distribution, and write the equation for the pmf for the following questions.

(a) What is the probability that the first morning the student arrives to a walk signal is the 3rd morning she/he approaches it?

(0.84)(0.84)(0.16) = 0.1129

Random Variable: The number of walks (if you said "mornings" that is OK too) before the commuter arrives at her/his first walk signal

Distribution: Geometric

pmf:
$$p(x) = \begin{cases} (0.84)^{x-1}(0.16) & x = 1 \quad 2 \quad 3\ldots \\ 0 & otherwise \end{cases}$$

(b) What is the probability that the student has to wait at the cross walk 11 days in a row starting from the first day, and does not need to wait at the 12th day?

Waiting 11 days, getting a walk signal on the 12th day: $p(12) = (0.84)^{11}(0.16) = 0.0235$

(c) What is the expected value and standard deviation for the random variable defined in part (b)?

Expected value of a geometric distributed RV: 1/p = 1/0.16 = 6.25

Variance of a geometric distributed RV: $q/p^2 = 0.84/(0.16^2) = 32.81 = (32.81)^{1/2} = 5.728$

Standard deviation = 5.728

4) *15 points, 5 each.* An automated container is used to dispense liquid soap in the bathrooms of a school. The amount of soap dispensed by the container is uniformly distributed with an average of 0.5 mL and a standard deviation of 25 microliters (1 milliliter = 1000 microliters).

(a) What is the interval over which the amount of soap dispensed from the container ranges?

(Hint, the uniform distribution is symmetric about its mean, which should help you calculate the range)

Solve using the values for standard deviation and the average (expected value):

$Var[X] = (25/1000)^2 = (0.025)^2 = 6.25E{-4} = (b-a)^2/12$

$(b-a) = ((6.25E{-4})12))^{1/2}$, slove for b in terms of a: b = 0.0866 + a

Substitute our equation for b into the equation for the expected value:

(b+a)/2 = (2a + 0.0866)/2 = 0.5

a = (1-0.0866)/2 = 0.4567

b = 0.0866+a = 0.5433

The interval = [0.4567, 0.5433]

(b) What is the probability the container will dispense between 0.46 and 0.53 mL?

Using the CDF for the uniform distribution, $F(y) = \frac{y}{b-a} - \frac{a}{b-a}$,

$P(0.46< Y <0.53) = F(0.53) - F(0.46) = \left(\frac{0.53}{0.0866} - \frac{0.4567}{0.0866}\right) - \left(\frac{0.46}{0.0866} - \frac{0.4567}{0.0866}\right) =$ 0.084642 – 0.038106 = **0.8083**

(c) A certain test is invalidated if more than 0.51 mL of liquid is dispensed. What is the probability of performing an invalid test using this container?

$P[Y > 0.51] = 1 - F[0.51] = 1- \left(\frac{0.51}{0.0866} - \frac{0.4567}{0.0866}\right) =$ **0.3845**

5) ***25 points, 5 each.*** **Assume X~N(13, 4). Recall, the notation X~N(μ, σ) says, "X is a Normally distributed random variable with a mean μ and standard deviation σ." Determine:**

(a) $Pr[X \le 9] = P[(X-13)/4 \le (9-13)/4] = P[Z \le -1] = 0.158655$

(b) $Pr[X > 25] = Pr[(X-13)/4 > (25-13)/4] = P[Z > 3] = 1 - P[Z \le 3] = 1 - 0.9987 = 0.0013$

(c) $Pr[3 \le X \le 14] =$

$P[(3-13)/4 \le Z \le (14-13)/4] = P[-2.5 \le z \le 0.25] = P[Z \le 0.25] - Pr[Z \le -2.5] = 0.598706 - 0.00621 = 0.592496$

(d) Calculate the value of x, given, $P[X<x] = 0.33$

First find the z-value for $P[Z<z] = 0.33$, $z = -0.44$, $z = (x-\mu)/\sigma = (x - 13)/4$

$x = (-.44(4))+13 = 11.24$

(e) Calculate the value of x, given, $P[x < X < 9] = 0.1$

$Pr[X < 9] - (Pr[X < x]) = 0.1 = 0.158655 - (Pr[X < x]) = 0.1$

$0.058655 = Pr[Z<z]$, $z = -1.57$, $z = -1.57 = (x - 13)/4$, $x = 6.72$

6) ***20 points, 5 each*****. For a continuous random variable Y, the PDF has the form:**

$$f(y) = \begin{cases} c\,y^2 & (0 \le y < 1) \\ 0.5\,y & (1 \le y < 2) \\ 0 & else \end{cases}$$

(a) Find the value of c the value c.

Recall: $\int_{-\infty}^{\infty} f(y)\,dy = 1$, so $\int_0^1 c\,y^2\,dy + \int_1^2 0.5\,y\,dy = 1$

$$\int_0^1 c\,y^2\,dy + \int_1^2 0.5\,y\,dy = 1 = \left.\frac{c\,y^3}{3}\right|_0^1 + \left.\frac{0.5\,y^2}{2}\right|_1^2 = c\,\frac{1}{3} + 0.75$$

$$\frac{c}{3} = 1 - 0.75 \Rightarrow c = 0.25*3 = 0.75$$

(b) Find E[Y] and Var[Y]

$$E\left[Y\right] = \int_{-\infty}^{\infty} y\,f(y)\,dy = \int_0^1 y\,(0.75\,y^2)\,dy + \int_1^2 y(0.5\,y)\,dy = 0.75\left.\frac{y^4}{4}\right|_0^1 + 0.5\left.\frac{y^3}{3}\right|_1^2 = \frac{0.75}{4} + \frac{0.5}{3}(7) = 1.354$$

Recall $Var[Y] = E[Y^2] - E[Y]^2$, first calculate $E[Y^2]$

$$E\left[Y^2\right] = \int_0^1 y^2\,(0.75\,y^2)\,dy + \int_1^2 y^2\,(0.5\,y)\,dy = \frac{0.75}{5}\left.y^5\right|_0^1 + 0.5\left.\frac{y^4}{4}\right|_1^2 = 0.15 + 1.875 = 2.025$$

$Var[Y] = 2.025 - 1.354^2 = 0.191684$

(c) Find $Pr[0.5 \le Y \le 1.25]$

$$\int_{0.5}^1 0.75\,y^2\,dy + \int_1^{1.25} 0.5\,y\,dy = \frac{0.75}{3}(1 - 0.5^3) + \frac{1}{4}(1.25^2 - 1) = .21875 + 0.140625 = 0.359375$$

(d) Construct the cumulative distribution function of Y, and use it to check your answer in (c).
(Drawing the F(y) as a plot figure also earn credits)

$$F(y) = \begin{cases} 0 & (y < 0) \\ \frac{1}{4} y^3 & (0 \le y < 1) \\ \frac{1}{4} y^2 & (1 \le y < 2) \\ 1 & (2 \le y < \infty) \end{cases}$$

F(1.25) – F(0.5) = 0.390625 – 0.03125 = 0.359375