HW 2-1 (Suggested solutions posted on 02/03/2017) **Name:**

HW 2.1. (Mean and Standard Deviation.) Trees are subjected to different levels of carbon dioxide atmosphere with 6% of them in a minimal growth condition at 350 parts per million (ppm), 10% at 450 ppm (slow growth), 47% at 550 ppm (moderate growth), and 37% at 650 ppm (rapid growth). What are the mean and standard deviation of the carbon dioxide atmosphere (in ppm) for these trees in ppm?

X = levels of carbon dioxide

$x =$	350	450	550	650
$f(x) =$	0.06	0.10	0.47	0.37
$x^2 =$	122500	202500	302500	422500

$$\mu = E(X) = \sum x f(x)$$

$$= 350(0.06) + 450(0.1) + 550(0.47) + 650(0.37)$$

$$= 565$$

$$E(X^2) = \sum x^2 f(x)$$

$$= 122500(0.06) + 202500(0.1) + 302500(0.47) + 422500(0.37)$$

$$= 326100 \text{ (in ppm)}$$

$$\sigma^2 = V(X) = E(X^2) - [E(X)]^2 = 6875 \text{ (in ppm}^2\text{)}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{V(X)} = \sqrt{6875} \approx 82.9156 \text{ (in ppm)}$$

HW 2.2. (Probability Distribution, Mean, and Standard Deviation.) An article in *Information Security Technical Report* ["Malicious Software–Past, Present and Future" (2004, Vol. 9, pp. 6–18)] provided the following data on the top 10 malicious software instances for 2002. The clear leader in the number of registered incidences for the year 2002 was the Internet worm "Klez," and it is still one of the most widespread threats. This virus was first detected on 26 October 2001, and it has held the top spot among malicious software for the longest period in the history of virology. The 10 most widespread malicious programs for 2002

Place	Name	% Instances
1	I-Worm.Klez	61.22%
2	I-Worm.Lentin	20.52%
3	I-Worm.Tanatos	2.09%
4	I-Worm.BadtransII	1.31%
5	Macro.Word97.Thus	1.19%
6	I-Worm.Hybris	0.60%
7	I-Worm.Bridex	0.32%
8	I-Worm.Magistr	0.30%
9	Win95.CIH	0.27%
10	I-Worm.Sircam	0.24%

(Source: Kaspersky Labs).

Suppose that 20 malicious software instances are reported. Assume that the malicious sources can be assumed to be **independent.**

(a) What is the probability that at least one instance is "Klez?"

(b) What is the probability that three or more instances are **not** "Klez?"

(c) What are the mean and standard deviation of the number of "Lentin" instances among the 20 reported?

Bonus! What are the mean and standard deviation of the number of **not** "Lentin" instances among the 20 reported? Comparing with (c), what is your observation?

(a) X = the number of instances "Klez" out of 20 malicious software instances.

$X \sim B(n=20, p=0.6122)$

$P(X \geq 1) = 1 - P(X=0) = 1 - \binom{20}{0}(0.6122)^0(1-0.6122)^{20}$

$= 1.$

(b) Y = the number of instances NOT "Klez" out of 20 malicious software instances.

$Y \sim B(n=20, p=1-0.6122)$

$Y \sim B(n=20, p=0.3878)$

$P(Y \geq 3) = 1 - P(Y<3) = 1 - P(Y \leq 2)$

$\approx 1 - 0.0049 = 0.9951.$

(c) Z = the number of instances "Lentin" out of 20 malicious software instances.

$Z \sim B(n=20, p=0.2052)$

$\mu = E(Z) = np = 4.104$

$\sigma^2 = V(Z) = np(1-p) \approx 3.2616$

$\sigma = \sqrt{\sigma^2} = \sqrt{V(Z)} = 1.8061.$

Bonus! W = the number of instances NOT "Lentin" out of 20 malicious software instances

$W \sim B(n=20, p=1-0.2052)$

$W \sim B(n=20, p=0.7948)$

$\mu = E(W) = np = 20(0.7948) = 15.896$

$\sigma^2 = V(W) = np(1-p) \approx 3.2616$

$\sigma = \sqrt{\sigma^2} = \sqrt{V(Z)} = 1.8061.$

Note that $W = 20 - Z$. (since $Z + W = 20$).

Then $E(W) = E(20-Z) = 20 - E(Z) = 20 - 4.104 = 15.896$

$V(W) = V(20-Z) = V(-Z) = V((-1)Z) = (-1)^2 V(Z)$

$= V(Z).$

HW 2.3. (Probability Distribution, Mean, and Standard Deviation.) Suppose that random variable X has a geometric distribution with a mean of 2.5. Determined the following:

(a) $P(X = 1)$

(b) $P(X = 4)$

(c) $P(X = 5)$

(d) $P(X \leq 3)$

(e) $P(X > 3)$

(f) $V(X)$

$X \sim Geom(p)$

Since $E(X) = 1/p = 2.5$, then $p = 0.4$.

(a) $P(X=1) = (1-0.4)^{1-1} \cdot (0.4) = 0.4$

(b) $P(X=4) = (1-0.4)^{4-1} \cdot (0.4) = (0.6)^3 (0.1) = 0.0864$

(c) $P(X=5) = (1-0.4)^{5-1} \cdot (0.4) = (0.6)^4 (0.1) = 0.05184$

(d) $P(X \leq 3) = 1 - (1-0.4)^3 = 0.784$

(e) $P(X>3) = 1 - P(X \leq 3) = 1 - [1-(1-0.4)^3] = 0.216$.

(f) $V(X) = (1-p)/p^2 = (1-0.4)/(0.4)^2 = 3.75$

HW 2.4. (**Probability Distribution, Mean and Standard Deviation.**) A player of a video game is confronted with a series of opponents and has an 80% probability of defeating each one. Success with any opponent is independent of previous encounters. Until defeated, the player continues to contest opponents.

(a) What is the probability mass function of the number of opponents contested in a game?

(b) What is the probability that a player defeats at least two opponents in a game?

(c) What is the probability that a player contest four or more opponents in a game?

(d) What is the expected number of opponents contested in a game?

X = the number of opponents until defeated in a game.

$X \sim \text{Geom}(p = 0.2)$

(a) $f(x) = \begin{cases} (1-0.2)^{x-1}(0.2), & x = 1, 2, 3, \cdots \\ 0, & \text{otherwise}. \end{cases}$

(b) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$

$= 1 - (1 - (1-0.2)^1) = 0.8$

(c) $P(X \geq 4) = 1 - P(X < 4) = 1 - P(X \leq 3)$

$= 1 - [1 - (1-0.2)^3] = (1-0.2)^3 = (0.8)^3 = 0.512$

(d) $E(X) = 1/p = 1/0.2 = 5.$

HW 2-2 **Name:**

HW 2.5.(Discrete Random Variables) Suppose that random variable X has negative binomial distribution with $p = .2$ and $r = 4$. Determined the following:

(a) $E(X)$

(b) $V(X)$

(c) $P(X = 19)$

(d) $P(X \leq 20)$

(e) $P(X > 21)$

(f) $P(19 < X < 22)$

$$X \sim NB(r=4, p=0.2)$$

(a) $E(X) = r/p = 4/0.2 = 20.$

(b) $V(X) = r(1-p)/p^2 = 4(1-0.2)/(0.2)^2 = 80$

(c) $P(X=19) = \binom{19-1}{4-1}(1-0.2)^{19-4}(0.2)^4 \approx 0.0459$

(d) $P(X \leq 20) \overset{(\text{by } R)}{\approx} 0.5886.$

(e) $P(X>21) = 1-P(X \leq 21) \approx 1-0.6296 = 0.3704.$

(f) $P(19<X<22) = P(X=20)+P(X=21)$

$\approx 0.0436 + 0.0411 = 0.0847.$

HW 2.6 (A review of Bernoulli, Binomial, Geometric, and Negative Binomial Distributions). For each of the following question, define your random variable, identify the type of its distribution (with correctly identifying the values of parameters), and then write down what probability the question is asking (NO NEED to calculate it, unless you prefer).

(1) (Example:) A research study uses 800 men under the age of 55. Suppose that 30% of them carry a marker on the male chromosome that indicates an increased risk for high blood pressure. If 10 men under the age of 55 are selected randomly and tested for the marker, what is the probability that more than 1 has the marker?

Define the random variable X:

Let X be the number of men who has the marker out of the 10 selected mean

Distribution of X:

$$X \sim \text{hyber}(N = 800, n = 10, K = 30\% \times 800 = 240)$$

Probability the question is asking:

$$P(X > 1)$$

(2) A research study finds out that each man under the age of 55, has 30% chance of carrying a marker on the male chromosome that indicates an increased risk for high blood pressure. If 10 men under the age of 55 are selected independently and tested for the marker, what is the probability that at least 1 has the marker?

Define the random variable X: X = the number of men have the marker.

Distribution of X: $X \sim B(n=10, p=0.3)$

Probability the question is asking: $P(X \geq 1)$

(3) A research study finds out that for each man under the age of 55, he has 30% chance of carrying a marker on the male chromosome that indicates an increased risk for high blood pressure. Suppose men under the age of 55 arrive for testing for the marker independently, what is the probability that the fifth men who is the third one carrying the maker?

Define the random variable X: X = the number of men to observe the third one carrying the marker.

Distribution of X: $X \sim NB(r=3, p=0.3)$

Probability the question is asking: $P(X=5)$

(4) A research study finds out that for each man under the age of 55, he has 30% chance of carrying a marker on the male chromosome that indicates an increased risk for high blood pressure. Suppose men under the age of 55 arrive for testing for the marker independently, what is the probability that the fifth men who is the first one carrying the maker?

Define the random variable X: X = the number of men to observe the first one carrying the marker.

Distribution of X: $X \sim NB(r=1, p=0.3)$

$X \sim Geom(p=0.3)$

Probability the question is asking: $P(X=5)$

(5) Heart failure is due to either natural occurrences (87%) or outside factors (13%). Outside factors are related to induced substances or foreign objects. Natural occurrences are caused by arterial blockage, disease, and infection. Assume that causes of heart failure for the individuals are independent. What is the probability that the third patient with heart failure who enters the emergency room is the first one due to outside factors?

Define the random variable X: the number of patients with heart failure until observing the <u>first</u> one due to outside factors.

Distribution of X: $X \sim NB(r=1, p=0.13)$

$X \sim Geom(p=0.13)$

Probability the question is asking:

$P(X=3)$.

(6) Suppose that a healthcare provider selects 20 patients randomly without replacement from among 500 to evaluate adherence to a medication schedule. Suppose that 10% of the 500 patients fail to adhere with the schedule. What is the probability that fewer than 10% of the patients in the sample fail to adhere.

Define the random variable X: the number of patients in the sample fail to adhere.

Distribution of X: $X \sim hyper(N=500,\ n=20,\ K=500\times 0.1)$

$X \sim hyper(N=500,\ n=20,\ K=50)$

Probability the question is asking:

$P(X < 20\times 0.1) = P(X<2)$.

(7) The time to recharge the flash is tested in three cell-phone cameras. The probability that a camera passes the test is 0.8, and the cameras perform independently. What is the probability that the second failure occurs in tests of four or fewer cameras.

Define the random variable X: the number of cameras to observe the second failure

Distribution of X: $X \sim NB(r=2,\ p=1-0.8)$

$X \sim NB(r=2,\ p=0.2)$

Probability the question is asking:

$P(X \leq 4)$

(8) The probability that a visitor to a Web site provides contact data for additional information is 0.01. Assume that 1000 visitors to the site behave independently. What is the probability that no visitor provides contact data.

Define the random variable X: the number of visitors provides contact data out of 1000 visitors.

Distribution of X: $X \sim B(n=1000,\ p=0.01)$

Probability the question is asking: $P(X=0)$.

HW 2.7.(Discrete Random Variables) Printed circuit cards are placed in a functional test after being populated with semiconductor chips. A lot contains 140 cards, and 20 are selected without replacement for functional testing. (You **should** clearly define your random variable, identify the type of its distribution with correctly parameters, and then write down what probability the question is asking, finally, **calculate the probability**)

(a) If 5 cards are defective, what is the probability that 2 defective cards appears in the sample?

(b) If 20 cards are defective, what is the probability that at least 1 defective card is in the sample?

X = the number of defective cards appears in the sample.

(a) If $K=5$ cards are defective, then $N-K=140-5=135$ cards are not defective.

$$X \sim hyper(N=140, n=20, K=5)$$

$$P(X=2) = \frac{\binom{5}{2}\binom{140-5}{20-2}}{\binom{140}{20}} \approx 0.1280.$$

(b) If $K=20$ cards are defective, then $N-K=140-20=120$ cards are not defective.

$$X \sim hyper(N=140, n=20, K=20)$$

$$P(X \geq 1) = 1 - P(X<1) = 1 - P(X=0) = 1 - \frac{\binom{20}{0}\binom{140-20}{20-0}}{\binom{140}{20}}$$

$$\approx 1 - 0.0356 = 0.9644.$$

HW 2.8.(Discrete Random Variables) Astronomers treat the number of stars in a given volume of space as a Poisson random variable. The density in the Milky Way Galaxy in the vicinity of our solar system is one star per 16 cubic light-years in average.

(a) What is the probability of two or more stars in 16 cubic light-years?

(b) What is the probability of one or more stars in 10 cubic light-years?

(c) How many cubic light-years of space must be studied so that the probability of one or more stars exceeds 0.95?

(a) X = the number of stars per $\underbrace{\text{16 cubic light-years}}_{\text{1 unit.}}$

$X \sim \text{Poisson}(\lambda)$

Since. $\lambda = E(X) = 1$ (one star per 16 cubic light-years)

then $X \sim \text{Poisson}(1)$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1)$$

$$\approx 1 - 0.7358 = 0.2642.$$

(b) Y = the number of stars per 10 cubic light-years

$Y \sim \text{Poisson}\left(\frac{10}{16} \cdot 1\right)$

$Y \sim \text{Poisson}\left(\frac{10}{16}\right)$

$$P(Y \geq 1) = 1 - P(Y < 1) = 1 - P(Y = 0) = 1 - \frac{\left(\frac{10}{16}\right)^0 \cdot e^{-\frac{10}{16}}}{0!}$$

$$= 1 - e^{-\frac{10}{16}} \approx 0.4647.$$

(c) Z = the number of stars per $t \cdot 16$ cubic light-years (t units).

$Z \sim \text{Poisson}(t \cdot 1)$

$Z \sim \text{Poisson}(t)$

$0.95 \leq P(Z \geq 1) = 1 - P(Z<1) = 1 - P(Z=0)$

$= 1 - \frac{(t)^0 \cdot e^{-t}}{0!} = 1 - e^{-t}$

i.e. $0.95 \leq 1 - e^{-t}$

$e^{-t} \leq 1 - 0.95 = 0.05$

$-t \leq \ln 0.05$

$t \geq -\ln 0.05 \approx 2.996$

So, at least $2.996(16) = 47.9317$ cubic light-years.