CSM 70 Discrete Mathematics and Probability Theory

Spring 2016 Distributions, Variance, Markov and Chebyshev **Worksheet 11**

I. Distributions

Geometric Distribution: Geom(p)

Number of trials required to obtain the first success. Each trial has probability of success equal to p. The probability of the first success happening at trial k is:

$$\Pr(X = k) = (1-p)^{k-1}p, \quad k > 0$$

The expectation of a geometric distribution is:

$$E(X) = \frac{1}{p}$$

Binomial Distribution: Bin(n, p)

Number of successes when we do n independent trials. Each trial has a probability p of success. The probability of having k successes:

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The expectation of a binomial distribution is:

$$E(X) = np$$

Poisson Distribution: Pois(λ)

This is an approximation to the binomial distribution. Let the number of trials approach infinity, let the probability of success approach 0, such that E(X) = $np = \lambda$. This is an accepted model for "rare events". The probability of having k successes:

$$\Pr(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

The expectation of a poisson distribution is:

$$E(X) = \lambda$$

You are Eve, and as usual, you are trying to break RSA. You are trying to guess the factorization of N, from Bob's public key. You know that N is approximately 1,000,000,000,000. To find the primes p and q, you decide to try random numbers from 2 to 1,000,000 ≈ sqrt(N), and see if they divide N.

To do this, you roll a 999,999-sided die to choose the number, and see if it divides N using your calculator, which takes five seconds. Of course, there will be one number in this range that does divide N—namely, the smaller of p and q.

1. What kind of distribution would you use to model this?

2. What is the expected *amount of time* until you guess the correct answer, if it takes five seconds per guess (you only have a calculator)? Answer in days.

3. What is the variance in the amount of time? (Answer in seconds, approximately.)

Now you are trying to guess the 6-digit factorization digit by digit. Let's assume that when you finish putting these digits together, you can figure out how many digits you got right. Use zeros for blank spaces. For example, to guess 25, you would put 000025

1. What kind of distribution would you use to model this?

2. What is the probability that you get exactly 4 digits right?

3. What is the probability that you get less than 3 correct?

You are Alice, and you have a high-quality RSA-based security system. However, Eve is often successful at hacking your system. You know that the number of security breaches averages 3 a day, but varies greatly.

1. What kind of distribution would you use to model this?

2. What is the probability you experience exactly seven attacks tomorrow? At least seven (no need to simplify your answer)?

3. What is the probability that, on some day in April, you experience exactly six attacks?

II. Variance

For a random variable X with expectation E(X) = μ, the **variance** of X is defined to be:

$$\text{Var}(X) = \text{E}((X - \mu)^2)$$

This can be rewritten as follows:

$$\text{Var}(X) = \text{E}(X^2) - \mu^2$$

1. Prove that for *independent* random variables X and Y, Var(X + Y) = Var(X) + Var(Y).

2. Is this true for non-independent random variables? Prove or give a counterexample.

3. Is it true that for a random variable X and a constant c, $\mathrm{Var}(cX) = c\,\mathrm{Var}(X)$? Prove or give a counterexample.

4. Prove that for a random variable X and a constant c, $\mathrm{Var}(cX) = c^2\,\mathrm{Var}(X)$.

5. Consider the random variable $X = X_1 + \ldots + X_n$, where X_i equals i with probability 1/i and 0 otherwise. What is the variance of X? (Assume that X_i and X_j are independent for $i \neq j$)

For what value of n does E(X) = Var(X)?

For what value of n does $E(X) = SD(X)*\sqrt{2} + 100$?

6. An urn contains n balls numbered 1, 2, . . . , n. We remove k balls at random (without replacement) and add up their numbers. Find the mean and variance of the total.

III. Markov and Chebyshev

Markov's Inequality
For a non-negative random variable X with expectation E(X) = μ, and any α > 0:

$$\Pr[X \geq \alpha] \leq \frac{\mathrm{E}(X)}{\alpha}$$

Chebyshev's Inequality
For a random variable X with expectation E(X) = μ, and any α > 0:

$$\Pr[|X - \mu| \geq \alpha] \leq \frac{\mathrm{Var}(X)}{\alpha^2}$$

Use Markov's to prove Chebyshev's Inequality:

A random variable X always takes on values greater than -60. Find the best bound possible for Pr[X>= -10] when E[X] = -35.

Tossing Coins

Consider a coin that comes up with head with probability 0.2 . Let us toss it n times. Use Markov's to bound the probability of getting 80 percent heads.

Squirrel Standard Deviation

As we all know, Berkeley squirrels are extremely fat and cute. The average squirrel is 40% body fat. The standard deviation of body fat is 5%. Provide an upper bound on the probability that a randomly trapped squirrel is either too skinny or too fat? A skinny squirrel has less than 27.5% body fat, and a fat squirrel has more than 52.5% body fat?

Bound It!!!

A random variable X is always strictly larger than -100. You know that $E[X] = -60$. Give the best upper bound you can on $P(X \geq -20)$.

Give a distribution for a random variable where the expectation is 1,000,000 and the probability that the random variable is zero is 99%.

Consider a random variable Y with expectation μ whose maximum value is $3\mu/2$, prove that the probability that Y is 0 is at most 1/3.