CSM 70 Discrete Mathematics and Probability Theory

Spring 2016 Markov and Chebyshev, Variance

Worksheet 11 Solutions

Note: In the first 30 minutes ask students what they are struggling with. Take a look at their homeworks to see where they could be confused. You could also prepare some of your own questions tailored to your group's understanding of the material.

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:

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

Can walk through the derivation of E(X): (Sinho)

The clever way to find the expectation of the geometric distribution uses a method known as the renewal method. E(X) is the expected number of trials until the first success. Suppose we carry out the first trial, and one of two outcomes occurs. With probability p, we obtain a success and we are done (it only took 1 trial until success). With probability 1 − p, we obtain a failure, and we are right back where we started. In the latter case, how many trials do we expect until our first success? The answer is 1 + E(X): we have already used one trial, and we expect E(X) more since nothing has changed from our original situation (the geometric distribution is memoryless). Hence $E(X) = p \cdot 1 + (1 - p) \cdot (1 + E(X))$

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:

$$E(X) = np$$

Can walk through the derivation of E(X): (Sinho)

We would have to compute this sum:

$$E(X) = \sum_k k \Pr(X = k) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k}$$

Instead of doing that just use Bernoulli variables:

$$X = X_1 + \cdots + X_n$$

And now use linearity of expectation:

$$\begin{aligned} E(X) &= E(X_1 + \cdots + X_n) \\ &= E(X_1) + \cdots + E(X_n) \end{aligned}$$

Since the probability of a success happening at each step is p, and there are n steps, we are just summing p n times.

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* = *λ*. 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:

$$E(X) = \lambda$$

Can walk through the derivation of P(X): (Sinho)

$$\begin{aligned} \Pr(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} \\ &= \frac{n!}{k!\,(n-k)!} p^k (1-p)^{n-k} \\ &\approx \frac{n^k p^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \\ &\approx \frac{\lambda^k e^{-\lambda}}{k!} \end{aligned}$$

Can walk through derivation of E(X): (Sinho)

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} \\ &= e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= e^{-\lambda}\lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \\ &= e^{-\lambda}\lambda e^{\lambda} \\ &= \lambda \end{aligned}$$

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?
   Geometric — probability of success each time is p = 1 / 999,999

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.
   E[x] = 1/p = 999,999 tries
   (999,999 * 5 sec) / (60 sec/min) / (60 min/hr) / (24 hr/day) ≈ 57.9 days

3. What is the variance in the amount of time? (Answer in seconds, approximately.)
   Var(x) = (1-p)/p^2 = (1-1/999,999) / (999,999)^2 = 999,999 * 999,998 ≈ 1 trillion tries
   When we scale a variable, we scale the variance by the square of that factor, so Var(x) = 25 trillion sec^2

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?
   Binomial, since this is multiple independent trials that can either succeed or fail.

2. What is the probability that you get exactly 4 digits right?
   (6 choose 4) * $(1/10)^4$ * $(9/10)^2$

3. What is the probability that you get less than 3 correct?
   (6 choose 2) * $(1/10)^2$ * $(9/10)^4$ + (6 choose 1) * (1/10) * $(9/10)^5$

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?
   Poisson! That's what we use to model the probably frequencies of rare events.

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

$$Pr[X = 7] = \frac{\lambda^7}{7!}e^{-\lambda} = \frac{3^7}{7!}e^{-3} \approx 0.0216$$

$$Pr[X \geq 7] = \sum_{i=7}^{\infty} Pr[X = i] = \sum_{i=7}^{\infty} \frac{3^i}{i!} e^{-3}$$

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

$$Pr[X = 6] = \frac{3^6}{6!} e^{-3} \approx 0.0504$$

$$1 - (1 - 0.0504)^{30} \approx 0.788 = 78.8\%$$

II. Variance

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

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

This can be rewritten as follows:

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

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

$$\begin{aligned} \mathrm{Var}(X+Y) &= \mathrm{E}((X+Y)^2) - \mathrm{E}(X+Y)^2 \\ &= \mathrm{E}(X^2) + \mathrm{E}(Y^2) + 2\mathrm{E}(XY) - (\mathrm{E}(X) + \mathrm{E}(Y))^2 \\ &= (\mathrm{E}(X^2) - \mathrm{E}(X)^2) + (\mathrm{E}(Y^2) - \mathrm{E}(Y)^2) + 2(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)) \\ &= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2(\mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)). \end{aligned}$$

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

No! One simple counterexample is $X = Y$. Then

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

$$= E\left(4X^2\right) - (2E(X))^2 = 4E\left(X^2\right) - 4\left(E(X)^2\right)$$

$$= 4\left(E\left(X^2\right) - E(X)^2\right) = 4\mathrm{Var}(X)$$

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

   No! The same counterexample as above works. There $c = 2$, and $\mathrm{Var}(cX) = 4\mathrm{Var}(X) \neq 2\mathrm{Var}(X) = c\mathrm{Var}(X)$.

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

   The proof is similar to what we worked out for the counterexample in #2. We have

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 $i \neq j$)

$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)$

$E(X_i^2) = \Pr(X_i = i)*i^2 + \Pr(X_i = 0)*0^2 = (1/i)*i^2 + 0 = i$

$(E(X_i))^2 = (\Pr(X_i = i)*i + \Pr(X_i = 0)*0 = (1/i)*i + 0 = 1$

$\mathrm{Var}(X_i) = E(X_i^2) - (E(X_i))^2 = i - 1$

$$\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$$

$\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) = \sum_i i\text{-}1 = -n + \sum_i i$

$= -n + (n(n+1)/2) = (n(n+1) -2n)/2 = (n^2 + n - 2n)/2 = (n^2 - n)/2 = n(n-1)/2$

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

$E(X_i) = \Pr(X_i = i)*i + 0 = (1/i)*i = 1$

$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = n$

$n = n(n-1)/2 \Rightarrow 1 = (n-1)/2 \Rightarrow 2 = n - 1 \Rightarrow 3 = n$

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

E(X) = n, SD(X) = sqrt(Var(X))

n = sqrt(n(n-1)/2)*sqrt(2) + 100

n = sqrt(n(n-1)) + 100

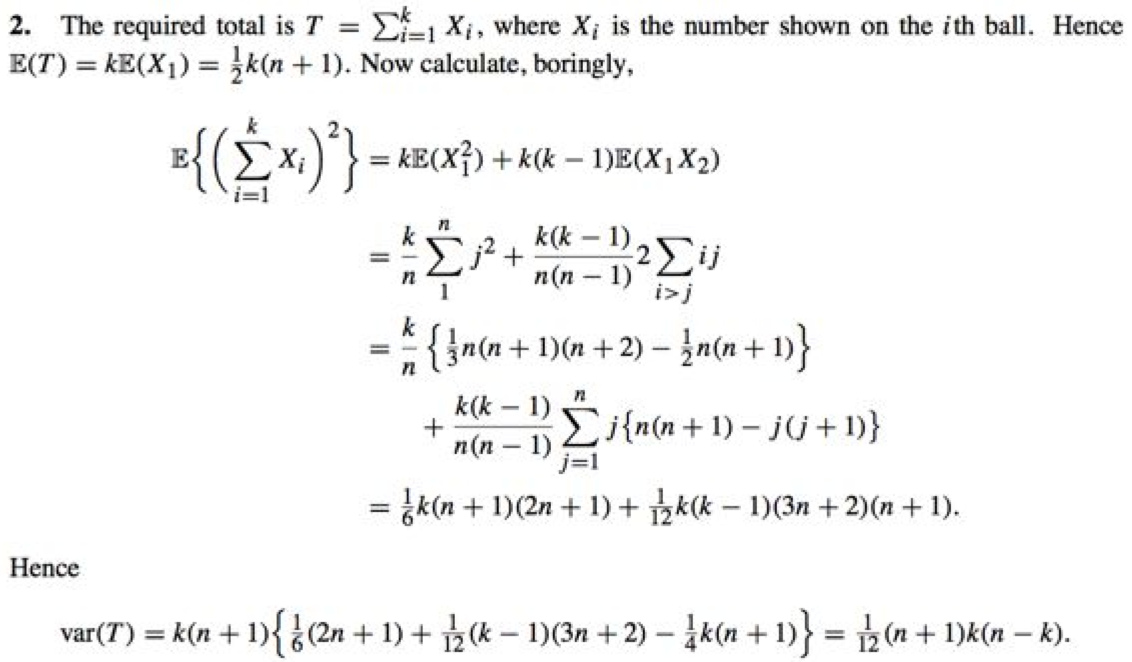
$(n - 100)^2 = n(n-1)$

$n^2 - 200n + 10000 = n^2 - n$

$10000 = 199n$

$n = 10000/199$

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.



2. The required total is $T = \sum_{i=1}^{k} X_i$, where X_i is the number shown on the ith ball. Hence $\mathbb{E}(T) = k\mathbb{E}(X_1) = \frac{1}{2}k(n+1)$. Now calculate, boringly,

$$\begin{aligned}\mathbb{E}\left\{\left(\sum_{i=1}^{k} X_i\right)^2\right\} &= k\mathbb{E}(X_1^2) + k(k-1)\mathbb{E}(X_1X_2)\\ &= \frac{k}{n}\sum_{1}^{n} j^2 + \frac{k(k-1)}{n(n-1)}2\sum_{i>j} ij\\ &= \frac{k}{n}\left\{\tfrac{1}{3}n(n+1)(n+2) - \tfrac{1}{2}n(n+1)\right\}\\ &\qquad + \frac{k(k-1)}{n(n-1)}\sum_{j=1}^{n} j\{n(n+1) - j(j+1)\}\\ &= \tfrac{1}{6}k(n+1)(2n+1) + \tfrac{1}{12}k(k-1)(3n+2)(n+1).\end{aligned}$$

Hence

$$\operatorname{var}(T) = k(n+1)\left\{\tfrac{1}{6}(2n+1) + \tfrac{1}{12}(k-1)(3n+2) - \tfrac{1}{4}k(n+1)\right\} = \tfrac{1}{12}(n+1)k(n-k).$$

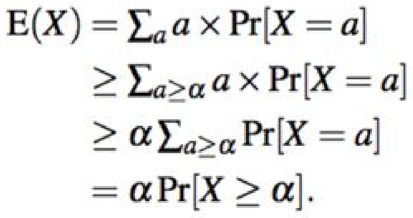
III. Markov and Chebyshev

Markov's Inequality

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

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

Proof as part of lesson plan:



$$\begin{aligned}\mathrm{E}(X) &= \textstyle\sum_a a \times \Pr[X = a]\\ &\geq \textstyle\sum_{a \geq \alpha} a \times \Pr[X = a]\\ &\geq \alpha \textstyle\sum_{a \geq \alpha} \Pr[X = a]\\ &= \alpha \Pr[X \geq \alpha].\end{aligned}$$

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:
Define the random variable $Y = (X - \mu)^2$. Note that $E(Y) = E((X - \mu)^2) = \mathrm{Var}(X)$
Also, notice that the event that we are interested in, $|X - \mu| \geq \alpha$ is exactly the same as the event $Y = (X - \mu)^2 \geq \alpha^2$.
Therefore, $\Pr[|X - \mu| \geq \alpha] = \Pr[Y \geq \alpha^2]$. Moreover, Y is obviously non-negative, so we can apply Markov's inequality to it to get

$$\Pr[Y \geq \alpha^2] \leq \frac{\mathrm{E}(Y)}{\alpha^2} = \frac{\mathrm{Var}(X)}{\alpha^2}.$$

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

Solutions: Take the Markov for Y which is X + 60. Notice that now, Y is positive always, so we can apply Markov to it, while before, X had the possibility of being negative. Taking the Markov Bound of this, we get that Pr[Y>=50] <= 25/50 = ½. The 50 came from adding 60 to -10, while the E[Y] is E[X] + 60.

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 more than 80 percent heads.

Let X be the amount of heads in n flips.

E[X] = .2n, by the binomial distribution.

So, P(X >= .8n) <= .2n / .8n, which is just .25.

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?

We use Chebyshev's inequality. We are looking for the probability we fall within 2.5 standard deviations of the mean. By Chebyshev's inequality, the probability we are within this range is 1/(2.5)^2, or 4/25 = 0.16.
If we were to use Markov's inequality, we would probabilities over 1, which yields a non-helpful value.

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 $P(X \geq -20)$.

Solution: Notice that we do not have the variance of X, so Chebyshev's bound is not applicable here. There is no upper bound on X, so Hoeffding's inequality cannot be used. We know nothing else about it's distribution so we cannot evaluate $E[esX]$ and so Chernoff bounds are not available. Since X is also not a sum of other random variables, other bounds or approximations are not available. This leaves us with just Markov's Inequality. But Markov Bound only applies on a nonnegative random variable, whereas X can take on negative values.

This suggests that we want to "shift" X somehow, so that we can apply Markov's Inequality on it. Define a random variable $Y = X + 100$, which means Y is strictly larger than 0, since X is always strictly larger than -100. Then, $E[Y] = E[X+100] = E[X]+100 = -60+100 = 40$. Finally, the upper bound on X that we want can be calculated via Y, and we can now apply Markov's Inequality on Y since Y is strictly positive.

$P(X \geq -20) = P(Y \geq 80) \leq E[Y]/80 = 40/80 = 1/2$

Hence, the best upper bound on $P(X \geq -20)$ is $1/2$.

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%.

X is 100, 000, 000 with probability 0.01, and 0 otherwise.

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

$\mu = \sum a\Pr[Y=a]\ _{a}$

$= \sum a\Pr[Y=a]\ _{a\neq 0}$

$= \sum 3\mu/2 \times \Pr[Y=a]\ _{a\neq 0}$

$= 3\mu/2 \times \sum \Pr[Y=a]\ _{a\neq 0}$

$= 3\mu/2 \times (1 - \Pr[Y=0])$

This implies Pr[Y=0] <= 1/3