Distributions, Variance, Inequalities, Confidence Intervals 8

Computer Science Mentors 70

October 23 - 27, 2017

1 Distributions

1.1 Introduction

Geometric Distribution: Geom(p)

With the geometric distribution, we count the number of failures until the first success. For example, we could count the number of rolls of a dice until we roll a 6. The probability that the first success occurs on trial k is:

$$\mathrm{P}[X = k] = (1 - p)^{k-1} * p, k > 0$$

In what way can we derive the geometric distribution from the binomial distribution?

Expectation:
We know that $E(X)$ is the number of trials until the first success occurs, including that first success. There are two cases:

1. The first success occurs, with probability p
2. We obtain a failure, with probability $1-p$, meaning that we are back where we started but already used one trial

Putting this together, we get:

$$E(X) = p * 1 + (1 - p) * (1 + E(X)) \implies E(X) = \frac{1}{p}$$

Variance:

$$var(X) = \frac{1-p}{p^2}$$

Binomial Distribution: Bin(n, p)

The binomial distribution counts the number of successes when we conduct n independent trials. Each trial has a probability p of success. For this reason, we can think of the binomial distribution as a sum of n independent Bernoulli trials, each with probability p.

The probability of having k successes:

$$\text{P}[X = k] = \binom{n}{k} * p^k * (1-p)^{n-k}$$

For example, if we flip a fair coin 10 times, the probability of 6 heads is

$$P(H = 6) = \binom{10}{6} \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4$$

Expectation:
If we were to compute the sum the traditional way, we would have to compute the sum

$$E(X) = \sum_{x \in X} x \cdot \binom{n}{x} p^x (1-p)^{n-x}$$

Instead of doing that, we can use the fact that the binomial distribution is the sum of n independent Bernoulli distributions:

$$X = X_1 + \ldots + X_n$$

And now use linearity of expectation:

$$E(X) = E(X_1 + \ldots + X_n) = E(X_1) + \ldots + E(X_n) = p + p + ... + p = np$$

Variance:
We know that variance is only separable when variables are mutually independent, i.e. $var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n)$ only when $X_1, X_2, ...X_n$ are mutually independent. Since our sum of Bernoulli trials is independent, we can do the following:

$$var(X) = var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n)$$

$$= p(1-p) + p(1-p) + ... + p(1-p) = np(1-p)$$

Poisson Distribution: Pois(λ) The Poisson distribution is an approximation of the binomial distribution under two conditions:

- n is very large
- p is very small

Let $\lambda = np$ represent the "rate" at which some event occurs. We usually use this distribution when these events are rare, such as a lightbulb failing.

The probability of k occurrences is

$$\mathrm{P}[X = k] = \frac{e^{-\lambda} * \lambda^k}{k!}$$

It turns out that the expectation and variance of the Poisson distribution are both equal to λ. This will be clear after we walk through the derivation of the Poisson distribution.

Derivation:
Recall, $\lambda = np$. Also, recall from calculus we have $\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n = e^x$, implying that $\lim_{n\to\infty}\left(1 - \frac{\alpha}{n}\right)^n = e^{-\alpha}$. We will also use the fact that for large n, $\frac{n!}{(n-k)!} \approx n^k$. We will use these facts below.

$$\begin{aligned}
\mathrm{P}[X = k] &= \binom{n}{k} * p^k * (1-p)^{n-k} && (1)\\
&= \frac{n!}{k! * (n-k)!} * p^k * (1-p)^{n-k} && (2)\\
&\approx \frac{n^k * p^k}{k!} * (1 - \frac{\lambda}{n})^{n-k} && (3)\\
&\approx \frac{\lambda^k * e^{-\lambda}}{k!} && (4)
\end{aligned}$$

Since we started with a binomial distribution, our expectation and variance should remain the same.

Expectation:
Since the expectation of a binomial is np, and we set $\lambda = np$, our expectation is also $E(X) = np$. We can also show this from scratch:

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

$$= e^{-\lambda} * \lambda * \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!}$$
$$= e^{-\lambda} * \lambda * \sum_{k=0}^{\infty} \frac{\lambda^{k}}{k!}$$
$$= e^{-\lambda} * \lambda * e^{\lambda}$$
$$= \lambda$$

Variance:
For variance, it is much easier to start with the binomial case and reason from there. The variance of a binomial is $np(1-p)$, which looks like $\lambda(1-p)$. However, we started with the assumption that p is very small, so we can assume $(1-p)$ is very close to 1 and thus $\lambda(1-p)$ is very close to λ. Therefore, $var(X) = \lambda$.

1.2 Questions

1. You are Eve, and as usual, you are trying to break RSA. You are trying to guess the factorization of N, from Bobs 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 $1,000,000 \approx \sqrt{N}$, 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 Nnamely, the smaller of p and q.

   (a) What kind of distribution would you use to model this?

   Solution: Geometric probability of success each time is $p = \frac{1}{999{,}999}$

   (b) 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.

   Solution:
   $$\mathrm{E}(x) = \frac{1}{p} = 999,999 \text{ tries}$$
   $$(999,999 * 5 \text{ sec}) * \frac{1 \text{ min}}{60 \text{ sec}} * \frac{1 \text{ hr}}{60 \text{ min}} * \frac{1 \text{ day}}{24 \text{ hr}} \approx 57.9 \text{ days}$$

2. Now you are trying to guess the 6-digit factorization digit by digit. Lets 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

   (a) What kind of distribution would you use to model this?

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

   (b) What is the probability that you get exactly 4 digits right?

   Solution: $\binom{6}{4} * \frac{1}{10}^4 * \frac{9}{10}^2$

   (c) What is the probability that you get less than 3 correct?

   Solution: P(less than 3 correct) = P(2 correct) + P(1 correct) + P(0 correct)
   = $\binom{6}{2} * \frac{1}{10}^2 * \frac{9}{10}^4 + \binom{6}{1} * \frac{1}{10} * \frac{9}{10}^5 + \binom{6}{0} * \frac{1}{10}^0 * \frac{9}{10}^6$

3. In this problem, we will explore how we can apply multiple distributions to the same problem.

   Suppose you are a professor doing research in *machine learning*. On average, you receive 12 emails a day from students wanting to do research in your lab, but this number varies greatly.

   (a) Which distribution would you use to model the number of emails you receive from students on any one day?

   Solution: Poisson with parameter $\lambda = 12$.

   (b) What is the probability that you receive 7 emails tomorrow? At least 7?

   Solution: The probability we receive exactly 7 emails tomorrow is

   $$P(X = 7) = \frac{e^{-12}12^7}{7!} \approx 0.0437$$

   The probability we receive at least 7 emails tomorrow is

   $$P(X \geq 7) = P(X = 7) + P(X = 8) + ... = e^{-12}\sum_{k=7}^{\infty} \frac{12^k}{k!}$$

Equivalently, we can calculate it as:

$$P(X \geq 7) = 1 - P(X \leq 6)$$

$$= 1 - P(X = 0) - P(X = 1) - ... - P(X = 5) - P(X = 6)$$

which gets rid of the infinite sum.

(c) Now, let's look at the month of April, in which lots of students are emailing you to secure a summer position. What is the probability that the first day in April that you receive exactly 15 emails is April 7th? *Hint: Break this problem down into parts, and assign your result to the first part to the variable p.*

Solution: It is worth mentioning that "receiving exactly 15 emails in one day" is an event, and either it happens or it does not. We will use the geometric distribution to model this. First, though, we need to find the probability p:

$$p = e^{-15}\frac{12^{15}}{15!} \approx 0.003604$$

Now, for days April 1, April 2, ... April 7, we know that we receive some number of emails that isn't 15, followed by receiving exactly 15 emails on April 8. This corresponds to 7 failures and 1 success in geometric:

$$P(\text{April 8th is first day with exactly 15 emails}) = (1-p)^7 p$$

$$\approx 0.996396^7 * 0.003604 \approx 0.003514$$

(d) Now, calculate the probability that April 8th is the first day that we receive **at least** 15 emails.

Solution: Our geometric model is the same, but we have a different p now.

$$p = e^{-12}\sum_{k=15}^{\infty}\frac{12^k}{k!} \approx 0.22798$$

$$P(\text{April 8th is first day with at least 15 emails}) = (1-p)^7 p$$

$$\approx (0.77202)^7 * 0.22798 \approx 0.03726$$

(e) What is the probability that you receive at least 15 emails on 10 different days in April?

Solution: We can take $p = 0.22798$ from the previous part. Now, we will count with the binomial distribution we have 30 trials (one for each day in April), and each is "successful" with probability p. We want the probability of exactly 10 "successes". Let X be the random variable that counts the number of days that we receive at least 15 emails.

$$P(X = 10) = \binom{30}{10} p^{10}(1-p)^{20}$$
$$\approx 0.06446$$

(f) What is the probability that you receive at least 15 emails on at least 15 days in April?

Solution:

$$P(X \geq 15) = P(X = 15) + P(X = 16) + ... + P(X = 30)$$
$$= \sum_{k=15}^{30} \binom{30}{k} 0.22798^k 0.77202^{30-k} \approx 0.00102$$

2 Variance

2.1 Introduction

Random variable: a function $X : \omega \to R$ that assigns a real number to every outcome ω in the probability space.

Expectation: The expectation of a random variable X is defined as

$$\mathrm{E}(X) = \sum_{\alpha \in A} a * \mathrm{P}[X = a]$$

where the sum is over all possible values taken by the random variable. Expectation is usually denoted with the symbol μ.

Linearity of Expectation: For any random variables $X_1, X_2, ...X_n$, expectation is linear, i.e.:

$$E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n)$$

This is true even when these random variables aren't independent.

Variance: The variance of a random variable X is defined as

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

The latter version of variance is the one we usually use in computations. The square root of Var(X) is called the standard deviation of X. It is usually denoted with the variable σ.

2.2 Questions

1. Let's consider the classic problems of flipping coins and rolling dice. Let X be a random variable for the number of coins that land on heads and Y be the value of the die roll.

   (a) What is the expected value of X after flipping 3 coins? What is the variance of X?

Solution:

$$\mathrm{E}(X) = 0 * \frac{1}{8} + 1 * \frac{3}{8} + 2 * \frac{3}{8} + 3 * \frac{1}{8} = \frac{3}{2}$$

$$\mathrm{E}(X^2) = 0^2 * \frac{1}{8} + 1^2 * \frac{3}{8} + 2^2 * \frac{3}{8} + 3^2 * \frac{1}{8} = \frac{24}{8} = 3$$

$$\mathrm{E}(X)^2 = \frac{9}{4}$$
$$\mathrm{Var}(X) = 3 - \frac{9}{4} = \frac{3}{4}$$

(b) Let Y be the sum of rolling a dice 1 time. What is the expected value of Y?

Solution: $\mathrm{E}(Y) = \frac{1}{6} * (1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$

(c) What is the variance of Y?

Solution: $\mathrm{E}(Y^2) = [\frac{1}{6}(1^2+2^2+3^2+4^2+5^2+6^2)] = \frac{91}{6}$ $\mathrm{Var}(Y) = \mathrm{E}(Y^2)-(\mathrm{E}(Y))^2 = \frac{91}{6}\frac{7}{2}^2 = \frac{35}{12}$

2. Say you're playing a game with a coin and die, where you flip the coin 3 times and roll the die once. In this game, your score is given by the number of heads that show multiplied with the die result. What is the expected value of your score? Whats the variance?

Solution: $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) = \frac{21}{4}$ since X and Y are independent. $\mathrm{Var}(XY) = \mathrm{E}(X^2Y^2) - \mathrm{E}(XY)^2 = \mathrm{E}(X^2)\mathrm{E}(Y^2) - \mathrm{E}(X)^2\mathrm{E}(Y)^2 = 3 * \frac{91}{6} - \frac{3}{2}^2 * \frac{7}{2}^2 = \frac{91}{2} - \frac{9}{4} * \frac{49}{4} = 17.9375 = \frac{287}{16}$

3. You are at a party with n people where you have prepared a red solo cup labeled with their name. Before handing red cups to your friends, you pick up each cup and put a sticker on it with probability $\frac{1}{2}$ (independently of the other cups). Then you hand back the cups according to a uniformly random permutation. Let X be the number of people who get their own cup back AND it has a sticker on it.

(a) Compute the expectation $\mathrm{E}(X)$.

Solution: Define $X_i = 1$ if the i-th person gets their own cup back and it has a sticker on it 0 otherwise Hence $\mathrm{E}(X) = \mathrm{E}(\sum i = 1^n(X_i) = \sum i = 1^n \mathrm{E}(X_i)$ $\mathrm{E}(X_i) = \mathrm{P}[X_i = 1] = \frac{1}{2n}$ since the i-th student will get his/her cup with probability $\frac{1}{n}$ and has a sticker on it with probability $\frac{1}{2}$ and stickers are put independently. Hence $\mathrm{E}(X) = n \cdot \frac{1}{2n} = \frac{1}{2}$.

(b) Compute the variance $\mathrm{Var}(X)$

Solution: To calculate $\text{Var}(X)$, we need to know $\text{E}(X^2)$

$$\text{E}(X^2) = \text{E}(X_1 + X_2 + \ldots + X_n)^2 = \text{E}(\sum_{i,j}(X_i * X_j)) = \sum_{i,j}(\text{E}(X_i * X_j)$$

(by linearity of expectation)

Then we consider two cases, either $i = j$ or $i \neq j$. Hence

$$\sum_{i,j} \text{E}(X_i * X_j) = \sum_i \text{E}(X_i^2) + \sum_{i \neq j} \text{E}(X_i * X_j)$$

$E(X_i^2) = \frac{1}{2n}$ for all i. To find $\text{E}(X_i * X_j)$, we need to calculate $\text{P}[X_i X_j = 1]$. $\text{P}[X_i * X_j = 1] = \text{P}[X_i = 1]\text{P}[X_j = 1 | X_i = 1] = \frac{1}{2n} * \frac{1}{2*(n-1)}$ since if student i has received his/her own cup, student j has $n - 1$ choices left. Hence

$$\text{E}(X^2) = n * \frac{1}{2n} + n * (n-1) * \frac{1}{2n} * \frac{1}{2 * (n-1)} = \frac{3}{4}$$

$\text{Var}(X) = \text{E}(X^2) - \text{E}(X)^2 = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$.

4. a. Prove that for independent random variables X and Y, $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$.

Solution:

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

b. Is the above result true for non-independent random variables? Prove or give a counterexample.

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

$$\begin{aligned}\text{Var}(X+Y) &= \text{Var}(2x) \\ &= \text{E}((2X)^2) - \text{E}(2X)^2 \\ &= \text{E}(4X^2) - (2 * \text{E}(X))^2 \\ &= 4\text{E}(X^2) - 4(\text{E}(X)^2) \\ &= 4(\text{E}(X^2) - \text{E}(X)^2) \\ &= 4 * \text{Var}(X)\end{aligned}$$

5. Consider the random variable $X = X_1 + \ldots + X_n$, where X_i equals i with probability $\frac{1}{i}$ and 0 otherwise.

(a) What is the variance of X? (Assume that X_i and X_j are independent for $i \neq j$)

Solution:

$$\begin{aligned}\text{Var}(X) &= \text{Var}(X_1) + \ldots + \text{Var}(X_n) \\ \text{E}(X_i^2) &= \text{P}[X_i = i] * i^2 + \text{P}[X_i = 0] * 0^2 = \frac{1}{i} * i^2 + 0 = i \\ (\text{E}(X_i))^2 &= (\text{P}[X_i = i] * i + \text{P}[X_i = 0] * 0 = \frac{1}{i} * i + 0 = 1 \\ \text{Var}(X_i) &= \text{E}(X_i^2) - (\text{E}(X_i))^2 = i - 1\end{aligned}$$

Recall,

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

$$\begin{aligned}
\text{Var}(X) &= \sum_i \text{Var}(X_i) \\
&= \sum_i i - 1 \\
&= -n + \sum_i i \\
&= -n + \frac{n*(n+1)}{2} \\
&= \frac{n*(n+1) - 2n}{2} \\
&= \frac{n^2 + n - 2n}{2} \\
&= \frac{n^2 - n}{2} \\
&= \frac{n*(n-1)}{2}
\end{aligned}$$

(b) For what value of n does $\text{E}(X) = \text{Var}(X)$?

Solution:

$$\begin{aligned}
\text{E}(X_i) &= \text{P}[X_i = i] * i + 0 \\
&= \frac{1}{i} * i = 1
\end{aligned}$$

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

$$\begin{aligned}
n = \frac{n*(n-1)}{2} &\to 1 = \frac{n-1}{2} \\
&\to 2 = n - 1 \\
&\to 3 = n
\end{aligned}$$

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

Solution: $\mathrm{E}(X) = n, SD(X) = \sqrt{\mathrm{Var}(X)}$

$$\begin{aligned}
n &= \sqrt{\frac{n*(n-1)}{2}} * \sqrt{2} + 100 \\
n &= \sqrt{n*(n-1)} + 100 \\
(n-100)^2 &= n*(n-1) \\
n^2 - 200n + 10000 &= n^2 - n \\
10000 &= 199n \\
n &= \frac{10000}{199}
\end{aligned}$$

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

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

$$\begin{aligned}
\mathrm{E}((\sum_{i=1}^{k} X_i)^2) &= k\mathrm{E}(X_1^2) + k*(k-1)*\mathrm{E}(X_1 * X_2) \\
&= \frac{k}{n}\sum_{1}^{n} j^2 + \frac{k*(k-1)}{n*(n-1)} * 2 * \sum_{i>j} i*j \\
&= \frac{k}{n}(\frac{1}{3}*n*(n+1)*(n+2) - \frac{1}{2}*n*(n+1)) \\
&+ \frac{k*(k-1)}{n*(n-1)} * \sum_{j=1}^{n} j*(n*(n+1) - j*(j+1)) \\
&= \frac{1}{6}*k*(n+1)*(2n+1) + \frac{1}{12}*k*(k-1)*(3n+2)*(n+1)
\end{aligned}$$

Hence,

$$\begin{aligned}
\mathrm{Var}(T) &= k(n+1)(\frac{1}{6}k(n+1)(2n+1) + \frac{1}{12}k(k-1)(3n+2)(n+1) - \frac{1}{4}k(n+1)) \\
&= \frac{1}{12}(n+1)k(n-k)
\end{aligned}$$

3 Weak Law of Large Numbers

1. As our number of fair coin flipping trials goes to infinity, what does the probability that the proportion of heads is not $\frac{1}{2}$ go to?

Solution: By the Weak Law of Large Numbers, 0.

4 Markov, Chebyshev

4.1 Introduction

Markov's Inequality
For a non-negative random variable X with expectation $\mathrm{E}(X) = \mu$, and any $\alpha > 0$:

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

Solution: Proof of Markov's Inequality

$$\begin{aligned} E(X) &= \sum_{a} a * Pr[X = a] \\ &\geq \sum_{a \geq \alpha} a * Pr[X = a] \\ &\geq \alpha \sum_{a \geq \alpha} Pr[X = a] \\ &= \alpha Pr[X \geq a] \end{aligned}$$

Chebyshev's Inequality
For a random variable X with expectation $\mathrm{E}(X) = \mu$, and any $\alpha > 0$:

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

4.2 Questions

1. Use Markov's to prove Chebyshev's Inequality:

> **Solution:** Define the random variable $Y = (X - \mu)^2$. Note that $\text{E}(Y) = \text{E}((X - \mu)^2) = \text{Var}(X)$ Also, notice that the event that we are interested in, $|X\mu|\alpha$ is exactly the same as the event $Y = (X\mu)^2\alpha^2$. Therefore, $\text{P}[|X\mu| \geq \alpha] = \text{P}[Y \geq \alpha^2]$. Moreover, Y is non-negative, so we can apply Markov's inequality to it to get:
>
> $$\text{P}[Y \geq \alpha^2] \leq \frac{\text{E}(Y)}{\alpha^2} = \frac{\text{Var}(X)}{\alpha^2}$$
>
> Substituting in $Y = (X - \mu)^2$ above and taking square roots yields the form we are used to.

2. **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?

> **Solution:** 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 $\frac{1}{2.5^2}, or \frac{4}{25} = 0.16$ If we were to use Markov's inequality, we would have probabilities over 1, which yields a non-helpful value.

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

> **Solution:** Notice that we do not have the variance of X, so Chebyshev's bound is not applicable here. Since X is also not a sum of other random variables, other bounds or approximations(chernoff, Hoeffding's inequality. Don't worry about them if they are not covered.) are not available. This leaves us with just Markov's Inequality. But Markov Bound only applies on a non-negative random variable, whereas X can take on negative values.
>
> This suggests that we want to shift X somehow, so that we can apply Markovs 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, $\text{E}(Y) = \text{E}(X +$

$100) = \mathrm{E}(X) + 100 = -60 + 100 = 40$. 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.

$\mathrm{P}[X \geq -20] = \mathrm{P}[Y \geq 80] \leq \frac{\mathrm{E}(Y)}{80} = \frac{40}{80} = \frac{1}{2}$

Hence, the best upper bound on $\mathrm{P}[X \geq -20]$ is $\frac{1}{2}$.

4. 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%.

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

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

Solution:

$$\begin{aligned}
\mu &= \sum_a a * \mathrm{P}[Y = a] \\
&= \sum_{a \neq 0} a * \mathrm{P}[Y = a] \\
&\leq \sum_{a \neq 0} \frac{3\mu}{2} * \mathrm{P}[Y = a] \\
&= \frac{3\mu}{2} * \sum_{a \neq 0} \mathrm{P}[Y = a] \\
&= \frac{3\mu}{2} * (1 - \mathrm{P}[Y = 0])
\end{aligned}$$

This implies that $\mathrm{P}[Y = 0] \leq \frac{1}{3}$

6. Let X be the sum of 20 i.i.d. Poisson random variables $X_1, \ldots, X_{20}$ with $\mathrm{E}(X_i) = 1$. Find an upper bound of $\mathrm{P}[X \geq 26]$ using,

(a) Markov's inequality:

Solution:

$$\mathrm{P}[X \geq a] \leq \frac{\mathrm{E}(X)}{a} \text{ for all } a > 0$$

$$\begin{aligned}
\mathrm{P}[X \geq 26] &\leq \frac{20}{26} \\
&\approx 0.769
\end{aligned}$$

(b) Chebyshev's inequality:

Solution:

$$\begin{aligned} \mathrm{P}[|X - \mathrm{E}(X) \geq c] &\leq \frac{\sigma_X^2}{c^2} \\ \mathrm{P}[|X - 20| \geq 6] &\leq \frac{20}{36} \\ &\approx 0.5556 \end{aligned}$$

5 Confidence Intervals

5.1 Questions

1. Define i. i. d. variables $A_k \sim \text{Bern}(p)$ where $k \in [1, n]$. Assume we can declare that $\mathrm{P}[|\frac{1}{n}\sum_k A_k - p| > 0.25] = 0.01$.

   (a) Please give a 99% confidence interval for p if given A_k.

   Solution: $[\frac{1}{n}\sum_i A_k - 0.25, \frac{1}{n}\sum_i A_k + 0.25]$

   (b) We know that the variables X_i, for i from 1 to n, are i.i.d. random variables and have variance. We also have a value (an observation) of $A_n = \frac{X_1 + \ldots + X_n}{n}$. We want to guess the mean, μ, of each X_i.

   Prove that we have 95% confidence μ lies in the interval $\left[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\right]$

   That is, $\mathrm{P}\big[\mu \in \left[A_n - 4.5\frac{\sigma}{\sqrt{n}}, A_n + 4.5\frac{\sigma}{\sqrt{n}}\right]\big] \geq 95\%$

   Solution: To do this, we use Chebyshev's. Because $\mathrm{E}[A_n] = \mu$ (A_n is the average of the X_is), we bound the probability that $|A_n - \mu|$ is more than the interval size at 5%:

   $$\mathrm{P}\big[|A_n - \mu| \geq 4.5\frac{\sigma}{\sqrt{n}}\big] \leq \frac{\text{Var}(A_n)}{(4.5\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{20\sigma^2}{n}} = \frac{1}{20} = 5\%$$

   .

   Thus, the probability that μ is *in* the interval is 95

   (c) Give the 99% confidence interval for μ:

Solution: Solution is similar to that of the 95% confidence interval.
$\left[A_n - 10\frac{\sigma}{\sqrt{n}}, A_n + 10\frac{\sigma}{\sqrt{n}}\right]$, because $\mathrm{P}\left[|A_n - \mu| \geq 10\frac{\sigma}{\sqrt{n}}\right] \leq \frac{\mathrm{Var}(A_n)}{(10\frac{\sigma}{\sqrt{n}})^2} \approx \frac{\frac{\sigma^2}{n}}{\frac{100\sigma^2}{n}} = \frac{1}{100} = 1\%$.

2. We have a die whose 6 faces are values of consecutive integers, but we dont know where it starts (it is shifted over by some value k; for example, if $k = 6$, the die faces would take on the values $7, 8, 9, 10, 11, 12$). If we observe that the average of the n samples (n is large enough) is 15.5, develop a 99% confidence interval for the value of k.

Solution: Consider the 99% confidence interval for the average of n rolls of a standard die, X. We can model it using the normal distribution with a mean $\mathrm{E}[\frac{nD}{n}]\frac{n\mathrm{E}[D]}{n} = \mathrm{E}[D] = 3.5$ and a variance $\mathrm{Var}[X] = \mathrm{Var}[\frac{nD}{n}] = (\frac{1}{n})^2\mathrm{Var}[nD] = (\frac{1}{n})^2 n\mathrm{Var}[D] = \frac{\mathrm{Var}[D]}{n} = \frac{35}{12n}$ where D is the distribution of a standard dice roll. We can now find the 99% confidence interval as $|\frac{x-\mathrm{E}[X]}{\sqrt{\mathrm{Var}[X]}}| = Z(99\%) = 2.576$. We solve this to find $x = \mathrm{E}[X] \pm Z(99\%)\sqrt{\mathrm{Var}[X]} = 3.5 \pm 2.576\sqrt{\frac{35}{12n}}$. We can now say that the confidence interval for k is $15.5 -$ confidence interval for X.