INEQUALITIES, DISTRIBUTIONS 10

COMPUTER SCIENCE MENTORS 70

November 5 - 9, 2018

1 Markov, Chebyshev

1.1 Introduction

Markov's Inequality:

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

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

Chebyshev's Inequality

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

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

Chebyshev's Inequality can be used to estimate the mean of an unknown distribution. Often times we do not know the true mean, so we take many samples $X_1, X_2, ..., X_n$. Our sample mean is thus $S_n = \frac{X_1+X_2+...+X_n}{n}$

We can upper-bound the probability of that our sample mean deviates too much from our true mean as:

$$P(|\hat{\mu} - \mu| > \epsilon) \leq \delta$$

where ϵ is known as our error and δ is known as our confidence.

1.2 Questions

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

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

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

> **Solution:** We do not have the variance of X, so Chebyshev's bound is not applicable here. This leaves us with just Markov's Inequality. But, the Markov bound only applies on a non-negative random variable, so we must shift X somehow. Define a random variable $Y = X + 100$; Y is strictly larger than 0. Then, $\mathrm{E}(Y) = \mathrm{E}(X+100) = \mathrm{E}(X) + 100 = -60 + 100 = 40$. The upper bound on X that we want can be calculated via Y:
> $\mathrm{P}[X \geq -20] = \mathrm{P}[Y \geq 80] \leq \frac{\mathrm{E}(Y)}{80} = \frac{1}{2}$
> Hence, the best upper bound on $\mathrm{P}[X \geq -20]$ is $\frac{1}{2}$.

3. **A**s 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 are unhelpful.

4. 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 $\mathrm{P}[Y = 0] \leq \frac{1}{3}$.

2 Distributions

2.1 Bernoulli Distribution

Bernoulli Distribution: Bernoulli(p)
Random variable X has the Bernoulli distribution if it takes on value 1 with probability p, and value 0 with probability $1-p$. With the Bernoulli distribution we can model a single countable event, i.e. a single coin flip.
Expectation:

$$E(X) = 0 * (1-p) + 1 * p = p$$

Variance:

$$var(X) = E(X^2) - E(X)^2 = 0^2 * (1-p) + 1^2 * p - p^2 = p(1-p)$$

2.2 Binomial Distribution

Binomial Distribution: Bin(n, p)
The binomial distribution is used to count the number of successes in 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:

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

Expectation:

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

Variance:
Since our sum of Bernoulli trials is independent, we can do the following:

$$var(X) = var(X_1 + X_2 + ... + X_n) = var(X_1) + var(X_2) + ... + var(X_n) = np(1-p)$$

1. Say that the Cleveland Browns have probability $p = 0.15$ chance of winning a football game. Assume that each game is independent from the last. The football regular season has 16 games.

(a) Write an expression for the probability that they win between 6 and 8 games during the course of a season.

Solution:

$$\binom{16}{6}(0.15)^6(0.85)^{10} + \binom{16}{7}(0.15)^7(0.85)^9 + \binom{16}{8}(0.15)^8(0.85)^8$$

(b) Find the probability that they win at least one game during the course of a season.

Solution: We could sum up the probabilities that they win 1 game, 2 games, etc. or, we could find the probability that they win 0 games and subtract that from 1.

$$p = 1 - 0.85^{16} = .92$$

So even if they just have a 15% chance of winning a single game, they still have a 92% chance of winning a game over the course of a season. Given that the Browns had 2 consecutive $0 - 16$, win-less seasons, think about just how bad they must have been ! :)

2.3 Poisson Distribution

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.

The probability of k occurrences is

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

Expectation:

Since $\mathrm{E}(X) = np$ for the binomial distribution, and we set $\lambda = np$, our expectation is $\mathrm{E}(X) = \lambda$.

Variance:

For a binomial distribution, Var$(X) = np(1-p)$, which looks like $\lambda(1-p)$. However, we started with the assumption that p is very small, so we can assume $(1-p) \approx 1$ and thus $\lambda(1-p) \approx \lambda$. Therefore, Var$(X) = \lambda$.

1. On the UC Berkeley meme page, on average, 3 good memes are posted a week. What is the probability that in a given week:

   - 6 good memes are posted?

     Solution: $\lambda = 3$, so $p = \frac{3^6}{6!}e^{-3} = e^{-3} = 0.051$

   - No good memes are posted?

     Solution: $\lambda = 3$, so $p = \frac{3^0}{0!}e^{-3} = e^{-3} = 0.05$

   - More than 1 good meme is posted?

Solution: Solution: $1 - e^{-3} = 0.95$

2.4 Geometric Distribution

Geometric Distribution: Geom(p)
With the geometric distribution, we count the number of failures until the first success. The probability that the first success occurs on trial k is:

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

Expectation:

We can derive the geometric distribution from the binomial distribution. 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:

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

1. Andy Go-es to class 20% of days. What is the probability that:
   - the first time he goes to class is the 5^{th} day of school?

     Solution: $(1 - 0.2)^4 0.2$

   - the first time he goes to class is after the 5^{th} day?

     Solution: Let $X \sim Geom(n, p)$.

     $$p = \sum_{i=6}^{\infty} (1 - 0.2)^{x-1} 0.2$$

     We can solve this using the sum of an infinite geometric series, OR we can consider the following approach: in order for him to go to class after the 5^{th}

day, all he has to do is fail to go to class for 5 days in a row. This probability is $(1-0.2)^5 = 0.32$. In general: $P(X > k) = (1-p)^k$.

- He goes to class on the 5^{th} day or before?

Solution: Use the complement of the last question. $p = 1-(1-0.2)^5 = 0.68$. In general: $P(X \leq k) = 1-(1-p)^k$.

2.5 Questions

1. 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 8th? *Hint: Break this problem down into parts, and assign your result to the first part to the variable p.*

> **Solution:** "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 the geometric distribution:
>
> $$P(\text{April 8th is first day with exactly 15 emails}) = (1-p)^7 p$$
> $$\approx 0.996396^7 * 0.003604 \approx 0.003514$$

(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. Using 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.

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