INEQUALITIES, DISTRIBUTIONS 10

COMPUTER SCIENCE MENTORS 70

April 9 – 11, 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 $\alpha > 0$:

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

Proof of Markov's Inequality

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

Alternate Proof of Markov's Inequality
Consider the indicator random variable Y which equals 1 if $X \geq a$ and 0 otherwise.
Now consider the relationship between X and aY

- If $X < a$, then $Y = 0$, which means $aY = a * 0 = 0$
  Because X is a non-negative random variable, $X \geq 0$, so $aY \leq X$ in this case.
- If $X \geq a$, then $Y = 1$, which means $aY = a * 1 = a \leq X$

Thus, we have $aY \leq X$.
We take expectation on both sides to get:

$$E[aY] \leq X$$

$$aE[Y] \leq E[X]$$

$$E[Y] \leq \frac{E[X]}{a}$$

Now we note that the expectation of an indicator random variable is the probability that it is equal to 1 and we have the proof:

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

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:

2. 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:

   (b) Chebyshev's inequality:

3. **Bound It**
   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 $\mathrm{P}[X \geq -20]$.

4. The citizens of the country USD (the United States of Drumpf) vote in the following manner for their presidential election: if the country is liberal, then each citizen votes for a liberal candidate with probability p and a conservative candidate with probability $1-p$, while if the country is conservative, then each citizen votes for a conservative candidate with probability p and a liberal candidate with probability $1-p$. After the election, the country is declared to be of the party with the majority of the votes.
   (a) Assume that $p = \frac{3}{4}$ and suppose that 100 citizens of USD vote in the election and that USD is known to be conservative. Provide a tight bound on the probability that it is declared to be a Liberal country.

   (b) Now let p be unknown; we wish to estimate it. Using Chebyshev's Inequality, determine the number of voters necessary to determine p within an error of 0.01, with probability at least 0.95.

5. **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?

6. 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%.

7. 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 $\frac{1}{3}$.

8. 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:

   (b) Chebyshev's inequality:

9. Let $X_1, X_2, ..., X_n$ be n iid Geometric random variables with parameter p. Using Chebyshev's inequality, provide an upper-bound on: $P(|\frac{X_1+X_2+...+X_n}{n} - \frac{1}{p})| \geq a)$
Recall the variance for a Geometric Distribution with parameter p is $\frac{1-p}{p^2}$

10. Suppose we have a sequence of iid random variables $X_1, X_2, ..., X_n$
Let $A_n = \frac{X_1+X_2+...+X_n}{n}$ be the sample mean.
Show that the true mean of $X_i = \mu$ is within the interval $[\mu - 4.5\frac{\sigma}{\sqrt{n}}, \mu + 4.5\frac{\sigma}{\sqrt{n}}]$ with 95% probability.

2 Distributions

2.1 Bernoulli Distribution

Bernoulli Distribution: Bernoulli(p)

We say 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 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:

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

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:

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

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

$$\begin{aligned} var(X) &= var(X_1 + X_2 + \ldots + X_n) = var(X_1) + var(X_2) + \ldots + var(X_n) \\ &= p(1-p) + p(1-p) + \ldots + p(1-p) = np(1-p) \end{aligned}$$

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, 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, $var(X) = \lambda$.

2.4 Geometric Distribution

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:

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

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?

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

   (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.*

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

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

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