DISTRIBUTIONS, VARIANCE, INEQUALITIES, CONFIDENCE INTERVALS 8

COMPUTER SCIENCE MENTORS 70

March 19 - 26, 2018

1 Distributions

1.1 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:

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

1.2 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:

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

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:

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

1.3 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:

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

$$= e^{-\lambda} * \lambda * e^{\lambda}$$
$$= \lambda$$

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, $var(X) = \lambda$.

1.4 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:

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

1.5 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?

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

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

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?

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

(c) What is the variance of Y?

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?

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 $\mathrm{E}(X)$.

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

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

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

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 $i \neq j$)

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.

3 Markov, Chebyshev

3.1 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

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

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

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

3.2 Questions

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

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?

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. 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%.

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 $\frac{1}{3}$.

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:

   (b) Chebyshev's inequality: