

CSE 103: Note Oct. 9 2007

October 11, 2007

1 Unbounded Expectations

Expectations can be infinite! For example, imagine a probability defined over the positive integers as follows:

$$p(i) = \frac{1}{Z} \frac{1}{i^2}$$

Where Z is a normalization factor that makes the sum over all $p(i)$ equal 1. Calculating the expectation gives

$$\begin{aligned} E[i] &= \frac{1}{Z} \sum_{i=1}^{\infty} i \frac{1}{i^2} \\ &= \frac{1}{Z} \sum_{i=1}^{\infty} \frac{1}{i} \end{aligned}$$

This quantity tends to ∞ . Therefore the expectation is infinite. We can also have undefined expectations. Take for example the same probability measure as defined above, but now over all non-zero integers. Here when we calculate $E[i]$ we can get a different sum depending on the ordering of the terms in the sum. Therefore this sum's value is undefined!

2 Covariance

We define the covariance between random variables X and Y to be:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

where $\mu_X = E[X]$ and $\mu_Y = E[Y]$. Covariance is closely related to variance it's easy to verify straight from the definitions that $Cov(X, X) = Var(X)$. Note that it's also easy to show that if X and Y are independent, then $Cov(X, Y) = 0$:

$$\begin{aligned} Cov(X, Y) &= E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - E[X]\mu_Y - E[Y]\mu_X + \mu_X\mu_Y \\ &= E[X]E[Y] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = 0 \end{aligned}$$

However, $Cov(X, Y) = 0$ DOES NOT imply X, Y are independent!

It is also often useful to the correlation coefficient, ρ , that allows us to compare different covariances.

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Note that $-1 \leq \rho \leq 1$ always. Two variables that are positively correlated exhibit a positive linear trend, whereas two variable that are negatively correlated exhibit a negative linear trend. Keep in mind the examples of the circle and square as shown in class:

- The circle example had the data distributed uniformly along the circumference of a unit circle, where our two random variables were the variation along the X axis and the other variable the variation along the Y axis. Here the two variables are highly dependent since knowing the value of one narrows down the distribution of the other greatly. For example, say that we know that $X = \frac{1}{\sqrt{2}}$, then we know that $Y = \frac{1}{\sqrt{2}}$ or $Y = \frac{-1}{\sqrt{2}}$ with equal probability. The same dependence holds when we consider data distributed uniformly inside the unit circle.
- The square example had the data distributed uniformly within a square with side length 1 centered at the origin. Here, knowing the value of one variable (say, X axis) does not help narrow down the distribution of the other (Y axis), so the two variables are independent.

3 Conditional Probability

We define the conditional probability of event A given that event B has happened as:

$$P(A | B) = \frac{P(A \text{ and } B)}{P(B)}$$

In words, considering that B has happened and is now our entire universe, what proportion of times does A occur in this reduced universe? If A and B are independent then it is easy to verify that $P(A | B) = P(A)$, or in words: that knowing that B happened does not affect our belief about how often A will happen. So if A and B are independent events then $P(A | B) = P(A)$ and if they are dependent then $P(A | B) \neq P(A)$

4 Simple Bounds

4.1 Markov Bound

Let X be a non-negative random variable, then

$$P(X > t) \leq \frac{E[X]}{t}$$

This follows from the definition of expectation, and was proven in class.

4.2 Chebyshev's Bound

Let X be any random variable, then

$$P(|X - \mu_X| \geq t) \leq \frac{\text{Var}(X)}{t^2}$$

This follows by applying Markov's bound to this left hand side, and was proven in class.

Eventually we will give much stronger bounds than these two, but Chebyshev's is the flavor of the type of bound we would like since we want to know how unlikely it is to deviate far from the mean.

5 Histograms

In the classes to follow we will illustrate the difficulty of designing histograms that represent the underlying distribution accurately. We will soon see why the cumulative distribution function (CDF) is much better.