Probability Review

Stock & Watson Ch 2

Andrew Bossie

NJCU

February 4, 2017

Probability Distribution (of a continuous variable)

- Notation: $Pr(X = x_i)$ or $Pr(X < x_i)$ or $Pr(X = x_i + x_j)$ etc sometimes just $P(X = x)$
- Probability Density Function (PDF): Shows distribution of probabilities of individual outcomes
- Cumulative Density Function (CDF): Shows probability that $Pr(X) < x$
- Figure 2.2 picture of PDF/CDF Table 2.1 Table of Probability example

Expectations (Means and Variance)

Expected Value is the weighted mean of a random variable

$$E(Y) = \sum_{i=1}^{k} y_i p_i$$

It is what you you would "expect" the result to be on average with many repeated events.

Expected value of a loan where PV=\$100, i=10% and a 1% chance of default.
Expected number of times your computer will crash.

Variance and Standard Deviation.

These are measures of the "dispersion" or precision of the expected value
An event that will happen with certainty will have a variance and standard deviation of 0.

$$\sigma^2 = var(Y) = E(Y - \mu_y)^2 = \sum_{i=1}^{k}(y_i - \mu_y)^2 p_i$$

Units for variance are weird. We uses the square of $(y_i - \mu_y)$ so that negative and positive values don't cancel each other out. If Y=\$s then variance is $\$^2$ which doesn't really mean anything.

So we will often use standard deviation (sd) $sd = \sqrt{\sigma^2} = \sigma$ This is in the original units (i.e. \$s instead of $\$^2$).

The variance and standard deviation of computer crashes.

Mean and variance of linear function

Here we can think of a basic equation we will come across in econometrics. Y as a linear function of X. Basically we ask questions about how is X correlated with Y?

- Y=Output X=Government Spending
- Y=Diabetes X=Weight
- Y=Vote for Trump X=Immigrant Population

Mean and variance of linear function

Post-tax and pre-tax income: Y=post-tax income X=pre-tax income. Assume a 20% tax and a $2000 "deduction".
If we know a random persons pre-tax income, we can make a statement about their post-tax income and/or the variance of post tax income.

$$Y = 2000 + 0.8X$$

$$E(Y) = \mu_y = 2000 + 0.8\mu_X$$

$$var(Y) = E[(Y - \mu_y)^2]$$

Mean and variance of linear function

Variance of a linear function more generally:

$$Y = a + bX$$

$$\mu_y = a + b\mu_x$$

$$\sigma^2 = b^2\sigma_X^2$$

Joint and Marginal Distributions

Joint Probability Distribution: The probably that X and Y take on specific values (jointly!) of x_i and y_j.
Probability of all possible x_i and $y_j = 1$

Written as $Pr(X = x_i, Y = y_j)$
Table 2.2 Joint distribution of weather and commuting

Joint and Marginal Distributions

Marginal Probability Distribution: The probability of Y (alone)

$$Pr(Y = y) = \sum_{i=1}^{l} Pr(X = x_i, Y = y)$$

Is is the probability that Y=y given the joint distribution of all X.
Table 2.2 Joint distribution of weather and commuting

Conditional Distributions

Conditional Distribution: The distribution of a random variable Y conditional on another random variable X

Written as $Pr(Y = y|X = x)$

$$PR(Y = y|X = x) = \frac{Pr(X = x, Y = y)}{Pr(X = x)} = \frac{JointDistribution}{Prob - of - X}$$

Table 2.2 and Table 2.3

Conditional Mean

This is just the mean of Y when X=x

$$E(Y|X = x) = \sum_{i=1}^{k} y_i Pr(Y = y_i|X = x)$$

Law of iterated expectations

The mean of Y is the weighted average of the conditional expectation of Y given X.

$$E(Y) = E[E(Y|X)]$$

Textbook example: the mean height of adults is the mean height of men and the mean height of women weighted by men and women. This is kind of a “duh” thing so I will just leave it. If it is important later we can look at it in more detail later.

Conditional Variance

The variance of Y conditional on X. That is the variance of Y given that X=x.

$$var(Y|X = x) = \sum_{i=1}^{k} [y_i - E(Y|X = x)]^2 Pr(Y = y_i|X = x)$$

Table 2.3.

Independence

Two random variables are independent if knowing one provides no information about the other.

- Rolling dice 2 times
- The Patriots win the Superbowl, Puxatony Phil sees his shadow

$$Pr(Y = y|X = x) = Pr(Y = y)$$

or the joint distribution is:

$$Pr(X = x, Y = y) = Pr(X = x)Pr(Y = y)$$

Since Y and X are unrelated the probability of Y is not conditional on X (see the conditional distribution equation)

Covariance and Correlation

This is important. The covariance is just like the variance, except we have two variables instead of one.

$$cov(X, Y) = \sigma_{XY}$$

$$E[(X - \mu_X)(Y - \mu_Y)] = \sum_{i=1}^{k} \sum_{j=1}^{l} (x_j - \mu_X)(y_i - \mu_Y) Pr(X = x_j, Y = y_i)$$

Unlike variance, covariance can be negative (if Y and X move in different directions).

Covariance and Correlation

Correlation solves a similar problem as standard deviation does for variance. The units of covariance are awkward.

So we can turn the covariance into something you can more directly interpret.

$$corr(X, Y) = \frac{cov(X, Y)}{\sqrt{var(X)var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

where

$$-1 \leq corr(X, Y) \leq 1$$

if corr(X,Y)=0 the X and Y are uncorrelated. If $E(Y|X) = \mu_Y$ then cov(X,Y)=corr(X,Y)=0. The expected mean of Y does not depend on X at all.

Mean and Variance of Sums of Random Variables

There is more detail on this in the textbook. Means and variance of random variables have certain rules.
For the moment the most important are:

$$E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$$

and

$$var(X+Y) = var(X) + var(Y) + 2cov(X, Y)$$

Normal, Chi-Squared, Student t and F Distributions

Normal Distribution

Notation:

$$N(\mu, \sigma^2)$$

We define the normal distribution by its mean and variance. 95% of the values fall between $\mu + / - 1.96\sigma$
We can also transform a normal distribution into a standard normal distribution N(0,1) by subtracting the mean and dividing it by the standard deviation.

Normal, Chi-Squared, Student t and F Distributions

Multivariate Normal Distribution (MVND): A distribution of two or more random variables. The distribution is a function of the two variables.
aX+bY has a distribution of:
$N([a\mu_X + b\mu_Y], [a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}])$

There are four properties of multivariate distributions to keep in mind

1. Linear combinations of normally distributed random variables are also normally distributed.
2. Variables with a combined MVND all have normal marginal distributions.
3. Zero covariances implies independence (special property of a normal dist)
4. Joint normality of X and Y implies linearity of conditional expectations $E(Y|X = x) = a + bx$

Normal, Chi-Squared, Student t and F Distributions

Chi-squared distribution

This is a distribution you get when you square m independent standard normal variables.
m is the "degrees of freedom"
For example $Z^2 + Y^2 + X^2$ has a chi-squared distribution with 3 degrees of freedom

Notation: χ^2_m

Normal, Chi-Squared, Student t and F Distributions

Student t distribution (or just t distribution)

We usually use this distribution for small samples (n¡30). For larger samples students $t \approx \textit{normal}$ closely enough to just assume a normal distribution.

A t distribution approximates a normal distribution, but can be used to compare two variables or to assess the accuracy of a prediction when you don't know the actual mean and variance of the population.

The F Distribution

The F distribution is a ratio of two chi-square distributions (divided by degrees of freedom). Like the chi square distribution, this will come up later when we want to talk about comparing the distributions of variables.

Random Sampling

Sampling is very important in econometrics. Often we will want to make statements about populations using samples.

n observations of Y are denoted $Y_1, Y_2 ... Y_n$.

and independent and identically distributed draws (i.i.d.)

Identically Distributed: Because $Y_1, ... Y_n$ are randomly drawn from the same population they all have the same marginal distribution, which is the marginal distribution of the population.

Independent knowing Y_1 gives us no information about Y_2 or Y_n etc.

i.i.d.: When all Y_i are drawn from the same distribution and are all independent.

Sample Average

The mean of the sample is also a random variable. That means the sample average has a distribution because different samples will have difference averages.

The sample mean:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

We can also talk about this as the expected value of $\bar{Y}$

$$E(\bar{Y}) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \mu_y$$

Sample Variance

Calculating the variance of the sample average is an example of how the i.i.d. assumption makes things easier.

$$var(\bar{Y}) = var\frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$var(\bar{Y}) = \frac{\sigma_Y^2}{n}$$

If the distribution of the population is normal $N(\mu_Y, \sigma_y^2)$ then the distribution of n iid draws is $N(\mu_Y, \frac{\sigma_Y^2}{n})$

Large Sample Approximations of Sample Distributions

As our sample grows larger it approximates a normal distribution better and better. This is called the **asymptotic distribution** because it becomes more and more exactly a normal distribution when $n \rightarrow \infty$ We often call a sample "large" when n=30.

The asymptotic normal distribution of $\bar{Y}$ does not depend on the distribution of Y. This is important because normal distributions have properties that make them easier/simpler to work with than other distributions.

Again, the asymptotic distribution is a function of the size of the random sample, not the population.

Large Sample Approximations of Sample Distributions

Law of Large Numbers: $\bar{Y}$ will be near μ_y with increasing probability as n increases. This is called **convergence in probability** or **consistency** Figure 2.8.

Central Limit Theorem: The distribution of $\bar{Y}$ approaches $N(\mu_Y, \sigma_y^2)$ as $n \to \infty$

Homework

Textbook questions: 2.2, 2.3, 2.6, 2.14