ECON 186 Class Notes: Probability and Statistics Part 2

Jijian Fan

Expected Value

- For a discrete random variable X, the expected value is

$$E(X) = \sum_{x \in X} xp(x)$$

- For a continuous random variable X, the expected value is

$$E(X) = \int_{x \in X} xf(x)dx$$

- A convenient notation that is often used is $E(X) = \mu$. Also referred to as the mean.
- We can think of the expected value as a weighted average of values where the probabilities are the weights. It is the central tendency of a distribution.

Expected Value

- If $Y = g(X)$, then for a discrete random variable X,

$$E(Y) = E(g(X)) = \sum_{x \in X} g(x)p(x)$$

- If $Y = g(X)$, then for a continuous random variable X,

$$E(Y) = E(g(X)) = \int_{x \in X} g(x)f(x)dx$$

- Properties of expected value: For random variables X and Y and constants a, b,

$$E(aX + b) = aE(X) + b$$

$$E(X + Y) = E(X) + E(Y)$$

Variance

- Although the expected value is a useful measure of the central tendency of a distribution, we also want to know about its variability. The variance of a random variable X is

$$Var(X) = E\left[(X - E(X))^2\right]$$

- Specifically, in the discrete case:

$$Var(X) = \sum_{x \in X} (x - E(X))^2 p(x)$$

- In the continuous case:

$$Var(X) = \int_{x \in X} (x - E(X))^2 f(x) dx$$

- Notation: $Var(X) = \sigma^2$

Properties of Variance of One Random Variable

- For random variable X and constants a, b, the properties of variance are:

$$Var(X) > 0$$

$$Var(X) = E\left(X^2\right) - (E(X))^2 = E\left(X^2\right) - \mu^2$$

$$Var(a + bX) = b^2 Var(X)$$

$$Var(a) = 0$$

Examples

- Suppose that X is a random variable that takes only two values 0 and 1 with $Pr(X = 1) = p$. Find the expected value and variance of X.

$$E(X) = 0 \times (1 - p) + 1 \times p = p$$

$$E(X^2) = 0^2 \times (1 - p) + 1^2 \times p = p$$

$$Var(X) = p - p^2 = p(1 - p)$$

- This is actually the Bernoulli distribution! We will come back to this soon.

Examples

- Suppose that an appliance has a maximum lifetime of one year. The time X until it fails is a random variable with a continuous distribution having pdf

$$f(x) = \begin{cases} 2x & \text{\textit{for} } 0 < x < 1 \\ 0 & \textit{otherwise} \end{cases}$$

- Find the expected value and variance of X.

$$E(X) = \int_0^1 x(2x)dx = 2\int_0^1 x^2 dx = \frac{2x^3}{3}\Big]_0^1 = \frac{2}{3}$$

$$E(X^2) = \int_0^1 x^2(2x)dx = 2\int_0^1 x^3 dx = \frac{2x^4}{4}\Big]_0^1 = \frac{1}{2}$$

$$Var(X) = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{9}{18} - \frac{8}{18} = \frac{1}{18}$$

Examples

- Suppose that three random variables X_1, X_2, X_3 form a random sample from a distribution for which the mean is 5 for each. Find the expected value of $2X_1 - 3X_2 + X_3 - 4$.

$$E(2X_1 - 3X_2 + X_3 - 4) = E(2X_1) - E(3X_2) + E(X_3) - E(4)$$

$$= 2E(X_1) - 3E(X_2) + E(X_3) - 4 = 2(5) - 3(5) + 5 - 4 = -4$$

- Suppose that X and Y are independent random variables for which $Var(X) = Var(Y) = 3$. Find the values of $Var(X - Y)$ and $Var(2X - 3Y + 1)$.

$$Var(X - Y) = 1^2 Var(X) + (-1)^2 Var(Y) = Var(X) + Var(Y) = 3 + 3 = 6$$

$$Var(2X - 3Y + 1) = 4Var(X) + 9Var(Y) + Var(1) = 4(3) + 9(3) + 0 = 39$$

Standard Deviation and Covariance

- The commonly used standard deviation is simply the square root of the variance of X. Specifically, for a random variable X with a finite variance,

$$Std(X) = \sqrt{Var(X)}$$

- What if two random variables are not actually independent, but actually move, or covary together (which is the case for nearly all economic variables)? In this case, we must define their covariance.
- For two random variables X and Y,

$$Cov(X, Y) = E\left[(X - E(X))\,(Y - E(Y))\right]$$

- We can also express covariance as

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

- Note that

$$Cov(X, X) = Var(X)$$

Properties of Variance and Covariance

- Now that we have defined covariance, we can properly define the properties of variance with more than one random variable. Given random variables X and Y and constants a, b,

$$Var(aX + bY) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y)$$

- Properties of covariance: Given random variables W, X, Y, and Z, and constants a, b,

$$Cov(X, a) = 0$$

$$Cov(aX, bY) = abCov(X, Y)$$

$$Cov(W + X, Y + Z)$$

$$= Cov(W, Y) + Cov(W, Z) + Cov(X, Y) + Cov(X, Z)$$

Variance/Covariance Example

- Example: Consider our previous example with random variables X and Y where $Var(X) = Var(Y) = 3$. However, now instead of X and Y being independent, they are dependent on each other and covary with each other, and $Cov(X, Y) = 2$. Find $Var(2X - 3Y + 1)$.

$$Var(2X - 3Y + 1) = 4Var(X) + 9Var(Y) + 2(2)(-3)Cov(X, Y)$$

$$= 4(3) + 9(3) - 12(2) = 15$$

Correlation

- A key statistic when analyzing economic data is the correlation between variables. The correlation between two random variables X and Y is
$$Corr(X,Y) = \frac{Cov(X,Y)}{Std(X)Std(Y)}$$
- Notation: $Corr(X,Y) = \rho(X,Y)$
- Property of the correlation statistic:
$$-1 \leq Corr(X,Y) \leq 1$$
- The rest of the properties follow directly from the properties of covariance and standard deviations.
- An important note about the correlation statistic is that this only represents a linear relationship. Two random variables could have a low correlation coefficient but still be related non-linearly.

Correlation Example

- Suppose that X and Y are negatively correlated. Is $Var(X+Y)$ larger or smaller than $Var(X-Y)$?

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

$$Var(X-Y) = Var(X) + Var(Y) - 2Cov(X,Y)$$

- Since X and Y are negatively correlated, it must be that $Cov(X,Y) < 0$, so

$$Var(X-Y) > Var(X+Y)$$

Moments

- For each random variable X and every positive integer k, the expectation $E\left(X^k\right)$ is called the *kth* moment of X.
- Theorem: If $E\left(|X|^k\right) < \infty$ for some positive integer k, then $E\left(|X|^j\right) < \infty$ for every positive integer j such that $j < k$.
- Suppose that X is a random variable for which $E(X) = \mu$. Then, for every positive integer k, the expectation $E\left[(X-\mu)^k\right]$ is called the *kth* central moment of X.
- So, by definition, the mean of X is the first moment of X, the first central moment is 0 for all distributions, and the variance is the second central moment of X.

Moment Generating Function

- Let X be a random variable. Then, for each real number t, the moment generating function (mgf) of X is

$$\psi(t) = E\left(e^{tx}\right)$$

- Let X be a random variable whose mgf $\psi(t)$ is finite for all values of t in some open interval around the point $t = 0$. Then, for each integer $n > 0$, the *nth* moment of X, $E\left(X^n\right)$ is finite and equals the *nth* derivative $\psi^{(n)}(t)$ at $t = 0$. That is, $E\left(X^n\right) = \psi^{(n)}(0)$ for $n = 1, 2, \ldots$

Moment Generating Function Example

- Suppose that X is a random variable for which the pdf is as following:

$$f(x) = \begin{cases} e^{-x} & \text{for } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

- Find the expected value and variance of X using moment generating functions.

$$\psi(t) = E\left(e^{tx}\right) = \int_0^\infty e^{tx} e^{-x} dx = \int_0^\infty e^{x(t-1)} dx = \lim_{b \to \infty} \int_0^b e^{x(t-1)} dx$$

$$= \lim_{b \to \infty} \left. \frac{e^{x(t-1)}}{t-1} \right|_0^b = \lim_{b \to \infty} \frac{e^{b(t-1)}}{t-1} - \frac{e^{0(t-1)}}{t-1} = \frac{1}{1-t} \qquad \text{for } t < 1$$

$$\psi'(t) = \frac{1}{(1-t)^2}$$

$$\psi''(t) = \frac{2}{(1-t)^3}$$

Moment Generating Function Example

- Therefore, we know that

$$E(X) = \psi^{'}(0) = \frac{1}{(1-0)^2} = 1, \quad E\left(X^2\right) = \frac{2}{(1-0)^3} = 2$$

$$Var(X) = E\left(X^2\right) - [E(X)]^2 = \psi^{''}(0) - \left[\psi^{'}(0)\right]^2 = 2 - 1 = 1$$

Bernoulli Distribution

- Bernoulli distribution - Only two outcomes. Event A happens with probability p. So if the random variable X is assigned if Event A happens and 0 otherwise, then the pmf of the Bernoulli distribution is

$$f(x) = Pr(X = x) = \begin{cases} p & for\ x = 1 \\ q = 1 - p & for\ x = 0 \\ 0 & otherwise \end{cases}$$

- A classic example of the Bernoulli distribution is the flipping of a coin.

$$E(X) = p$$

$$Var(X) = p(1 - p)$$

Binomial Distribution

- Binomial distribution - n Bernoulli experiments repeated independently with probability of success p.
- The random variable X is defined to be the number of successes in n Bernoulli trials. Then, the pmf is

$$f(x) = Pr(X = x) = \begin{cases} \begin{pmatrix} n \\ x \end{pmatrix} p^x (1-p)^{n-x} & \text{for } x = 0, 1, ..., n \\ 0 & \text{otherwise} \end{cases}$$

- Frequently used to model the number of successes in a sample of size n drawn with replacement from a population of size N.
- One example is if a surveyor picks a random $50,000$ people out of the population and asks them a yes or no question each year for 10 years. The answer to this question could be modeled as a binomial distribution.

$$E(X) = np$$

$$Var(X) = np(1-p)$$

Hypergeometric Distribution

- Hypergeometric distribution - A box has A red balls and B blue balls. n balls are drawn without replacement.
- The random variable X is the number of red balls. Then, the pmf is

$$f(x) = \begin{cases} \left(\dfrac{\begin{pmatrix} A \\ x \end{pmatrix} \begin{pmatrix} B \\ n-x \end{pmatrix}}{\begin{pmatrix} A+B \\ n \end{pmatrix}} \right) & \textit{for } max\{0, n-B\} \leq X \leq min\{n, A\} \\ 0 & \textit{otherwise} \end{cases}$$

- The hypergeometric distribution represents the number of successes in n trials *without* replacement. That is, it is the binomial distribution without replacement. Common uses are testing which sub-populations are over or underrepresented in a sample.

$$E(X) = \frac{nA}{A+B} \qquad Var(X) = \frac{nAB}{(A+B)^2} \frac{A+B-n}{A+B-1}$$

Poisson Distribution

- Poisson Distribution - Suppose that the random variable X represents how many customers enter a store in a given time period and λ is the intensity parameter, then

$$f(x) = Pr(X = x) = \begin{cases} \frac{e^{-\lambda}\lambda^{x}}{x!} & \text{for } x = 0, 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$$

- The Poisson distribution, assuming the "event flow" is random, models how likely it is for an event to occur in a given period of observation, given a mean intensity λ.

$$E(X) = Var(X) = \lambda$$

- Interestingly, the mean and variance of the Poisson distribution is equal. In econometrics, Poisson regression is sometimes used to model the frequency of an event for a binary dependent variable, but note that it only makes sense to use if the mean and variance of the dependent variable are the same.

Negative Binomial Distribution

- Negative binomial distribution - Suppose we perform independent and identically distributed Bernoulli trials until r successes are observed, where each Bernoulli trials has success with probability p. The random variable X is the number of failures that occur before r successes. Then,

$$f(x|r,p) = \begin{cases} \begin{pmatrix} r+x-1 \\ x \end{pmatrix} p^r (1-p)^x & \textit{for } x = 0,1,2,..., \\ 0 & \textit{otherwise} \end{cases}$$

- For example, suppose that we want to know how many tails ("failure") we will get before we get 10 heads("success"). The negative binomial will model this.
- When $r = 1$, the distribution is called the Geometric distribution.

Negative Binomial Distribution

- In econometrics, similarly to Poisson, negative binomial can be used to model a binary dependent variable, but the mean and variance can differ.

$$E(X) = \frac{pr}{1-p}$$

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

Uniform Distribution (Continuous)

- Continuous uniform distribution - A distribution with minimum value a and maximum value b where all intervals of the same length on the distribution's support are equally probable. Then the pdf is

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

- One useful application is that when a p-value is used as a test statistic for a simple null hypothesis, and the distribution of the test statistic is continuous, then the p-value is uniformly distributed between 0 and 1 if the null hypothesis is true.
- Another application is drawing random numbers from a computer program, which will draw from the uniform distribution.

$$E(X) = \frac{1}{2}(a+b)$$

$$Var(X) = \frac{1}{12}(b-a)^2$$

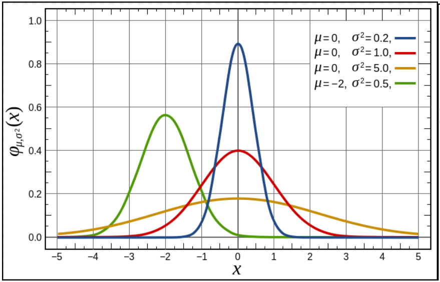
Normal Distribution

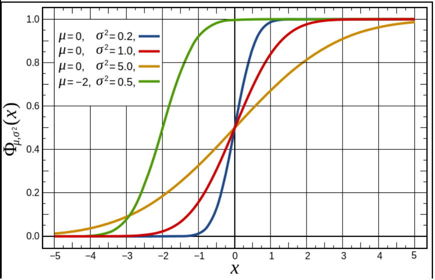
- The normal distribution, also known as the Gaussian distribution or informally as the bell curve, is a very commonly occurring and thus widely used distribution.
- Examples include the price of homes in a given area, as well as height, weight, and heart rate of a given population. Due to the Central Limit Theorem (which we will return to later), the normal distribution is crucial in statistical inference.
- The normal distribution is fully characterized by its mean μ and its standard deviation σ. The pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2\sigma^2}(x-\mu)^2}$$

- Notation: If X is normally distributed with mean μ and standard deviation σ, then $X \sim N(\mu, \sigma)$.

Normal Distribution





Standard Normal Distribution

- Using the properties of expected values and variances, we get that if $X \sim N(\mu, \sigma)$, then

$$a + bX \sim N\left(a + b\mu, b^2\sigma^2\right)$$

- Now, suppose that $a = -\frac{\mu}{\sigma}$ and $b = \frac{1}{\sigma}$. Then we have

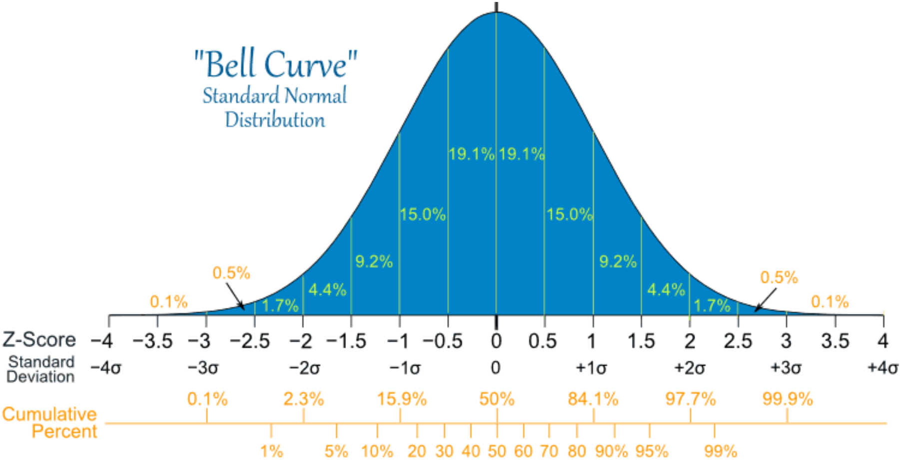
$$\left(-\frac{\mu}{\sigma}\right) + \left(\frac{1}{\sigma}\right) X \sim N(-\frac{\mu}{\sigma} + \left(\frac{1}{\sigma}\right)\mu, \left(\frac{1}{\sigma}\right)^2 \sigma^2) \sim N(0,1)$$

- Then, we can define the standard normal distribution as

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

- So, to get the standard normal distribution, we take the normally distributed X and divide by its mean and standard deviation, which is known as normalizing.
- Each value of the distribution is now the number of standard deviations the value is from the mean.

Standard Normal Distribution



Standard Normal Distribution Example

- Suppose that $X \sim N(5, 2)$. Then, $Z = \frac{X-5}{2}$. Then Z has the standard normal distribution.

$$Pr(1 < X < 8) = Pr\left(\frac{1-5}{2} < \frac{X-5}{2} < \frac{8-5}{2}\right) = Pr\left(-2 < Z < 1.5\right)$$

$$= Pr\left(Z < 1.5\right) - Pr\left(Z < -2\right) = Pr\left(Z < 1.5\right) - \left(1 - Pr\left(Z < 2\right)\right)$$

$$= 0.9332 - 1 + .9772 = 0.9104$$

Chi-Square Distribution

- If $Z \sim N(0,1)$ then $Z^2 \sim \chi^2(1)$, where $\chi^2(1)$ is a chi-square distribution with 1 degree of freedom (the number of values in the final calculation of a statistic that are free to vary).
- If $X_1, \ldots, X_n$ are n independent $\chi^2(1)$ random variables, then

$$\sum_{i=1}^{n} X_i \sim \chi^2(n)$$

- This is a very useful distribution since we often consider the transformation and sum of many normally distributed random variables.
- Some uses are the commonly used chi-squared test as well as the definition of the important Student's t distribution.

F Distribution

- If X_1 and X_2 are independent chi-squared random variables with degrees of freedom n_1 and n_2, respectively, then

$$F(n_1, n_2) = \frac{X_1/n_1}{X_2/n_2}$$

- has an F-distribution with (n_1, n_2) degrees of freedom.
- The primary use of this distribution is the F-test, which is often used to compare statistical models. Specifically, the F-test most often thought of by econometricians tests how well the regression model fits the data overall, a standard statistic reported in any regression output.

Samples

- A random sample is a sample of n observations of one or more variables drawn independently from the same population of probability distributions, $f(x_i;\theta)$.
- These n observations are independent and identically distributed and we write them as $X_1, X_2, ..., X_n$.
- We have previously discussed various statistics such as the mean and variance for a distribution. However, since we very rarely observe the entire population, we need to know these descriptive statistics for samples.

Sample Statistics

- The sample mean is specified as

$$\bar{X} = \frac{\sum X_i}{n}$$

- The median is the middle observation if all observations are ranked by their value.
- The sample variance is

$$s^2 = \frac{1}{n-1} \sum \left(X_i - \overline{X}\right)^2$$

Example of Sample Statistics

- Suppose we have a list of prices for one good over a 5 month period $p_1, ..., p_5 = 15, 22, 45, 46, 111$. Calculate the sample mean, median, and standard deviation.

$$\overline{X} = \frac{\sum X_i}{n} = \frac{15 + 22 + 45 + 46 + 111}{5} = 47.8$$

$$Median = 45$$

$$s_x = \sqrt{\frac{\sum \left(X_i - \overline{X}\right)^2}{n-1}} =$$

$$\sqrt{\frac{(15 - 47.8)^2 + (22 - 47.8)^2 + (45 - 47.8)^2 +}{4}}$$

$$\sqrt{\frac{+ (46 - 47.8)^2 + (111 - 47.8)^2}{4}} = 37.9$$

Estimation

- A point estimate is a statistic computed from a sample that gives a single value for the parameter we are trying to estimate, θ.
- The standard error is the standard deviation of the sampling distribution of the statistic.
- An interval estimate is the range of values that will contain the true parameter with a preassigned probability.
- An estimator is a rule or strategy for using data to estimate the parameter.