Eco 4306 Economic and Business Forecasting

Appendix A: Review of Probability and Statistics

Outline

Review

- population and sample
- random variable
- probability density function and cumulative distribution function
- moments of the probability distribution
  - measures of centrality: mean, median
  - measures of dispersion: variance, standard deviation
  - measures of shape: skewness, kurtosis

1. Population and Sample

- suppose that we want to forecast the income tax revenue in some state, e.g. California
- we would be interested in leaarning as much as posisble about the distirbution of income
- mean (or median) household income, how many households are below the poverty threshold, how many households are in the highest bracket of income, . . .
- we can proceed in two ways
  (1) collect **population information**: interview every single household in California and ask for their level of income so that we collect information on *all* households in the state
  (2) obtain **sample information**: choose a subset of households in California and ask for the level of income collecting a subset of information
- ideally, we would like to have information on the full population
- this kind of extensive data collection would however be very expensive and time consuming
- because of this constraint, we are forced to work with sample information

1. Population and Sample

- statistical inference provides the necessary tools to infer the properties of the population based on sample information
- let Y be random variable “household income” - this is a continuous random variable
- suppose that we choose a random sample of 100 households in California, for each we have a realization or outcome so that our sample information is denoted as $\{y_1, y_2, \ldots, y_{100}\}$
- there are also discrete random variables for which the outcomes are associated with only non-negative integers, e.g., the “number of cars in a household” $y_i =$ 0, 1, 2, or 3
- notation: we will be using uppercase letters to denote the random variable and lowercase to denote a particular numerical value or outcome of the random variable

1. Population and Sample

- we can characterize a random variable Y by
  (1) **cumulative distribution function** (cdf) and **probability density function** (pdf)
  (2) **moments** of Y such as the **mean**, **median**, **variance**, **skewness**, **kurtosis**, etc.
- first route is superior, if we have cdf or pdf, we can calculate any moment
- in some cases we will be interested in a particular features of Y and knowing only some moments will be sufficient

2. Cumulative Distribution Function (CDF) and Probability Density Function (PDF)

- for discrete random variable pdf provides the probability associated with each outcome
- e.g. if we throw a fair die random variable Y has six possible outcomes $\{1, 2, 3, 4, 5, 6\}$ and each outcome has an equal probability of occurrence, $P(Y = 1) = P(Y = 2) = \ldots = P(Y = 6) = 1/6$
- we define the pdf of a discrete random variable as the function $f(y)$ that assigns to each outcome y_i for $i = 1, 2, \ldots, k$ a probability p_i

$$f(y_i) = P(Y = y_i) \equiv p_i \qquad \text{for } i = 1, 2, \ldots, k$$

- probabilities satisfy $0 \leq p_i \leq 1$ and $\sum_{i=1}^{k} p_i = 1$
- the cdf $F(y)$ provides answers to questions such as what the probability is that on throwing a die we get at most a y
- for example if $y = 3$ we are considering just three possible outcomes $\{1, 2, 3\}$, and the probability that we are concerned about is

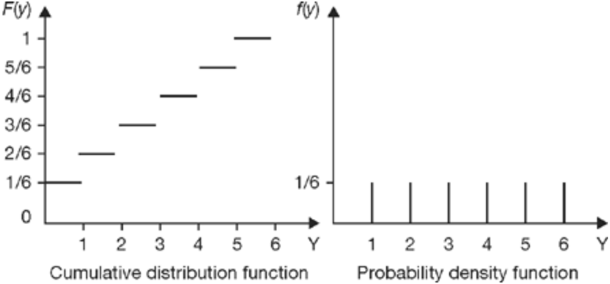
$$P(Y \leq 3) = P(Y = 1) + P(Y = 2) + P(Y = 3) = 1/6 + 1/6 + 1/6 = 1/2$$

- cdf thus accumulates the probabilities in pdf of single events y_j for all j such that $y_j \leq y$

$$F(y) = P(Y = y_i) \equiv p_i \qquad \text{for } i = 1, 2, \ldots, k$$

2. Cumulative Distribution Function (CDF) and Probability Density Function (PDF)

- cdf and pdf for the throw of a die

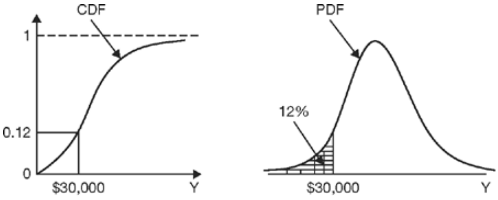


2. Cumulative Distribution Function (CDF) and Probability Density Function (PDF)

- for continuous random variable, cdf is defined in a similar fashion

$$F(y) = P(Y \leq y)$$

- example: probability that a household in California has a level of income of at most \$30,000 is $F(30,000) = P(Y \leq 30,000)$
- pdf of a continuous variable is a continuous function $f(y)$ such that the area under $f(y)$ up to the point y is the probability for a range of outcomes $(-\infty, y]$



Cumulative distribution function

Probability density function

3. Moments of a Random Variable

- random variable can also be characterized by its moments, measures based on the probability density function of the random variable
- we will review three types of measures
  (1) measures of centrality
  (2) measures of dispersion
  (3) measures of shape
- need to distinguish between **population moments** and **sample moments**

3.1 Measures of Centrality

- measure of centrality is the **mean** or **expected value**, which is also known as the *first central moment* of Y
- mean is a weighted average of all possible values of Y, with weights given by probabilities
- for a discrete random variable Y with pdf $f(y)$

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

- for a continuous random variable Y with pdf $f(y)$

$$E(Y) = \int_{-\infty}^{\infty} y f(y) dy$$

3.1 Measures of Centrality

- expected value is also known as the **population mean** and is usually denoted as $\mu_Y = E(Y)$
- properties of the mean
  1. expectation of a constant c is the constant itself: $E(c) = c$
  2. for any constants a and b, the expectation of a linear combination $aY + b$ is $E(aY + b) = aE(Y) + b$.
  3. expectation of a linear combination of random variables $Y_1, Y_2, \ldots, Y_k$ is the sum of the expectations, so for any constants $c_1, c_2, \ldots, c_k$ it holds $E(c_1Y_1 + c_2Y_2 + \ldots + c_kY_k) = c_1E(Y_1) + c_2E(Y_2) + \ldots + c_kE(Y_k)$

3.1 Measures of Centrality

- in practice, we work with sample rather than population information
- we thus compute sample statistics, which will are approximations to population statistics
- for a sample of n observations $\{y_1, y_2, \ldots, y_n\}$, **sample mean** $\bar{y}$ is defined as

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$$

- sample mean is an estimator of the population mean $E(Y)$, it can be proven that $E(\bar{y}_n) = E(Y)$, thus sample mean is an unbiased estimator of the population mean

3.1 Measures of Centrality

- another measure of centrality: **median** y_m
- this is the value of Y for which one half of the observations are below y_m and the other half are above it
- using cdf, median y_m is the value such that $F(y_m) = P(Y \leq y_m) = 0.50$
- if the pdf is symmetric around μ_Y, then the median and the mean are identical

3.2 Measures of Dispersion

- **variance** $var(Y)$ measures how far on average values of Y are from the mean μ_Y
- it is defined as the expected value of the squared deviations of Y from μ_Y

$$var(Y) = E(Y - \mu_Y)^2$$

- variance $var(Y)$ is also denoted as σ_Y^2, and it is also known as the *second central moment* of Y
- properties
  - variance of a random variable is always a positive number; variance of a constant is zero
  - variance of a linear combination: $var(aY + b) = a^2 var(Y)$, for any constants a and b

3.2 Measures of Dispersion

- units of variance are different from units of mean because in variance we take the square of the values of Y; so for example, if the values are measured in dollars, then the variance is in squared dollars
- to have a measure of dispersion with the same units as the mean, we define the **standard deviation** σ_Y of Y as the square root of the variance

$$\sigma_Y = \sqrt{var(Y)}$$

- corresponding sample moments are the **sample variance** $\hat{\sigma}_n^2$ and the **sample standard deviation** $\hat{\sigma}_n$; for a random sample of n observations $\{y_1, y_2, \ldots, y_n\}$

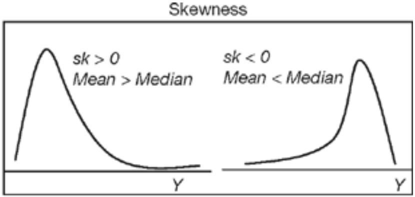
$$\hat{\sigma}_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad \hat{\sigma}_n = \sqrt{\hat{\sigma}_n^2}$$

3.3 Measures of Shape

- define a measure of **skewness** as

$$sk = E\left[\left(\frac{Y - \mu_Y}{\sigma_Y}\right)^3\right]$$

- if the pdf of a random variable Y is symmetric around μ_Y, then the skewness is zero
- positive values of sk: distribution is skewed to the right, median is smaller than mean
- negative values of sk: distribution is skewed to the left, median is larger than mean



3.3 Measures of Shape

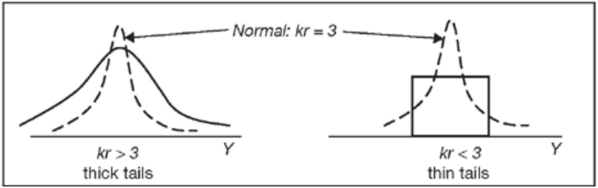
- define measure of **kurtosis** as

$$kr = E\left[\left(\frac{Y-\mu_Y}{\sigma_Y}\right)^4\right]$$

  which provides information about the "thickness" of the tails of the distribution

- kurtosis is always positive, and because it is scaled by σ_Y^4, it is also a unit-free measure
- higher the value of kr mean that tails are thicker
- we measure kurtosis in relation to the kurtosis of a normal distribution - for a normal pdf, $kr = 3$
- if $kr > 3$, we say that there is **leptokurtosis** or **fat tails**; this means that there is more probability in the tails than in the density of normal distribution
- if $kr < 3$, we say that the distribution has **thin tails**, less probability in tails than normal distribution

Kurtosis



4. Common Probability Density Functions

- review of four common probability density functions
  - normal distribution
  - chi-square distribution
  - the student-t distribution
  - F distribution

4.1 Normal and Log-Normal Distribution

- random variable Y is normally distributed $Y \sim N(\mu, \sigma^2)$
- pdf $f(y)$ is bell-shaped, it is centered in the mean μ, variance is σ^2, it is symmetric thus skewness is $sk = 0$, and the kurtosis is $kr = 3$
- if $Y \sim N(\mu, \sigma^2)$ then $Z = \frac{Y-\mu}{\sigma} \sim N(0, 1)$ is the standard normal distribution
- customary to write the pdf of standard normal as $\phi(z)$ and the cdf as $\Phi(z) = P(Z \leq z)$

4.1 Normal and Log-Normal Distribution

- random variable Y is log-normally distributed $Y \sim N(\mu, \sigma^2)$ if $X = \log Y$ is a normally distributed random variable $X \sim N(\mu, \sigma^2)$
- Y only takes on positive values
- shape of the pdf is asymmetric and right skewed, so the coefficient of skewness is positive

4.2 Chi-Square Distribution

- chi-square distribution with ν degrees of freedom is defined as the sum of ν independent standard normal random variables

$$\chi^2_\nu = \sum_{i=1}^{\nu} Z_i^2$$

  where $Z_i \sim N(0, 1)$
- shape of the pdf is asymmetric and skewed to the right, so the coefficient of skewness is positive
- the larger the degrees of freedom, the less skewed the density becomes

4.3 Student-t Distribution

- Student-t distribution with ν degrees of freedom is defined as the ratio of a standard normal random variable to the squared root of a chi-square random variable

$$t_\nu = \frac{Z}{\sqrt{\chi^2_\nu/\nu}}$$

  where Z and χ_ν are independent random variables
- degrees of freedom of the Student-t are the same as those of the chi-square random variable in the denominator
- Student-t distribution has a pdf symmetric around zero, skewness coefficient is thus equal to zero
- main difference with the normal density is that there is more probability mass in the tails - Student-t has fat tails
- lower the degrees of freedom are, the fatter the tails become, making the kurtosis coefficient larger than 3
- as the degrees of freedom increase, Student-t converges toward the standard normal

4.4 F-Distribution

- F-distribution is the ratio of two independent chi-square distributions

$$F_{\nu_1,\nu_2} = \frac{\chi^2_{\nu_1}/\nu_1}{\chi^2_{\nu_2}/\nu_2}$$

- this is a ratio of two strictly positive random variables, so F-distribution is also strictly positive
- it is right skewed density

Example: S&P 500 weekly returns

- weekly returns of the S&P 500 - the percentage gain or loss that you would obtain if you were to buy the S&P 500 index and sell it one week later
- histogram that is a frequency distribution
- when we smooth the histogram, we obtain an estimation of the pdf
- mean weekly return is 0.16%, median 0.28%, standard deviation is 2.06%
- density is very peaked, centered approximately in zero, though there is some mild negative skewness $sk = -0.35$
- tails are quite fat with a coefficient of kurtosis of $kr = 7.94$, which is much larger than 3 (the value for the normal density)
- because kurtosis is very large density is closer to a Student-t than to the normal

Example: S&P 500 weekly returns

- EViews output also shows the Jarque-Bera statistic JB based on the kurtosis coefficient kr and the skewness coefficient sk

$$JB = \frac{n}{6}\left(sk^2 + \frac{(kr-3)^2}{4}\right)$$

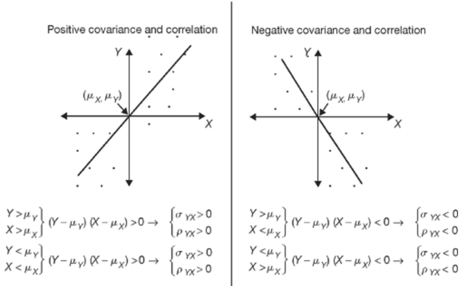
  with n being the number of observations
- the JB statistic is used to construct a test for normality
- null hypothesis is H_0: density is normal, and the alternative H_1: density is not non-normal
- under H_0, $JB = 0$ because $sk = 0$ and $kr = 3$
- Jarque-Bera test is distributed as a chi-square with 2 degrees of freedom
- we will reject normality whenever $JB > \chi^2_{2,\alpha}$ where $\chi^2_{2,\alpha}$ is the critical value of the chi-square distribution at 100α% level
- for weekly S&P 500 returns $JB = 3637.15 > 5.991$ and so we reject normality very strongly

5 Measures of Association: Covariance and Correlation

- mean, median, variance, standard deviation, skewness, and kurtosis are all **univariate moments**, they refer to only one random variable
- in economics and business we are very often interested in finding relations between two or more variables
- for example,
  - we might want to know how income and consumption are related to each other
  - how the stock market and the real economy interact with each other
- **covariance**: a measure of linear relationship between two random variables Y and X defined as

$$\sigma_{YX} = cov(X, Y) = E[(Y - \mu_Y)(X - \mu_X)]$$

- when $\sigma_{XY} > 0$ then Y and X tend to move in the same direction
- when $\sigma_{XY} < 0$ then Y and X tend to move in opposite directions



5 Measures of Association: Covariance and Correlation

- for any two random variables Y_1, Y_2 and two constant c_1, c_2

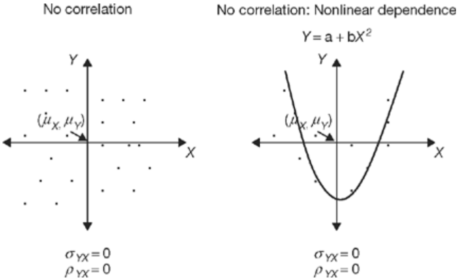
$$var(c_1Y_1 + c_2Y_2) = c_1^2 var(Y_1) + c_2^2 var(Y_2) + 2c_1c_2 cov(Y_1, Y_2)$$

- more generally

$$var(c_1Y_1 + c_2Y_2 + \ldots + c_kY_k) = \sum_{i=1}^{k} c_i^2 var(Y_i) + 2\sum_{i=1}^{k-1}\sum_{j>i}^{k} c_ic_j cov(Y_i, Y_j)$$

5 Measures of Association: Covariance and Correlation

- covariance measures only linear dependence
- there may be a nonlinear relation between Y and X, for instance a quadratic relation $Y = a + bX^2$, but the covariance will be zero because sometimes Y and X move in the same direction and sometimes in opposite directions
- thus when Y and X are independent than $\sigma_{YX} = 0$, but the opposite is not -true - covariance may be equal to zero and the random variables may still be dependent, albeit in a nonlinear fashion



5 Measures of Association: Covariance and Correlation

- for samples of n observations, $\{y_1, y_2, \ldots, y_n\}$ and $\{x_1, x_2, \ldots, x_n\}$, **sample covariance** is calculated as

$$\hat{\sigma}_{YX} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)(x_i - \bar{x}_n)$$

where $\bar{y}_n$ and $\bar{x}_n$ are sample means of Y and X

5 Measures of Association: Covariance and Correlation

- **correlation coefficient**: another measure of linear relationship between two random variables
$$\rho_{YX} = \frac{\sigma_{YX}}{\sigma_Y \sigma_X}$$
- advantage of correlation coefficient over covariance: easy to interpret, does not have units, is bounded $-1 \leq \rho_{YX} \leq 1$