FN241: Session 2-3

Part 2: Probability and Statistical Concepts

Winai Homsombat

Bachelor of Economics, International Program

Thammasat University

Common Probability Distributions

Probability distribution

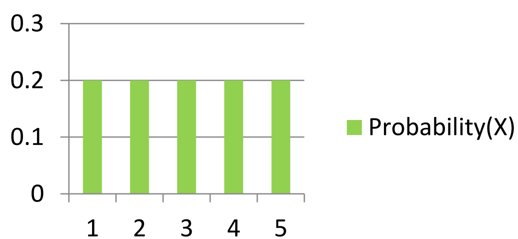
The set of probabilities for the possible outcomes of a random variable is called a "probability distribution."

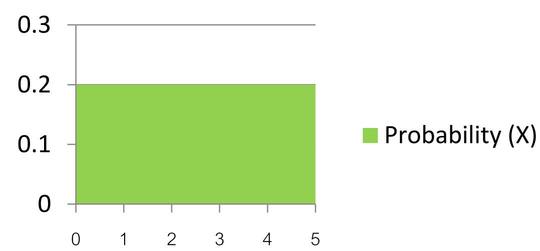
- The underlying foundation of most inferential statistical analysis is the concept of a probability distribution.
- The focus in the investments arena is on four probability distributions.
  1. Uniform
  2. Binomial
  3. Normal
- An understanding of probability distributions is critical to using further quantitative methods such as hypothesis testing, regression, and time-series analysis.

Discrete and Continuous Random Variables

A random variable is a variable whose future values are uncertain.

- **Discrete random variables** have a theoretically countable number of outcomes.
  - There may be an infinite number of them, but they are countable.
  - Price is a discrete random variable.
- **Continuous random** variables have a theoretically uncountable number of outcomes.
  - Rate of return is a continuous random variable.
  - Temperature is a continuous random variable.





The probability density function (pdf)

The mathematical expression that describes the individual probabilities that a random variable will take on each of a set of specified values is known as its probability density function.

- For **a discrete distribution**, the pdf has discrete, countable, nonzero probabilities for every possible outcome.
- For **a continuous distribution**, the pdf has continuous, uncountable probabilities for each possible specified outcome in the set of infinite, uncountable outcomes. Hence, the probability of any specific outcome is zero.
  - For continuous distributions, this result means that the cumulative distribution function will be more useful and, to some extent, more meaningful.

$$\text{pdf(Div)} = \begin{cases} P(\text{Div} = \$1) = 0.05 \\ P(\text{Div} = \$5) = 0.30 \\ P(\text{Div} = \$7) = 0.50 \\ P(\text{Div} = \$10) = 0.10 \\ P(\text{Div} = \$11) = 0.05 \end{cases}$$

The cumulative distribution function (cdf)

The mathematical expression that describes the probability that a random variable will be less than or equal to a specific value for all possible values of that variable is known as its cumulative distribution function.

- The cumulative distribution function, denoted *F(x)*, is represented as the sum of the probabilities of the specified outcome and all prior outcomes in the distribution for each and every possible outcome.
  - By analogy, this is very similar to the concept of cumulative relative frequency from Chapter 3 on Statistical Concepts and Market Returns.
- The cdf has the same properties as the pdf, in addition to
  1. All values of the cdf are between 0 and 1;
  2. As we increase the value of the specified outcome, the cdf must increase or remain constant.

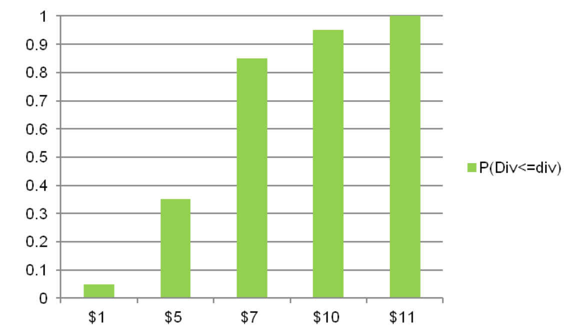
$$\text{pdf(Div)} = \begin{cases} P(\text{Div} = \$1) = 0.05 \\ P(\text{Div} = \$5) = 0.30 \\ P(\text{Div} = \$7) = 0.50 \\ P(\text{Div} = \$10) = 0.10 \\ P(\text{Div} = \$11) = 0.05 \end{cases}$$

$$\text{cdf(Div)} = \begin{cases} P(\text{Div} \le \$1) = 0.05 \\ P(\text{Div} \le \$5) = 0.35 \\ P(\text{Div} \le \$7) = 0.85 \\ P(\text{Div} \le \$10) = 0.95 \\ P(\text{Div} \le \$11) = 1.00 \end{cases}$$

The cdf in action

- **Example:** Returning to the special dividend example, the cdf can be written and depicted as:

$$\text{cdf(Div)} = \begin{cases} P(\text{Div} \leq \$1) = 0.05 \\ P(\text{Div} \leq \$5) = 0.35 \\ P(\text{Div} \leq \$7) = 0.85 \\ P(\text{Div} \leq \$10) = 0.95 \\ P(\text{Div} \leq \$11) = 1.00 \end{cases}$$



- What is the probability of receiving at most a $10 dividend?

- What is the probability of receiving more than a $7 dividend?

The discrete uniform distribution

A basic distribution wherein the probably of every possible countable outcome is equally likely.

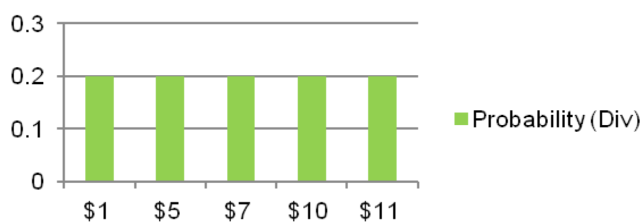
Consider again our special dividend with five possible year-end outcomes of \$1, \$5, \$7, \$10, and \$11, except that **now the probability of each outcome is 0.2, so this is a discrete uniform random variable.**

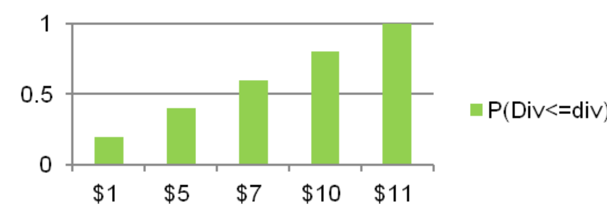
- The outcomes are countable.
- Each possible outcome has a probability between 0 and 1.
- The sum of the probabilities of the outcomes is 1.0.
- AND, the probability of each outcome is ....
  - Our new special dividend is a discrete uniformly distributed random variable.

The discrete uniform distribution

Example: What is the probability we will receive a $5 dividend? A dividend of **at least** $5?

$$\text{pdf(Div)} = \begin{cases} P(\text{Div} = \$1) = 0.20 \\ P(\text{Div} = \$5) = 0.20 \\ P(\text{Div} = \$7) = 0.20 \\ P(\text{Div} = \$10) = 0.20 \\ P(\text{Div} = \$11) = 0.20 \end{cases} \qquad \text{cdf(Div)} = \begin{cases} P(\text{Div} \leq \$1) = 0.20 \\ P(\text{Div} \leq \$5) = 0.40 \\ P(\text{Div} \leq \$7) = 0.60 \\ P(\text{Div} \leq \$10) = 0.80 \\ P(\text{Div} \leq \$11) = 1.00 \end{cases}$$





- A dividend that is greater than $9?

The continuous uniform distribution

Characterizing the Distribution

- The pdf and cdf for a continuous uniform distribution are written as

$$f(x) = \begin{cases} \dfrac{1}{b-a} \, when \ a < x < b \\ 0 \ otherwise \end{cases}$$

$$F(x) = \begin{cases} 0 \ when \ x \leq a \\ \dfrac{x-a}{b-a} \ when \ a < x < b \\ 1 \ when \ x \geq b \end{cases}$$

- Probabilities are calculated from $P(a \leq x \leq b) = \int_a^b f(x)dx$

- The mean and variance of a continuous uniformly distributed variable are

$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}$$

Binomial Random Variables

A binomial random variable has only two possible outcomes, termed "success" and "failure" by convention.

- The basic building block of the binomial distribution is a Bernoulli random variable.
  - A Bernoulli random variable is one for which there are only two possible outcomes, and the probability of these outcomes satisfies the conditions for a valid pdf. That is, each probability is between 0 and 1 and they sum to 1.
  - A single observation of the outcome of a Bernoulli random variable is called a "trial" when the random variable can repeat.
- The sum of a series of Bernoulli trials is distributed as a binomial random variable.
- In order to use the binomial distribution, we must satisfy two conditions:
  1. The probability of each outcome must be constant for all trials; and
  2. The trials must be independent.

Binomial random variables

Characterizing the Distribution

- The pdf for a discrete binomial distribution is written as:

$$p(x) = P(X = x) = \frac{n!}{(n-x)!\,x!} p^x (1-p)^{n-x}$$

- We indicate that a random variable is binomially distributed as: $X \sim B(n, p)$
- The mean and variance of a binomially distributed variable

	Mean	Variance
Bernoulli, $B(1,p)$	p	$p(1-p)$
Binomial, $B(n,p)$		

Binomial Random Variables

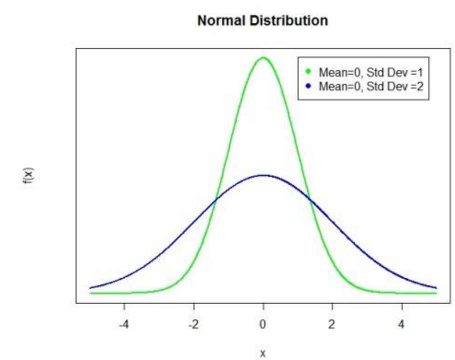
Example:

- You decide to assess an analyst's ability to forecast the sufficiency of earnings over a 20-quarter period. Over that time, the analyst correctly predicted earnings 13 times and incorrectly predicted earnings 7 times. You decide to model the "correctness" of his predictions using the binomial distribution.
  - What is your estimate of the probability of a successful prediction by this analyst?
- Assuming the estimated probability is the actual probability, answer the following questions:
  - What is the probability that the analyst will be correct for the next four quarters?
  - What is the expected number of quarters the analyst will be correct over the next three years?
  - What is the standard deviation of "correctness" for that period?

The Normal distribution

A continuous, symmetrical distribution that is completely described by its mean and variance.

- Mean, median, and mode are equal.
- The normal distribution has skewness of zero.
  - Option returns are skewed; hence, they are not normally distributed.
- Kurtosis of 3 or excess kurtosis of 0 (3 – 3 = 0).
  - k > 3 → fat tails → underestimated probability of extreme values (the blue distribution has excess kurtosis).
  - This area is one for which the normal distribution is a poor approximation for stock returns, which have "fat tails."



The Normal distribution

Characterizing the Normal Distribution

- The pdf for a normal distribution is written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \quad \text{for } -\infty < x < \infty$$

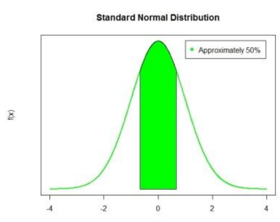
- We indicate that a random variable is normally distributed as: $X \sim N(\mu, \sigma^2)$

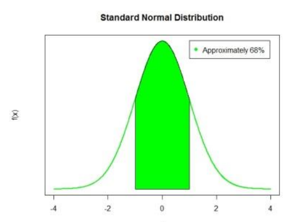
- The mean and variance of a normally distributed variable are

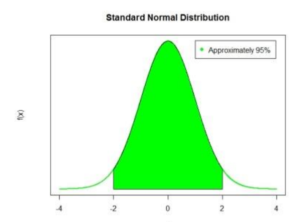
$$\mu \approx \bar{x} \qquad\qquad \sigma^2 = s^2$$

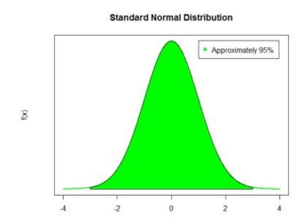
The normal distribution

- Approximately 50% of all observations fall in the interval $\mu \pm (2/3)\sigma$.
- Approximately 68% of all observations fall in the interval $\mu \pm \sigma$.
- Approximately 95% of all observations fall in the interval $\mu \pm 2\sigma$.
- Approximately 99% of all observations fall in the interval $\mu \pm 3\sigma$.









- We generally don’t observe population mean and variance (μ and σ), but we can estimate them with sample mean and variance.
  - When we do, the same intervals apply, with the sample mean and variance used in place of their population analogs.

The normal distribution

Example: Confidence Interval Calculations

- Your client's portfolio has a mean monthly return of 1.2% with a standard deviation of 3.7%. You assume for now that returns are normally distributed.
- Your client's return can be expected to fall in what range 50% of the time?
- 68% of the time?
- 95% of the time?

Standard Normal

A normal distribution with a mean of 0 and standard deviation of 1 is called "standard normal."

- The prevalence of the normal distribution has led to a process whereby probability tables that have been calculated for a standard normal distribution can be used to make probability statements for any normally distributed variable.
- This process is known as "standardizing" and is accomplished by:
  1. Taking the observation(s) of interest and subtracting the mean of that observation's observed distribution;
  2. Dividing the result by the observed distribution's standard deviation.

$$Z = \frac{X - \mu}{\sigma} \qquad z = \frac{x - \bar{x}}{s}$$

Homework: The Standard normal distribution

Example:

- Your client's portfolio has a mean monthly return of 1.2% with a standard deviation of 3.7%. You assume for now that returns are normally distributed.
  - What is the chance that returns will be between –2.5% and 4.9%?
  - What is the chance that returns will be negative?
- A stop-loss order automatically sells the stock if the price is below a set amount. You can set a stop-loss so that the portfolio is liquidated when it is triggered. How often will such a stop-loss be triggered if you set it so that it triggers when losses are below 1%?

Homework: Expected Number of Defaults Example

- Determine the number of expected defaults in a bond portfolio with 25 issues.
  - The estimated annual default rate is .107.

1. Over the next year, what is the expected number of defaults in the portfolio, assuming a binomial model for defaults?

2. Estimate the standard deviation of the number of defaults over the coming year.

3. Critique the use of the binomial probability model in this context.

Homework: Common Stocks and Normality

You have a portfolio with

- Weighted average forecast mean of 0.12.
- Forecast standard deviation of 0.22.

1. Calculate and interpret a one-standard-deviation confidence interval for portfolio return, with a normality assumption for returns.
2. Calculate and interpret a 90% confidence interval for portfolio return, with a normality assumption for returns.
3. Calculate and interpret a 95% confidence interval for portfolio return, with a normality assumption for returns.

Common Stocks and Normality

- What is the probability that portfolio return will exceed 20%?
- What is the probability that portfolio return will be between 12% and 20%? In other words, what is P(12% ≤ Portfolio return ≤ 20%)?
- You can buy a one-year T-bill that yields 5.5%. This yield is effectively a one-year risk-free interest rate. What is the probability that your portfolio's return will be equal to or less than the risk-free rate?

Correlation and Regression

Sample covariance

- **Recall that covariance is the weighted average of the cross-product of each variable's departure from its mean.**
- Sample covariance is calculated by using the same process as sample variance; however, rather than squaring the deviation of each observation from its mean, we take the product of two different variables' deviations from their respective means.

Sample covariance

Example:

- Lending rates and current borrower burden are generally believed to be related. The following data cover the debt-to-income ratio for 10 borrowers and the interest rate they are being charged on five-year loans.
- What is the sample covariance between loan rate (*Y*) and debt-to-income ratio (*X*)?

Client	*Y*	*X*	Y-Yhat	X-Xhat	Product
1	0.1595	0.1952	0.0070	0.0323	0.0002
2	0.1171	0.1239	–0.0354	–0.0390	0.0014
3	0.1171	0.1229	–0.0354	–0.0400	0.0014
4	0.1269	0.1625	–0.0256	–0.0004	0.0000
5	0.1343	0.1078	–0.0182	–0.0551	0.0010
6	0.1523	0.1470	–0.0002	–0.0159	0.0000
7	0.1523	0.1823	–0.0002	0.0194	0.0000
8	0.2295	0.2599	0.0770	0.0970	0.0075
9	0.1112	0.1384	–0.0413	–0.0245	0.0010
10	0.2247	0.1890	0.0722	0.0261	0.0019
Mean	**0.1525**	**0.1629**		**Sum =**	**0.0144**
StDev	**0.0427**	**0.0454**		**Cov =**	

Correlation Coefficient

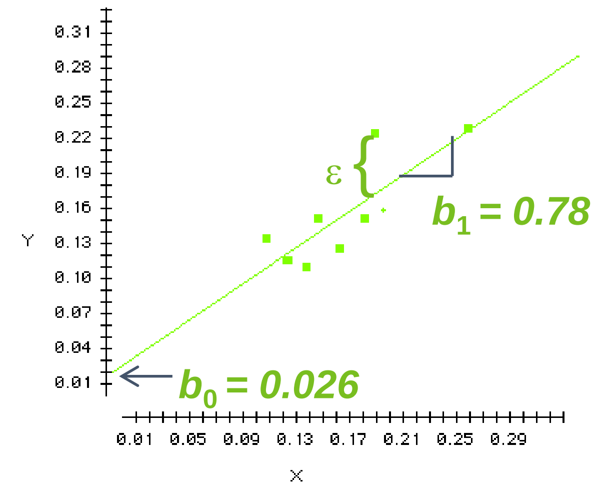
The correlation coefficient measures the extent and direction of a linear association between two variables.

- If the sample covariance is denoted as $s_{x,y}$, then the sample correlation coefficient is the sample covariance divided by each sample standard deviation or
- Continuing with our example, the sample correlation coefficient is then
- From this result, we can conclude that there is a strong linear relationship between the debt-to-income ratio of the borrowers and the loan rate they are charged. Furthermore, we can conclude that the relationship has a positive sign, indicating that an increase in the debt-to-income ratio is associated with a higher loan rate.

The Basics of Linear regression

Linear regression allows us to describe one variable as a linear function of another variable.

- The **independent variable (X_i)** is the variable you are using to explain changes in the **dependent variable (Y_i),** the variable you are attempting to explain.
- The linear regression estimation process chooses parameter estimates to minimize the sum of the squared departures of the predicted values from the observed values.
  - b_0 is known as the intercept and b_1 is known as the slope coefficient.
  - If the value of the independent variable increases by one unit, the value of the dependent variable changes by b_1 units.



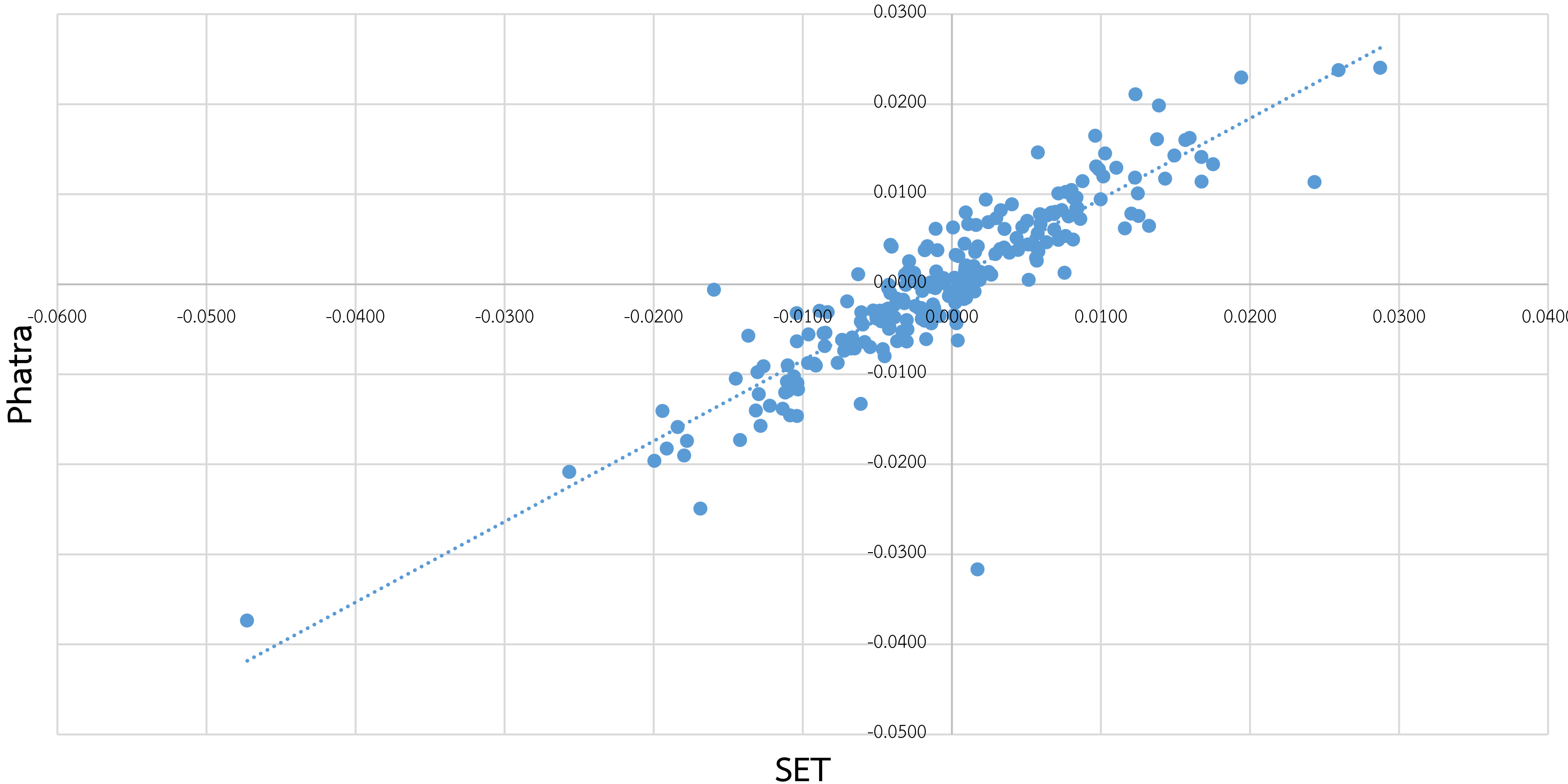
Assumptions underlying linear regression

$$Y_i = b_0 + b_1 X_i + \varepsilon_i$$

1. The relationship between the dependent variable, *Y*, and the independent variable, *X*, is linear in the parameters b_0 and b_1.
2. The independent variable, *X*, is not random.
3. The expected value of the error term is 0 → $E(\varepsilon) = 0$.
4. The variance of the error term is the same for all observations.
5. The error term, ε, is uncorrelated across observations.

   Consequently, $E(\varepsilon_i, \varepsilon_j) = 0$ for all *i* not equal to *j*.
6. The error term, ε, is normally distributed.

The Basics of Linear regression: Phatra vs. SET



The Basics of Linear regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8962
R Square	0.8033
Adjusted R Square	0.8024
Standard Error	0.0040
Observations	236

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.0150	0.0150	955.3962	0.0000
Residual	234	0.0037	0.0000		
Total	235	0.0187			

The Basics of Linear regression

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0005	0.0003	1.9478	0.0526
X Variable 1	0.8956	0.0290	30.9095	0.0000

	Lower 95%	*Upper 95%*
Intercept	0.0000	0.0010
X Variable 1	0.8385	0.9527

Question?