ECON 257D1 Final Review Session

26 November 2021

Tiffany Yong

1 Arithmetic

Theorem 1.1. $\log(x^a) = a\log(x)$

Theorem 1.2. $\log(x) + \log(y) = \log(xy)$

Theorem 1.3. $k^x \cdot k^y = k^{x+y}$

2 Probability

Definition 2.1. (Random experiment) An experiment where you don't know the outcome beforehand.

Definition 2.2. (Complement) $\overline{A}$ is everything that is not in A.

Definition 2.3. (Union) $A \cup B$ is everything in either A or B.

Definition 2.4. (Intersection) $A \cap B$ is everything in both A and B.

Theorem 2.1. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

Theorem 2.2. (Rules regarding Set Operations) I'll illustrate using the intersection, but the same applies for the union. Just swap all the ∩s and ∪s.

1. (Commutativity) $A \cap B = B \cap A$
2. (Associativity) $A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C)$
3. (Distributivity) $A \cap (B \cup C) = A \cap B \cup A \cap B$
4. (De Morgan's Laws) $\overline{(A \cap B \cap C)} = \overline{A} \cup \overline{B} \cup \overline{C}$

Definition 2.5. (Mutually Exclusive) A and B are mutually exclusive if $A \cap B = \emptyset$. This means that they never happen at the same time.

Definition 2.6. (Exhaustive) A sequence of events $A_1, A_2, \ldots, A_n$ are exhaustive if they cover all of the possible outcomes. The probability of an exhaustive sequence of events happening must be 1.

Theorem 2.3. $\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(\overline{A} \cap B)$.

Example 2.1. A traffic light is either Red, Yellow, or Green. The probability that a car drives through the intersection and the light is Red is 0.01. The probability that a car drives through the intersection and the light is Yellow is 0.09. The probability that a car drives through the intersection and the light is Green is 0.9. What is the probability that the car drives through the intersection?

The light can only be Red, Yellow or Green, so they form an exhaustive set of events. Let D be the event that the car drives through. Then,

$$\mathbb{P}(Red \cap D) + \mathbb{P}(Green \cap D) + \mathbb{P}(Yellow \cap D) = \mathbb{P}(D)$$

Definition 2.7. (Conditional Probability)

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Theorem 2.4. $\mathbb{P}(\overline{A}|B) = 1 - \mathbb{P}(A|B)$.

Theorem 2.5. $\mathbb{P}(A) = \mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}(A|\overline{B})\mathbb{P}(\overline{B})$

Definition 2.8. (Bayes' Theorem)

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

Definition 2.9. (Independence, 1) A and B are independent if $\mathbb{P}(A|B) = \mathbb{P}(A)$, or equivalently, $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Definition 2.10. (Independence, 2) A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

3 Combinatorics

Definition 3.1. (Factorial) $n! = n \times (n-1) \times \cdots \times 2 \times 1$

Theorem 3.1. The number of ways to order n distinct items is $n!$.

Definition 3.2. (Combinations) ${}^nC_k = \binom{n}{k} = \frac{n!}{(n-k)!k!}$

Theorem 3.2. The number of ways to select k items from a group of n distinct items is ${}^nC_k$.

Definition 3.3. (Permutations) ${}^nP_k = \frac{n!}{(n-k)!}$

Theorem 3.3. The number of ways to select k items from a group of n distinct items where the order you select them matters is ${}^nP_k$.

4 Statistics

Definition 4.1. (Random Variable) A random variable (RV) is a real-valued number that depends on a random event.

Definition 4.2. (Cumulative Distribution Function) The cumulative distribution function (CDF) of a random variable X is the function $F_X(\cdot)$ such that $F_X(x) = \mathbb{P}(X \leq x)$.

The difference between the empirical CDF and the regular CDF is that the empirical CDF is constructed by running the experiment many times and plotting the distribution of the data.

Definition 4.3. (Probability Density Function) The probability distribution function (PDF) of a random variable X is the function $f_X(x)$ where $F_X(x) = \int_{-\infty}^{x} f_X(x)dx$. If the CDF has a derivative at point x, then $f_X(x)$ is that derivative. It's basically saying how much weight is assigned to that particular point, without being $\mathbb{P}(X = x)$.

For a continuous function, $\mathbb{P}(X = x) = 0$, for any value of x. It's like asking how far is the distance between me and the blackboard right now. It's not sufficient to say 10cm, you have to say 10.10935...cm, which is basically impossible for it to happen.

Definition 4.4. (Expectation of a Discrete RV) The values of the random variable is weighted by the probability that it occurs, giving you the arithmetic mean. Also known as the first moment of a distribution.

$$\mathbb{E}(X) = \sum_{i=1}^{k} p_i x_i$$

Definition 4.5. (Expectation of a Continuous RV) Since $\mathbb{P}(X = x) = 0$ for all values of x, we need a different definition for a continuous RV.

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx$$

Theorem 4.1. (Linearity of Expectation) $\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y)$

Theorem 4.2. (Jensen's Inequality) This inequality is used in microeconomic theory as well, for expected utility. If $g(\cdot)$ is a convex function, then $\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))$.

Definition 4.6. (Variance) Also known as the second central moment of the distribution.

$$\mathbb{V}\text{ar}(X) = \sigma^2 = E(X - \mu_X)^2$$

.

Definition 4.7. (Standard deviation) $\sigma = \sqrt{\sigma^2} = \sqrt{\mathbb{V}\text{ar}(X)}$

Theorem 4.3. (Mean of Square - Square of Mean, MS-SM) $\mathbb{V}\text{ar}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$

Theorem 4.4. $\mathbb{V}\text{ar}(aX) = a^2\mathbb{V}\text{ar}(X)$

Theorem 4.5. If X and Y are independent, $\mathbb{V}\text{ar}(X - Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y)$

Theorem 4.6. (Chebyshev Inequality) The probability of a random variable being further than l standard deviations from its mean is less than $\frac{1}{l^2}$. This applies for any random variable, not just one with a normal distribution.

$$P(|X - \mu| \geq l\sigma) \leq \frac{1}{l^2}$$

Definition 4.8. The joint CDF of X_1 and X_2 is

$$F_{X_1,X_2}(x_1, x_2) = P(X_1 \leq x_1, X_2 \leq x_2)$$

This can be generalised to a greater number of RVs, $X_1, X_2, \ldots, X_n$.

Definition 4.9. (Conditional Distribution Function, Discrete) The conditional distribution function for discrete RVs is given by

$$P_{Y|X} = \frac{P_{X,Y}(x, y)}{P_X(x)}$$

This follows the definition of the conditional probability, and is fairly intuitive. We still need a different definition for the continuous case, as explained before.

Definition 4.10. (Conditional Distribution Function, Continuous)

$$f_{Y|X} = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Definition 4.11. (Independence of continuous RVs) Two continuous random variables X and Y with joint PDF $f_{X,Y}(x, y)$ are independent if and only if $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$.

The if and only if just means it goes both ways. If it's independent, you can conclude that the joint PDF is the product of the two PDFs. If the joint PDF is the product of the two PDFs, it is independent.

Definition 4.12. (Covariance) $\mathbb{C}\text{ov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$

Definition 4.13. (Correlation) When $\sigma_X, \sigma_Y > 0$,

$$\text{corr}(X, Y) = \frac{\mathbb{E}[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Theorem 4.7. $\mathbb{C}\text{ov}(X, X) = \mathbb{V}\text{ar}(X)$

Theorem 4.8. $\mathbb{V}\text{ar}(X + Y) = \mathbb{V}\text{ar}(X) + \mathbb{V}\text{ar}(Y) + 2\mathbb{C}\text{ov}(X, Y)$

5 Distributions

Definition 5.1. (Bernoulli) A random variable $X \sim \text{Bernoulli}(p)$ if X has two outcomes, x_1 and x_2, and x_1 occurs with probability p, while x_2 occurs with probability $1 - p$. A coin flip is an example of a Bernoulli($\frac{1}{2}$).

Definition 5.2. (Binomial) A random variable $X \sim \text{Binomial}(n, p)$ if Y has two outcomes, x_1 and x_2, and X is n trials of Y, such that $X = Y_1 + Y_2 + \cdots + Y_n$.

Definition 5.3. (Poisson) Used to model data that are counts of the number of times that something occurs.

Definition 5.4. (Exponential) Used as a model for continuous, positive functions like waiting times in a room.

Definition 5.5. (Normal) The normal distribution appears a lot in real life, and its most common application is that the distribution of many sample means of any distribution has a normal distribution. Takes two parameters, mean and variance.

Definition 5.6. (Standard Normal) The standard normal RV is an RV $X \sim N(0, 1)$.

Definition 5.7. (Student-t) The t-distribution is used when we don't know the population variance of the distribution, and we can only estimate the sample variance from running an experiment. There, we must use the t-distribution to get a more conservative estimate. The t-distribution has 3 parameters, mean, variance and degrees of freedom. As the degrees of freedom approaches infinity, the t-distribution converges to the normal distribution.

Definition 5.8. (Chi-squared) Sums of independent squared standard Normal RVs have the χ^2 distribution, with the degrees of freedom being the number of RVs that are added.

Theorem 5.1. (Linear combinations of Normal random variables are Normal) If $X_1, X_2, \ldots, X_n$ are independent normal random variables, each with mean 0 and variance σ_i^2, then the linear combination of these random variables $Y = a_1X_1 + a_2X_2 + \cdots + a_nX_n \sim N(0, \sum_{i=1}^{n} a_i^2\sigma_i^2)$.

6 Sample Statistics

Definition 6.1. (Sample Mean) Also known as the arithmetic mean.

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}(X_1 + X_2 + \cdots + X_N)$$

Definition 6.2. (Weak Law of Large Numbers) Let $x_i, i = 1, \ldots, n$, be independent random draws from a distribution with cumulative distribution function $F_X(x)$, such that the distribution has a mean μ and variance $\sigma^2 > 0$. Then, for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X} - \mu| < \epsilon) = 1$$

This means that as the sample size gets larger and larger, the sample mean gets closer and closer to μ with a large probability.

We also have that when the sample size $n > \frac{\sigma^2}{\epsilon^2\delta}$,

$$\mathbb{P}(|\overline{X} - \mu| < \epsilon) \geq 1 - \delta$$

Theorem 6.1. Let $Z \sim N(0, 1))$ and let $W \sim \chi_r^2$. Then if Z and W are independent, the ratio $\frac{Z}{\sqrt{W/r}}$ has a t distribution with r degrees of freedom.

Theorem 6.2. (Confidence Interval for μ) We know that the distribution of the sample mean follows a normal distribution, where $\overline{X} \sim N(\mu, \frac{s^2}{N}) \Rightarrow \overline{X} - \mu \sim N(0, \frac{s^2}{N})$. We also know that $s_2 \sim \chi_N^2$. Then, by Theorem 6.1 above, we have that

$$\frac{\overline{X} - \mu}{s/\sqrt{N}} \sim t_{N-1}$$

This gives us a confidence interval for the value of μ.

$$\begin{aligned}
\mathbb{P}(-q_{\alpha/2} < \frac{\overline{X} - \mu}{s/\sqrt{N}} < q_{\alpha/2}) &= 1 - \alpha \\
\mathbb{P}(-\frac{s}{\sqrt{N}} \cdot q_{\alpha/2} < \overline{X} - \mu < \frac{s}{\sqrt{N}} \cdot q_{\alpha/2}) &= 1 - \alpha \\
\mathbb{P}(-\frac{s}{\sqrt{N}} \cdot q_{\alpha/2} < \mu - \overline{X} < \frac{s}{\sqrt{N}} \cdot q_{\alpha/2}) &= 1 - \alpha \\
\mathbb{P}(\overline{X} - \frac{s}{\sqrt{N}} \cdot q_{\alpha/2} < \mu < \overline{X} + \frac{s}{\sqrt{N}} \cdot q_{\alpha/2}) &= 1 - \alpha
\end{aligned}$$

$\Rightarrow \mu \in \overline{X} \pm \frac{s}{\sqrt{N}} \cdot q_{\alpha/2}$ with probability $1 - \alpha$.

Definition 6.3. (Central Limit Theorem) Let $x_i, i = 1, \ldots, n$, be independent random draws from a distribution with cumulative distribution function $F_X(x)$, such that the distribution has a mean μ and variance $\sigma^2 > 0$. Then,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0, 1)$$

This means that as the sample size gets larger and larger, the distribution of the standardised sample mean gets closer and closer to the standard normal distribution.

However, this assumes that you know what σ is. When you don't know what σ is, Theorem 6.2 above gives you the correct method to solve for a confidence interval. If you know what σ is, follow the same steps to get a better confidence interval for μ.

Definition 6.4. (Trimmed Mean) The trimmed mean $\overline{X}_{[k,l]}$ is just the sample mean that leaves out the k smallest X_i's and the l largest X_i's. Used if you know that there's some strong outliers that happen.

Definition 6.5. (Sample Variance) A measure of the dispersion around the sample mean.

$$s^2 = (N-1)^{-1} \sum_{i=1}^{N} (x_i - \overline{X})^2$$

Theorem 6.3. (Relation between Variance and Sample Variance) If the RV X has a normal distribution with an unknown mean and variance (μ, σ^2), then with the definition of s^2 above with N sample size, we have that $\mathbb{E}(s^2) = \sigma^2$, and also

$$(N-1)\frac{s^2}{\sigma^2} \sim \chi^2_{N-1}$$

Theorem 6.4. (Confidence Interval for σ^2) We know that $(N-1)\frac{s^2}{\sigma^2} \sim \chi^2_{N-1}$, so we do the exact same procedure as above for the confidence interval of μ.

Definition 6.6. (Standard error) $s = \sqrt{s^2} = \sqrt{(N-1)^{-1} \sum_{i=1}^{N} (x_i - \overline{X})^2}$

Definition 6.7. (Coefficient of Skewness) A measure of the asymmetry of the data.

$$\frac{(N-1)^{-1} \sum_{i=1}^{N} (x_i - \overline{X})^3}{s^3}$$

Definition 6.8. (Coefficient of Kurtosis) A measure of the spread of the data.

$$\frac{(N-1)^{-1} \sum_{i=1}^{N} (x_i - \overline{X})^4}{s^4}$$

Definition 6.9. (Sample Covariance)

$$\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \overline{X})(y_i - \overline{Y})$$

Definition 6.10. (Sample Correlation) When correlation is positive, the two variables move together. When correlation is negative, the two variables move opposite to each other. When correlation is zero, there is no association in their movement.

$$\frac{(N-1)^{-1} \sum_{i=1}^{N} (x_i - \overline{X})(y_i - \overline{Y})}{s_X s_Y}$$