Review for Stat 515

- **Probability : Chapter 2**
  - Definition of probability and Basic set theory (using Venn diagram)
  - Basic definitions of probabilistic model (Def 2.1 - Def 2.6)
    * Experiment, sample space \( S \), sample point (compound/simple) event : each simple event corresponds to one and only one sample point
    * Event on discrete sample space
    * Definition of probability (three axioms) and properties of probability
  - First method to find the probability of an event on discrete sample space : sample point method
    (Example) If the sample space, \( S \) contains \( N \) equiprobable sample points and an event \( A \) contains exactly \( n_A \) sample points, then \( P(A) = n_A/N \). When \( n_A \) and \( N \) are large, we need to know how to count the number of sample points in \( n_A \) and \( N \), respectively.
  - Counting sample points
    * MN rule and its generalization (Thm 2.1)
    * Permutation (ordered arrangement under sampling without replacement) and Combination (unordered arrangement under sampling without replacement) (Def 2.7, Thm 2.2, Def 2.8, Thm 2.4)
    * Extension of combination (Thm 2.3)
  - Conditional probability
    * Conditional probability (Def 2.9)
    * Independence between two events (Def 2.10), mutual independence among three events and its extension for \( n \) events
  - Laws of probability
    * Multiplicative law of probability (Thm 2.5)
    * Additive law of probability (Thm 2.6)
    * Probability of complement (Thm 2.7)
  - Second method to find the probability of an event on discrete sample space : event composition method
  - Bayes’ rules (Def 2.11, Thm 2.8, Thm 2.9)
    * Law of total probability : partition of sample space
    * Bayes’ rule

1
• Random variable (r.v.) and its probability distribution - Chapter 3 and 4

- Definition of random variable (Def 2.12)
  
  A random variable \( Y \) is a real-valued function from the sample space, \( S \) to a set of real number, \( R \), \( Y: S \rightarrow R \) where \( y = Y(s) \) and \( s \in S \)

(Example) In an opinion poll, one might determine to ask four people whether they agree or disagree with a certain issue. Suppose one might be interested in the number of people who agree out of four. The sample space \( S \) contains 16 sample points: if 1=agree and 0= disagree, then \( E_1 = \{1111\}, E_2 = \{0111\}, E_3 = \{1011\}, E_4 = \{1101\}, E_5 = \{1110\}, E_6 = \{1100\}, E_7 = \{1010\}, E_8 = \{1001\}, E_9 = \{0110\}, E_{10} = \{0101\}, E_{11} = \{0011\}, E_{12} = \{0100\}, E_{13} = \{0100\}, E_{14} = \{0010\}, E_{15} = \{0001\}, E_{16} = \{0000\}. Let \( Y \) be the number of people who agree out of four. Then the space of \( Y \), \( R \) is \( R = \{0,1,2,3,4\} \).

Here, \( Y \) is a real-valued function such that \( Y(\{E_i\}) = 4 \) if \( i = 1 \), \( Y(\{E_i\}) = 3 \) if \( i = 2,\ldots,5 \), \( Y(\{E_i\}) = 2 \) if \( i = 6,\ldots,11 \), \( Y(\{E_i\}) = 1 \) if \( i = 12,\ldots,15 \), \( Y(\{E_i\}) = 0 \) if \( i = 16 \). We call this \( Y \) a random variable.

- Cumulative Distribution Function (CDF), \( F(y) \equiv P(Y \leq y) \) for any (discrete or continuous) r.v. and its properties (Def 4.1 and Thm 4.1)

- Definition of discrete r.v. and its probability mass function (PMF) (Def 3.1, Def 3.2, Def 3.3)
  
  * Discrete r.v. and its probability mass function (PMF), \( P(Y = y) = p(y) \)
  
  * Properties for any \( p(y) \) (Thm 3.1)
  
  * Definition of Cumulative Distribution Function (CDF) for a discrete r.v.,
    \( F(y) \equiv P(Y \leq y) = \sum_{\ell \leq y} p(\ell) \)
  
  * Relationship between CDF \( F(y) \) and \( p(y) \) for a discrete r.v. : \( P(Y = a) = P(Y \leq a) - P(Y < a) = P(Y \leq a) - P(Y \leq a - 1) = F(a) - F(a-1) \)

  (Note) When \( Y \) is a discrete r.v., are \( P(a \leq Y \leq b) \), \( P(a < Y \leq b) \), \( P(a \leq Y < b) \) and \( P(a < Y < b) \) same? Not necessarily

- Definition of continuous r.v. (Def 4.2 and Def 4.3) and its probability density function (PDF)
  
  * Continuous r.v. and its \( f(y) \)
  
  * Properties of \( f(y) \) for any continuous r.v. (Thm 4.2)
  
  * Definition of Cumulative Distribution Function (CDF) for a continuous r.v.,
    \( F(y) \equiv P(Y \leq y) = \int_{-\infty}^{y} f(t)dt \)
  
  * Relationship between CDF \( F(y) \) and \( f(y) \) for a continuous r.v. : \( F(y) = \int_{-\infty}^{y} f(t)dt \) and \( \frac{dF(y)}{dy} = f(y) \)

  * \( P(Y = y) = 0 \) for any real number \( y \) if \( Y \) is a continuous r.v.
* Calculation of $P(a \leq Y \leq b)$ for a continuous r.v., $Y$: $P(a \leq Y \leq b) = \int_a^b f(y) dy$ (Thm 4.3).

(Note) When $Y$ is a continuous r.v., $P(Y = y_0) = 0$ and $P(a \leq Y \leq b)$, $P(a < Y \leq b)$, $P(a \leq Y < b)$ and $P(a < Y < b)$ are same.

- Comparisons between discrete r.v. and continuous r.v. in terms of their CDFs and probability distribution functions, $p(y)$ or $f(y)$

- Expected value of a discrete/continuous r.v. or a function of a discrete/continuous r.v.

  * Mean of $g(Y)$, $E(g(Y))$
    - $E(g(Y)) = \sum_y g(y)p(y)$ for discrete $Y$
    - $E(g(Y)) = \int_{-\infty}^{\infty} g(y)p(y)$ for continuous $Y$
    - For $g(Y) = Y$, Mean of $Y$ is $E(Y)$
    - For $g(Y) = (Y - E(Y))^2$, Variance of $Y$ is $V(Y) = E((Y - E(Y))^2) = E(Y^2) - E(Y)^2$ and its standard deviation, $\sigma = \sqrt{Var(Y)}$

  * Four useful expectation theorems
    - $E(c) = c$ for a constant $c$
    - $E(cg(Y)) = cE(g(Y))$ for any (discrete/continuous) r.v. $Y$
    - $E(g_1(Y) + \ldots + g_k(Y)) = E(g_1(Y)) + \ldots + E(g_k(Y))$ for any functions of $Y$, $g_1(Y), \ldots, g_k(Y)$
    - For constants, $a$ and $b$, $E(aY + b) = aE(Y) + b$ and $V(aY + b) = a^2V(Y)$

- Well-known discrete r.v.s and their PMFsp($y$)

  * Binomial r.v., $Y \sim b(n, p)$
    - if $n = 1$, $Y \sim b(1, p)$ is Bernoulli r.v.
  * Geometric r.v., $Y \sim geo(p)$
  * Poisson r.v., $Y \sim poisson(\lambda)$

- Well-known continuous r.v.s and their PDFs, $f(y)$

  * Continuous uniform r.v., $Y \sim uniform(\theta_1, \theta_2)$
  * Normal r.v., $Y \sim N(\mu, \sigma^2)$
  * Gamma r.v., $Y \sim gamma(\alpha, \beta)$
    - if $\alpha = \nu/2$ and $\beta = 2$, $Y$ is a Chi-squared r.v., $Y \sim \chi^2(\nu)$
    - if $\alpha = 1$, $Y$ is an Exponential r.v., $Y \sim exponential(\beta)$

[Note 1] For well-known discrete/continuous r.v.s, you should know the following information

  * Definition and (experiment) conditions for each discrete/continuous r.v., and its $p(y)/f(y)$ (as a formula)
Parameters in \( p(y)/f(y) \)
- Possible values of each discrete/continuous r.v., and of the parameters
- Mean and variance
- Moment generating function (MGF)

[Note 2] For normal and Chi-squared r.v., you should know the following information
- Normal r.v. : properties of normal r.v., standardized normal r.v., \( Z \sim N(0,1) \), and reading Table 4 (standard normal probability distribution in right tail)
- Tschebysheff’s theorem for a (discrete/continuous) r.v. (Section 3.11 and Section 4.10)

- **Bivariate (Multivariate) probability distributions - Chapter 5**
  - Bivariate joint CDF for any r.v. (Def 5.2, Thm 5.2)
  - Bivariate joint probability distribution \( p(y_1, y_2) \) of discrete r.v.s (Def 5.1, Thm 5.1), and \( f(y_1, y_2) \) of continuous r.v.s (Def 5.3, Thm 5.3)).
  - Marginal and conditional probability distributions of discrete and continuous r.v.s. (Def 5.4, Def 5.5, Def 5.6, Def 5.7)
  - Independent r.v.s (Def 5.8, Thm 5.4, Thm 5.5, Thm 5.9)
  - Expected value of a function of r.v.s (Thm 5.9)
  - Special theorems for r.v.s (Thm 5.6, Thm 5.7, Thm 5.8 and Thm 5.9)
  - Covariance of two random variables (Def 5.10, Thm 5.10)
    - Covariance between two r.v.s and its meaning
    - Correlation between two r.v.s and its meaning
    - Easy calculation for covariance
    - General relationship between independence and zero covariance, and one exception for bivariate normal distribution (Thm 5.11 and Section 5.10)
  - **Moments of linear combination of r.v.s (Thm 5.12)**
  - Multinomial probability distribution (Section 5.9)
  - Bivariate normal distribution (Section 5.10)
    - Relationship between independence and zero covariance
  - Conditional expectation (for discrete and continuous r.v.s, X and Y) (Section 5.11)
    - Unconditional and conditional mean
- $E(X) = E[E(X \mid Y)]$. Here, $E(X \mid Y)$ is a function of $Y$.
- Unconditional and conditional variance
  \[ V(X) = V[E(X \mid Y)] + E[V(X \mid Y)] \geq V[E(X \mid Y)], \] as $E[V(X \mid Y)]$ is always positive.

**Moment Generating Function (MGF)** for a r.v., $Y$ (Section 3.9, and Section 4.9)

- k-th moment of $Y$, $E(Y^k)$ and k-th central moment of $Y$, $E((Y - \mu)^k)$
- Usage of moments (provide information on mean, variance, skewness and kurtosis)
- Definition of MGF for $Y$, $m_Y(t) = E(e^{tY})$:
  : generation of any moments for $Y$ from $m_Y(t)$, i.e., $m_Y(k) \equiv \frac{d^k m_Y(t)}{dt^k} \bigg|_{t=0} = E(Y^k)$
- Uniqueness MGF for $Y$, $m_Y(t)$ as long as it exists.
  : MGF, $m_Y(t)$ for well-known (discrete/continuous) r.v.s
  : If the MFG, $m_Y(t)$ for a r.v. $Y$ is given, you should be able to obtain its $p(y)$ (or $f(y)$)

**Functions of Random Variables**: Chapter 6

- Finding the probability distribution, $p(g(Y_1, \ldots, Y_n))$ and PDF, $f(g(Y_1, \ldots, Y_n))$ of a function of r.v.s, $g(Y_1, \ldots, Y_n)$
  - Method of moment generating function (MGF)
    - uniqueness theorem
    - finding the distribution of sums of independent r.v.s (Thm 6.2)
  - Useful facts
    - $Z = \frac{Y - \mu}{\sigma} \sim \chi^2(1)$ if $Y \sim N(\mu, \sigma^2)$
    - $\sum_{i=1}^n a_i Y_i \sim \chi^2 \left( \sum_{i=1}^n a_i^2 \right)$ if $Y_1, \ldots, Y_n$ are independent and $Y_i \sim N(\mu_i, \sigma_i^2)$ where $i = 1, \ldots, n$
    - If $a_i = \frac{1}{n}$ for all $i$, $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i \sim N(\mu, \sigma^2/n)$ and $\frac{Y - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$
    - $Z^2 \sim \chi^2(1)$ if $Z \sim N(0, 1)$
    - $\sum_{i=1}^n Y_i \sim gamma(n, \beta)$ if $Y_1, \ldots, Y_n$ are independent and $Y_i \sim exponential(\beta)$ where $i = 1, \ldots, n$
    - $\sum_{i=1}^n Y_i \sim b(n, p)$ if $Y_1, \ldots, Y_n$ are independent and $Y_i \sim b(1, p)$
    - $\sum_{i=1}^n Y_i \sim b(\sum_{i=1}^n n_i, p)$ if $Y_1, \ldots, Y_n$ are independent and $Y_i \sim b(n_i, p)$ where $i = 1, \ldots, n$
\[ \sum_{i=1}^{n} Y_i \sim \text{poisson} \left( \sum_{i=1}^{n} \lambda_i \right) \] if \( Y_1, \ldots, Y_n \) are independent and \( Y_i \sim \text{poisson} (\lambda_i) \) where \( i = 1, \ldots, n \)

- **Sampling distribution and the Central Limit Theorem: Chapter 7**

  - Introduction

  * Assumptions: \( Y_1, \ldots, Y_n \) are independent samples from a discrete/continuous distribution. Note that “\( Y_1, \ldots, Y_n \) are IID (random) samples” means that \( Y_1, \ldots, Y_n \) are independent, identically distributed random variables with a common discrete/continuous distribution.

  * Definition of a statistic

    - (Def 7.1) A statistic is a function of the observable random variables in a sample and known constants for a parameter of our interest.

  * Definition of sampling distribution of a statistic

    - The sampling distribution of a statistic is the probability distribution, under repeated sampling of the population, of a given statistic. In other words, the sampling distribution of a statistic is the population of all possible values for that statistic. For example, consider a large normal population. Assume we repeatedly take samples of a given size from the population and calculate the sample mean (\( \bar{y} \), the sample mean of the data values) for each sample. Different samples will lead to different sample means. The distribution of these means is the “sampling distribution of the sample mean” (for the given sample size).

    - We need to know the sampling distribution of a statistic because the goodness of a statistic for the parameter of interest depends on the probability distribution of a statistic

    - The following example illustrates how the sampling distribution for \( \bar{Y} \) can be obtained.

      (Example) The sample \( \bar{Y} \) is to be calculated from a random sample of size 2 taken from a population consisting of ten values \( (2,3,4,5,6,7,8,9,10,11) \). Find the sampling distribution of \( \bar{Y} \), based on a random sample of size 2.

      There are 45 possible samples of two items selected from the ten items (see Table 1). Assuming each sample of size 2 is equally likely, Table 2 shows the sampling distribution for \( \bar{Y} \) based on \( n = 2 \) observations selected from the population \( (2,3,4,5,6,7,8,9,10,11) \).

      Note that the shape of this sampling distribution is symmetric (bell-shaped). This sampling distribution provides a way to make statistical inferences about \( \bar{Y} \) in the example: calculate the following probability: \( P(3.5 \leq \bar{Y} \leq 9.5) = 41/45 \).
- The mathematical derivation of sampling distribution is one of the basic problems of the mathematical statistics.

**Table 1**

<table>
<thead>
<tr>
<th>Sample</th>
<th>Value of $\bar{Y}$</th>
<th>Sample</th>
<th>Value of $\bar{Y}$</th>
<th>Sample</th>
<th>Value of $\bar{Y}$</th>
<th>Sample</th>
<th>Value of $\bar{Y}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2,3</td>
<td>2.5</td>
<td>2,4</td>
<td>3</td>
<td>2,5</td>
<td>3.5</td>
<td>2,6</td>
<td>4</td>
</tr>
<tr>
<td>2,7</td>
<td>4.5</td>
<td>2,8</td>
<td>5</td>
<td>2,9</td>
<td>5.5</td>
<td>2,10</td>
<td>6</td>
</tr>
<tr>
<td>2,11</td>
<td>6.5</td>
<td>3,4</td>
<td>3.5</td>
<td>3,5</td>
<td>4</td>
<td>3,6</td>
<td>4.5</td>
</tr>
<tr>
<td>3,7</td>
<td>5</td>
<td>3,8</td>
<td>5.5</td>
<td>3,9</td>
<td>6</td>
<td>3,10</td>
<td>6.5</td>
</tr>
<tr>
<td>3,11</td>
<td>7</td>
<td>4,5</td>
<td>4.5</td>
<td>4,6</td>
<td>5</td>
<td>4,7</td>
<td>5.5</td>
</tr>
<tr>
<td>4,8</td>
<td>6</td>
<td>4,9</td>
<td>6.5</td>
<td>4,10</td>
<td>7</td>
<td>4,11</td>
<td>7.5</td>
</tr>
<tr>
<td>5,6</td>
<td>5.5</td>
<td>5,7</td>
<td>6</td>
<td>5,8</td>
<td>6.5</td>
<td>5,9</td>
<td>7</td>
</tr>
<tr>
<td>5,10</td>
<td>7.5</td>
<td>5,11</td>
<td>8</td>
<td>6,7</td>
<td>6.5</td>
<td>6,8</td>
<td>7</td>
</tr>
<tr>
<td>6,9</td>
<td>7.5</td>
<td>6,10</td>
<td>8</td>
<td>6,11</td>
<td>8.5</td>
<td>7,8</td>
<td>7.5</td>
</tr>
<tr>
<td>7,9</td>
<td>8</td>
<td>7,10</td>
<td>8.5</td>
<td>7,11</td>
<td>9</td>
<td>8,9</td>
<td>8.5</td>
</tr>
<tr>
<td>8,10</td>
<td>9</td>
<td>8,11</td>
<td>9.5</td>
<td>9,10</td>
<td>9.5</td>
<td>9,11</td>
<td>10</td>
</tr>
<tr>
<td>10,11</td>
<td>10.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2**

<table>
<thead>
<tr>
<th>$\bar{y}$</th>
<th>2.5</th>
<th>3</th>
<th>3.5</th>
<th>4</th>
<th>4.5</th>
<th>5</th>
<th>5.5</th>
<th>6</th>
<th>6.5</th>
<th>7</th>
<th>7.5</th>
<th>8</th>
<th>8.5</th>
<th>9</th>
<th>9.5</th>
<th>10</th>
<th>10.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(\bar{y})$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Sampling distributions related to one normal distribution

: For $n$ random samples, $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$,

* When $\mu$ is a parameter of interest, $\bar{Y}$ is a statistic and its sampling distribution is $\bar{Y} \sim N(\mu, \sigma^2/n)$

* Inference about $\mu$ (population mean) when $\sigma^2$ (population variance) is known

  - $P(a \leq \bar{Y} - \mu \leq b) = P \left( \frac{a - \mu}{\sigma / \sqrt{n}} \leq \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \leq \frac{b - \mu}{\sigma / \sqrt{n}} \right)$ where $\frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$

  - read Table 4 for calculating $P(a \leq \bar{Y} - \mu \leq b)$

- Central Limit Theorem (CLT)

  * $\bar{Y} \sim N(\mu, \sigma^2/n)$ for $n$ random samples, $Y_1, \ldots, Y_n \sim N(\mu, \sigma^2)$

  * $E(\bar{Y}) = \mu$ and $V(\bar{Y}) = \sigma^2/n$ for $n$ random samples, $Y_1, \ldots, Y_n$ from any distribution with $E(Y) = \mu$ and $V(Y) = \sigma^2$.

  But, the distribution for $\bar{Y}$ is unknown for finite sample size $n$.

  * The Central Limit Theorem (CLT) (Thm 7.4) provides us an approximation for the sampling distribution of $\bar{Y}$ regardless of the distribution of the population from which the sample is taken

7
- Let \( Y_1, Y_2, \ldots, Y_n \) be independent random variables with \( E(Y_i) = \mu \) and \( V(Y_i) = \sigma^2 < \infty \). Define

\[
U_n = \sqrt{n} \left( \frac{\bar{Y} - \mu}{\sigma} \right) \quad \text{where} \quad \bar{Y} = \frac{1}{n} \sum_{i=1}^{n} Y_i
\]

Then the distribution function of \( U_n \) converges to a standard normal distribution function as \( n \to \infty \) : \( U_n \sim N(0, 1) \).

- \( P(a \leq \bar{Y} - \mu \leq b) = P \left( \frac{a}{\sqrt{\sigma^2/n}} \leq U_n \leq \frac{b}{\sqrt{\sigma^2/n}} \right) \approx P \left( \frac{a}{\sqrt{\sigma^2/n}} \leq Z \leq \frac{b}{\sqrt{\sigma^2/n}} \right) \) for large \( n \) where \( Z \sim N(0, 1) \) and \( \sigma^2 \) is known. Note that one can

\[
P \left( \frac{a}{\sqrt{\sigma^2/n}} \leq Z \leq \frac{b}{\sqrt{\sigma^2/n}} \right) \] using Table 4.

- The central limit theorem can be applied to a random sample \( Y_1, Y_2, \ldots, Y_n \) from any distributions, so long as \( E(Y_i) = \mu \) and \( V(Y_i) = \sigma^2 \) are both finite and the sample size is large.