Summary of Topics for the Midterm Exam
STA 371G, Fall 2019

Listed below are the major topics covered in class that are likely to be in the Midterm Exam. Good Luck!

- Mean (expectation), variance and standard deviation of a discrete random variable.

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i P(X = x_i), \quad \mathrm{Var}[X] = \sum_{i=1}^{n} (x_i - \mathbb{E}[X])^2 P(X = x_i), \quad \mathrm{sd}[X] = \sqrt{\mathrm{Var}[X]}$$

- Add a constant to a random variable, multiply a random variable by a constant.
  If $Y = a + bX$, then

$$\mathbb{E}[Y] = a + b\mathbb{E}[X], \quad \mathrm{Var}[Y] = b^2 \mathrm{Var}[X], \quad \mathrm{sd}[Y] = |\, b \,| \times \mathrm{sd}[X].$$

- Conditional, joint and marginal probabilities.

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

$$P(Y = y, X = x) = P(Y = y \,|\, X = x) P(X = x)$$

$$P(Y = y) = \sum_{x} P(Y = y, X = x)$$

- Independent random variables, sum of independent random variables.
  - Two random variables X and Y are independent if $P(Y = y \,|\, X = x) = P(Y = y)$ for all possible x and y.
  - If X and Y are independent , then $P(Y = y, X = x) = P(Y = y)P(X = x)$.
  - If $Y = a_0 + a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$, then

$$\mathbb{E}[Y] = a_0 + a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2] + \cdots + a_n \mathbb{E}[X_n].$$

    If X_i and X_j are independent for $i \neq j$, then we further have

$$\mathrm{Var}[Y] = a_1^2 \mathrm{Var}[X_1] + a_2^2 \mathrm{Var}[X_2] + \cdots + a_n^2 \mathrm{Var}[X_n].$$

  - If $Y = a_0 + a_1 X_1 + a_2 X_2$, then

$$\mathbb{E}[Y] = a_0 + a_1 \mathbb{E}[X_1] + a_2 \mathbb{E}[X_2].$$

    If X_1 and X_2 are independent, then we have

$$\mathrm{Var}[Y] = a_1^2 \mathrm{Var}[X_1] + a_2^2 \mathrm{Var}[X_2]$$

- Decision Making Under Uncertainty
  - Frequency interpretation and subjective interpretation of probability.
  - Probabilities and betting odds
  - Payoff table
  - Payoffs and losses, loss table
  - Nonprobabilistic decision criteria
    * maximin
    * maximax
    * minimax loss
  - Probabilistic decision criteria: expected payoff (ER) or expected loss (EL)
  - Utility
    * Risk avoider, risk neutral, risk taker
    * A typical utility function: $U(x) = 1 - e^{-x/R}$
    * Expected utility
  - Bayes' theorem

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + \cdots + P(B|A_p)p(A_p)}$$

  - Decision tree
    * Represent a payoff table with a decision tree
    * Time proceeds from left to right
    * Folding back procedure
    * Risk profile
    * Sensitivity analysis
    * Decision making and Bayes' theorem
  - The value of information
    * Value of perfect information
    * Expected value of perfect information (EVPI)
    * Value of sample information
    * Expected value of sample information (EVSI)
    * Bayes' theorem and the value of information
- Normal distribution $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ is the mean, σ^2 is the variance, and σ is the standard deviation.
  - Probability density function: area under the curve represents probability.
  - Standard normal distribution $Z \sim \mathcal{N}(0, 1)$.
  - Standardizing a normal random variable $Z = \frac{X-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.
  - $P(X < x) = P(\frac{X-\mu}{\sigma} < \frac{x-\mu}{\sigma}) = P(Z < \frac{x-\mu}{\sigma})$.

- $P(-2 < Z < 2) \approx 0.95$; $P(\mu - 2\sigma < X < \mu + 2\sigma) \approx 0.95$.

- Estimate μ and σ^2 when $X \sim \mathcal{N}(\mu, \sigma^2)$.
  - Use the sample mean $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$ to estimate μ.
  - Use the sample variance $s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ to estimate σ^2.

- Sampling distribution of a sample mean $\bar{X}$:
  - $\bar{X} \sim \mathcal{N}(\mu, \sigma^2_{\bar{X}} = \frac{\sigma^2}{n})$.
  - The sampling distribution of $\bar{X}$ is useful in determining the quality of $\bar{X}$ as an estimator for the population mean μ.
  - As the population variance σ^2 is usually unknown, we use the sample variance $s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$ to estimate σ^2 and hence s^2/n to estimate $\sigma^2_{\bar{X}}$.
  - 95% confidence interval of μ (approximately): $\bar{X} \pm 2\sqrt{\frac{s^2}{n}}$.

- Binomial distribution and its normal approximation
  - $X \sim \text{Binomial}(n, p)$ can be approximated with $X \sim \mathcal{N}(np, np(1-p))$ if n is large enough and p is not too close to 0 or 1.
  - Estimate the population proportion p when $X \sim \text{Binomial}(n, p)$, where n is the sample size.
    * Use the sample proportion $\hat{p} = \frac{X}{n}$ to estimate p.
    * Approximately, we have $\hat{p} \sim \mathcal{N}(p, \frac{\hat{p}(1-\hat{p})}{n})$.
    * 95% confidence interval of p: $\hat{p} \pm 2\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$.

- Introduction to Monte Carlo Simulation
  - Uniform random numbers
  - Flip a coin, toss a die, flip two coins, toss tow dice
  - Normal random numbers, Student's t random numbers
  - Understand how to simulate from a discrete distribution
  - Understand how to use simulation to estimate $P(X < x)$, $\mathbb{E}[X]$ and $\text{Var}[X]$, where X is a random variable following some distribution.
  - Understand how to use simulation to demonstrate Law of Large Numbers
  - Understand how to use simulation to demonstrate the sampling distribution of sample mean
  - Understand how to use simulation to demonstrate the Central Limit Theorem
  - Simulation and decision making
  - Understand how to construct a random experiment and find relevant answers by simulating the same experiment repeatedly under identical conditions

- Correlated random variables, sum of correlated random variables
  - If $Y = a_0 + a_1 X_1 + a_2 X_2$, then

    $$\mathbb{E}[Y] = a_0 + a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2].$$

    If X_1 and X_2 are not independent, then we have

    $$\text{Var}[Y] = a_1^2\text{Var}[X_1] + a_2^2\text{Var}[X_2] + 2a_1a_2\text{Cov}(X_1, X_2)$$

    and the strength of linear relationship between X_1 and X_2 can be measured by the correlation between them, defined as

    $$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}[X_1]\text{Var}[X_2]}}$$