CAS CS 365
Lab 3

1. Continuous RVs practice problem
    a. Let X and Y be two jointly continuous random variables with join PDF

$$f_{XY}(x,y) = \begin{cases} x + cy^2 & 0 \le x \le 1,\, 0 \le y \le 1 \\ 0 & otherwise \end{cases}$$

Find the constant c

$$\begin{aligned} 1 &= \int_0^1 \int_0^1 x + cy^2 \, dx \, dy \\ &= \int_0^1 \frac{1}{2}x^2 + cy^2 x|_0^1 \, dy \\ &= \int_0^1 \frac{1}{2} + cy^2 \, dy \\ &= \frac{1}{2}y + \frac{c}{3}y^3|_0^1 = \frac{1}{2} + \frac{c}{3} \end{aligned}$$

        i.
    b. Suppose a straight stick is broken in three at two points chosen independently at random along its length. What is the chance that the three sticks so formed can be made into the sides of a triangle?

$$\begin{cases} X + Y - X > 1 - Y \longrightarrow & Y > \frac{1}{2} \\ X + 1 - Y > Y - X \longrightarrow & X + \frac{1}{2} > Y \\ Y - X + 1 - Y > X \longrightarrow & X < \frac{1}{2} \end{cases}$$

        i.
        ii. The boundary conditions above form a triangle of size ⅛. The sample space is a larger triangle bounded by 0 < X, Y < 1, and X < Y, whose size is ½. Since x, Y are sampled uniformly at random, the probability that a pair of sampled X, Y satisfy conditions 1 is equal to the ratio of the size of two triangles, which is ¼.

c. X, Y be independent random variables that have N~(0,1). U= min(X, Y). V = max(X, Y). Find E[U] and calculate the Cov(U, V)

   i.

$$E[U] = E[\frac{1}{2}(X+Y-|X-Y|)] = \frac{1}{2}(E[X]+E[Y]-E[|X-Y|])$$

Let $Z = X - Y \sim N(0,2)$,

$$E[U] = -\frac{1}{2}E[|Z|] = -\frac{1}{2}(\int_{-\infty}^{0} -z \cdot Pr[Z=z]\,dz + \int_{0}^{\infty} z \cdot Pr[Z=z]\,dz)$$

$$= -\int_{0}^{\infty} z \cdot \frac{1}{\sqrt{2\pi\sigma_z^2}} exp(\frac{-z^2}{2\sigma_z^2})\,dz) = \frac{1}{\sqrt{\pi}}\int_{0}^{\infty} -\frac{z}{2}exp(\frac{-z^2}{4})\,dz$$

$$= \frac{1}{\sqrt{\pi}}\int_{0}^{-\infty} exp(\frac{-z^2}{4})\,d(\frac{-z^2}{4}) = \frac{1}{\sqrt{\pi}}exp(\frac{-z^2}{4})|_0^{\infty} = -\frac{1}{\sqrt{\pi}}$$

By symmetry, $E[V] = \frac{1}{\sqrt{\pi}}$.

$$Cov(U,V) = E[\frac{1}{4}(X+Y-|X-Y|)(X+Y+|X-Y|)] - E[U]E[V]$$

$$= \frac{1}{4}(E[(X+Y)^2] - E[(X-Y)^2]) + \frac{1}{\pi}$$

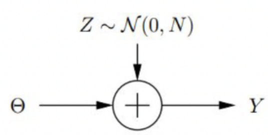
Since $X+Y$ *and* $X-Y$ *both are random variables with distribution* $N(0,2)$, $E[(X+Y)^2] - E[(X-Y)^2] = 0$. *Therefore,* $Cov(U,V) = \frac{1}{\pi}$.

2. Bayes Practice Problem

   a.

Example: Additive Gaussian Noise Channel

- Consider the following communication channel model



where the signal sent

$$\Theta = \begin{cases} +1, & \text{with probability } p \\ -1, & \text{with probability } 1-p, \end{cases}$$

the signal received (also called observation) $Y = \Theta + Z$, and Θ and Z are independent

Given $Y = y$ is received (observed), find the *a posteriori* pmf of Θ, $p_{\Theta|Y}(\theta|y)$

   b. Give the equation and describe each of these terms in your own words
      i. Prior: P(Theta = 1) = p, P(Theta = -1) = 1-p
      ii. Likelihood: P(Y | Theta)
      iii. Posterior: P(Theta | Y)

3. Application of central limit theorem
   a. Conditions for CLT:
      i. Distributions with finite variance
      ii. Independent and identically distributed random variables

4. Confidence intervals
    a. For a confidence interval from (a,b) with confidence x%:
        i. There is a x% chance that the true value lies between (a,b)
    b. CI according to CLT for population mean and population SD:
        i.
$$\Pr(-\sigma < X - \mu < \sigma) \approx 0.68$$
$$\Pr(-2\sigma < X - \mu < 2\sigma) \approx 0.95$$
$$\Pr(-3\sigma < X - \mu < 3\sigma) \approx 0.99$$
    c. For computations on samples:
        i.
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \qquad \bar{X} \pm Z * \frac{\sigma}{\sqrt{n}}$$
    d. Problem 1 - Taxes
        i. Suppose that an accounting firm does a study to determine the time needed to complete one person's tax forms. It randomly surveys 100 people. The sample mean is 23.6 hours. There is a known standard deviation of 7.0 hours. The population distribution is assumed to be normal
            1. Compute the 95% confidence interval for the example above
                a. (23.6)+2*(7.0/10) < 23.6 < (23.6)+2*(7.0/10)
                   (22.228, 24.972)
            2. Does the sample size w.r.t population size affect the confidence interval
                a. No. Population sizes that are large enough do not affect confidence intervals. A sample of 100 of 10,000 is just as good as a sample of 100 in 100,000
            3. If I wanted to increase the confidence, would the interval I report back be: Smaller, Equal to , or larger w.r.t the initial interval
                a. Larger
    e. Problem 2 - Cookies
        i. Suppose I am looking at grams of fat in a randomly selected set of 36 different cookie brands, and I report a confidence interval of (7.95, 9.05) grams of fat with a known standard deviation of 2
            1. What is the confidence of this interval?
                a. Mean = 8.5, sample std = ⅓
                b. Confidence = 90%
                c. 8.5 - 1.645(⅓) = 7.95, 8.5 + 1.645(⅓) = 9.05

2. If I wanted to keep the confidence the same as the original problem, but decrease the interval, what would I do?
    a. Sample more cookie brands