1 CSE103 Final Practice Problems, set 3

1. Let X be a discrete random variable that is uniformly distributed over the set of integers in the range $[a, b]$, where a and b are integers with $a < 0 < b$. Find the PMF of the random variables $max\{0, X\}$ and $min\{0, X\}$.

   Solution (Same as HW4 Prob2)

   $\Omega : \{$ set of integers in the range $[a, b]\}$.

   $|\Omega| = b - a + 1$

$$max\{0, X\} = \begin{cases} X & \text{if } X > 0 \\ 0 & \text{if } X \leq 0 \end{cases}$$

$$P(X = x) = \begin{cases} \frac{1}{b-a+1} & \text{if } x > 0 \\ \frac{1-a}{b-a+1} & \text{if } x \leq 0 \end{cases}$$

$$min\{0, X\} = \begin{cases} X & \text{if } X < 0 \\ 0 & \text{if } X \geq 0 \end{cases}$$

$$P(X = x) = \begin{cases} \frac{1}{b-a+1} & \text{if } x < 0 \\ \frac{b-1}{b-a+1} & \text{if } x \geq 0 \end{cases}$$

2. The MIT football team wins any one game with probability p, and loses it with probability $1-p$. Its performance in each game is independent of its performance in other games. Let L_1 be the number of losses before its first win, and let L_2 be the number of losses after its first win and before its second win. Find the joint PMF of L_1 and L_2.

   Solution Same as HW4 Prob5

   $P(L_1 = n) = (1-p)^n p$

   $P(L_2 = k) = (1-p)^k p$

   Since L_1 and L_2 are independent.

   $P(L_1 = n, L_2 = k) = P(L_1 = n) * P(L_2 = k) = ((1-p)^n p)((1-p)^k p)$

3. Consider $f(x, y) = a + bx + cy + dxy$. Given a set of n data points where the i'th data point is given by : (x_i, y_i, z_i), find the system of linear equations to solve for a, b, c, d such that

$$\frac{1}{n}\sum_{i=1}^{n}(f(x_i, y_i) - z_i)^2$$

is minimized.

Solution (Similar to HW7 Prob5)

$MSE = \frac{1}{n}\sum_{i=1}^{n}(a + bx_i + cy_i + dx_iy_i - z_i)^2$

$$\frac{\partial}{\partial a}MSE(a, b, c, d) = \frac{1}{n}\sum_{i=1}^{n}(1)(a + bx_i + cy_i + dx_iy_i - z_i)$$

$$\frac{\partial}{\partial a}MSE(a, b, c, d) = \frac{1}{n}\sum_{i=1}^{n}(x)(a + bx_i + cy_i + dx_iy_i - z_i)$$

$$\frac{\partial}{\partial a}MSE(a, b, c, d) = \frac{1}{n}\sum_{i=1}^{n}(y)(a + bx_i + cy_i + dx_iy_i - z_i)$$

$$\frac{\partial}{\partial a}MSE(a, b, c, d) = \frac{1}{n}\sum_{i=1}^{n}(xy)(a + bx_i + cy_i + dx_iy_i - z_i)$$

(Set up the system of linear equation similar to HW7 Prob 5)

4. For the set of n data points where i'th data point is given by : (x_i, y_i), answer the following questions:

   (a) If $corr(x, y) = 1$, what is the MSE of the regression line?
   $y = \mu_y + \frac{\sigma_y}{\sigma_x}(x - \mu_x)$

   (b) Let p be a distribution over (x, y) such that the mean and SD of x are $\mu x, \sigma x$ and that of y are $\mu y, \sigma y$. If $corr(x, y) = r$ such that $r > 0$, find the slope of the SD line. Also, write the equation of the regression line.
   $y = \mu_y + r\frac{\sigma_y}{\sigma_x}(x - \mu_x)$