10-751 Assignment 3

Due: Tuesday October 7

Problem 1.

1. Suppose that X, Y, and Z are independent and uniformly distributed in $[0,1]$. Derive the density function of $X+Y+Z$.
2. If $X_1, X_2, \ldots, X_r$ are independent and uniformly distributed on $[0,1]$, show that the density $f_r(x)$ of $\sum_{i=1}^{r} X_i$ is a polynomial in x of degree $r-1$.

Solution. For two uniform random variables X, Y we have

$$\begin{aligned} f_2(z) = f_{X+Y}(z) &= \int_0^z dx \\ &= z \end{aligned}$$

if $0 \le z \le 1$ and

$$\begin{aligned} f_2(z) = f_{X+Y}(z) &= \int_{z-1}^1 dx \\ &= 2 - z \end{aligned}$$

if $1 \le z \le 2$. Therefore

$$\begin{aligned} f_3(w) &= \int_0^1 f_2(w-z)\, dz \\ &= \int_{w-1}^1 (w-z)\, dz + \int_0^{w-1} (2-w+z)\, dz \\ &= \frac{3}{4} - (x - \frac{3}{2})^2 \end{aligned}$$

Part 2. follows from induction.

Problem 2. There are $\binom{n}{m}$ items, each assigned a cost (with no equal costs). You seek to find the best item item—the item with the least cost. Suppose that a random walk among the items is constructed so that from the j-th best item (that is, the j-th item ranked according to cost), you step to any one of the $j-1$ less costly items with uniform probability.

1. Let r_j be the expected number of steps to reach the best item from the j-th best item. Show that $r_j = \sum_{k=1}^{j-1} k^{-1}$.
2. Given an asymptotic expression for the expected time to reach the best item starting from the worst item, for large n and m.

(Note, this problem is related to the expected running time of the simplex algorithm.)

Solution. Letting p_{jk} denote the transition probabilities of the random walk, we have that $p_{jk} = \frac{1}{j-1}$ for $k = 1, 2, \ldots j-1$. By conditional expectation,

$$r_j = 1 + \frac{1}{j-1} \sum_{k=1}^{j-1} r_k$$

with $r_1 = 0$. Induction now gives part 1. For part 2, note that since $\log k = \int_0^k \frac{1}{x}\,dx \approx \sum_{j=1}^k \frac{1}{j}$, we have that $r_j \approx \log j$. The worst-case expectation is thus $O\left(\log \binom{n}{m}\right) = O\left(n \log n\right)$, using Stirling's formula for the second bound.

Problem 3. Let X and Y be independent Bernoulli random variables with parameter $\frac{1}{2}$. Show that $X+Y$ and $|X-Y|$ are dependent but uncorrelated.

Solution. $X+Y$ and $|X-Y|$ are uncorrelated since

$$\begin{aligned} \operatorname{cov}(X+Y, |X-Y|) &= E\left((X+Y)|X-Y|\right) - E(X+Y)E(|X-Y|) \\ &= \frac{1}{4} + \frac{1}{4} - 1 \cdot \frac{1}{2} \\ &= 0 \end{aligned}$$

However,

$$\begin{aligned} \frac{1}{4} &= \Pr(X+Y=0, |X-Y|=0) \\ &\neq \Pr(X+Y=0)\Pr(|X-Y|=0) \\ &= \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8} \end{aligned}$$

Problem 4. Suppose that A, B and C are such that A and B are independent, $\Pr(ABC) = 0.04$, $\Pr(C \mid AB) = 0.25$, and $\Pr(B) = 4\Pr(A)$. What is $\Pr(A \cup B)$?

Solution.

$$\begin{aligned} \Pr(AB) &= \frac{\Pr(ABC)}{\Pr(C \mid AB)} \\ &= \frac{0.04}{0.25} = 0.16 \end{aligned}$$

But by independence

$$\begin{aligned} \Pr(AB) &= \Pr(A)\ \Pr(B) \\ &= 4\ \Pr(A)^2 \end{aligned}$$

So, $\Pr(A) = 0.2$. Therefore,

$$\begin{aligned}
\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(AB) \\
&= 0.2 + 4(0.2 - 0.16 \\
&= 0.84
\end{aligned}$$

Problem 5. (DeGroot 3.7.12) Let Y be the rate (calls per hour) at which calls arrive at a switchboard. Let X be the number of calls during a two-hour period. Suppose that the marginal p.d.f. of Y is

$$f_2(y) = \begin{cases} e^{-y} & \text{if } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

and that the conditional p.f. of X given $Y = y$ is

$$g_1(x \mid y) = \begin{cases} \frac{(2y)^x}{x!} e^{-2y} & \text{if } x = 0, 1, \ldots, \\ 0 & \text{otherwise} \end{cases}$$

1. Find the marginal p.f. of X.
2. Find the conditional p.d.f. $g_2(y \mid 0)$ of Y given $X = 0$.
3. Find the conditional p.d.f. $g_2(y \mid 1)$ of Y given $X = 1$.
4. For what values of y is $g_2(y \mid 1) > g_2(y \mid 0)$? Does this agree with the intuition that the more calls you see, the higher you should think the rate is?

Solution.

1. The joint distribution is

$$f(x, y) = \begin{cases} \frac{(2y)^x}{x!} \exp(-3y) & \text{if } y > 0 \text{ and } x = 0, 1, \ldots \\ 0 & otherwise \end{cases}$$

   The marginal distribution of X is thus

$$\begin{aligned}
f_1(x) &= \int_0^\infty \frac{(2y)^x}{x!} \exp(-3y)\, dy \\
&= \frac{1}{3} \left(\frac{2}{3}\right)^x
\end{aligned}$$

2. The conditional p.d.f. of Y given $X = 0$ is

$$\begin{aligned}
g_2(y \mid 0) &= \frac{(2y)^0 \exp(-3y)/0!}{(1/3)(2/3)^0} \\
&= 3 \exp(-3y)
\end{aligned}$$

   for $y > 0$.

3. The conditional p.d.f. of Y given $X = 1$ is

$$\begin{aligned} g_2(y \mid 1) &= \frac{(2y)^1 \exp(-3y)/1!}{(1/3)(2/3)^1} \\ &= 9y \exp(-3y) \end{aligned}$$

for $y > 0$.

4. The ratio is

$$\begin{aligned} \frac{g_2(y \mid 1)}{g_2(y \mid 0)} &= \frac{9y \exp(-3y)}{3 \exp(-3y)} \\ &= 3y \end{aligned}$$

Greater if $y > 1/3$; yes, we expect the rate is higher is we observe more calls.

Problem 6. Let $\beta > 1$, let $p_1, p_2, \ldots$ denote the prime numbers, and let $N(1), N(2), \ldots$ be independent random variables having probability mass function

$$\Pr(N(i) = k) = \left(1 - p_i^{-\beta}\right) p_i^{-\beta k}$$

Show that $M = \prod_{i=1}^{\infty} p_i^{N(i)}$ is a random integer having probability mass fuction $\Pr(M = m) = Cm^{-\beta}$ for $m \geq 1$, where C is the constant given by

$$\begin{aligned} C &= \prod_{i=1}^{\infty} \left(1 - \frac{1}{p_i^{\beta}}\right) \\ &= \left(\sum_{m=1}^{\infty} \frac{1}{m^{\beta}}\right)^{-1} \end{aligned}$$

Solution. Write the prime factorization of m as $m = \prod_i p_i^{m(i)}$. Then

$$\begin{aligned} \Pr(M = m) &= \prod_i \Pr(N(i) = m(i)) \\ &= \prod_i \left(1 - \frac{1}{p_i^{\beta}}\right) \frac{1}{p_i^{\beta m(i)}} \\ &= C \left(\prod_i p_i^{-m(i)}\right)^{\beta} \\ &= \frac{C}{m^{\beta}} \end{aligned}$$

where $C = \prod_i \left(1 - p_i^{-\beta}\right)$. So, $C^{-1} = \sum_m m^{-\beta}$.