MA 413 - Statistics for Data Science

Solutions to Exercise 2

1. The CDF of Y is

$$F_Y(y) = \begin{cases} 0 & -\infty < y \le 0 \\ y & 0 < y < 1 \\ 1 & 1 \le y < \infty, \end{cases}$$

and the CDF of X is given by $\int_{-\infty}^{x} f_X(u)\ du$, i.e.,

$$F_X(x) = \begin{cases} 0 & -\infty < x \le 0 \\ 1 - e^{-\lambda x} & x > 0. \end{cases}$$

We would like $F_X(x) = F_Y(y)$ for some function $r : X \to Y$, or $X = r(Y)$. Particularly, when $0 < y < 1$ and $x > 0$, we let $y = 1 - e^{-\lambda x}$, solving for x leading to

$$X = -\frac{1}{\lambda} \ln(1 - Y). \tag{1}$$

Eq. (1) is used to generate X from Y, and the resulting X satisfies the given distribution of $f_X(x)$.

Remark. As $1-Y \sim U(0,1)$, we can replace $(1-Y)$ with Y for simplicity, and X can be alternatively generated with $X = -\ln(Y)/\lambda$.

2. (a) The second axiom of probability in the continuous case needs $\int\int f_{X,Y}(x,y)\ dx\ dy = 1$, which is equivalent to

$$C \int_0^1 \left(\int_0^y x\ y\ dx \right) dy = \frac{C}{8} = 1,$$

so we have $C = 8$.

(b) We calculate $f_X(x)$ and $f_Y(y)$,

$$f_X(x) = \int f_{X,Y}(x,y)\ dy = 8x \int_x^1 y\ dy = 4x(1 - x^2),$$
$$f_Y(y) = \int f_{X,Y}(x,y)\ dx = 8y \int_0^y y\ dx = 4y^3,$$

so $f_{X,Y}(x,y) \neq f_X(x) f_Y(y)$ and X, Y are not independent.

3. (a) In the discrete case, the second axiom states $\sum_{x=0}^{\infty} f_X(x) = 1$, so we have

$$\sum_{x=0}^{\infty} f_X(x) = C \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = Ce^{\lambda} = 1,$$

yielding $C = e^{-\lambda}$. Note that the Taylor series for the exponential function e^{λ} is used. As a result, $f_X(x)$ is PMF of the Poisson distribution.

(b) We compute the probability as follows

$$\Pr(X \le 5) = e^{-\lambda} \sum_{x=0}^{5} \frac{\lambda^x}{x!}.$$

When $\lambda = 2$, we compute

$$\begin{aligned} \Pr(X \le 5) &= e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!} + \frac{2^4}{4!} + \frac{2^5}{5!} \right) \\ &= e^{-2} \left(1 + 2 + \frac{4}{2} + \frac{8}{6} + \frac{16}{24} + \frac{32}{120} \right) \\ &= \frac{109}{15e^2} \approx 0.9834. \end{aligned}$$

(c) We compute the mean as follows,

$$\mathrm{E}(X) = e^{-\lambda}\sum_{x=0}^{\infty} x\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty} x\frac{\lambda^x}{x!} == \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda e^{-\lambda}e^{\lambda} = \lambda.$$

For $\mathrm{Var}(X)$, we use formula $\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$. Note that $\mathrm{E}(X^2) = \mathrm{E}(X(X-1)+X) = \mathrm{E}(X(X-1)) + \mathrm{E}(X)$. We compute $\mathrm{E}(X(X-1))$ as follows,

$$\mathrm{E}(X(X-1)) = e^{-\lambda}\sum_{x=0}^{\infty} x(x-1)\frac{\lambda^x}{x!} = \lambda^2 e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} = \lambda^2 e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = \lambda^2.$$

To sum up, we obtain the variance,

$$\mathrm{Var}(X) = \mathrm{E}(X(X-1)) + \mathrm{E}(X) - (\mathrm{E}(X))^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

The dispersion of X is $\mathrm{Var}(X)/\mathrm{E}(X) = 1$.

4. We compute its mean as

$$\begin{aligned}
\mathrm{E}(X) &= \sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=1}^{n} x\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
&= np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1}(1-p)^{n-x} \\
&= np\sum_{x=1}^{n}\frac{(n-1)!}{(x-1)!((n-1)-(x-1))!}p^{x-1}(1-p)^{(n-1)-(x-1)} \\
&= np\sum_{x=0}^{n-1}\frac{(n-1)!}{x!(n-1-x)!}p^{x}(1-p)^{n-1-x} \\
&= np(p+(1-p))^{n-1} = np.
\end{aligned}$$

Remark. Alternatively, as X is a Binomial random variable, it represents the number of heads in n tosses of a coin, so it can be written as $X = \sum_{x=0}^{n} X_i$, where $X_i = 1$ if the i^{th} toss is heads and $X_i = 0$ otherwise. Then $\mathrm{E}(X_i) = p \times 1 + (1-p) \times 0 = p$, and $\mathrm{E}(X) = \mathrm{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n}\mathrm{E}(X_i) = np$.

The following is the Matlab code to compare the mean and variance between Binomial and Poisson.

```
lambda = 1/50;
n_trial = 100;
p_success = lambda / n_trial;
pmf_B = makedist('Binomial','N',n_trial,'p',p_success);
E_B = mean(pmf_B)
Var_B = var(pmf_B)

pmf_P = makedist('Poisson','lambda',lambda);
E_P = mean(pmf_P)
Var_P = var(pmf_P)
```

5. Let $X \sim \mathrm{Poisson}(\lambda)$. We would like the CDF of X matches that of Y, i.e., $F_X(x) = F_Y(y)$, but we notice that X is discrete whereas Y is continuous, so given a particular $y \in (0,1)$, we choose a corresponding x ($x \geq 1$) such that $F_X(x-1) < F_Y(y) \leq F_X(x)$, or equivalently,

$$e^{-\lambda}\sum_{n=0}^{x-1}\frac{\lambda^n}{n!} < y \leq e^{-\lambda}\sum_{n=0}^{x}\frac{\lambda^n}{n!}. \tag{2}$$

In other words, X is given by Eq. (2) when the uniform random variable $Y > F_X(0) = e^{-\lambda}$, and $X = 0$ otherwise.

The following is the Matlab code to generate a Poisson random variable from a uniform one.

```
lambda = 2; % Poisson parameter, can be other value
Y = rand % uniform random variable in (0,1)
x = 0; % initialization of the target Poisson random variable
pmf = exp(-lambda);
FX = pmf;
while FX < Y
    x = x + 1;
    pmf = pmf * lambda / x;
    FX = FX + pmf;
end
x
```

6. We first compute $\Gamma(2)$ as follows,

$$\Gamma(2) = \int_0^\infty y e^{-y}\, dy = -ye^{-y}\Big|_0^\infty + \int_0^\infty e^{-y}\, dy = 0 - e^{-y}\Big|_0^\infty = 1.$$

To find the distribution of $Z = X/Y$, we need a joint density function, so we assume X and Y are independent. As a result, we have

$$f_{X,Y}(x,y) = \begin{cases} f_X(x)f_Y(y) = \frac{y}{8}e^{-(x+y)/2} & x > 0,\ y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Given $z > 0$, we get the CDF of Z as follows,

$$\begin{aligned}
F_Z(z) &= \Pr(Z \le z) = \Pr(X/Y \le z) \\
&= \int\int_{\{(x,y):x\le zy\}} f_{X,Y}(x,y)\, dx\, dy = \int_0^\infty \left(\int_0^{zy} \frac{y}{8} e^{-(x+y)/2}\, dx\right)\, dy \\
&= \frac{1}{8}\int_0^\infty ye^{-y/2}\left(\int_0^{zy} e^{-x/2}\, dx\right) dy = \frac{1}{8}\int_0^\infty ye^{-y/2}\left(-2e^{-x/2}\Big|_0^{zy}\right) dy \\
&= \frac{1}{4}\int_0^\infty ye^{-y/2}\left(-e^{-zy/2}+1\right) dy = \frac{1}{4}\int_0^\infty y\left(-e^{-(z+1)y/2}+e^{-y/2}\right) dy \\
&= \frac{y}{4}\left(\frac{2}{z+1}e^{-(z+1)y/2} - 2e^{-y/2}\right)\Bigg|_0^\infty - \frac{1}{4}\int_0^\infty \frac{2}{z+1}e^{-(z+1)y/2} - 2e^{-y/2}\, dy \\
&= -\frac{1}{4}\int_0^\infty \frac{2}{z+1}e^{-(z+1)y/2} - 2e^{-y/2}\, dy = -\frac{1}{4}\left(-\frac{4}{(z+1)^2}e^{-(z+1)y/2} + 4e^{-y/2}\right)\Bigg|_0^\infty \\
&= 1 - \frac{1}{(z+1)^2},
\end{aligned}$$

and $F_Z(z) = 0$ when $z \le 0$.

We compute the mean of X as

$$\mathrm{E}(X) = \int_0^\infty x\frac{1}{2}e^{-x/2}\, dx = -xe^{-x/2}\Big|_0^\infty + \int_0^\infty e^{-x/2}\, dx = \int_0^\infty e^{-x/2}\, dx = -2e^{-x/2}\Big|_0^\infty = 2,$$

from which we notice $\int_0^\infty ue^{-u/2}\, du = 4$. We will use it in the following computations. To compute the variance of X, we first compute $\mathrm{E}(X^2)$,

$$\mathrm{E}(X^2) = \int_0^\infty x^2\frac{1}{2}e^{-x/2}\, dx = -x^2e^{-x/2}\Big|_0^\infty + \int_0^\infty 2xe^{-x/2}\, dx = 2\int_0^\infty xe^{-x/2}\, dx = 2\times 4 = 8,$$

and thus

$$\mathrm{Var}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = 8 - 2^2 = 4.$$

In addition, we have $\int_0^\infty u^2 e^{-u/2}\, du = 16$. We use it to compute the mean and variance of Y,

$$\mathrm{E}(Y) = \int_0^\infty y \frac{y}{4} e^{-y/2}\, dy = \frac{1}{4}\int_0^\infty y^2 e^{-y/2}\, dy = \frac{1}{4} \times 16 = 4,$$

$$\mathrm{E}(Y^2) = \frac{1}{4}\int_0^\infty y^3 e^{-y/2}\, dy = -\frac{1}{2} y^3 e^{-y/2}\Big|_0^\infty + \frac{1}{2}\int_0^\infty 3y^2 e^{-y/2}\, dy = \frac{3}{2}\int_0^\infty y^2 e^{-y/2}\, dy = \frac{3}{2} \times 16 = 24,$$

and

$$\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2 = 24 - 4^2 = 8.$$