2019 Probability and Statistic Examination Solution

1. a. i) The maximum likelihood estimation for Poisson(λ) is

$$L(\lambda) = \prod_{i=1}^{n} p(x_i) = \frac{e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)}$$

$$l(\lambda) = \log e^{-\lambda n}\lambda^{\sum_{i=1}^{n} x_i} - \log \prod_{i=1}^{n}(x_i!)$$

$$= -\lambda n + \sum_{i=1}^{n} x_i \log \lambda - (\sum_{i=1}^{n} \log (x_i!))$$

$$l'(\lambda) = -n + \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

$$l'(\lambda) = 0 \Rightarrow n = \frac{\sum_{i=1}^{n} x_i}{\lambda}$$

$$\lambda = \frac{\sum_{i=1}^{n} x_i}{n}$$

   ii) The maximum likelihood estimate for Uniform(α, β) is

$$L(\alpha, \beta) = \prod_{i=1}^{n} p(x_i) = (\frac{1}{\beta - \alpha})^n$$

$$l(\alpha, \beta) = n \log \frac{1}{\beta - \alpha} = -n \log (\beta - \alpha)$$

$$\frac{d}{d\alpha} l(\alpha, \beta) = \frac{n}{\beta - \alpha}$$

$$\frac{d}{d\beta} l(\alpha, \beta) = -\frac{n}{\beta - \alpha}$$

   We can see that the derivative with respect to both α and β are monotonically increasing or descreasing, thus we can obtain the maximum value of $l(\alpha, \beta)$ by letting $\alpha = \max(x_1, \ldots x_n)$ and $\beta = \min(x_1, \ldots, x_n)$.

   b. We construct a confidence interval as follows. Since we do not know the population variance, we will calculate the bias-corrected variance first

$$\mu = \frac{29.9 + 35.1 + 10 + 55.1 + 25.1 + 33.9 + 42.3 + 31.9 + 60.5}{9} = 35.98$$

$$s = \sqrt{\frac{1}{n-1}(\sum_i x_i^2 - \frac{1}{n}(\sum_i x_i)^2) = \frac{1}{8}(13508.4 - \frac{104846.44}{9})} = \sqrt{232.35} = 15.24$$

   then we need can find the confidence interval as follows

$$[\mu - t_{8,0.995}\frac{s}{\sqrt{n}}, \mu + t_{8,0.995}\frac{s}{\sqrt{n}}] \Rightarrow [35.98 - 3.36 \cdot \frac{15.24}{3}, 35.98 + 3.36 \cdot \frac{15.24}{3}] \Rightarrow [18.91, 53.05]$$

   If we draw a large number of samples, 99% of them will have mean value that lies under this interval.

c. To test if type A and type B batteries perform equally well, we will have the following null and alternative hypothesis.

$$H_0 : \mu_A - \mu_B = 0 \qquad H_1 : \mu_A - \mu_B \neq 0$$

We now calculate the test statistics as follows

$$T = \frac{(\bar{A} - \bar{B}) - (\mu_A - \mu_B)}{S_{n_a+n_b-2}\sqrt{1/n_a + 1/n_b}}$$

For null hypothesis, this becomes

$$S_{n_a+n_b-2} = \sqrt{\frac{11-1}{11+11-2} \cdot 900 + \frac{11-1}{11+11-2} \cdot 700} = \sqrt{800} = 28.28$$

$$t = \frac{300 - 270}{28.28\sqrt{1/11 + 1/11}} = 2.49$$

From t-distribution with 5% significance level and degree of freedom 20, we can acquire the rejection region as

$$(-\infty, -2.09) \cup (2.09, \infty)$$

The test statistic t falls under this range and hence we reject the null hypothesis and we are 95% confident that type A and type B are NOT performing equally well.

d. If there are no correlation between political affliation and opinion on ULEZ, then the expected number of people look like this

	favour	**indifferenet**	**opposeed**	**total**
Labour	114	84.36	86.64	285
Conservative	86	63.64	65.36	215
Total	200	148	152	500

We then perform a χ^2-independence test

$$\chi^2 = \frac{(137-114)^2}{114} + \frac{(63-86)^2}{86} + \frac{(82-84.36)^2}{84.36} + \frac{(66-63.64)^2}{63.64} + \frac{(66-86.64)^2}{86.64} + \frac{(86-65.36)^2}{65.36} = 22.38$$

2. a. i) Since all games aree mutually independent, we can calculate the mean of draws in all divisions as follow

$$\mu = 380 \cdot 0.2 + 552 \cdot (0.25 + 0.3 + 0.35) = 572.8$$

Since the number of draws follows a binomial distribution, we can calculate the standard deviation as follow

$$\sigma = \sqrt{380 \cdot 0.2 \cdot 0.8 + 552 \cdot 0.25 \cdot 0.75 + 552 \cdot 0.3 \cdot 0.7 + 552 \cdot 0.35 \cdot 0.65} = 20.14$$

b. i) The characteristic function of a continuous random variable is

$$\phi_X(t) = E(e^{itX}) = \int_{-\infty}^{\infty} e^{itx} f_X(x)dx$$

According to this definition, the characteristic function of a Bernoulli random variable would be

$$\phi_{\text{Bernoulli}}(t) = E(e^{itB}) = e^{it\cdot 0}(1-p) + e^{it\cdot 1}p = 1 - p + pe^{it}$$

ii) The proof is as follow

$$\phi_{S_n}(t) = E(e^{itS_n}) = E(\prod_n e^{itX_n}) = \prod_n E(e^{itX_n}) = \prod_n \phi_{X_n}(t)$$

iii) According to the definition of mean on random variable, we can obtain the mean of $\tilde{B}$ as follows

$$\begin{aligned} \mu &= p\sqrt{\frac{1-p}{p}} + (1-p)(-\sqrt{\frac{p}{1-p}}) \\ &= \sqrt{p^2 \cdot \frac{1-p}{p}} - \sqrt{(1-p)^2 \cdot \frac{p}{1-p}} \\ &= \sqrt{p(1-p)} - \sqrt{(1-p)p} \\ &= 0 \end{aligned}$$

According to the definition of variance on random variables, we can obtain the variance of $\tilde{B}$ as follows

$$\begin{aligned} \sigma^2 &= E(X^2) - E(X)^2 \\ &= (\frac{1-p}{p} \cdot p + \frac{p}{1-p} \cdot (1-p)) - 0 \\ &= (1 - p + p) - 0 \\ &= 1 \end{aligned}$$

iv) The characteristic function of $\tilde{B}$ is

$$\phi_B(t) = E(e^{itB}) = e^{it\cdot\sqrt{\frac{1-p}{p}}}p + e^{-it\cdot\sqrt{\frac{p}{1-p}}}(1-p)$$

The characteristic function of $S = \sum_{i=1}^{n} \frac{\tilde{B}_i}{\sqrt{n}}$ is

$$\phi_S(t) = E(e^{it\sum_{i=1}^{n}\frac{\tilde{B}}{\sqrt{n}}}) = E(\prod_{i=1}^{n} e^{it\frac{\tilde{B}}{\sqrt{n}}}) = \prod_{i=1}^{n} E(e^{it\frac{\tilde{B}}{\sqrt{n}}}) = \prod_{i=1}^{n} \phi_B(\frac{t}{\sqrt{n}}) = (\phi_B(\frac{t}{\sqrt{n}}))^n$$

v) According to Central Limit Theorem, $Z = \lim_{n\to\infty} \frac{S-n\mu}{\sqrt{n\sigma^2}} \sim N(0,1)$. From iii) we know that $\mu = 0$ and $\sigma^2 = 1$ in this equation, hence it becomes

$$Z = \lim_{n\to\infty} \frac{S}{\sqrt{n}} \sim N(0,1)$$

which means that as $n \to \infty$, the distribution of S will approach to a normal distribution with its characteristic function being the normal distribution characteristic function as well, which is $e^{-\frac{t^2}{2}}$.

c. i) From the nature of exponential distribution, the mean is calculated as $\mu = \frac{1}{\lambda}$.The probability of professor being m minutes late is

$$\begin{aligned} P(m) &= P(Z < -1)(1 - F_X(m)) + P(X > -1)(1 - F_Y(m)) \\ &= 0.159 \cdot (1 - 1 + e^{-0.1\cdot m}) + 0.841 \cdot (1 - 1 + e^{-0.05\cdot m}) \\ &= 0.159e^{-0.1m} + 0.841e^{-0.05m} \end{aligned}$$

ii) The probability that the professor will arrive at Victoria before 07:30 is

$$P(Z < -7) = 1.28 \times 10^{-12}$$