1. (*14pts, 2pts for each*)
   (a) False (b) False (c) True (d) True (e) True (f) False (g) True

2. (*18pts, 6pts for each*)

   (a) **(1)** $X \sim \text{binomial}(1919, p)$, where p is the parameter. (Note that the total number of deaths, i.e., 1919, is treated as a fixed number, rather than a random variable.)

   (2) $\Omega = \{p|\ 0 < p \leq 1/2\}$ (**Note.** It is either there is no holiday effect, i.e., $p = 1/2$, or there is a holiday effect, i.e., $0 < p < 1/2$. We do not assume there could be an "anti-holiday" effect, i.e., higher chance to die before holiday than after holiday.)

   (3) $\Omega_0 = \{1/2\}$, i.e., no holiday effect. Notice that a statistical test uses a small significance level to offer protection against rejecting the null hypothesis, and in scientific investigations, the null hypothesis is often an explanation that must be discredited in order to demonstrate the presence of some effect. In this case, it is better to use "no holiday effect" as the null hypothesis, rather than "there is a holiday effect."

   (b) **(1)** $X_1, \ldots, X_{20}$ are i.i.d. from exponential(λ), where λ is the parameter.

   (2) $\Omega = \{\lambda|\ 0 < \lambda < \infty\}$

   (3) From the customers viewpoint, we would like to protect the hypothesis "the mean time to failure is *less* than 2 years" and therefore use it as the null hypothesis, i.e., $\Omega_0 = \{\lambda|\ \lambda \geq \frac{1}{2\times 365} = \frac{1}{730}\}$. (**Note**. $E(X_i) = 1/\lambda$.)

   (c) **(1)** $(X_0, X_1, \ldots, X_{10}) \sim \text{multinomial}(280, p_0, p_1, \ldots, p_{10})$

   (2) $\Omega = \{(p_0, p_1, \ldots, p_{10})|\ \sum_{i=0}^{10} p_i = 1\}$ (Note that $\dim(\Omega) = 11 - 1 = 10$.)

   (3) $\Omega_0 = \{(p_0, p_1, \ldots, p_{10})|\ p_i = P(Z = i),\ i = 0, 1, \ldots, 10$, where $Z \sim \text{binomial}(10, p)$ and $0 < p < 1$ } (Note that $\dim(\Omega_0)$=1.)

3. (*15pts*)

   (a) (*3pts*) Because $0 < Y_1, \ldots, Y_n < \theta$, it is straightforward that $P(Y_{(n)} \leq y) = 0$ for $y < 0$, and $P(Y_{(n)} \leq y) = 1$ for $y \geq \theta$. For $0 \leq y < \theta$,

$$\begin{aligned} F_{Y_{(n)}}(y) &= P(Y_{(n)} \leq y) = P(Y_1 \leq y, \ldots, Y_n \leq y) \\ &= \prod_{i=1}^{n} P(Y_i \leq y) = \left(\int_0^y \frac{1}{\theta}\, dy\right)^n = \left(\frac{y}{\theta}\right)^n. \end{aligned}$$

   (b) (*5pts*) Because $0 < Y_{(n)} < \theta$, it is straightforward that $0 < U = Y_{(n)}/\theta < 1$, and $P(U \leq u) = 0$ for $u < 0$; $P(U \leq u) = 1$ for $u \geq 1$. For $0 \leq u < 1$, the cdf of U is:

$$P(U \leq u) = P\left(Y_{(n)}/\theta \leq u\right) = P\left(Y_{(n)} \leq u\theta\right) = \left(\frac{u\theta}{\theta}\right)^n = u^n,$$

   which indicates that $U \sim \text{gamma}(n, 1)$. Because (i) U is a function of the parameter θ and the data $Y_1, \ldots, Y_n$, and (ii) the distribution of U is irrelevant to all parameters (only θ in the case), U is a pivotal quantity for θ.

(c) (*7pts*) Let $0 < a < 1$ and $0 < b < 1$ satisfy

$$\begin{aligned} \frac{0.05}{2} &= P(U \le a) = a^n \qquad \Rightarrow \qquad a = \sqrt[n]{0.025}, \\ \frac{0.05}{2} &= P(U > b) = 1 - b^n \qquad \Rightarrow \qquad b = \sqrt[n]{0.975}. \end{aligned}$$

Then, from (b), we know

$$1 - \alpha = P\left(a \le U = Y_{(n)}/\theta \le b\right) = P\left(\frac{Y_{(n)}}{b} \le \theta \le \frac{Y_{(n)}}{a}\right).$$

Therefore,

$$\left[\frac{Y_{(n)}}{\sqrt[n]{0.975}}, \frac{Y_{(n)}}{\sqrt[n]{0.025}}\right]$$

is a 95% confidence interval for θ.

4. (*15pts*)

(a) (*12pts*) The joint pdf of $X_1, \ldots, X_n$ is

$$f(\mathbf{x}|\sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-x_i^2/(2\sigma^2)\right) = \left(\sqrt{2\pi}\sigma\right)^{-n} \exp\left(-\frac{\sum_{i=1}^n x_i^2}{2\sigma^2}\right).$$

Therefore, the likelihood ratio is

$$\Lambda = \frac{f(\mathbf{x}|\sigma_0)}{f(\mathbf{x}|\sigma_1)} = \left(\frac{\sigma_1}{\sigma_0}\right)^n \exp\left[\left(\sum_{i=1}^n x_i^2\right)\left(\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2}\right)\right],$$

where $\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2} < 0$ when $\sigma_1 > \sigma_0$. Because $\frac{1}{2\sigma_1^2} - \frac{1}{2\sigma_0^2} < 0$, Λ decreases as $\sum_{i=1}^n X_i^2$ increases. We should reject H_0 when $\sum_{i=1}^n X_i^2$ is large (i.e., larger $\sum_{i=1}^n X_i^2$ is more extreme). The test function based on the likelihood ratio is then given by:

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \sum_{i=1}^n X_i^2 > c, \\ 0, & \text{if } \sum_{i=1}^n X_i^2 \le c. \end{cases}$$

The c is determined by

$$\mathrm{E}_{\sigma_0}(\phi) = P_{\sigma_0}\left(\sum_{i=1}^n X_i^2 > c\right) = P_{\sigma_0}\left(\sum_{i=1}^n X_i^2/\sigma_0^2 > c/\sigma_0^2\right) = \alpha.$$

Because $\sum_{i=1}^n X_i^2/\sigma_0^2 \sim \chi_n^2$ under H_0, we can get $c = \sigma_0^2\chi_n^2(1-\alpha)$, where $\chi_n^2(1-\alpha)$ is the $(1-\alpha)$-quantile of the χ_n^2 distribution.

(b) (*3pts*) By Neyman-Pearson lemma, the test in (a) is the most powerful test for any particular simple alternative $H_A : \sigma = \sigma_1$, where $\sigma_1 > \sigma_0$. Furthermore, because the rejection region and the critical value (i.e., c) of the test does not depend on σ_1, the test is UMP for $H_0 : \sigma = \sigma_0$ versus $H_A : \sigma > \sigma_0$.

5. (*20pts*)

(a) (*2pts*) $\dim(\Omega_0) = 1$ and $\dim(\Omega) = m$

(b) (*12pts*) Because

$$\Lambda(x_1,\ldots,x_m) = \frac{\max_{\theta\in\Omega_0}\mathcal{L}(\theta,\mathbf{x})}{\max_{\theta\in\Omega}\mathcal{L}(\theta,\mathbf{x})} = \frac{\prod_{i=1}^{m}\begin{pmatrix} n_i \\ x_i \end{pmatrix}\hat{p}^{x_i}(1-\hat{p})^{n_i-x_i}}{\prod_{i=1}^{m}\begin{pmatrix} n_i \\ x_i \end{pmatrix}\hat{p}_i^{x_i}(1-\hat{p}_i)^{n_i-x_i}}$$

$$= \prod_{i=1}^{m}\left(\frac{\hat{p}}{\hat{p}_i}\right)^{x_i}\left(\frac{1-\hat{p}}{1-\hat{p}_i}\right)^{n_i-x_i},$$

we can get

$$\begin{aligned} -2\log\Lambda &= -2\sum_{i=1}^{m}\left[x_i\log\left(\frac{\hat{p}}{\hat{p}_i}\right)+(n_i-x_i)\log\left(\frac{1-\hat{p}}{1-\hat{p}_i}\right)\right] \\ &= -2\sum_{i=1}^{m}\left[n_i\hat{p}_i\log\left(\frac{n_i\hat{p}}{n_i\hat{p}_i}\right)+n_i(1-\hat{p}_i)\log\left(\frac{n_i(1-\hat{p})}{n_i(1-\hat{p}_i)}\right)\right] \\ &= -2\sum_{i=1}^{m}\sum_{j=1}^{2}O_{ij}\log\left(\frac{E_{ij}}{O_{ij}}\right) = 2\sum_{i=1}^{m}\sum_{j=1}^{2}O_{ij}\log\left(\frac{O_{ij}}{E_{ij}}\right). \end{aligned}$$

(c) (*2pts*) Because $\dim(\Omega)-\dim(\Omega_0) = m-1$, the large sample distribution of $-2\log\Lambda$ is Chi-square distribution with degrees of freedom $m-1$, i.e., χ^2_{m-1}.

(d) (*4pts*) The rejection region is $2\sum_{i=1}^{m}\sum_{j=1}^{2}O_{ij}\log\left(\frac{O_{ij}}{E_{ij}}\right) > c$, where $c = \chi^2_{m-1}(1-\alpha)$ is determined by $P(\chi^2_{m-1} > c) = \alpha$.

6. (*18pts*)

(a) (*5pts*) The posterior pdf is

$$\begin{aligned} h(\theta|x_1,\ldots,x_n) &\propto f(x_1,\ldots,x_n|\theta)\cdot g(\theta) = \left[\prod_{i=1}^{n}f(x_i|\theta)\right]\cdot g(\theta) \\ &\propto \left[\prod_{i=1}^{n}\theta^{x_i}e^{-\theta}\right]\cdot\theta^{\alpha-1}e^{-\lambda\theta} \\ &= \theta^{\sum_{i=1}^{n}x_i}e^{-n\theta}\cdot\theta^{\alpha-1}e^{-\lambda\theta} \\ &= \theta^{(n\overline{X}+\alpha)-1}e^{-(n+\lambda)\theta}, \end{aligned}$$

which follows the form of the pdf of gamma distribution with shape parameter $n\overline{X}+\alpha$ and scale parameter $n+\lambda$, i.e., $\Theta|x_1,\ldots,x_n \sim \Gamma(n\overline{X}+\alpha, n+\lambda)$.

(b) (*5pts*) Because $\Theta|x_1,\ldots,x_n \sim \text{gamma}(n\overline{X}+\alpha, n+\lambda)$, the Bayes estimator is

$$\hat{\theta}_B = E\left(\Theta|x_1,\ldots,x_n\right) = \frac{n\overline{X}+\alpha}{n+\lambda} = \frac{n}{n+\lambda}\cdot\overline{X}+\frac{\lambda}{n+\lambda}\cdot\frac{\alpha}{\lambda}, \qquad \text{(I)}$$

where $\frac{\alpha}{\lambda}$ is the prior mean, and sum of the weights is one, i.e., $\frac{n}{n+\lambda}+\frac{\lambda}{n+\lambda} = 1$.

(c) (*3pts*) When n is large, the weights in (I) will approximate 1 and 0 respectively, i.e.,

$$\frac{n}{n+\lambda}\approx 1 \quad \text{and} \quad \frac{\lambda}{n+\lambda}\approx 0.$$

Therefore, $\hat{\theta}_B \approx \overline{X}$, which is a function of sample (data) only and is irrelevent to the prior.

(d) (*5pts*) When $\alpha = \lambda = 1$, the Bayes estimator is

$$\hat{\theta}_B = \frac{n}{n+1}\overline{X} + \frac{1}{n+1}.$$

Because $X_1, \ldots, X_n$ are i.i.d. $\sim$ Poisson(θ) given $\Theta = \theta$, we have

$$E_{\mathbf{X}|\theta}(\hat{\theta}_B) = \frac{n}{n+1}E_{\mathbf{X}|\theta}(\overline{X}) + \frac{1}{n+1} = \frac{n}{n+1}\theta + \frac{1}{n+1} = \frac{n\theta+1}{n+1},$$

and

$$Var_{\mathbf{X}|\theta}(\hat{\theta}_B) = \left(\frac{n}{n+1}\right)^2 Var_{\mathbf{X}|\theta}(\overline{X}) = \left(\frac{n}{n+1}\right)^2 \frac{\theta}{n} = \frac{n\theta}{(n+1)^2}.$$

The risk function (i.e., MSE) of $\hat{\theta}_B$ is

$$\begin{aligned} \text{MSE}_{\mathbf{X}|\theta}(\hat{\theta}_B) &= Var_{\mathbf{X}|\theta}(\hat{\theta}_B) + \left(\text{bias}_{\mathbf{X}|\theta}(\hat{\theta}_B)\right)^2 \\ &= \frac{n\theta}{(n+1)^2} + \left(\frac{n\theta+1}{n+1} - \theta\right)^2 = \frac{n\theta + (1-\theta)^2}{(n+1)^2} = \frac{\theta^2 + (n-2)\theta + 1}{(n+1)^2}. \end{aligned}$$

Because $\overline{X}$ is an unbiased estimator, the MSE of the MLE $\overline{X}$ is $Var_{\mathbf{X}|\theta}(\overline{X}) = \theta/n$. The Bayes estimator has s smaller risk than $\overline{X}$ if and only if

$$\begin{aligned} &\frac{\theta^2 + (n-2)\theta + 1}{(n+1)^2} < \frac{\theta}{n} \quad \Leftrightarrow \quad n\theta^2 - \left[(n+1)^2 - n(n-2)\right]\theta + n < 0 \\ &\Leftrightarrow \quad n\theta^2 - (4n+1)\,\theta + n < 0 \\ &\Leftrightarrow \quad \frac{(4n+1) - \sqrt{(4n+1)^2 - 4n^2}}{2n} < \theta < \frac{(4n+1) + \sqrt{(4n+1)^2 - 4n^2}}{2n} \\ &\Leftrightarrow \quad \frac{(4n+1) - \sqrt{12n^2 + 8n + 1}}{2n} < \theta < \frac{(4n+1) + \sqrt{12n^2 + 8n + 1}}{2n} \end{aligned}$$