Final Exam - solutions

22 June 2018

Ex. 1 **a)** [*3 marks*] The sample space is $\Omega = \{\{R,R\},\{R,W\},\{W,W\}\}$, since there is no mention of order in the sampling.
The elements of Ω are not equiprobable:

$$\Pr\{\{R,R\}\} = \frac{\binom{3}{0}\binom{2}{2}}{\binom{5}{2}} = \frac{1}{10},\quad \Pr\{\{R,W\}\} = \frac{\binom{3}{1}\binom{2}{1}}{\binom{5}{2}} = \frac{6}{10},\quad \Pr\{\{R,R\}\} = \frac{\binom{3}{2}\binom{2}{0}}{\binom{5}{2}} = \frac{3}{10}.$$

Let A and B denote the events 'both balls red' and 'at least one red' respectively. So, $A = \{\{R,R)\}\}$ and $B = \{\{R,W\},\{R,R\}\}$, and since $A \subset B$,

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\Pr(A)}{\Pr(B)} = \frac{1/10}{1/10 + 6/10} = \frac{1}{7}.$$

b) [*2 marks*] For $y \in (0,1]$, $F_Y(y) = \Pr(Y \le y) = \Pr(1/X \le y) = \Pr(X \ge 1/y) = y^2$. So,

$$f_Y(y) = \tfrac{\partial F_Y(y)}{\partial y} = 2y, \quad 0 < y \le 1.$$

c) [*2 marks*] Let $x_{0.5}$ be the median. Then

$$F(x_{0.5}) = \tfrac{1}{2} \Rightarrow 1 - \exp\Big\{ -\Big(\tfrac{x_{0.5}}{\mu}\Big)^{\alpha}\Big\} = \tfrac{1}{2} \Rightarrow x_{0.5} = \mu(\ln 2)^{1/\alpha}.$$

d) [*2 marks*] Firstly, note that

$$\Pr(X > 6) = 1 - \Pr\Big(\tfrac{X-2}{\sigma} < \tfrac{6-2}{\sigma}\Big) = 1 - \Phi(4/\sigma).$$

Using the table for the standard normal distribution, we can easily see that

$$\Pr(X > 6) = 0.1 \Rightarrow \Phi^{-1}(0.9) = 4/\sigma \Rightarrow 1.282 = 4/\sigma \Rightarrow \sigma \approx 3.12.$$

Hence

$$\Pr(X < 0) = \Pr\Big(\tfrac{X-2}{3.12} < \tfrac{0-2}{3.12}\Big) = \Phi(-2/3.12) \approx \Phi(-0.64) \approx 0.26.$$

e) [*2 marks*] Given observations x_1 and x_2 from the random variables X_1 and X_2 respectively, the likelihood function is

$$L(\mu; x_1, x_2) = \prod_{i=1}^{2} \tfrac{1}{\sqrt{2\pi}} \exp\big\{\tfrac{-(x_i-\mu)^2}{2}\big\}.$$

Hence, the log-likelihood is

$$\ell(\mu; x_1, x_2) = \log L(\mu; x_1, x_2) = \sum_{i=1}^{2} \Big\{ -\tfrac{(x_i-\mu)^2}{2} - \log\big(\sqrt{2\pi}\big)\Big\}.$$

To find the maximum likelihood estimate $\widehat{\mu}$, we differentiate the log-likelihood with respect to μ and then equate the derivative to zero. This gives

$$\tfrac{2(x_1-\widehat{\mu})}{2} + \tfrac{2(x_2-\widehat{\mu})}{2} = 0,$$

which gives

$$\widehat{\mu} = \tfrac{x_1+x_2}{2}.$$

For a bonus point, check that the second derivative is negative, so this is the (unique) maximum.

f) [*2 marks*] Let $W = \min(X_1, X_2, X_3)$. Then, for all $t > 0$,

$$\begin{aligned}\Pr(W \geq t) &= \Pr(X_1 \geq t, X_2 \geq t, X_3 \geq t)\\ &= \Pr(X_1 \geq t)\Pr(X_2 \geq t)\Pr(X_3 \geq t)\\ &= \exp(-t\lambda_1)\exp(-t\lambda_2)\exp(-t\lambda_3)\\ &= \exp\Big\{-t(\lambda_1+\lambda_1+\lambda_1)\Big\},\end{aligned}$$

where the second equality holds from the independence assumption. So,

$$\Pr(W \leq t) = 1 - \exp\Big\{-t(\lambda_1+\lambda_1+\lambda_1)\Big\}, \quad t > 0.$$

Hence, W is exponentially distributed with parameter $\lambda_1 + \lambda_2 + \lambda_3$.

g) [*2 marks*] Since all the entries of the table must sum to 1, we must have $c = 1/12$. Hence

$$\mathrm{E}(X) = 1 \cdot \tfrac{4}{12} + 3 \cdot \tfrac{4}{12} + 5 \cdot \tfrac{4}{12} = \tfrac{36}{12} = 3,$$

and the conditional expectation is

$$\mathrm{E}(X \mid Y = 4) = \frac{1 \times 3/12 + 3 \times 2/12 + 5 \times 1/12}{6/12} = \frac{14}{6}.$$

The random variables X and Y are clearly dependent, since $\Pr(X = 1 \mid Y = 2) \neq \Pr(X = 1)$. We accept other similar arguments.

h) [*2 marks*] The density of X is $f(x) = 1/2$, for $x \in [-1, 1]$, so

$$M_X(t) = \mathrm{E}\{\exp(tx)\} = \tfrac{1}{2}\int_{-1}^{1} \exp(tx)\mathrm{d}x = \left[\frac{\exp(tx)}{2t}\right]_{x=-1}^{x=1} = \tfrac{\exp(t)-\exp(-t)}{2t} = t^{-1}\sinh t.$$

i) [*2 marks*] The null hypothesis $\mathcal{H}_0$ represents the theory/model we want to test. The test statistic T is chosen such that large values of T provide evidence against $\mathcal{H}_0$. The p-value $= p_{obs} = \Pr_0(T \geq t_{obs})$, where t_{obs} is the observed t-statistic and $\Pr_0$ is the probability under the null. Thus the p-value is the probability that we observe a value of T bigger than or equal to t_{obs}, when the null hypothesis is true. For the given situation, we reject H_0 at level α for all $\alpha > 0.001$, which suggests that H_0 is rejected at the conventional levels 0.05, 0.01.

j) [*3 marks*] Since $X_1, \ldots, X_n \overset{i.i.d}{\sim} \mathrm{Poiss}(\lambda)$, each indicator variable $I(X_j = 0)$ is Bernoulli with success probability $p = e^{-\lambda}$. So, $R = nT \sim \mathrm{Bin}(n, e^{-\lambda})$, as it is just a sum of i.i.d Bernoulli random variables. Hence,

$$\mathrm{E}(T) = \tfrac{1}{n}\mathrm{E}\left\{\sum_{j=1}^{n} I(X_j = 0)\right\} = \tfrac{1}{n}np = e^{-\lambda}$$

and (quickly) $\mathrm{var}(T) = \mathrm{var}(R/n) = n^{-2}\mathrm{var}(R) = np(1-p)/n^2 = e^{-\lambda}(1-e^{-\lambda})/n$, or (in more detail)

$$\mathrm{var}(T) = \tfrac{1}{n^2}\mathrm{var}\left\{\sum_{j=1}^{n} I(X_j = 0)\right\} = \tfrac{1}{n}\mathrm{var}\Big(I(X_1 = 0)\Big) = \frac{e^{-\lambda}(1-e^{-\lambda})}{n},$$

where the second equality holds because of the independence assumption. If the X_j's become dependent, the expectation of T remains the same because the expectation is a linear operator.

Ex. 2 **a)** [*2 marks*] Here is one possible formulation (we accept other equivalent ones). Let W_i be the Bernoulli random variable (with success probability 0.5) that takes the value 1 if it rains on the i-th day, and 0 otherwise. Given that the days are independent, the probability of seven consecutive days with no rain is $\Pr(W_i = 0)^7 = 0.5^7$.

b) [*4 marks*] Let $W \sim \text{Ber}(0.5)$ be the random variable indicating whether tomorrow is wet, and let A be a random variable representing the amount of rainfall. Using the hints given, the amount of rainfall tomorrow can be written as the product $R = WA$. We then have

$$\mathrm{E}(R) = \mathrm{E}\{\mathrm{E}(R \mid W)\} = \mathrm{E}\{\mathrm{E}(A)W\} = \mathrm{E}(8W) = 8\mathrm{E}(W) = 8 \cdot 0.5 = 4\text{mm}.$$

For the variance, first note that

$$\mathrm{var}(R \mid W) = \mathrm{var}(AW) = W^2\mathrm{var}(A) = 16W^2 = 16W,$$

where the second equality holds because $W^2 = W$, since W only takes the values 0 or 1. With this preliminary result, we have

$$\begin{aligned}\mathrm{var}(R) &= \mathrm{E}\{\mathrm{var}(R \mid W)\} + \mathrm{var}\{\mathrm{E}(R \mid W)\} = \mathrm{E}(16W) + \mathrm{var}\{\underbrace{\mathrm{E}(A)}_{=8}W\} \\ &= 16\mathrm{E}(W) + 8^2(0.5)(1 - 0.5) = 16 \cdot 0.5 + 16 = 24\text{mm}^2.\end{aligned}$$

Alternatively we note that $\mathrm{var}(AW) = \mathrm{E}\{(AW)^2\} - \mathrm{E}(AW)^2$, that independence of A and W gives $\mathrm{E}(AW)^2 = \{\mathrm{E}(A)\mathrm{E}(W)\}^2 = 16\text{mm}^2$, and that as $W^2 = W$,

$$\mathrm{E}\{(AW)^2\} = \mathrm{E}(WA^2) = \mathrm{E}(W)\mathrm{E}(A^2) = \mathrm{E}(W)\{\mathrm{var}(A) + \mathrm{E}(A)^2\} = 0.5 \times (4^2 + 8^2) = 40\text{mm}^2,$$

so $\mathrm{var}(R) = \mathrm{var}(AW) = 40 - 16 = 24\text{mm}^2$.

c) [*3 marks*] Let $R = R_1 + \cdots + R_{30}$ be the total amount of rainfall over 30 days, with R_i the amount of rainfall on day i. Using the central limit theorem (CLT), we can approximate R as a normal variable with mean $30 \cdot 4 = 120$ mm and variance $30 \cdot 24 = 720$ mm^2. Being careful to always have the same units (in mm or cm), we then have

$$\Pr(R > 160) = \Pr\left(\frac{R - 120}{\sqrt{720}} > \frac{160 - 120}{\sqrt{720}}\right) = 1 - \Phi(1.491) \approx 0.068.$$

Ex. 3 **a)** [*4 marks*] Each mark should be reserved for

i) the description of the whiskers: the maximum is higher and the minimum is lower in group 1

ii) the description of the inter-quartile range (IQR): group 1 has a higher IQR

iii) the description of the variability of both groups: group 1 has higher variability while group 2's marks are more concentrated

iv) the median: group 1 has a slightly lower median.

b) [*4 marks*] 2 marks should be reserved for the description of group 1's Q-Q plot. The graph is not close to a straight line, which seems to suggest that the observations may not be well fitted by a normal model. The slope and the intercept of the best fitted straight line at $x = 0$ give estimates of σ and μ respectively, which should be close to 1.06 and 3.66. The Q-Q plot drawn for group 2 should have a best fitted straight line with a slope of approximately 0.75 and an intercept of 4 at $x = 0$.

c) [*3 marks*] Using the assumptions given in the question, we have

$$\sigma^2_{\text{diff}} = \text{var}(\overline{x}_1 - \overline{x}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{1.06^2}{100} + \frac{0.75^2}{80} \approx 0.018267 \approx 0.135^2$$

Since the estimated mean difference in marks is $3.66 - 4.05 = -0.39$, an approximate 95% confidence interval for the mean difference in marks is

$$-0.39 \pm z_{0.975} \cdot \sigma_{\text{diff}} = [-0.39 - 1.96 \cdot 0.135, -0.39 + 1.96 \cdot 0.135] = [-0.645, -0.125],$$

where $z_{0.975}$ is the 97.5% quantile of the standard normal distribution. The 95% confidence interval for the mean difference in marks does not contain 0, so we reject the null hypothesis of equal group means at the 5% level.

d) [*3 marks*] The assumptions for the calculation in (b) are between-group independence (to argue that $\text{var}(\overline{x}_1 - \overline{x}_2) = s_1^2/n_1 + s_2^2/n_2$) and within-group independence, so that a normal approximation from the CLT applies to the group means. If every mark from the exam is independent, both of these assumptions will hold. This is not an unrealistic assumption for exam marks, which clearly have a finite mean and variance.

[A very critical student might question from what populations the two groups are drawn: is this test result applicable only to these particular students, and if so, what does it tell us more generally, if anything? If it applies only to these students, then what CLT applies (there are no larger population parameters for the averages and sample variances to tend to ...)? Good questions for which a bonus should be given.]

Ex. 4 **a)** [*4 marks*] A mark should only be given in each subsection if the answer is correct (appropriate/not appropriate) and a correct justification is given.

i) The uniform distribution is not appropriate since its support is finite (the possible values it can take are bounded).

ii) The normal distribution is not appropriate since negative values are possible with the normal distribution while λ should always be positive.

iii) The gamma distribution is a good choice, since its' support is positive and infinite. Furthermore, it is the conjugate prior for the Poisson likelihood, which makes Bayesian computation for the posterior easier.

iv) The Bernoulli distribution is a discrete random variable that is not appropriate since the support of the distribution is $\{0, 1\}$.

b) [*3 marks*] Since $\pi(\lambda)$ is an exponential density function with parameter c, the prior mean of λ is $\frac{1}{c}$. Full marks can be given if this result for the mean of the exponential density is correctly quoted. Otherwise, the full derivation of the mean should be given, using integration by parts:

$$\mathrm{E}(\lambda) = \int_0^\infty \lambda \cdot c \exp(-c\lambda) \mathrm{d}\lambda = [-\lambda \exp(-c\lambda)]_0^\infty + \int_0^\infty \exp(-c\lambda) \mathrm{d}\lambda = 0 + [-\exp(-c\lambda)/c]_0^\infty = 1/c.$$

From the information given in the question, we know that the weekly crash rate based on prior knowledge is $\frac{1}{2}$ (once in two weeks). Thus a suitable value for c is $c = 2$.

c) [*3 marks*] There are many equivalent approaches. One is to directly calculate the posterior density

$$\begin{aligned}
\pi(\lambda \mid N = k) &= \frac{\Pr(N=k|\lambda) \cdot \pi(\lambda)}{\Pr(N=k)} \\
&= \frac{\Pr(N=k|\lambda) \cdot \pi(\lambda)}{\int \Pr(N=k) \pi(\lambda) \mathrm{d}\lambda} \\
&= \frac{2 \cdot \frac{\lambda^k}{k!} \cdot \exp(-\lambda) \cdot \exp(-2\lambda)}{\int_0^\infty 2 \cdot \frac{\lambda^k}{k!} \cdot \exp(-\lambda) \cdot \exp(-2\lambda) \mathrm{d}\lambda} \\
&= \frac{\frac{\lambda^k}{k!} \cdot \exp(-3\lambda)}{\left(\frac{1}{3}\right)^{k+1} \int_0^\infty 3^{k+1} \frac{\lambda^{(k+1)-1}}{\Gamma(k+1)} \cdot \exp(-3\lambda) \mathrm{d}\lambda} \\
&= 3^{k+1} \cdot \frac{\lambda^{(k+1)-1}}{\Gamma(k+1)} \cdot \exp(-3\lambda), \quad \lambda > 0.
\end{aligned}$$

When $k = 3$, this is the density for the Gamma(4,3) distribution. An alternative approach is to notice that the posterior density is

$$\pi_\lambda(\lambda \mid N = k) \propto \pi_\lambda(\lambda) \Pr(N = k \mid \lambda) \propto e^{-c\lambda} \lambda^k e^{-\lambda} = \lambda^{(k+1)-1} e^{-(c+1)\lambda}, \quad \lambda > 0. \qquad (1)$$

Since this has to integrate to 1 for all $k \geq 0$ and $c > 0$, the constant of proportionality should be $\frac{(c+1)^{(k+1)}}{\Gamma(k+1)}$, so the posterior density is for the Gamma($k+1$,$c+1$) distribution, with $k = 3$ and $c = 2$. This gives us a Gamma(4,3) posterior density.

d) [*3 marks*] The MAP estimate is the mode of the posterior density. To find the mode we could differentiate the logarithm of (1) with respect to λ and equate the derivative to 0. This gives

$$k/\widehat{\lambda}_{MAP} - (c + 1) = 0,$$

so $\widehat{\lambda}_{MAP} = k/(c + 1) = 3/(2 + 1) = 1$. The second derivative is negative, so this is the (unique) maximum. Hence crashes have become more frequent, since the rate λ has increased from $\frac{1}{2}$ to 1.

Alternatively the mode of the Gamma(α, β) distribution is $\frac{\alpha-1}{\beta}$, for $\alpha \geq 1$. Thus, since we know from (b) that the posterior density of λ is a Gamma(4,3) density, we get $\widehat{\lambda}_{MAP} = 1$.