Solutions to 2006 First Year Exam

January 29, 2008

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

1. (a) i.

$$\begin{aligned} f(x) &= \int_0^{4-2x} (4-2x-y)dy \\ &= \frac{3}{16}\left[(4-2x)y|_0^{4-2x} - \frac{1}{2}y^2|_0^{4-2x}\right] \\ &= \frac{3}{8}(2-x)^2 \\ \Rightarrow f(y|x) &= \frac{f(y,x)}{f(x)} \\ &= \frac{1}{2-x} - \frac{y}{x(2-x)^2} \text{ for } y \in [0, 4-2x] \end{aligned}$$

ii.

$$\begin{aligned} Pr(Y \geq 2|X = .5) &= \int_2^3 \left(\frac{1}{1.5} - \frac{y}{2(1.5^2)}\right) dy \\ &= \frac{1}{1.5} - \frac{1}{2 \times 1.5^2}.5y^2|_2^3 \\ &= \frac{1}{9} \end{aligned}$$

(b) Solving for x in terms of y gives,

$$x = \begin{cases} 1 - \sqrt{1-y} & \text{if } 0 < x < 1 \\ 1 + \sqrt{1-y} & \text{if } 1 < x \leq 2 \end{cases}$$

In either case the Jacobian is $1/\sqrt{1-y}$. Now using Theorem 2.1.8 of Casella and Berger,

$$\begin{aligned} f(y) &= f_x(1 - \sqrt{(1-y)})\frac{1}{2\sqrt{1-y}} + f_x(1 + \sqrt{1-y})\frac{1}{2\sqrt{1-y}} \\ &= \frac{1}{2\sqrt{1-y}} \end{aligned}$$

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

2. (a) In order to find the minimum value of n for which rejecting is possible, set your test statistic (x) to be as extreme as possible. Specifically, solve

$$\begin{aligned} \sup_{\theta \in H_0} \binom{n}{n} \theta^n (1-\theta)^0 &< \alpha \\ \Rightarrow \binom{n}{n} 0.5^n &< 0.01 \\ \Rightarrow 0.5^n &< 0.01 \\ \Rightarrow n &> 6 \end{aligned}$$

(b) Recalling the definition of p-value to be the probability under H_0 of observing a test statistic as or more extreme than the observed value, the p-value is then,

$$\begin{aligned} P(x) &= \binom{10}{9} 0.5^9 (1-0.5)^1 + \binom{10}{10} 0.5^{10} (1-0.5)^0 \\ &= \frac{10}{2^{10}} + \frac{1}{2^{10}} \\ &= \frac{11}{2^{10}} \\ &> 0.01 \\ &\Rightarrow \text{Fail to Reject } H_0 \end{aligned}$$

(c) It is trivial to show that for $\theta \sim Unif(0,1)$, the posterior distribution of $\theta | x = 5, n = 5 \sim Beta(6,1)$. Therefore, the posterior probability of H_0 is then,

$$\begin{aligned} Pr(\theta \leq 0.5 | X = 5, n = 5) &= \int_0^{0.5} \frac{\Gamma(7)}{\Gamma(6)\Gamma(1)} \theta^5 d\theta \\ &= 6 \times \frac{1}{6} \theta^6 |_0^{0.5} \\ &= \frac{1}{2^6} \end{aligned}$$

(d) Finding a one-sided *confidence* interval is non-trivial but details can be seen in Casella and Berger Example 9.2.5. An alternative would be to find the Bayesian 95% *credible* interval as

$$\begin{aligned} \int_L^1 \frac{\Gamma(7)}{\Gamma(6)\Gamma(1)} \theta^5 &= 0.95 \\ \Rightarrow L &= 0.05^{1/6} \end{aligned}$$

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

3. (a)

$$\begin{aligned} Pr(X < Y) &= \int_0^2 \int_x^\infty \frac{1}{2} e^{-y} dy dx \\ &= \int_0^2 \frac{1}{2} e^{-x} dx \\ &= \frac{1}{2}(-e^{-x})|_0^2 \\ &= \frac{(1-e^{-2})}{2} \end{aligned}$$

(b) Notice that when $Z = 1.0$ the space is restricted to the line $Y = 1.0$ if $Y > X$ or the line $X = 1.0$ if $Y < X$. Since these regions are on sets of measure zero, we must take a limit.

$$\begin{aligned} Pr(X < Y|Z = 1.0) &= \lim_{\epsilon\to 0} \frac{\int_1^{1+\epsilon}\int_0^1 \frac{1}{2}e^{-y}dxdy}{\int_1^{1+\epsilon}\int_0^1 \frac{1}{2}e^{-y}dxdy + \int_1^{1+\epsilon}\int_0^1 \frac{1}{2}e^{-y}dydx} \\ &= \lim_{\epsilon\to 0} \frac{\int_1^{1+\epsilon} \frac{1}{2}e^{-y}dy}{\int_1^{1+\epsilon} \frac{1}{2}e^{-y}dy + \int_1^{1+\epsilon} \frac{1}{2}(1-e^{-1})dx} \\ &= \lim_{\epsilon\to 0} \frac{e^{-1} - e^{-(1+\epsilon)}}{e^{-1} - e^{-(1+\epsilon)} + (1-e^{-1})\epsilon} \\ &= \lim_{\epsilon\to 0} \frac{e^{-(1+\epsilon)}}{e^{-(1+\epsilon)} + 1 - e^{-1}} \text{ by L'Hopital's rule} \\ &= e^{-1}. \end{aligned}$$

(c) Consider the p.d.f. of $Z|X = 1.0$ given by,

$$f_Z(z) = \begin{cases} 0 & \text{if } z < 1 \\ \int_0^1 e^{-y}dy & \text{if } z = 1 \\ e^{-z} & \text{if } z > 1 \end{cases}$$

Now integrate the above p.d.f. with respect to Z for each case separately. First, if $z < 1$ then the c.d.f. is 0 because the maximum can not be less than 1. Second, if $z = 1$ there is simply a point mass because this can only occur if $Y < X$. Finally, if $z > 1$ then the c.d.f. of Z must include the point mass at 1 as well as integrate the p.d.f. of Y from 1 to z. In conclusion, the c.d.f. of Z is defined by,

$$\begin{aligned} F_Z(z) &= \begin{cases} 0 & \text{if } z < 1 \\ \int_0^1 e^{-y}dy & \text{if } z = 1 \\ \int_0^1 e^{-y}dy + \int_0^z e^{-t}dt & \text{if } z > 1 \end{cases} \\ &= \begin{cases} 0 & \text{if } z < 1 \\ 1 - e^{-1} & \text{if } z = 1 \\ 1 - e^{-z} & \text{if } z > 1 \end{cases} \end{aligned}$$

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

4. (a)

$$E(e^{tX}) = \int_0^\infty e^{tx}\lambda e^{-\lambda x} = -\frac{\lambda e^{-x(\lambda-t)}}{\lambda - t}|_0^\infty = \frac{\lambda}{\lambda - t}$$

(b)

$$E(e^{tX}) = e^t\theta + e^0(1-\theta) = \theta e^t + (1-\theta)$$

(c) Let $Z_1, \ldots, Z_n$ be independent $Ber(\theta)$ trials. Then $X = \sum_i Z_i$. Hence,

$$\begin{aligned} E(e^{tX}) &= E(e^{t\sum_i Z_i}) \\ &= E(e^{tZ_1}\cdots e^{tZ_n}) \\ &= E(e^{tY_1})\cdots E(e^{tY_n}) \text{ (by independence)} \\ &= \left[\theta e^t + (1-\theta)\right]^n \end{aligned}$$

(d)

$$\begin{aligned} E(e^{tY}) &= \sum_{y=0}^\infty \frac{e^{-\lambda}\lambda^y e^{tY}}{y!} \\ &= e^{-\lambda}\sum_{y=0}^\infty \frac{(e^t\lambda)^y}{y!} \\ &= e^{-\lambda}e^{e^t\lambda} \\ &= \exp\{\lambda(e^t - 1)\} \end{aligned}$$

(e)

$$\begin{aligned} \lim_{n\to\infty}[(c/n)e^t + (1-(c/n))]^n &= \lim_{n\to\infty}\left[1 + \frac{c(e^t-1)}{n}\right]^n \\ &= e^{c(e^t-1)} \end{aligned}$$

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

5. (a) Notice that this is a mixture of 2 exponential distributions. The first component is an $Ex(\mu)$ which occurs with probability 0.75. The second component is an $Ex(10\mu)$ which should occur with probability 0.25. Therefore, $c = 10 \times 0.25 = 2.5$. Recognizing the form of $p(t|\mu)$ is much easier than trying to integrate the whole thing and solving for c (which you can do if you want).

   (b) This model places 0.75 probability of sampling from an $Ex(\mu)$ distribution and 0.25 probability of sampling from $Ex(10\mu)$. This is consistent with the problem description which states that type A is 3 times as common as type B. Therefore, this model captures all information given in the problem.

   (c)

$$\begin{aligned}
p(\mu|t) &= \frac{p(t|\mu)p(\mu)}{p(t)} \text{ (by Bayes rule)} \\
&= \frac{0.75\mu \exp\{-\mu t\} A\mu^{a-1} \exp\{-am\mu\}}{p(t)} + \frac{0.25 \times 10\mu \exp\{-10\mu t\} A\mu^{a-1} \exp\{-am\mu\}}{p(t)} \\
&= c_1(t,a,m)\mu^a \exp\{-\mu(t+am)\} + c_2(t,a,m)\mu^a \exp\{-\mu(10t+am)\} \\
&= q_1(t,a,m)p_A(\mu|t) + q_2(t,a,m)p_B(\mu|t)
\end{aligned}$$

   where $p_A(\mu|t)$ and $p_B(\mu|t)$ are two gamma densities. Now, because $p(\mu|t)$ is also a density, it must integrate to 1. Furthermore, because each of the gamma densities integrate to 1, $q_1(t,a,m) + q_2(t,a,m) = 1$ which shows $q_1(t,a,m) = 1 - q_2(t,a,m)$. In conclusion,

$$p(\mu|t) = q(t,a,m)p_A(\mu|t) + (1 - q(t,a,m))p_B(\mu|t)$$

   (d) $q(t)$ is then the posterior probability that the source of particles produces particles of type A.

   (e)

$$\begin{aligned}
E(\mu|t) &= E(q(t,a,m)p_A(\mu|t) + (1 - q(t,a,m))p_B(\mu|t)) \\
&= q(t,a,m)E(p_A(\mu|t)) + (1 - q(t,a,m))E(p_B(\mu|t)) \\
&= q(t,a,m)\frac{a+1}{t+am} + (1 - q(t,a,m))\frac{a+1}{10t+am}
\end{aligned}$$

These solutions were prepared by students studying to take the 2008 FYE. These solutions have *not* been reviewed by faculty members. Therefore, the solutions presented may contain errors. Use at your own risk.

6. (a) The problem states that β follows a multivariate normal distribution. Therefore, according to the theory of the multivariate normal distribution, any linear combination of a normal random variable also has a normal distribution resulting in μ have a normal distribution with moments,

$$\begin{aligned} E(\mu|Y) &= E(X\beta|Y) \\ \tilde{\mu} &= XE(\beta|Y) \\ &= \frac{g}{1+g}X(X'X)^{-1}X'Y \\ V(\mu|Y) &= XV(\beta)X' \\ \Sigma &= \frac{g}{1+g}X(X'X)^{-1}X' \end{aligned}$$

(b) No, because $E(E(\mu|Y)) = E(\frac{g}{1+g}X(X'X)^{-1}X'Y) = \frac{g}{1+g}XB$. The g screws up the unbiasedness.

(c) From above, but with $X'X = I$, we have

$$\begin{aligned} Var(\mu|Y) &= \frac{g}{1+g}X(X'X)^{-1}X' \\ &= \frac{g}{1+g}XX' \end{aligned}$$

which is not a diagonal matrix and, hence, μ_j and μ_k are *not* independent.

(d) By the expected value of a quadratic form we calculate,

$$\begin{aligned} E((\mu-\tilde{\mu})'(\mu-\tilde{\mu})|Y) &= \text{tr}(\Sigma) + E(\mu-\tilde{\mu})'E(\mu-\tilde{\mu}) \\ &= \text{tr}(\Sigma) \\ &= \frac{g}{1+g}\text{tr}((X'X)^{-1}X'X) \\ &= \frac{gp}{1+g} \end{aligned}$$

where Σ is as given above, but that the vector is centered, so has an expectation of zero.