CSC 591 HW5

Answer Key

December 7, 2015

Q1

(a)

(1)

$$
\begin{aligned}
f(x \mid p) &= \binom{2}{x} p^x (1-p)^{2-x} \\
f(p \mid x) &= \frac{f(p)p(x \mid p)}{f(x)} = \frac{f(p)p(x \mid p)}{\sum_{p'} f(p')f(x \mid p')} = \frac{f(p)p(x \mid p)}{0.7f(x \mid p=0.2) + 0.3f(x \mid p=0.3)} \\
&= \frac{\binom{2}{x} f(p)p^x(1-p)^{2-x}}{\binom{2}{x}(0.7 \times 0.2^x \times 0.8^{2-x} + 0.3 \times 0.3^x \times 0.7^{2-x})} \\
&= \frac{f(p)p^x(1-p)^{2-x}}{0.7 \times 0.2^x \times 0.8^{2-x} + 0.3 \times 0.3^x \times 0.7^{2-x}}
\end{aligned}
$$

(2)

$$
\begin{aligned}
f(p \mid x=2) &= \frac{f(p)p^2(1-p)^0}{0.7 \times 0.2^2 \times 0.8^0 + 0.3 \times 0.3^2 \times 0.7^0} \\
&= \frac{f(p)p^2}{0.055} \\
f(0.2 \mid x=2) &= \frac{0.7 * 0.2^2}{0.055} = 0.509 \\
f(0.3 \mid x=2) &= 1 - f(0.2 \mid x=2) = 1 - 0.509 = 0.491
\end{aligned}
$$

(b)

Let $\bar{x}$ be the mean acceleration of a sample of 10 cars.

$$
\begin{aligned}
P(\bar{x} \mid \mu) &= \mathcal{N}(\bar{x}; \mu, 0.8^2/10) \\
P(\mu \mid \bar{x}) &= \frac{P(\mu)P(X \mid \mu)}{P(\bar{x})} \\
&\propto \mathcal{N}(\mu; 8, 0.2)\mathcal{N}(\bar{x}; \mu, 0.8^2/10) \\
&= \mathcal{N}(\mu; 8, 0.2)\mathcal{N}(\mu; \bar{x}, 0.8^2/10) \\
&\propto \mathcal{N}\left(\mu; \left(\frac{1}{0.2} + \frac{1}{0.8^2/10}\right)^{-1}\left(\frac{8}{0.2} + \frac{\bar{x}}{0.8^2/10}\right), \left(\frac{1}{0.2} + \frac{1}{0.8^2/10}\right)^{-1}\right) \\
&= \mathcal{N}\left(\mu; 1.9394 + 0.7576\bar{x}, 0.0485\right)
\end{aligned}
$$

For $\bar{x} = 9$, this gives us a posterior mean of 8.7578 and standard deviation of $\sqrt{0.0485} = 0.2202$. Using a z table we see that the middle 95% of the standard normal distribution falls between $z = \pm 1.96$, so our 95% Bayesian interval is given by $\mu \pm 1.96\sigma = 8.7578 \pm 1.96 \times 0.2202 = [8.3262, 9.1894]$.

Q2

(1)

The missing data follows a univariate pattern because the only variable that is missing for some samples is "mt2".

(2)

MT1, complete:

$$\text{Mean} = \frac{\sum_i \text{mt1}_i}{10} = \frac{508}{10} = 50.8$$
$$\text{S.E.} = \sqrt{\frac{\sum_i (\text{mt1}_i - 50.8)^2}{10}} = 14.9$$

MT2, complete:

$$\text{Mean} = \frac{\sum_i \text{mt2}_i}{10} = \frac{507}{10} = 50.7$$
$$\text{S.E.} = \sqrt{\frac{\sum_i (\text{mt2}_i - 50.7)^2}{10}} = 12.7$$

MT1, missing:

$$\text{Mean} = \frac{74 + 70 + 66 + 55 + 52 + 47 + 33}{7} = 56.7$$
$$\text{S.E.} = 13.3$$

MT2, missing:

$$\text{Mean} = \frac{66 + 58 + 74 + 47 + 61 + 38 + 41}{7} = 55$$
$$\text{S.E.} = 12.3$$

(3)

It appears that MT2 values are either MNAR (with lower values missing) or possibly MAR (missing for samples with a lower value of MT1). It does not seem likely that the data is MCAR, so the estimates obtained using list-wise deletion (part (ii)) will be biased.

(4)

$$\text{MT2} = 0.715\text{MT1} + 14.452$$

The predicted values of MT2 for MT1 = 45, 38 and 28 are 46.63, 41.62 and 34.47, respectively.

(5)

The mean of MT2 including the imputed values is 50.77, and the (uncorrected) standard error is 12.52. The mean is an unbiased estimator while the S.E. is biased.

Q3

(a)

Patterns: univariate, unit nonresponse, monotone, general, planned missing, latent variable

Mechanisms: MCAR, MAR, MNAR

For details, see W15C1 slides.

(b)

See W15C1 slides.

Q4

Part 1

Independence	Holds?	Active Path
$A \perp C$	Yes	
$A \perp C \mid E$	Yes	
$A \perp C \mid I$	No	A–D–G–C
$D \perp I$	No	D–G–I
$D \perp I \mid G$	No	D–F–I
$D \perp I \mid F, G, J$	No	D–A–E–J–I
$D \perp I \mid F, G, J, A$	Yes	
$F \perp H$	No	F–D–H
$F \perp H \mid A$	No	A–D–G–C
$F \perp H \mid D$	Yes	
$F \perp H \mid D, J$	No	F–I–J–H
$F \perp H \mid D, I, J$	Yes	

Part 2

(a)

$$
\begin{aligned}
P(A=1, B=0, C=0, D=1, E=0, F=1) =& P(A=1)P(B=0)P(C=0 \mid A=1)P(D=1 \mid A=1, B=0) \\
& P(E=0 \mid C=0, D=1)P(F=1 \mid E=0) \\
=& 0.7 \times 0.6 \times 0.3 \times 0.6 \times 0.4 \times 0.9 \\
=& 0.0272
\end{aligned}
$$

(b)

$$
\begin{aligned}
P(A=1,E=0) &= \sum_b \sum_c \sum_d \sum_f P(A=1,b,c,d,E=0,f) \\
&= \sum_b \sum_c \sum_d \sum_f P(A=1)P(b)P(c \mid A=1)P(d \mid A=1,b)P(E=0 \mid c,d)P(f \mid E=0) \\
&= \sum_b \sum_c \sum_d P(A=1)P(b)P(c \mid A=1)P(d \mid A=1,b)P(E=0 \mid c,d) \sum_f P(f \mid E=0) \\
&= \sum_b \sum_c \sum_d P(A=1)P(b)P(c \mid A=1)P(d \mid A=1,b)P(E=0 \mid c,d) \\
&= \sum_b \sum_c P(A=1)P(b)P(c \mid A=1) \sum_d P(d \mid A=1,b)P(E=0 \mid c,d) \\
&= \sum_b P(A=1)P(b) \sum_c P(c \mid A=1) \sum_d P(d \mid A=1,b)P(E=0 \mid c,d) \\
&= P(A=1) \sum_b P(b) \sum_c P(c \mid A=1) \sum_d P(d \mid A=1,b)P(E=0 \mid c,d) \\
&= P(A=1) \sum_b P(b) \sum_c P(c \mid A=1) \times \\
&\quad [P(D=0 \mid A=1,b)P(E=0 \mid c,D=0) + P(D=1 \mid A=1,b)P(E=0 \mid c,D=1)] \\
&= P(A=1) \sum_b P(b)[P(C=0 \mid A=1)[P(D=0 \mid A=1,b)P(E=0 \mid C=0,D=0)+ \\
&\quad P(D=1 \mid A=1,b)P(E=0 \mid C=0,D=1)]+ \\
&\quad P(C=1 \mid A=1)[P(D=0 \mid A=1,b)P(E=0 \mid C=1,D=0)+ \\
&\quad P(D=1 \mid A=1,b)P(E=0 \mid C=1,D=1)]] \\
&= 0.7 \sum_b P(b)[0.3[0.2P(D=0 \mid A=1,b) + 0.4P(D=1 \mid A=1,b)]+ \\
&\quad 0.7[0.9P(D=0 \mid A=1,b) + 0.5P(D=1 \mid A=1,b)]] \\
&= 0.7P(B=0)[0.3[0.2P(D=0 \mid A=1,B=0) + 0.4P(D=1 \mid A=1,B=0)]+ \\
&\quad 0.7[0.9P(D=0 \mid A=1,B=0) + 0.5P(D=1 \mid A=1,B=0)]]+ \\
&\quad 0.7P(B=1)[0.3[0.2P(D=0 \mid A=1,B=1) + 0.4P(D=1 \mid A=1,B=1)]+ \\
&\quad 0.7[0.9P(D=0 \mid A=1,B=1) + 0.5P(D=1 \mid A=1,B=1)]] \\
&= 0.7 \times 0.6[0.3[0.2 \times 0.4 + 0.4 \times 0.6] + 0.7[0.9 \times 0.4 + 0.5 \times 0.6]]+ \\
&\quad 0.7 \times 0.4[0.3[0.2 \times 0.8 + 0.4 \times 0.2] + 0.7[0.9 \times 0.8 + 0.5 \times 0.2]] \\
&= 0.415
\end{aligned}
$$