HW5: Bayesian Inference, Missing Data Analysis

Q1. Bayesian Inferencing (30 points)

(a) *Following table give prior distribution for the proportion of defective parts produced by a machine*

p (proportion of defects)	0.2	0.3
f(p) : prior probability	0.7	0.3

Let x denote the number of defectives among a random sample of size 2.

(1) *Find the posterior probability distribution of p, given that x is observed.* (***10 points)***

(Answer)
The random variable X follows a binomial distribution.

Marginal distribution is computed as follows:
$f(x \mid 0.2)\, f(0.2) + f(x \mid 0.3)\, f(0.3)$ -> (1)

Since random variable X follows a binomial distribution,
$f(x \mid p) = b(x;2;p) = {}^2C_x\, p^x\, q^{2-x}$ where x = 0, 1, 2 -> (2)

Substituting (1) in (2) we get,
$f(x \mid 0.2)\, f(0.2) + f(x \mid 0.3)\, f(0.3) = {}^2C_x\, (0.2)^x\, (0.8)^{2-x}\, (0.7) + {}^2C_x\, (0.3)^x\, (0.7)^{2-x}\, (0.3)$
$= {}^2C_x\, [(0.2)^x\, (0.8)^{2-x}\, (0.7) + (0.3)^x\, (0.7)^{2-x}\, (0.3)]$ -> (3)

Substituting the value x = 0, 1, 2 in 3 we get,

$$
\begin{aligned}
F(0) &= {}^2C_0\, [(0.2)^0\, (0.8)^{2-0}\, (0.7) + (0.3)^0\, (0.7)^{2-0}\, (0.3)] \\
&= 1[(0.8)^2\, (0.7) + (0.7)^2 (0.3)] \\
&= (0.7)\, [0.64 + 0.21] \\
&= (0.7)\, (0.85) \\
&= \mathbf{0.595}
\end{aligned}
$$

$$
\begin{aligned}
F(1) &= {}^2C_1\, [(0.2)^1\, (0.8)^{2-1}\, (0.7) + (0.3)^1\, (0.7)^{2-1}\, (0.3)] \\
&= 2[(0.2)\, (0.8)\, (0.7) + ((0.09)(0.7)] \\
&= (1.4)\, [0.16 + 0.09] \\
&= (1.4)\, (0.25) \\
&= \mathbf{0.35}
\end{aligned}
$$

F(2) $= {}^2C_2\ [(0.2)^2\ (0.8)^{2-2}\ (0.7) + (0.3)^2\ (0.7)^{2-2}\ (0.3)]$
$= 1\ [(0.2)^2\ (0.7) + ((0.09)(0.3)]$
$= (1)\ [0.028 + 0.027]$
$= (1)\ (0.055)$
= **0.055**

The marginal distribution is summarized as below:

x	0	1	2
F(x)	0.595	0.35	0.055

Now that we have calculated the marginal distribution, we can write the formula to compute the posterior probability for p = 0.2 and p = 0.3

$$f(0.2 \mid x) = \frac{f(x \mid 0.2)\ f(0.2)}{{}^2C_x\ [(0.2)^x\ (0.8)^{2-x}\ (0.7) + (0.3)^x\ (0.7)^{2-x}\ (0.3)]}$$

$$f(0.3 \mid x) = 1 - f(0.2 \mid x)$$

If x = 0, then

$$f(0.2 \mid 0) = \frac{{}^2C_0\ (0.2)^0\ (0.8)^{2-0}\ (0.7)}{{}^2C_0\ [(0.2)^0\ (0.8)^{2-0}\ (0.7) + (0.3)^0\ (0.7)^{2-0}\ (0.3)]}$$

$f(0.2 \mid 0) = (1)\ (0.8)^2\ (0.7) / 0.595 = 0.448 / 0.595$

$f(0.2 \mid 0)$ = **0.753**

$f(0.3 \mid 0) = 1 - f(0.2 \mid 0) = 1 - 0.753$

$f(0.3 \mid 0)$ = **0.247**

If x = 1, then

$$f(0.2 \mid 1) = \frac{{}^2C_1\ (0.2)^1\ (0.8)^{2-1}\ (0.7)}{{}^2C_0\ [(0.2)^0\ (0.8)^{2-0}\ (0.7) + (0.3)^0\ (0.7)^{2-0}\ (0.3)]}$$

$f(0.2 \mid 1) = (2)\ (0.2)\ (0.8)\ (0.7) / 0.595 = 0.224 / 0.595$

$f(0.2 \mid 1)$ = **0.376**

$f(0.3 \mid 1) = 1 - f(0.2 \mid 1) = 1 - 0.376$

$f(0.3 \mid 1)$ = **0.624**

(2)** Estimate the proportion of defectives being produced by the machine if the random sample of size 2 yields 2 defects. **(10 points)

$$f(0.2 \mid 2) = \frac{{}^{2}C_{2}\,(0.2)^{2}\,(0.8)^{2-2}\,(0.7)}{{}^{2}C_{0}\,[(0.2)^{0}\,(0.8)^{2-0}\,(0.7) + (0.3)^{0}\,(0.7)^{2-0}\,(0.3)]}$$

$f(0.2 \mid 2)$ = (1) $(0.2)^2$ (0.7) / 0.595 = 0.028 / 0.595

$f(0.2 \mid 2)$ = **0.047**

$f(0.3 \mid 2)$ = 1 - $f(0.2 \mid 2)$ = 1 - 0.047

$f(0.3 \mid 2)$ = **0.953**

Hence, the proportion of defectives being produced by the machine is 0.3.

Q2. The following table summarizes two exam scores. Left half of the table gives complete scores and right half gives an example of missing data. (35 points)

Complete Data		Missing Data	
mt1	**mt2**	**mt1**	**mt2**
74	66	74	66
70	58	70	58
66	74	66	74
55	47	55	47
52	61	52	61
47	38	47	38
45	32	45	-
38	46	38	-
33	41	33	41
28	44	28	-

(1) Based on the missing data, determine missing data pattern and justify your answer (5 points)

The missing data pattern resembles **Univariate** pattern because the missing values are isolated to a single variable.

(2) Compute Mean and Standard Error (SE) for (i) complete data, and (ii) missing data using list-wise deletion. (10 points)

(Answer)
(i) Complete data

mt1		
x	**x - mean(x)**	**(x - mean(x))**
74	23.2	538.24
70	19.2	368.64
66	15.2	231.04
55	4.2	17.64
52	1.2	1.44000000000001
47	-3.8	14.44
45	-5.8	33.64
38	-12.8	163.84
33	-17.8	316.84
28	-22.8	519.84

mean(mt1) = $(\sum_i x_i) / n$ = 508 / 10 = 50.8

standard error s = $\sqrt{[\sum(x - mean(x))^2 / (n - 1)]}$
s = $\sqrt{(2205.6)}/(10-1)$
= $\sqrt{245.067}$
= **15.65**

mt2		
x	**x - mean(x)**	**(x - mean(x))**
66	15.3	234.09
58	7.3	53.29
74	23.3	542.89
47	-3.7	13.69
61	10.3	106.09
38	-12.7	161.29
32	-18.7	349.69

mt2		
x	**x - mean(x)**	**(x - mean(x))**
46	-4.7	22.09
41	-9.7	94.0900000000001
44	-6.7	44.89

mean(mt2) = $(\sum_i x_i) / n$ = 507 / 10 = 50.7

standard error s = $\sqrt{[\sum (x - mean(x))^2 / (n - 1)]}$
s = $\sqrt{(1622.1)/(10-1)}$
= $\sqrt{180.23}$
= **13.42**

(ii) Missing data (list-wise deletion)

mt1		
x	**x - mean(x)**	**(x - mean(x))**
74	17.29	298.9441
70	13.29	176.6241
66	9.29	86.3041
55	-1.71	2.9241
52	-4.71	22.1841
47	-9.71	94.2841
33	-23.71	562.1641

mean(mt1) = $(\sum_i x_i) / n$ = 397 / 7 = 56.71

standard error s = $\sqrt{[\sum (x - mean(x))^2 / (n - 1)]}$
s = $\sqrt{(1243.43)/(7-1)}$
= $\sqrt{207.24}$
= **14.40**

mt2		
x	**x - mean(x)**	**(x - mean(x))**
66	11	121
58	3	9
74	19	361
47	-8	64
61	6	36
38	-17	289
41	-14	196

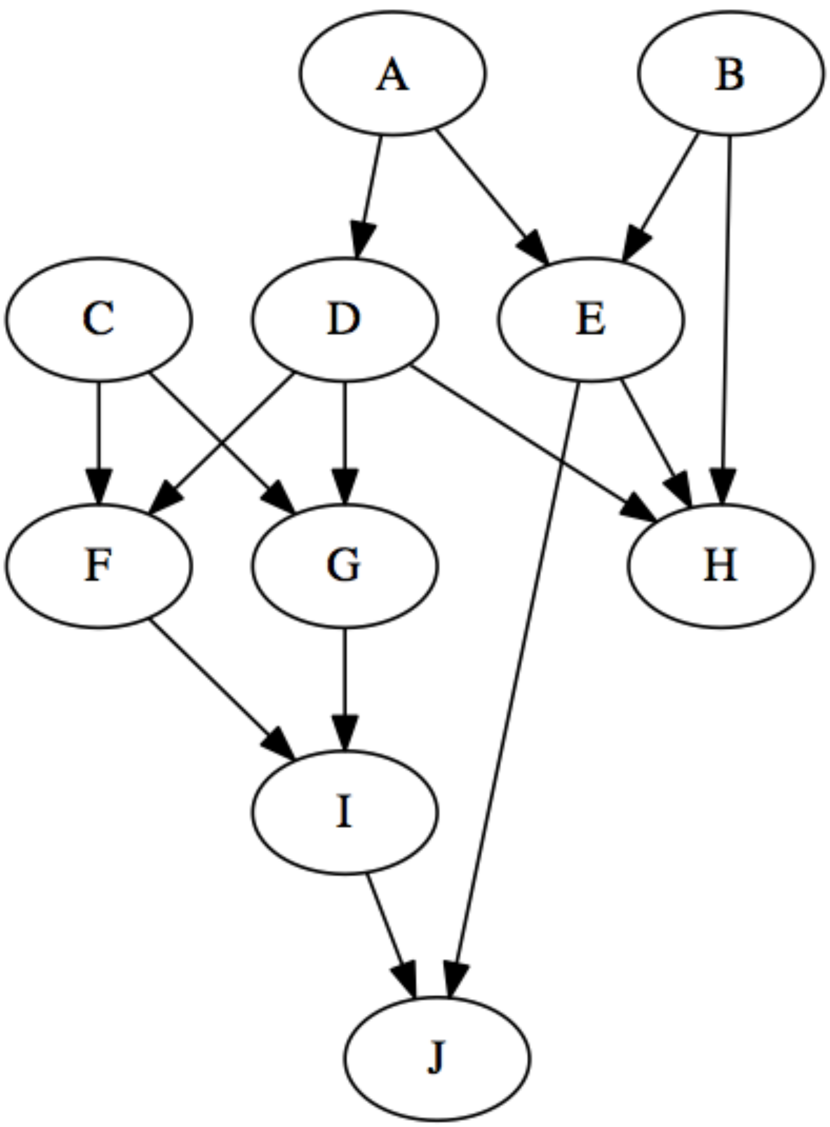
mean(mt2) = $(\sum_i x_i) / n = 385 / 7 = 55$

standard error $s = \sqrt{[\sum (x - mean(x))^2 / (n - 1)]}$
$s = \sqrt{(1076)/(7-1)} = \sqrt{179.33}$ = **13.39**

(3) Comment on bias of the estimates of (2.ii) as compared to estimates from complete data (2.i). (5 points)

	mt1		mt2	
	Complete Data	**Missing Data**	**Complete Data**	**Missing Data**
Mean	50.8	56.71	50.7	55
SE	15.65	14.40	13.42	13.39

Q4. Bayes Networks (25 points)
(a) *For each independence statement, state whether the independence is implied by the given Bayes net. If there are any active trails between the given variables, name one.*



(Answer)
A ⊥ C ? d-separated, no active trails
A ⊥ C | E ? d-separated, no active trails
A ⊥ C | I ? not d-separated, there is active trail
D ⊥ I ? not d-separated, there is active trail
D ⊥ I | G ? not d-separated, there is active trail
D ⊥ I | F,G,J ? d-separated, no active trails
D ⊥ I | F,G,J,A ? d-separated, no active trails
F ⊥ H ? not d-separated, there is active trail
F ⊥ H | A ? d-separated, no active trail
F ⊥ H | D ? d-separated, no active trail
F ⊥ H | D,J ? not d-separated, there is active trail
F ⊥ H | D,J,I ? d-separated, no active trail

(b) *Given the following Bayes net (all variables are binary)*

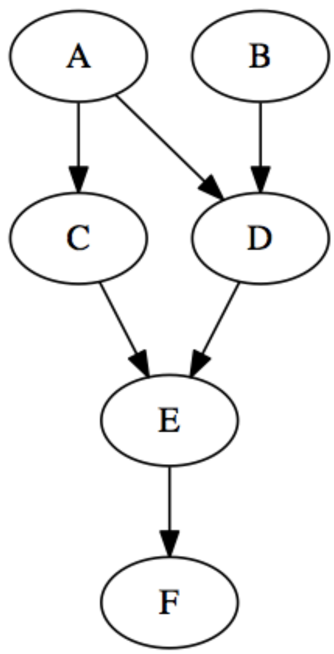
P(A = 1) = 0.7
P(B = 1) = 0.4

P(C = 1 | A = 0) = 0.2
P(C = 1 | A = 1) = 0.7

P(D = 1 | A = 0, B = 0) = 0.3
P(D = 1 | A = 0, B = 1) = 0.7
P(D = 1 | A = 1, B = 0) = 0.6
P(D = 1 | A = 1, B = 1) = 0.2

P(E = 1 | C = 0, D = 0) = 0.8
P(E = 1 | C = 0, D = 1) = 0.6
P(E = 1 | C = 1, D = 0) = 0.1
P(E = 1 | C = 1, D = 1) = 0.5

P(F = 1 | E = 0) = 0.9
P(F = 1 | E = 1) = 0.6



(Answer)