November 15, 2013

Question 1

Part A

Important Information

- Asked to compute $P(D)$: Defective
- Suppliers: Company A : $P(A) = 0.70$ / Company B : $P(B) = 0.30$
- Conditional Probability formule

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}$$

  Re-arranged

$$P(A \text{ and } B) = P(A|B) \times P(B)$$

- $P(D|A) = 0.02$
- $P(D|B) = 0.03$

Using Total Probability Law

$$P(D) = P(D \text{ and } A) + P(D \text{ and } B)$$

$$P(D) = P(D|A) \times P(A) + P(D|B) \times P(B)$$

$$P(D) = 0.02 \times 0.70 + 0.03 \times 0.30 = 0.014 + 0.0009 = 0.023$$

Part 2

$$P(A|D) = \frac{P(D|A) \times P(A)}{P(D)} = \frac{0.02 \times 0.70}{0.023} = 0.6086$$

Part B

- Mean
$$\bar{x} = \frac{4+18+2+7+18+3+4}{7} = 56/7 = 8$$
- Median
Firstly: put the data set in order.
Median is middle value (4)
- Variance
$$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$$
Ans $= \frac{294}{6} = 49$
- Standard deviation is the square root of the variance. $s = 7$

Part C : Discrete RVs

- Find k
$P(X = 2) = 0.25$. The sum of the probabilities must sum up to 1.
- $E(X) = \sum x_i p(x_i) = 9.5$
$$\sum x_i p(x_i) = (2\times 0.25)+(5\times 0.25)+(10\times 0.15)+(15\times 0.25)+(25\times 0.10)$$
- $E(X^2) = \sum x_i^2 p(x_i)$
$$\sum x_i p(x_i) = (4\times 0.25)+(25\times 0.25)+(100\times 0.15)+(225\times 0.25)+(625\times 0.10)$$
$E(X^2) = 141$
- $V(X) = E(X^2) - E(X)^2 = 141 - 9.5^2 = 50.75$

Part D : Sampling without replacement

- P(At least one is white) = 1-P(neither is white)
- $1 - [(4/10) \times (3/9)] = 78/90 = 0.8666$
- P(Exactly one) - disjoint events
- probability is sum of components
- First is white $4/10 \times 6/9 = 20/90$
- Second is white $6/10 \times 4/9 = 20/90$also
- P(Exactly One white)=40/90

Part E : Probability Laws

- $P(A \cap B) = P(A) \times P(B) = 0.5 \times 0.4 = 0.2$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.7$

Question 2

Part A : Poisson Dsitribution

Parameter: Poisson Mean m=3
Use statistical tables : Murdoch Barnes 2

- From tables $P(X \geq 2) = 0.8009$
- From tables $P(X = 0) = 1 - P(X \geq 1) =$ 1-0.9502 = 0.0498
- From tables $P(X = 1) = P(X \geq 1) - P(X \geq 2) =$ 0.9502 - 0.8009 = 0.1493
- Poisson Mean $m = 3$ and Poisson Standard Deviation $m = 3$

Part B : Binomial Distribution

n=100, p=0.05

- from tables: $P(X = 5) = P(X \geq 5) - P(X \geq 6) = 0.5640 - 0.3840 = 0.18$
- from tables: $P(X \geq 10) = 0.0282$
- from tables: $P(X \leq 12) = 1 - P(X \geq 13) = 1 - 0.0015 = 0.9985$

Part C : Exponential Distribution

Exponential Mean =10
Rate Parameter Lambda (λ) =1/10
Therefore $P(X \geq k) = exp(-k/10)$

- $P(X \geq 10) = exp(-10/10) = exp(-1) = 0.3678$
- $P(X \geq 10) = exp(-20/10) = exp(-2) = 0.1353$

Part D : Poisson Approximation

- Bin(n,p)
- Poisson(m)
- Appropropriate when n is greater than 50 and p is less than 0.05.
- Let $m = np$ and compute as poisson distribution.
- For large values of n and really small vaues of p, Poisson Approximation is much simpler computationally, with negligible error.

Question 3

Part A : Normal Distribution

Parameter Mean $\mu = 1000$ Standard Deviation $\sigma = 50$ Normal Distribution

$$X \sim N(\mu = 1000, \sigma^2 = 50^2)$$

- Compute $P(X \geq 975)$
- Zscore
$$z = frac975 - 100050 = -25/50 = -0.5$$
- using Z identity and Symmetry rule $P(X \geq 975) = P(Z \geq -0.5) = P(Z \leq -0.5) = 0.6914$
- Values derived from statistical tables
- Compute $P(X \leq 950)$
- Zscore
$$z = frac950 - 100050 = -50/50 = -1$$
- using Z identity and Symmetry rule $P(X \leq 950) = P(Z \leq -0.1) = P(Z \geq 1) = 0.8413$
- Values derived from statistical tables

Part B : Short Theory Definitions

From Notes.

Part C : Confidence interval

- Point estimate = 0.81 (89/110)
- Large Sample : 95% Quantile = 1.96
- Standard Error
$$S.E.(\hat{p}) = \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}}$$
$$S.E.(\hat{p}) = \sqrt{\frac{0.81 \times 0.19}{110}} = \sqrt{0.001411}$$
$$S.E.(\hat{p}) = 0.0374$$
- Margin of Error = $1.96 \times 0.0374 = 0.07345$
- Answer : (0.81 - 0.07345, 0.81 + 0.07344) = (0.7356,0.8825)
- 0.90 in interval - can contradict claim.

Part D : Shapiro Wilk Test

Null: Data Set is normally distributed
Alt : Data set is not normally distributed
conclusion : Large p-value. Fail to reject null.

Part E : Normal Probability Plot

- If the points on the plot follow the trendline, then the data set can be assumed to be normally distributed. (Marks for sketch, counter example etc)
- Data is normally distributed.

Question 4

Accuracy, predicion recall

Accuracy, Precision and recall are defined as

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

The F measure is computed as

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

	Predict Negative	Predict Positive
Observed Negative	9530	10
Observed Positive	300	160

Accuracy, Precision and recall are defined as

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn} = 0.9790$$

$$\text{Precision} = \frac{tp}{tp + fp} = \frac{160}{170} = 0.9411$$

$$\text{Recall} = \frac{tp}{tp + fn} = \frac{300}{460} = 0.6521$$

The F measure is computed as

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

$$= \frac{2 \times 0.6521 \times 0.9411}{0.6521 + 0.9411}$$

Two sample mean

- Hypotheses (brief written description required)

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$

-

- point estimate: diffference in means (950-910=) 40 marks
- Standard Error (from formulae)

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

- Critical Value = 1.96 (Large Sample.)

Two Sample proportion

- Hypotheses (brief written description required)

$$H_0 : \pi_1 = \pi_2$$

$$H_1 : \pi_1 \neq \pi_2$$

- point estimate: diffference in proportions
- Critical Value = 1.96 (Large Sample.)

Question 5

Part A : Huffman Coding

Part B : Binary Channels

Part C : Rate of Information

$$R = rH(X)$$

- $H(X) = -\sum_{i=1}^{16} \frac{1}{16}\log_2\frac{1}{16}$

- i.e. $H(X) = [-\frac{1}{16}\log_2\frac{1}{16}] + [-\frac{1}{16}\log_2\frac{1}{16}] \ldots [-\frac{1}{16}\log_2\frac{1}{16})]$

- Sixteen identical terms. Compute one and multiply by 16.

$$H(X) = 16 \times [-\frac{1}{16}\log_2\frac{1}{16}] = -\log_2\frac{1}{16} = -(-4) = 4$$

- $H(X) = 4$ b
- $r = 2(10^6)(32) = 64(10^6)$ elements/sec

- $R = rH(X) = 64(10^6)(4) = 256(l0^6)$ b/sec $= 256$ Mb/sec