Statistics 2009-2010

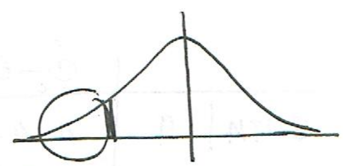
1 i) $P(\bar{B}|A) = \dfrac{P(A \cap \bar{B})}{P(A)} = \dfrac{P(A|\bar{B})P(\bar{B})}{P(A)} = \dfrac{(1-P(\bar{A}|\bar{B}))(1-P(B))}{P(A)}$

$$= \frac{(1-0.8)(1-0.3)}{0.2} = 0.7$$

$\therefore$ e

ii) $X \sim N(100, \sigma^2)$, ~~then~~ and the standardising equation $Z = \dfrac{X-\mu}{\sigma}$ then $Z \sim N(0,1)$ (standardised normal dist.)

$\therefore$ ~~For~~ $P(X<90)$, $Z = \dfrac{90-100}{\sigma}$



rejection region of size 0.1

$$\Phi^{-1}(0.1) = -\Phi^{-1}(1-0.1)$$
$$= -\Phi^{-1}(0.9)$$
$$= -1.282 \quad \text{(from table)}$$

$$\therefore \quad -1.282 = \frac{90-100}{\sigma}$$
$$\sigma = \frac{-10}{-1.282}$$
$$\sigma \approx 7.8 \quad \therefore \underline{d}$$

iii) $\bar{X} \sim N(\mu, \frac{\sigma^2}{n}) \therefore \underline{e}$

iv) MLE for $\lambda = \dfrac{n}{\sum_{i=1}^{n} x_i}$

$$= \frac{3}{0.1+0.5+0.4}$$
$$= 2 \quad \therefore \underline{f}$$

v)

observed count (O)

	1.9 GHz	2.1 GHz	
SATA	26	27	53
SSD	48	81	129
	74	108	182

expected count (E)

	1.9 GHz	2.1 GHz	
SATA	$\frac{74 \times 53}{182} = 21.55$	$\frac{108 \times 53}{182} = 31.45$	53
SSD	$\frac{74 \times 129}{182} = 52.45$	$\frac{108 \times 129}{182} = 76.55$	129
	74	108	182

	$O_i - E_i$	$(O-E_i)^2/E_i$
SATA | 1.9	4.45	$\frac{(4.45)^2}{21.55} = 0.94$
SATA | 2.1	-4.55	$\frac{(-4.55)^2}{31.45} = 0.64$
SSD | 1.9	-4.55	$\frac{(-4.55)^2}{52.45} = 0.39$
SSD | 2.1	4.45	$\frac{(4.45)}{76.55} = 0.26$

$X^2 = 2.23$ $\therefore$ b

2 i a) A geometric distribution has a fixed number of successes and counts the number of trials needed to obtain the first success.

A binomial distribution has a fixed number of trials before the experiment begins, and X (the random variable) counts the number of successes obtained in that fixed number of trials.

By looking at above definitions and problem statement the appropriate distribution is geometric, since the r.v. X denotes the number of deliveries examined until discovering one / first delivery which was late.

$\therefore \quad X \sim Geo(0.2)$

$E(X) = \frac{1}{p} = \frac{1}{0.2} = 5$

b) $P(X \leq 2) = P(X=1) + P(X=2)$

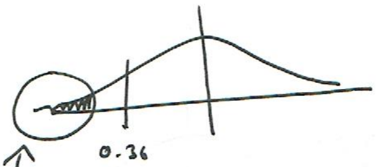
$= (0.2)(1-0.2)^0 + (0.2)(1-0.2)^1$

$= 0.36$

(Not $P(X=0)$ since we are asking: what is the probability of lateness on first, second etc. pick. No sense ask for probability on zeroth pick).
(also look at formula) $p(1-p)^{n-1}$

c) $H_0: \lambda = 0.2$

$H_1: \lambda > 0.2$



Typical rejection region ≥ 0.1

Since 0.36 is larger than any typical rejection region, there is very little evidence to reject the null hypothesis of 80% next day deliveries

ii a) Due to previous definition this is now a binomial distribution.

$Y \sim B(20, 0.2)$

$E(Y) = np = 20 \times 0.2 = 4$

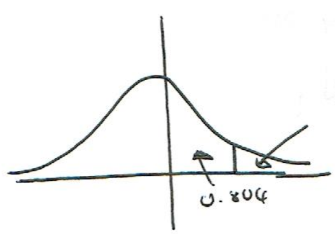
now legal to ask for 'no' picks & look at formula.

b) $P(Y \leq 5) = P(Y=0) + P(Y=1) + P(Y=2) + P(Y=3) + P(Y=4) + P(Y=5)$

$= C^{20}_0 0.2^0 \times 0.8^{20} + C^{20}_1 0.2^1 \times 0.8^{19} + C^{20}_2 0.2^2 \times 0.8^{18} +$
$C^{20}_3 \times 0.2^3 \times 0.8^{17} + C^{20}_4 \times 0.2^4 \times 0.8^{16} + C^{20}_5 \times 0.2^5 \times 0.8^{15}$

$= 0.804$

c)



$1 - 0.804 = 0.196$

i.e: The probability for delivering 6 or more packages late is 20%

$0.196 < 0.36$ ∴ more evidence against the null hypothesis than before, but there is insufficient evidence to reject the null hypothesis at any typical significance level.

iii) $Y \sim B(1000, 0.8)$ where Y is a random variable denoting the probability of deliveries that were on time

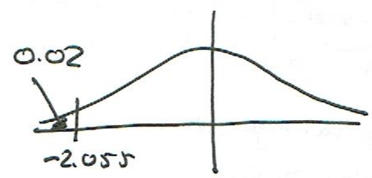
Since n is large we can approximate normal distribution due to CLT.

$E(Y) = np = 1000 \times 0.8 = 800$

$Var(Y) = npq = 1000 \times 0.8 \times 0.2 = 160$

$\therefore \quad Z \sim N(800, 160)$

$P(Z \leq 774) = \Phi\left(\frac{774-800}{\sqrt{160}}\right)$



$= \Phi(-2.055)$

$= 1 - \Phi(2.055)$

$= 1 - 0.9798$

$= 0.02$ (from tables)

With this larger sample there is now much stronger evidence to reject the null hypothesis, since the p-value of 0.02 is smaller than a typical significance level of 5/10 %.

3 i) Properties of a pdf

1) $f(x) \geq 0 \quad \forall x \in \mathbb{R}$

2) $\int_{-\infty}^{\infty} f(x) = 1$

$$1 = \int_{-\infty}^{\infty} \frac{c}{x^3}\, dx = \int_{1}^{\infty} c x^{-3} dx = \left[\frac{-c}{2} x^{-2}\right]_1^{\infty}$$

$$1 = 0 - -\frac{c}{2}$$

$$2 = c$$

ii) $$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{1}^{\infty} 2x^{-2} dx = \left[-2x^{-1}\right]_1^{\infty}$$

$$= 0 - -2$$

$$= 2$$

iii) $F(x) = \int_{-\infty}^{x} f(u)\, du = \int_{1}^{x} 2u^{-3}\, du = \left[-u^{-2}\right]_{1}^{x}$

$$= \frac{-1}{x^2} - \frac{-1}{1^2}$$

$$F(x) = 1 - \frac{1}{x^2}$$

Median is when $F(x) = \frac{1}{2}$

$\frac{1}{2} = 1 - \frac{1}{x^2}$

$\frac{1}{x^2} = \frac{1}{2}$

$2 = x^2$

$x = \pm\sqrt{2} \qquad \therefore\ x = \sqrt{2}$

iv) Since the median is less than the mean suggest a positive skew

4 i a) $\text{bias}(T) = E(T|\theta) - \theta$

b) unbiased mean estimate $= E(X) = \frac{1013 + 997 + 1013 + 1013 + 1004 + 985 + 991 + 997}{8}$

$= 1001.25$

unbiased estimate for verience =>

$$S^2 = \frac{n}{n-1}\left(\frac{\sum x^2}{n} - x^2\right)$$

$$= \frac{8}{7}\left(\frac{1013^2 + 997^2 + 1013^2 + 1013^2 + 1004^2 + 985^2 + 991^2 + 997^2}{} - (1001.25 \times 8)^2\right)$$

$= 117.98$

c) Let X be a r.v. denoting the weight of a randomly chosen bag of potatoes.

$X \sim N(1001.25, 117.98)$

$$P(X \leqslant 990) = \Phi\left(\frac{990 - 1001.25}{\sqrt{117.982}}\right) = \Phi(-1.070)$$

$= 1 - \Phi(1.070)$

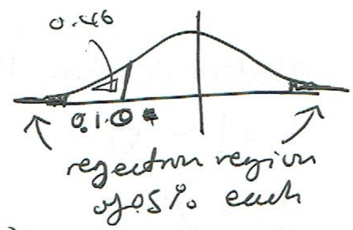
$= 1 - 0.8577$

$= 0.14$

ii a) $H_0: \mu = 1000g$ 1% sig. lev.

$H_1: \mu \neq 1000g$

Test statistic $Z = \dfrac{\bar{x}-\mu}{\sqrt{\dfrac{\sigma^2}{n}}} = \dfrac{1001.65-1000}{\sqrt{\dfrac{10^2}{8}}} = 0.460$



Either $0.46 > 0.05$

or $\Phi^{-1}(0.46) > \Phi^{-1}(0.05)$

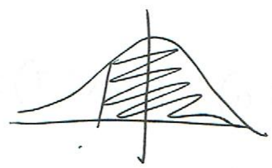
∴ There is insufficient evidence to reject the null hypothesis of the mean being 1000g (1kg).

b) $X \sim N(1000, 10^2)$

Then the sum S_5 of the weights of 5 independent bags is

$S_5 \sim N(5000, 5\times 10^2)$

$P(S_5 > 4990) = 1-\Phi\left(\dfrac{4990-5000}{\sqrt{500}}\right)$



$= 1 - \Phi(-0.447)$

$=$ ~~[illegible]~~

$= 0.673$ (from tables)