MAT3377

Fall 2015

Solution to Assignment 1

1. (i) Since

$$\sum_{k=0}^{4} P(x=k) = 1 = c + 2c + c + c/c + c/2 = 5c$$

we get $c = 1/5$.

(ii)

$$\mu = E(x) = \sum_{k=0}^{4} kP(x=k) = 0(1/5) + 1(2/5) + 2(1/5) + 3(1/10) + 4(1/10) = 15/14 = 1.5.$$

and

$$E(x^2) = \sum_{k=0}^{4} k^2 P(x=k) = \frac{37}{14} = 3.7$$

therefore

$$\sigma^2 = 3.7 - 2.25 = 1.45.$$

This gives $\sigma = 1.204159$

(iii) We have $\mu - 2.5\sigma = 1.5 - 2.5(1.204159) = -1.510397$, $\mu + 2.5\sigma = 1.5 + 2.5(1.204159) = 4.510397$, $\mu - 2.7\sigma = -1.751229$ and $\mu + 2.7\sigma = 4.751229$. Therefore

$$P(|x-\mu| < 2.5\sigma) = P(\mu - 2\sigma < x < \mu + 2\sigma) = P(x=0) + P(x=1) + P(x=2)$$

$$+P(x=3) + P(X=4)$$

$$= 1$$

and

$$P(|x-\mu| < 2.7\sigma) = P(\mu - 3\sigma < x < \mu + 3\sigma) = P(x=0) + P(x=1) + P(x=2)$$

$$+P(x=3) + p(x=4) = 1.$$

For the chebyshev's inequality the minimum probabailites are $1 - 1/(2.5^2) = 0.84$ and $1 - 1/(2.7^2) = 0.8628258$ while for the normal distribution these values are 0.988 and 0.993.

2. (i) We have

$$\sum_{i=1}^{n}(x_i-\bar{x}))(y_i-\bar{y})=\sum_{i=1}^{n}(x_iy_i-\bar{x}y_i-\bar{y}x_i+\bar{x}\bar{y})=\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}\bar{x}y_i-\sum_{i=1}^{n}\bar{y}x_i+\sum_{i=1}^{n}\bar{x}\bar{y}$$

$$=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}.$$

Similarly

$$\sum_{i=1}^{n}(x_i-\bar{x})y_i=\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}\bar{x}y_i=\sum_{i=1}^{n}x_iy_i-n\bar{x}\bar{y}.$$

Notice that

$$\sum_{i=1}^{n}\bar{x}y_i=\bar{x}\sum_{i=1}^{n}y_i,\sum_{i=1}^{n}\bar{x}\bar{y}=n\bar{x}\bar{y}.$$

(ii) We have

$$\gamma=\sqrt{n-1}\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{(\sum_{i=1}^{n}(x_i-\bar{x})^2)^{3/2}}.$$

Therefore expanding both summations gives:

$$\gamma=\sqrt{n-1}\frac{\sum_{i=1}^{n}x_i^3-3\bar{x}\sum_{i=1}^{n}x_i^2+3\bar{x}^2\sum_{i=1}^{n}x_i-n\bar{x}^3}{(\sum_{i=1}^{n}x_i^2-2\bar{x}\sum_{i=1}^{n}x_i+n\bar{x}^2)^{3/2}}$$

$$=\sqrt{n-1}\frac{nm_3-3nm_1m_2+3nm_1^3-nm_1^3}{(nm_2-2nm_1^2+nm_1^2)^{3/2}}.$$

This can be simplified to

$$\gamma=\frac{\sqrt{n-1}\,(nm_3-3nm_1m_2+2nm_1^3)}{(nm_2-nm_1^2)^{3/2}}=\sqrt{\frac{n-1}{n}}\left\{\frac{m_3-3m_1m_2+2m_1^3}{(m_2-m_1^2)^{3/2}}\right\}.$$

3. We have

$$w=\text{average}=\frac{\sum_{i=1}^{n}x_i+\sum_{i=1}^{m}y_i}{n+m}=\frac{n\bar{x}+m\bar{y}}{n+m}.$$

Therefore, for the total variance V, we can write

$$(n+m-1)V=\sum_{i=1}^{n}(x_i-w)^2+\sum_{i=1}^{n}(y_i-w)^2=\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-w)^2+\sum_{i=1}^{n}(y_i\bar{y}+\bar{y}-w)^2.$$

Since

$$\bar{x}-w=\frac{m(\bar{x}-\bar{y})}{n+m},\bar{y}-w=\frac{n(\bar{y}-\bar{x})}{n+m}$$

and

$$\sum_{i=1}^{n}(x_i-\bar{x})=0=\sum_{i=1}^{n}(y_i-\bar{y}),$$

we get

$$(n+m-1)V = \sum_{i=1}^{n}(x_i - \bar{x})^2 + \frac{nm^2(\bar{x}-\bar{y})^2}{(n+m)^2} + \sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{mn^2(\bar{x}-\bar{y})^2}{(n+m)^2}$$

$$= (n-1)s_1^2 + (m-1)s_2^2 + \frac{nm(\bar{x}-\bar{y})^2}{n+m}.$$

This will give

$$V = \frac{(n-1)s_1^2 + (m-1)s_2^2 + \frac{nm(\bar{x}-\bar{y})^2}{n+m}}{n+m-1}.$$

4. There are 6 possible distinct samples

$$S_1 = \{2,4\}, S_2 = \{2,0\}, S_3 = \{0,6\}, S_4 = \{2,6\}, S_5 = \{6,4\}, S_6 = \{0,4\}.$$

(i) We have

$$\mu = \frac{0+2+4+6}{4} = 3$$

and

$$\sigma^2 = \frac{1}{4}\sum_{i=1}^{4}(Y_i - 3)^2 = \frac{9+1+1+9}{4} = 5.$$

(ii) We calculate the mean and the variance of each sample from the formula

$$\bar{y} = \sum_{i=1}^{2} y_i/2, s^2 = \frac{1}{2-1}\sum_{i=1}^{2}(y_i - \bar{y})^2 = \sum_{i=1}^{2}(y_i - \bar{y})^2.$$

and tabulate them for each sample.

i	s_1	s_2	s_3	s_4	s_5	s_6
$\bar{x}$	3	1	3	4	5	2
s^2	2	2	18	8	2	8

Now the average of all the $\bar{x}$s is

$$\frac{3+1+3+4+5+2}{6} = 3$$

which is the population mean μ. Calculating the average of the variances gives

$$\frac{3+1+3+4+5+2}{6} = \frac{2+2+18+8+2+8}{6} = \frac{40}{6} = 6.67 \neq \sigma^2.$$

Calculating

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2 = \frac{(2-3)^2 + (0-3)^2 + (4-3)^2 + (6-3)^2}{3} = 6.67.$$

This shows that

$$E(s^2) = S^2, E(s^2) \neq \sigma^2.$$

Therefore s^2 is not an unbiased estimator for σ^2 but it is an unbiased estimator for S^2. Also notice that $E(\bar{y}) = 3 = \mu$. In other words $\bar{y}$ is an unbiased estimate for $\mu = \bar{Y}$.