MATH10282 Introduction to Statistics
Semester 2, 2019/2020
Example Sheet 2 - Solutions

1. Only method (iii) is likely to lead to a representative sample. All the other methods are likely only to sample a restricted part of the population and lead to bias.

2. (i) No, because passengers on a luxury cruise liner are not likely to be a cross-section of the population. They will likely tend to be more affluent individuals who can afford to spend more than average on a vacation.

(ii) No, because people earning different amounts may be less likely to answer. For example, the rich may not want to answer for privacy or security reasons, while the poor may be embarrassed to answer. Unless we can guarantee that all surveys are returned and answered honestly the sample may not be representative. Also it may be harder to locate wealthy individuals, or poor individuals who are not employed, in which case the questionnaires were less likely to reach those people.

(iii) No. Here the wording of the question (describing the practice as 'unfair') may bias the responses. A more neutral wording should be used.

3. The population is comprised of the values {2,3,6,8,11}, all equally likely.

(i) The population mean is given by

$$\mu = \frac{1}{5}(2+3+6+8+11) = \frac{30}{5} = 6\,.$$

(ii) First note that X_1 is equally likely to take value $2, 3, 6, 8$ or 11. Thus,

$$\mathrm{E}(X_1) = \sum_x x p_X(x) = \frac{1}{5}(2+3+6+8+11) = 6$$

$$\mathrm{E}(X_1^2) = \sum_x x^2 p_X(x) = \frac{1}{5}(2^2+3^2+6^2+8^2+11^2) = \frac{234}{5}\,.$$

Hence,

$$\mathrm{Var}(X_1) = \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2 = \frac{234}{5} - 6^2 = \frac{234-180}{5} = \frac{54}{5} = 10.8\,.$$

(iii) The possible (ordered) samples of size two that can be drawn with replacement are as follows:

(2,2)	(2,3)	(2,6)	(2,8)	(2,11)
(3,2)	(3,3)	(3,6)	(3,8)	(3,11)
(6,2)	(6,3)	(6,6)	(6,8)	(6,11)
(8,2)	(8,3)	(8,6)	(8,8)	(8,11)
(11,2)	(11,3)	(11,6)	(11,8)	(11,11)

The corresponding sample means are:

2.0	2.5	4.0	5.0	6.5
2.5	3.0	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

The sampling distribution of $\bar{X}$ is thus as follows:

$\bar{x}$	2.0	2.5	3.0	4.0	4.5	5.0	5.5	6.0	6.5
$\mathrm{P}(\bar{X} = \bar{x})$	1/25	2/25	1/25	2/25	2/25	2/25	2/25	1/25	2/25
$\bar{x}$	7.0	8.0	8.5	9.5	11.0				
$\mathrm{P}(\bar{X} = \bar{x})$	4/25	1/25	2/25	2/25	1/25				

We can compute the expectation and variance of $\bar{X}$ using the following:

$$\mathrm{E}(\bar{X}) = \sum_{\bar{x} \in R_{\bar{X}}} \bar{x}\, \mathrm{P}(\bar{X} = \bar{x}) = 6$$

$$\mathrm{E}(\bar{X}^2) = \sum_{\bar{x} \in R_{\bar{X}}} \bar{x}^2\, \mathrm{P}(\bar{X} = \bar{x}) = 41.4$$

$$\mathrm{Var}(\bar{X}) = \mathrm{E}(\bar{X}^2) - \mathrm{E}(\bar{X})^2 = 41.4 - 36 = 5.4.$$

These calculations can be checked by hand using a calculator, or more quickly using R:

```
> xbar <- c( 2, 2.5, 3, 4, 4.5, 5, 5.5, 6,
    6.5, 7, 8, 8.5, 9.5, 11)
> prob <- (1/25)*c(1,2,1,2,2,2,2,1,2,4,1,2,2,1)
> sum(xbar*prob)
[1] 6
> sum(xbar^2 * prob) - 6^2
[1] 5.4
```

Theorem 1.4 states that if $X_1, \ldots, X_n$ are i.i.d., then $\mathrm{E}(\bar{X}) = \mu$ and $\mathrm{Var}(\bar{X}) = \sigma^2/n$. Under sampling with replacement, $X_1, \ldots, X_n$ are i.i.d., and so the above calculations agree with this general theory.

4. The 10 possible samples of size two drawn without replacement are:

(2,3) (2,6) (2,8) (2,11) (3,6) (3,8) (3,11) (6,8) (6,11) (8,11)

Note that now, for example, the selection (2,3) and (3,2) are considered the same. The corresponding sample means are: 2.5, 4.0, 5.0, 6.5, 4.5, 5.5, 7.0, 7.0, 8.5, 9.5 so that the sampling distribution of $\bar{X}$ is:

$\bar{x}$	2.5	4.0	4.5	5.0	5.5	6.5	7.0	8.5	9.5
$\mathrm{P}(\bar{X} = \bar{x})$	1/10	1/10	1/10	1/10	1/10	1/10	2/10	1/10	1/10

We can compute the expectation and variance of $\bar{X}$ under sampling without replacement as follows

$$\begin{aligned}
\mathrm{E}(\bar{X}) &= \sum_{\bar{x}\in R_{\bar{X}}} \bar{x}\,\mathrm{P}(\bar{X}=\bar{x}) = 6 \\
\mathrm{E}(\bar{X}^2) &= \sum_{\bar{x}\in R_{\bar{X}}} \bar{x}^2\,\mathrm{P}(\bar{X}=\bar{x}) = 40.05 \\
\mathrm{Var}(\bar{X}) &= \mathrm{E}(\bar{X}^2) - \mathrm{E}(\bar{X})^2 = 40.5 - 36 = 4.05.
\end{aligned}$$

These calculations can be checked quickly using R:

```
> xbar <- c(2.5,4,4.5,5,5.5,6.5,7,8.5,9.5)
> probs <- (1/10) * c(1,1,1,1,1,1,2,1,1)
> sum(xbar*probs)
[1] 6
> sum(xbar^2*probs) - 6^2
[1] 4.05
```

The variance is smaller when sampling without replacement is used (4.05 rather than 5.4).

[*The (non-examinable) theory in lectures stated that under sampling without replacement* $\mathrm{E}(\bar{X}) = \mu$ *and* $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}\frac{N-n}{N-1}$. *Here the f.p.c. is* $\frac{N-n}{N-1} = \frac{5-2}{5-1} = 0.75$, *and f.p.c.* $\times\ \frac{\sigma^2}{n} = 0.75 \times 5.4 = 4.05$, *which is the same as the value of* $\mathrm{Var}(\bar{X})$ *calculated above. Thus the example calculations agree with the general theory.*]

5. (i) The population is $\{3, 7, 9, 11, 15\}$. The population mean is $\mu = (1/5)(3 + 7 + 9 + 11 + 15) = 9$. Observe that

$$\sum_{j=1}^{N} v_j^2 = (9 + 49 + 81 + 121 + 225) = 485\,,$$

and so $\sigma^2 = \frac{1}{N}\sum_{j=1}^{N} v_j^2 - \mu^2 = 485/5 - 81 = 97 - 81 = 16$.

(ii) The set of possible samples of size three that can be drawn without replacement is:

(3,7,9) (3,7,11) (3,7,15) (3,9,11) (3,9,15) (3,11,15)
(7,9,11) (7,9,15) (7,11,15)
(9,11,15)

Each of the above samples has probability 1/10 of being selected. The corresponding sample means are:

$$\begin{array}{llllll} 19/3 & 21/3 & 25/3 & 23/3 & 27/3 & 29/3 \\ 27/3 & 31/3 & 33/3 & & & \\ 35/3 & & & & & \end{array}$$

Thus the mean of the sampling distribution of $\bar{X}$ is

$$\mathrm{E}(\bar{X}) = \frac{19+21+25+23+27+29+27+31+33+35}{3\times 10} = 9\,.$$

This agrees with the theory, which states that $\mathrm{E}(\bar{X}) = \mu$. To calculate the variance of the sampling distribution note that

$$\begin{aligned} \mathrm{E}(\bar{X}^2) &= \frac{19^2+21^2+25^2+23^2+27^2+29^2+27^2+31^2+33^2+35^2}{3^2\times 10} \\ &= \frac{7530}{90} \approx 83.667\,, \end{aligned}$$

and so

$$\mathrm{Var}(\bar{X}) = \mathrm{E}(\bar{X}^2) - (\mathrm{E}\,\bar{X})^2 \approx 83.667 - 81 = 2.667\,.$$

(iii) The sample medians $\hat{Q}(0.5)$ corresponding to the above samples are:

$$\begin{array}{llllll} 7 & 7 & 7 & 9 & 9 & 11 \\ 9 & 9 & 11 & & & \\ 11 & & & & & \end{array}$$

Each of these occurs with probability 1/10. The mean of this sampling distribution is

$$\mathrm{E}(\hat{Q}(0.5)) = \frac{7+7+7+9+9+11+9+9+11+11}{10} = 9\,.$$

To compute the variance, note

$$\begin{aligned} \mathrm{E}([\hat{Q}(0.5)]^2) &= \frac{7^2+7^2+7^2+9^2+9^2+11^2+9^2+9^2+11^2+11^2}{10} \\ &= \frac{834}{10} = 83.4\,, \end{aligned}$$

and so

$$\mathrm{Var}(\hat{Q}(0.5)) = \mathrm{E}([\hat{Q}(0.5)]^2) - (\mathrm{E}\,\hat{Q}(0.5))^2 = 83.4 - 81 = 2.4\,.$$

(iv) In this scenario, both sampling distributions have a mean equal to the population mean $\mu = 9.0$. The sampling variance of the sample means is slightly higher than that for the sample medians.

6. Since $X_1, \ldots, X_{10}$ are assumed to be a random sample from $U[0,1]$, they are identically and independently distributed. Hence we may use Theorem 1.4 with $n = 10$ to show that

$$\mathrm{E}(\bar{X}) = \mu = \int_{-\infty}^{\infty} x f_X(x) dx$$
$$\mathrm{Var}(\bar{X}) = \sigma^2/10 = \mathrm{Var}(X_1)/10\,.$$

Note from MATH10141 Probability I that the p.d.f. of a $U[0,1]$ random variable is

$$f_X(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases}$$

and so

$$\begin{aligned} \mu &= \int_{-\infty}^{\infty} x f_X(x) dx = \int_{-\infty}^{0} x f_X(x) dx + \int_0^1 x f_X(x) dx + \int_1^{\infty} x f_X(x) dx \\ &= \int_{-\infty}^{0} x \times 0 dx + \int_0^1 x \times 1 dx + \int_1^{\infty} x \times 0 dx \\ &= \int_0^1 x dx = \left[\frac{x^2}{2}\right]_0^1 = 1/2\,. \end{aligned}$$

To find the variance, note that

$$\mathrm{E}(X_1^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \int_0^1 x^2 dx = \left[\frac{x^3}{3}\right]_0^1 = 1/3$$

and so $\sigma^2 = \mathrm{Var}(X_1) = \mathrm{E}(X_1^2) - \mathrm{E}(X_1)^2 = 1/3 - (1/2)^2 = 1/(12)$. Hence

$$\mathrm{Var}(\bar{X}) = \sigma^2/10 = \frac{1}{120}\,.$$

7. If $X_1, \ldots, X_6 \sim \mathrm{Po}(\lambda)$ then

$$\begin{aligned} &\mathrm{P}(X_1 = x_1,\, X_2 = x_2,\, X_3 = x_3,\, X_4 = x_4,\, X_5 = x_5,\, X_6 = x_6) \\ &= \prod_{i=1}^{6} P(X_i = x_i) \text{ by independence} \\ &= \prod_{i=1}^{6} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum x_i} e^{-6\lambda}}{\prod_{i=1}^{6} x_i!}\,. \end{aligned}$$

Note that $(x_1, x_2, x_3, x_4, x_5, x_6) = (9, 13, 6, 8, 10, 13)$ and so $\sum_{i=1}^{6} x_i = 59$.

For $\lambda = 10$, we have

$$\begin{aligned} &\mathrm{P}(X_1 = 9,\, X_2 = 13,\, X_3 = 6,\, X_4 = 8,\, X_5 = 10,\, X_6 = 13) \\ &= \frac{10^{59} e^{-6\times 10}}{9!\,13!\,6!\,8!\,10!\,13!} = 5.907 \times 10^{-7}\,. \end{aligned}$$

For $\lambda = 12$, we have

$$\begin{aligned} &\mathrm{P}(X_1 = 9,\, X_2 = 13,\, X_3 = 6,\, X_4 = 8,\, X_5 = 10,\, X_6 = 13) \\ &= \frac{12^{59} e^{-6\times 12}}{9!\,13!\,6!\,8!\,10!\,13!} = 1.704 \times 10^{-7}\,. \end{aligned}$$

These can be evaluated quickly in R via

```
> prod( dpois(x=c(9,13,6,8,10,13),lambda=10) )
[1] 5.907332e-07
> prod( dpois(x=c(9,13,6,8,10,13),lambda=12) )
[1] 1.70432e-07
```

The value $\lambda = 10$ makes the joint probability of the observed data larger.