ST102 Outline solutions to Exercise 15

1.* The probability function of the Bernoulli distribution is:

$$p(x;\pi) = \begin{cases} \pi^x (1-\pi)^{1-x} & \text{for } x = 0, 1 \\ 0 & \text{otherwise.} \end{cases}$$

Taking the logarithm, we have:

$$\log p(x;\pi) = x \log \pi + (1-x)\log(1-\pi).$$

Hence:

$$\frac{\mathrm{d}}{\mathrm{d}\pi} \log p(x;\pi) = \frac{x}{\pi} - \frac{1-x}{1-\pi} \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}\pi^2} \log p(x;\pi) = -\frac{x}{\pi^2} - \frac{1-x}{(1-\pi)^2}.$$

Fisher information can then be computed by:

$$\begin{aligned} I(\pi) = -\sum_{x=0}^{1} p(x;\pi) \frac{\mathrm{d}^2}{\mathrm{d}\pi^2} \log p(x;\pi) &= \frac{1}{\pi^2} \sum_{x=0}^{1} x\, p(x;\pi) + \frac{1}{(1-\pi)^2} \sum_{x=0}^{1} (1-x) p(x;\pi) \\ &= \frac{1}{\pi^2} \times \mathrm{E}(X) + \frac{1}{(1-\pi)^2} \times (1 - \mathrm{E}(X)) \\ &= \frac{\pi}{\pi^2} + \frac{1-\pi}{(1-\pi)^2} \\ &= \frac{1}{\pi} + \frac{1}{1-\pi} \\ &= \frac{1}{\pi(1-\pi)}. \end{aligned}$$

We are told that the maximum likelihood estimator of π is $\widehat{\pi} = \bar{X}$. Hence the asymptotic distribution is:

$$\bar{X} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right).$$

Note the result is also a consequence of the central limit theorem as the estimator is the sample mean.

2. We have $\bar{x} = 19.1$ and $s = 3.8$, so a 90% confidence interval is of the form:

$$\bar{x} \pm t_{n-1,\,0.05} \times \frac{s}{\sqrt{n}}.$$

(a) When $n = 22$, $t_{21,\,0.05} = 1.721$. Hence a 90% confidence interval is:

$$19.1 \pm 1.721 \times \frac{3.8}{\sqrt{22}} \quad \Rightarrow \quad (17.71, 20.49).$$

(b) When $n = 121$, $t_{120,\,0.05} = 1.658$. Hence a 90% confidence interval is:

$$19.1 \pm 1.658 \times \frac{3.8}{\sqrt{121}} \quad \Rightarrow \quad (18.53, 19.67).$$

In spite of the same sample mean and sample standard deviation, the sample of size $n = 121$ offers a much more accurate estimate as the interval width is merely $19.67 - 18.53 = 1.14$ hours, in contrast to the interval width of $20.49 - 17.71 = 2.78$ hours with the sample size of $n = 22$.

Note that to derive a confidence interval for μ with σ^2 unknown, the formula used in the calculation involves both n and $n - 1$. We then refer to the Student's t distribution with $n - 1$ degrees of freedom.

Also, note that $t_{120,\,\alpha} \approx z_\alpha$, where $P(Z > z_\alpha) = \alpha$ for $Z \sim N(0, 1)$. Therefore, it would be acceptable to use $z_{0.05} = 1.645$ as an approximation for $t_{120,\,0.05} = 1.658$.

3.* The population is a Bernoulli distribution on two points: 1 (agree) and 0 (disagree). We have a random sample of size $n = 475$, i.e. $\{X_1, X_2, \ldots, X_{475}\}$. Let $\pi = P(X_i = 1)$, hence $\mathrm{E}(X_i) = \pi$ and $\mathrm{Var}(X_i) = \pi(1 - \pi)$ for $i = 1, 2, \ldots, 475$. The sample mean and variance are:

$$\bar{x} = \frac{1}{475} \sum_{i=1}^{475} x_i = \frac{410}{475} = 0.8632$$

and:

$$s^2 = \frac{1}{474} \left(\sum_{i=1}^{475} x_i^2 - 475\bar{x}^2 \right) = \frac{1}{474} \left(410 - 475 \times (0.8632)^2 \right) = 0.1183.$$

(a) Based on the central limit theorem for the sample mean, an approximate 95% confidence interval for π is:

$$\begin{aligned} \bar{x} \pm z_{0.025} \times \frac{s}{\sqrt{n}} &= 0.8632 \pm 1.96 \times \sqrt{\frac{0.1183}{475}} \\ &= 0.8632 \pm 0.0309 \\ &\Rightarrow (0.8323, 0.8941). \end{aligned}$$

(b) For a 99% confidence interval, we use $z_{0.005} = 2.576$ instead of $z_{0.025} = 1.96$ in the above formula. Therefore, the confidence interval becomes wider.

Note that the width of a confidence interval is a random variable, i.e. it varies from sample to sample. The comparison in (b) above is with the understanding that the same random sample is used to construct the two confidence intervals.

Be sure to pay close attention to how we interpret confidence intervals in the context of particular practical problems. Try to interpret the confidence intervals obtained in Question 2 above.

4. (a) Based on the central limit theorem for the sample mean, an approximate 90% confidence interval is:

$$\bar{x} \pm z_{0.05} \times \frac{\sigma}{\sqrt{n}} = 36{,}000 \pm 1.645 \times \frac{12{,}000}{\sqrt{400}}$$
$$= 36{,}000 \pm 987$$
$$\Rightarrow (\pounds 35{,}013, \pounds 36{,}987).$$

We may interpret this result as follows. According to the assumption made by the economist and the survey results, we may conclude at the 90% confidence level that the average of all first-time home buyers' incomes is between £35,013 and £36,987.

Note that it is *wrong* to conclude that 90% of all first-time home buyers' incomes are between £35,013 and £36,987.

(b) Replacing $\sigma = 12{,}000$ by $s = 17{,}000$, we obtain an approximate 90% confidence interval of:

$$\bar{x} \pm z_{0.05} \times \frac{s}{\sqrt{n}} = 36{,}000 \pm 1.645 \times \frac{17{,}000}{\sqrt{400}}$$
$$= 36{,}000 \pm 1{,}398$$
$$\Rightarrow (\pounds 34{,}602, \pounds 37{,}398).$$

Now, according to the survey results (only), we may conclude at the 90% confidence level that the average of all first-time home buyers' incomes is between £34,602 and £37,398.

(c) The interval estimates are different. The first one gives a smaller range by £822. This was due to the fact that the economist's assumed σ of £12,000 is much smaller than the sample standard deviation, s, of £17,000. With a sample size as large as 400, we would think that we should trust the data more than an assumption by an economist!

The key question is whether σ being £12,000 is a reasonable assumption. This issue will be properly addressed using statistical hypothesis testing.