STAT2001 Tutorial 11 Solutions

Problem 1

(a) Let Y be your gain in dollars on a single game.

Then: $\mu = EY = 10(18/37) + (-10)(19/37) = -10/37$

$EY^2 = 10^2(18/37) + (-10)^2(19/37) = 100$

$\sigma^2 = VarY = 100 - (-10/37)^2$ = 99.927.

Now let X be your net gain in dollars after n = 50 games.

Then by the central limit theorem, $X \mathrel{\dot\sim} \mathrm{N}(n\mu, n\sigma^2)$.

So $P(X \geq 0) \approx P\left(Z > \dfrac{0 - 50(-10/37)}{\sqrt{50(99.927)}}\right) = P(Z > 0.19)$ = **0.4247**.

Solution with continuity correction

$$P(X \geq 0) \approx P\left(Z > \frac{(0-10) - 50(-10/37)}{\sqrt{50(99.927)}}\right) = P(Z > 0.05) = \mathbf{0.4801}.$$

Note: After an even number of games, you will be up or down by a multiple of \$20. Therefore the appropriate continuity correction is less half this amount, ie –10.

Alternative working

Let U be the number of times you will win.
Then U ~ Bin(n,p), where n = 50 and p = 18/37.

Note that by the central limit theorem, $U \sim \mathrm{N}(a, b^2)$,
where a = np = 24.324 and $b^2 = np(1-p)$ = 12.4909.

You will not lose any money overall so long as you win at least half the time.
Therefore $P(X \geq 0) = P(U \geq 25) \approx P\left(Z > \dfrac{25 - 0.5 - 24.324}{\sqrt{12.4909}}\right)$

= $P(Z > 0.05)$ = 0.4801, as before.

Note: The appropriate continuity correction here is –0.5, since after 50 games you will have won a number of games which could be any integer from 0 to 50.

(b) The probability that you will gain some money overall is

$$P(X \geq 20) = P(U \geq 26) \approx P\left(Z > \frac{26 - 0.5 - 24.324}{\sqrt{12.4909}}\right)$$

$$= P(Z > 0.33) = 0.3707.$$

Therefore the probability that you will come out even is

$$P(U \geq 25) - P(U \geq 26) = 0.4801 - 0.3707 = \mathbf{0.1094}.$$

Alternative working

$$P(U = 25) = p_{\text{Bin}(n,p)}(25) \approx f_{\text{N}(np,np(1-p))}(25) = 1 \times \frac{1}{\sqrt{12.4909}\sqrt{2\pi}} e^{-\frac{1}{2(12.4909)}(25-24.324)^2}$$

$$= \mathbf{0.1108}.$$

Of course we can in this case work out the probability exactly:

$$P(Y = 25) = \binom{50}{25}\left(\frac{18}{37}\right)^{25}\left(\frac{19}{37}\right)^{25} = \mathbf{0.1102}.$$

Also, $P(Y \geq 25) = \sum_{y=25}^{50} \binom{50}{y}\left(\frac{18}{37}\right)^{y}\left(\frac{19}{37}\right)^{50-y}$ = **0.4797**,

which is closer to 0.4801 than to 0.4247.

Thus the continuity correction in (a) did indeed improve the approximation.

Problem 2

A 95% CI for μ is $(\bar{y} \pm t_{0.025}(14)s/\sqrt{15}) = (9.8 \pm 2.145(0.5)/\sqrt{15})$

$$= (9.8 \pm 0.277)$$

$$= (9.52, 10.1).$$

Problem 3

A 90% CI for p is $(\hat{p} \pm z_{0.05}\sqrt{\hat{p}(1-\hat{p})/n}) = (0.75 \pm 1.645\sqrt{0.75(1-0.75)/60})$

$$= (0.75 \pm 0.092)$$

$$= (0.658, 0.842).$$

Problem 4

Let the sample observations be denoted $Y_1, \ldots, Y_n$,
where *n* = 3 and the realised values are $y_1 = 2.6,\ y_2 = 1.2,\ y_3 = 4.9$

Then define $\tilde{Y} = \sum_{i=1}^{n} Y_i^2$.

The realised value of this random variable is $\tilde{y} = \sum_{i=1}^{n} y_i^2 = 2.6^2 + 1.2^2 + 4.9^2 = 32.21$.

Now $Y_i / \sqrt{v} \sim \mathrm{N}(0,1) \Rightarrow Y_i^2 / v \sim \chi^2(1) \Rightarrow \frac{1}{v}\sum_{i=1}^{n} Y_i^2 = \frac{\tilde{Y}}{v} \sim \chi^2(n)$.

Therefore $1-\alpha = P\left(\chi^2_{1-\alpha/2}(n) < \frac{\tilde{Y}}{v} < \chi^2_{\alpha/2}(n)\right) = P\left(\frac{\tilde{Y}}{\chi^2_{\alpha/2}(n)} < v < \frac{\tilde{Y}}{\chi^2_{1-\alpha/2}(n)}\right)$.

So a 99% confidence interval for *v* is

$$\left(\frac{\tilde{y}}{\chi^2_{0.005}(n)}, \frac{\tilde{y}}{\chi^2_{0.995}(n)}\right) = \left(\frac{32.21}{12.8381}, \frac{32.21}{0.0717212}\right)$$ = **(2.51,449)**.

A second approach

$\frac{(n-1)S^2}{v} \sim \chi^2(n-1)$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ is the sample variance.

So $1-\alpha = P\left(\chi^2_{1-\alpha/2}(n-1) < \frac{(n-1)S^2}{v} < \chi^2_{\alpha/2}(n-1)\right)$, leading to the 99% CI

$$\left(\frac{(n-1)s^2}{\chi^2_{0.005}(n-1)}, \frac{(n-1)s^2}{\chi^2_{0.995}(n-1)}\right) = \left(\frac{2(3.49)}{10.5966}, \frac{2(3.49)}{0.0100251}\right)$$ = **(0.659,696)**.

A third approach

$\bar{Y} \sim \mathrm{N}(0, v/n) \Rightarrow \bar{Y}\sqrt{n/v} \sim \mathrm{N}(0,1) \Rightarrow \bar{Y}^2 n/v \sim \chi^2(1)$.

So $1-\alpha = P\left(\chi^2_{1-\alpha/2}(1) < \frac{n\bar{Y}^2}{v} < \chi^2_{\alpha/2}(1)\right)$, leading to the 99% CI

$$\left(\frac{n\bar{y}^2}{\chi^2_{0.005}(1)}, \frac{n\bar{y}^2}{\chi^2_{0.995}(1)}\right) = \left(\frac{3(2.9)^2}{7.87944}, \frac{3(2.9)^2}{0.0000393}\right)$$ = **(3.20,642000)**.

Discussion

Observe that the first of the above three CI's is the narrowest
(and therefore the best), and the third is the widest (and therefore the worst).

The third CI is bad because it ignores the variability amongst the sample observations. It is based on the sample mean $\bar{Y}$, which is not a measure of dispersion.

The second CI is better than the third because it makes use of the sample variance S^2, a measure of dispersion. However it does not perform so well as the first. This is because it ignores some of the available information, namely the fact that the normal mean is zero.

The second interval is the one used most often for estimating a normal variance. This is because it is also valid when the normal mean is unknown, which will typically be the case. The first CI is better, but can only be used when the normal mean is known to be zero. If the normal mean is not known, one should use the second interval.

A generalisation

Suppose that the normal mean μ is not necessarily zero but still known.

Then $(Y_i - \mu)/\sqrt{v} \sim \mathrm{N}(0,1) \Rightarrow (Y_i - \mu)^2 / v \sim \chi^2(1) \Rightarrow \frac{1}{v}\sum_{i=1}^{n}(Y_i - \mu)^2 \sim \chi^2(n)$.

This leads us to define $Y' = \sum_{i=1}^{n}(Y_i - \mu)^2$, so that $\frac{Y'}{v} \sim \chi^2(n)$ (an easier notation).

Then, $1-\alpha = P\left(\chi^2_{1-\alpha/2}(n) < \frac{Y'}{v} < \chi^2_{\alpha/2}(n)\right) = P\left(\frac{Y'}{\chi^2_{\alpha/2}(n)} < v < \frac{Y'}{\chi^2_{1-\alpha/2}(n)}\right)$.

Therefore a 99% confidence interval for *v* is $\left(\frac{y'}{\chi^2_{0.005}(n)}, \frac{y'}{\chi^2_{0.995}(n)}\right)$.

Non-central confidence intervals

Suppose now that we wish to find an *upper range* 99% CI for *v*.

$1-\alpha = P\left(0 < \frac{Y'}{v} < \chi^2_{\alpha}(n)\right) = P\left(\frac{Y'}{\chi^2_{\alpha}(n)} < v\right)$,

and so the required CI is $\left(\frac{y'}{\chi^2_{0.01}(n)}, \infty\right)$ = $\left(\frac{32.21}{11.3449}, \infty\right) = (2.84, \infty)$ if μ = 0.