Question 1.

1. If the price of unleaded petrol at UK filling stations is a random variable with mean 120.8p per litre, and standard deviation 4.9p, use the Central Limit theorem to determine the probability that the average price in a random sample of 50 filling stations is below 122p.

So we have that $\mu_F = 120.8$, $\sigma_F = 4.9$.

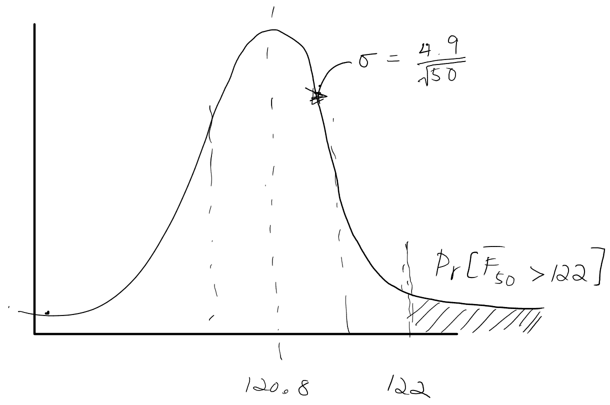
Let $\bar{F}_{50}$ be $\frac{1}{50}\sum_{i=1}^{50}(F_i)$.

$\mathbb{E}[\bar{F}_{50}] = \mu_F$, $Var(\bar{F}_{50}) = \frac{\sigma_F}{50}$

And so we are asking $Pr[\bar{F}_{50} > 122]$?

By the CLT,

$\bar{F}_{50} \dot{\sim} N(120.8, \frac{4.9^2}{50})$, and so that looks like the following:



Now we simply standardise both sides:

$$Pr[\bar{F}_{50} > 122] = Pr\left[\frac{\bar{F}_{50} - 120.8}{4.9/\sqrt{50}} > \frac{122 - 120.8}{4.9/\sqrt{50}}\right]$$

$$= Pr[Z > 1.73]$$

2. Suppose that students' marks on the economics prelims paper are normally distributed with mean 61 and standard deviation 9.5.
(Assume that the number of colleges is sufficiently large that individual observations may be considered i.i.d.)

(a) What is the distribution of the sample mean for a random sample of size n?

Denote students' marks on the paper as $M \sim N(61, 9.5^2)$.

Let the sample mean be $\bar{M}_n = \frac{1}{n}(\sum^{n} M_i)$

The distribution of the sample mean, has

$$E(\bar{M}_n) = \mu_M = 61 \quad \text{and}$$

$$Var(\bar{M}_n) = \frac{\sigma_M^2}{n} = \frac{9.5^2}{n}$$

Is this true?

YES, the sum of indep. normal variables is itself normal.

Given that the population is normally distributed, the sample mean is also normally distributed as $\{M_i\}$ are i.i.d from the population (what property is this?)

The sampling distribution is therefore

$$\bar{M}_n \sim N\left(61, \frac{9.5^2}{n}\right).$$

(b) In a random sample of 10 students, what is the probability that their average mark exceeds 63?

$$M_{10} \sim N\left(61, \frac{9.5^2}{10}\right)$$

We want $Pr[M_{10} > 63]$, Standardising, we obtain

$$Pr\left[\frac{M_{10} - 61}{9.5/\sqrt{10}} > \frac{63-61}{9.5/\sqrt{10}}\right],$$ which can be simplified to give

$$Pr\left[Z > \frac{2\sqrt{10}}{9.5}\right]$$ Checking the standard normal table, we obtain

$$Pr[Z > 0.66574] \approx 0.25.$$

(c) Suppose that you have a sample of 10 students that is selected by choosing a college at random, and then choosing 10 students at random from that college.
i. What is the expected value of their average mark?
ii. Explain why you cannot determine the variance of their average mark. Is it likely to be higher or lower than the variance of the sample mean in random sample of 10 students? Explain the intuition for your answer.

(i) The expected value of their average mark is also 61.

(ii) The variance of their average mark differs from college to college; the variance of the average mark is likely to be higher in this case compared to the variance of the sample mean in a random sample of all students. This is because ... why?

Consider drawing 10 students from a particular college ....

5. The 1165 Oxford PPE applicants in 2007 achieved an average score of 60.86 on the TSA test, with a standard deviation of 8.02. Construct a 95% confidence interval for the population mean score.

The population mean score M : $\mu_M = 60.86$, $\sigma_M = 8.02$.

What is the question asking us actually? Suppose we started taking samples of size n and seeing the sample mean score $\bar{M}_n$. If $\bar{M}_n$ were large enough, then $\bar{M}_n \sim N\left(60.86, \frac{8.02}{\sqrt{n}}\right)$ by the CLT.

The confidence interval is a random variable: it is the low and the high values such that μ_M would fall within in 95% of samples. (Doesn't this depend on n, though?)

$$95\%\ CI = \left\{ \mu_M \pm 1.96\, \sigma(\bar{M}_n) \right\}.$$

$$= \left\{ 60.86 \pm 1.96 \frac{SE(\bar{M}_n)}{\sqrt{n}} \right\} \rightarrow \text{unbiased estimator of}$$

$$= \left\{ 60.86 \pm 1.96 \cdot \frac{8.02}{\sqrt{n}} \right\}.$$

6.(a) Consider a random sample of size n from a Bernoulli distribution with parameter p. If the sample mean is X̄, show that the sample variance is given by s2 = n
X̄(1 − X̄). Compare the sample mean and variance with the
n−1
population mean and variance.

Sample mean $\bar{X} = \frac{1}{n}\sum^n X_i$

$\mathbb{E}(\bar{X}) = p.$

Show: $s^2 = \frac{n}{n-1}\bar{X}(1-\bar{X}).$

We have that $s^2 = \frac{1}{n-1}\sum_i^n (X_i - \bar{X})^2$

$$\Rightarrow s^2 = \frac{1}{n-1}\sum_i^n (X_i^2 - \bar{X}X_i - \bar{X}^2)$$

$$= \frac{1}{n-1}\left[\sum_i^n X_i^2 - \sum^n \bar{X}X_i - \sum^n \bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum^n X_i^2 - \bar{X}\sum^n X_i - n\bar{X}^2\right]$$ Given that $\bar{X} = \frac{1}{n}\sum^n X_i$,

$$= \frac{1}{n-1}\left[\sum^n X_i^2 - n\bar{X}^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[\sum^n X_i^2 - n\bar{X}^2\right]$$ Since the variable X is Bern, $X_i^2 = X_i$

$$= \frac{1}{n-1}\left[\sum^n X_i - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[n\bar{X} - n\bar{X}^2\right]$$

$$= \frac{n}{n-1}\bar{X}(1-\bar{X})$$ Shown. □

The population mean and variance of a Bernoulli random variable X:

$$\mu_X = \mathbb{E}[X] \qquad Var(X) = \mathbb{E}[(X - \mathbb{E}(X))^2]$$

$$= \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$= \mathbb{E}[X] - \mathbb{E}[X]^2$$

$$= \mathbb{E}[X](1 - \mathbb{E}[X])$$

Compare the pop. mean and variance with the sample mean and variance: don't know what they are asking me to compare...

(b) In an opinion poll of 300 voters, 140 say that they will vote for the incumbent, and 160 for the rival candidate. Estimate the proportion of votes that will be obtained by the incumbent in the election. Calculate the sample variance. Find 95% and 99% confidence intervals for the incumbent's proportion of votes in the election.

Let the proportion of votes obtained by the incumbent be X.

The sample mean $\bar{X} = \frac{1}{300}\sum_{i=1}^{300} X_i$

By the CLT, $\bar{X} \sim N\left(\frac{140}{300}, \sigma^2/n\right)$

The sample variance is an unbiased estimator of the pop. variance:

$$s^2 = \frac{1}{n-1}\sum_{i}^{n}(X_i - \bar{X})^2$$

In part a) we showed that for a Bernoulli var,

$$s_X^2 = \frac{n}{n-1}\bar{X}(1-\bar{X})$$

$$s^2_{X\,actual} = \frac{\cancel{300}}{299}\left(\frac{140}{\cancel{300}}\right)\left(\frac{160}{300}\right)$$

To find the estimate of the pop variance we plug in the results.

$$= \frac{140 \times 160}{299 \cdot 300}$$

$$= 0.250$$

The 95% and 99% confidence intervals are:

95% CI: $\left\{ \bar{X} + 1.96\,\sigma_{\bar{Y}} \right\}$

99% CI: $\left\{ \bar{X} + 2.58\,\sigma_{\bar{Y}} \right\}$

ESTIMATORS & THEIR ESTIMATES

$s_X^2 \to \sigma^2$, $s_X \to \sigma$

$SE(\bar{X}) \to \frac{\sigma}{\sqrt{n}}$

We know that s_X^2 is an unbiased estimator of σ^2, and $\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}}$. Therefore, s_X is an unbiased estimator of $\sqrt{n}(s_X)$.