Session 6: Solutions for Practical Exercise

Question 1

$X_1, \ldots, X_n \overset{iid}{\sim} N(3, \sigma^2)$. H_0: $\sigma^2 = 1/4$ vs H_1: $\sigma^2 = 1$.

Part (a)

The best (most powerful) test statistic rejects H_0 for small values of the likelihood ratio L_{H_0}/L_{H_1}, or of the log of this ratio, $l_{H_0} - l_{H_1}$. Below we denote the null and alternative values of σ^2 as σ_0^2 and σ_1^2 respectively.

$$L(\sigma^2 | \underline{x}, \mu = 3) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\left(\frac{x_i - 3}{\sigma}\right)^2\right)$$

$$\Rightarrow l(\sigma^2) = -\frac{1}{2}\sum_{i=1}^{n} \log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - 3)^2$$

ignoring terms not in σ^2

$$l(\sigma^2) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - 3)^2$$

$$l_{H_0} - l_{H_1} = l(\sigma_0^2) - l(\sigma_1^2) = \frac{n}{2}\log\sigma_1^2 + \frac{1}{2\sigma_1^2}\sum_{i=1}^{n}(x_i - 3)^2 - \frac{n}{2}\log\sigma_0^2 - \frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(x_i - 3)^2$$

$$= \frac{n}{2}(\log\sigma_1^2 - \log\sigma_0^2) + \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^{n}(x_i - 3)^2$$

$$= \frac{n}{2}\log\frac{\sigma_1^2}{\sigma_0^2} + \frac{1}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\sum_{i=1}^{n}(x_i - 3)^2$$

On repeated sampling, this varies only through the term involving the data, $\sum_{i=1}^{n}(x_i - 3)^2$, other terms remaining constant. When $\sigma_0^2 = \frac{1}{4}, \sigma_1^2 = 1$, then $\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right) < 0$, so $l_{H_0} - l_{H_1}$ is only 'very small' (very negative) if $\sum_{i=1}^{n}(x_i - 3)^2$ is very large. Thus this term, or any constant multiple of it, is our best test statistic.

Part (b)

In the above $\sigma_1^2 = 1$. However, if this alternative value is replaced by any value $\sigma_1^2 > 1/4$, $\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)$ remains negative and the above argument yields the same test statistic: so this test is uniformly most powerful for the composite alternative hypothesis H_1: $\sigma^2 > 1/4$.

Part (c)

$$H_0: X_1, \ldots, X_n \overset{iid}{\sim} N(3, 1/4) \quad \Rightarrow \quad \frac{X_i - 3}{\sqrt{1/4}} \sim N(0,1) \quad \Rightarrow \quad \left(\frac{X_i - 3}{\sqrt{1/4}}\right)^2 \sim \chi_1^2$$

$$\Rightarrow \quad \sum_{i=1}^{10}\left(\frac{X_i - 3}{\sqrt{1/4}}\right)^2 \sim \chi_{10}^2 \quad \Rightarrow \quad 4\sum_{i=1}^{10}(X_i - 3)^2 \sim \chi_{10}^2$$

So $4T \sim \chi^2_{10}$, where $T = \sum_{i=1}^{10}(X_i - 3)^2$.

A 5% rejection region is given by threshold c where $\text{Prob}(T \geq c) = 0.05$, and since

$$\text{Prob}\left(4T \geq \chi^2_{10,0.95}\right) = 0.05 \quad \Rightarrow \quad \text{Prob}\left(T \geq \frac{1}{4}\chi^2_{10,0.95}\right) = 0.05$$

The threshold value defining the rejection region is therefore a quarter of the 95% point of the χ^2_{10} distribution.

Part (d)

From fact (ii) $E(X^2_{10}) = 10, Var(X^2_{10}) = 20$ so from fact (iii), approximately, $X^2_{10} \sim N(10, 20)$

$$\Rightarrow \quad 4T \dot{\sim} N(10, 20) \quad \Rightarrow \quad \frac{4T - 10}{\sqrt{20}} \dot{\sim} N(0,1)$$

Part (e)

Critical region can now be defined (approximately) by $\text{Prob}\left(\frac{4T-10}{\sqrt{20}} \geq Z_{0.95}\right) = 0.05$.
The 95% point of the standard Normal distribution is 1.64. Thus we reject when

$$\frac{4T - 10}{\sqrt{20}} \geq 1.64 \quad \Rightarrow \quad T \geq \frac{1}{4}\left(10 + 1.64 \times \sqrt{20}\right) = \frac{1}{4} \times 17.33 = 4.334$$

From the data, $T = 11.5$, so we reject the null hypothesis and conclude that there is strong evidence at the 5% level that taking the drug orally gives a greater variability in the 24-hour concentration of the drug than the hypothesised value of $1/4$.

Note: Since the test in e) is essentially a one-sided test, we have no option, in rejecting the null, but to find evidence that the population variance is greater than ¼. If we had performed a two-sided test, and rejected the null based on these data, where the observed variance is in fact greater than ¼, we would still have found evidence that the population variance is greater than ¼. Sometimes students are reluctant, when rejecting the null hypothesis in a two-sided test, to say more than that the null is rejected. The null can only be rejected in the observed direction, so don't be reluctant to state that there is evidence (at the appropriate significance level) that the true value is greater (or smaller) than the null, if that's what the observed result suggests. Similarly, when comparing, e.g., means in two groups, if you observe mean A > mean B and obtain a significant two-sided test, that is then evidence that population mean A is greater than population mean B; even though your two-sided test considered the reverse to be possible.

Question 2 (Optional)

Suppose $X^2 \sim \chi_1^2$ and $X_m^2 \sim \chi_m^2$.
Want to show that a) $E[X^2] = 1$, b) $Var[X^2] = 2$, c) $E[X_m^2] = m$, d) $Var[X_m^2] = 2m$

a) $E[X^2] = Var[X] + (E[X])^2$ from definition of variance.
And if $X^2 \sim \chi_1^2$ then $X \sim N(0,1)$ so $E[X^2] = 1 + 0 = 1$.

b) By definition, and using part a), $Var[X^2] = E[X^4] - (E[X^2])^2 = E[X^4] - 1$
By definition $E[X^4] = \int_{-\infty}^{\infty} x^4 \, f(x) \, dx$
$X \sim N[0,1]$ so

$$E[X^4] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^4 \, e^{-\frac{x^2}{2}} \, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 . x e^{-\frac{x^2}{2}} \, dx$$
$$= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^3 . \frac{d}{dx}\left(e^{-\frac{x^2}{2}}\right) dx$$

Integrating by parts:

$$= -\frac{1}{\sqrt{2\pi}} \left[x^3 . e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} 3x^2 . e^{-\frac{x^2}{2}} \, dx$$
$$= -\frac{1}{\sqrt{2\pi}} \left[x^3 . e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} - \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x . \frac{d}{dx}\left(e^{-\frac{x^2}{2}}\right) dx$$

Looking at the part in the square brackets: $x^3/e^{x^2/2} \to 0$ as $x \to \infty$ or $-\infty$, so

$$E[X^4] = -\frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x . \frac{d}{dx}\left(e^{-\frac{x^2}{2}}\right) dx$$
$$= -\frac{3}{\sqrt{2\pi}} \left[x . e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} + \frac{3}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx$$

We can write the last term as:

$$3 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 3 \int_{-\infty}^{\infty} f(x) dx$$

where $f(x)$ is the standard normal probability density. We know that $f(x)$ has to integrate to 1. So

$$E[X^4] = -\frac{3}{\sqrt{2\pi}} \left[x . e^{-\frac{x^2}{2}}\right]_{-\infty}^{\infty} + 3$$

Now let's look at the bit in square brackets: $x/e^{x^2/2} \to 0$ as $x \to \infty$ or $-\infty$, so

$$E[X^4] = 3$$

And hence

$$Var[X^2] = 3 - 1 = 2$$

c) Suppose $X_i^2 \overset{iid}{\sim} \chi_1^2$ for $i = 1, 2, \ldots$, then $X_n^2 = \sum_{i=1}^{n} X_i^2 \sim \chi_n^2$ using the property (i) of χ^2 distributions from practical sheet. So

$$\sum_{i=1}^{m} X_i^2 \sim \chi_m^2 \quad \text{and} \quad E\left[\sum_{i=1}^{m} X_i^2\right] = \sum_{i=1}^{m} E[X_i^2] = m.1 = m$$

d) Since the X_i^2 are independent:

$$Var\left[\sum_{i=1}^{m} X_i^2\right] = \sum_{i=1}^{m} Var[X_i^2] = m.2 = 2m$$