Dr. ~~Bate~~ Ruklisa Dace

IB Statistics
Example Sheet 2

Samuel Lee

Question 1

The likelihood ratio of an observed data point x with respect to H_0 and H_1 is

$$\Lambda_X(H_0, H_1) = \frac{f(x \mid H_1)}{f(x \mid H_0)} = \frac{2/(x+2)^2}{1/(x+1)^2} = \frac{2(x+1)^2}{(x+2)^2}$$

Yes.

We want to find $k > 0$ such that

$$\Pr[\Lambda_X(H_0, H_1) > k \mid \theta = 1] = \Pr\left[\frac{2(x+1)^2}{(x+2)^2} > k \,\middle|\, \theta = 1\right] = 0.05$$

Yes.

Since $x, k > 0$, k can also be found by equating the following to 0.05:

$$\Pr\left[\frac{x+1}{x+2} > \sqrt{\frac{k}{2}} \,\middle|\, \theta = 1\right] = \Pr\left[x+1 > (x+2)\sqrt{k/2} \,\middle|\, \theta = 1\right]$$
$$= \Pr\left[\left(1 - \sqrt{k/2}\right)x > \sqrt{2k} - 1 \mid \theta = 1\right]$$

Yes.

If $0 < k \leq \frac{1}{2}$, then the right-hand side is non-positive and the probability is 1. Also, k must be less than 2 since $\frac{x+1}{x+2} < 1$. Therefore, $\frac{1}{2} < k < 2$, and the above can be further simplified:

$$\Pr\left[\left(1 - \sqrt{k/2}\right)x > \sqrt{2k} - 1 \mid \theta = 1\right] = \Pr\left[x > \frac{\sqrt{2k}-1}{1-\sqrt{k/2}} \,\middle|\, \theta = 1\right]$$

The cumulative density function of X given θ is

$$\int_0^x \frac{\theta}{(t+\theta)^2}\mathrm{d}t = \left[-\frac{\theta}{t+\theta}\right]_0^x = \frac{x}{x+\theta}$$

$$-\frac{\theta}{X+\theta} + 1 = \frac{X+\theta-\theta}{X+\theta}$$

Yes.

So we have

$$\Pr[\Lambda_X(H_0, H_1) \mid \theta = 1] = \Pr\left[x > \frac{\sqrt{2k}-1}{1-\sqrt{k/2}}\right] = 1 - F_X\left(\frac{\sqrt{2k}-1}{1-\sqrt{k/2}} \,\middle|\, \theta = 1\right)$$
$$= 1 - \frac{\frac{\sqrt{2k}-1}{1-\sqrt{k/2}}}{\frac{\sqrt{2k}-1}{1-\sqrt{k/2}} + 1} = \frac{1}{\frac{\sqrt{2k}-1}{1-\sqrt{k/2}} + 1} = \frac{1 - \sqrt{k/2}}{\sqrt{2k} - \sqrt{k/2}}$$

Yes.

The required k satisfies

$$\frac{1-\sqrt{k/2}}{\sqrt{2k}-\sqrt{k/2}} = 0.05 \implies 0.05\sqrt{2k}+0.95\sqrt{k/2} = 1$$

$$\implies \sqrt{k} = \frac{1}{0.05\sqrt{2}+0.95\sqrt{\frac{1}{2}}}$$

This section can be simplified by noticing that $\Lambda_X(H_0;H_1)$ is increasing in X.

$$\implies k = \frac{1}{0.55125} = \frac{800}{441}$$

Yes.

and therefore the required test rejects the null hypothesis when $\Lambda_X(H_0,H_1) > \frac{800}{441}$. Following similar steps from before, the probability of a Type II error is

Then look for a test $C_2=\{X:X>k\}$.

$$\Pr\left[\Lambda_X(H_0,H_1) \le \frac{800}{441} \,\middle|\, \theta=2\right] = \Pr\left[x \le \frac{\sqrt{2k}-1}{1-\sqrt{k/2}} \,\middle|\, \theta=2\right]$$

$$= F_X\left[\cancel{x\le}\frac{\sqrt{2k}-1}{1-\sqrt{k/2}} \,\middle|\, \theta=2\right]$$

$$= \frac{\frac{\sqrt{2k}-1}{1-\sqrt{k/2}}}{\frac{\sqrt{2k}-1}{1-\sqrt{k/2}}+2} = \frac{\frac{\sqrt{\frac{1600}{441}}-1}{1-\sqrt{\frac{400}{441}}}}{\frac{\sqrt{\frac{1600}{441}}-1}{1-\sqrt{\frac{400}{441}}}+2} = \frac{\frac{\frac{40}{21}-1}{1-\frac{20}{21}}}{\frac{\frac{40}{21}-1}{1-\frac{20}{21}}+2} = \frac{19}{21}$$

Yes.

Question 2

In both parts, the two hypotheses are simple hypotheses, so the likelihood ratio test is the most powerful test of any size. Given an observation x, the likelihood ratio is

Yes. Both hypotheses could also be stated explicitly.

$$\Lambda_X(H_0,H_1) = \frac{f(x\mid H_1)}{f(x\mid H_0)} = \frac{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu_1)^2}{2}}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu_0)^2}{2}}} = \exp\left[-\frac{(x-\mu_1)^2-(x-\mu_0)^2}{2}\right]$$

$$= \exp\left[(\mu_1-\mu_0)x - \frac{\mu_1^2-\mu_0^2}{2}\right]$$

Yes.

For a given size α, we can solve for the critical value $k>0$ given μ_0 and μ_1:

$$\Pr\left[\exp\left[(\mu_1-\mu_0)x-\frac{\mu_1^2-\mu_0^2}{2}\right] > k \,\middle|\, H_0\right] = \Pr\left[(\mu_1-\mu_0)x-\frac{\mu_1^2-\mu_0^2}{2} > \ln k \,\middle|\, H_0\right]$$

Again, note that for $\mu_1>\mu_0$, $\Lambda_X(H_0,H_1)$ is an increasing function in X.

$$= \Pr\left[(\mu_1-\mu_0)x > \ln k + \frac{\mu_1^2-\mu_0^2}{2} \,\middle|\, H_0\right]$$

$$= \begin{cases} \Pr\left(x-\mu_0 > \frac{\ln k}{\mu_1-\mu_0}+\frac{\mu_1-\mu_0}{2} \,\middle|\, H_0\right) & \text{if } \mu_1>\mu_0 \\ \Pr\left(x-\mu_0 < \frac{\ln k}{\mu_1-\mu_0}+\frac{\mu_1-\mu_0}{2} \,\middle|\, H_0\right) & \text{if } \mu_1<\mu_0 \end{cases} = \alpha$$

Yes.

We can see that the likelihood ratio test is equivalent to testing whether $x-\mu_0$ is greater/less than some critical value if μ_1 is greater/less than μ_0. So we can use $x-\mu_0$ as a test statistic instead, with a critical value C such that $1-\Phi(C)=\alpha$ if $\mu_1>\mu_0$ or $\Phi(C)=\alpha$ if $\mu_1<\mu_0$.

A test $\{X: X>K\}$ is suitable here. $P(X\in C\mid H_0)=\alpha$ (hypothesis is important here)

(a)

We reject the null hypothesis when the test statistic $x > C$ where $1 - \Phi(C) = \Phi(-C) = \alpha$. So $C = 1.645$ for the test of size 0.05 and $C = 2.326$ for the test of size 0.01. Yes.

But H_0 for evaluation is different here.

(b)

We reject the null hypothesis when the test statistic $x - 4 < C$ where $\Phi(C) = \alpha$. So $C = -1.645$ for the test of size 0.05 and $C = -2.326$ for the test of size 0.01.

When $X(\omega) = 2.1$, we find that the test in (a) rejects the null hypothesis at the 5% level but not the 1% level. However, this is also true when we apply the test in (b), where the test statistic becomes $2.1 - 4 = -1.9$. There is no contradiction between the two; it is simultaneously true that the probability one would get $x = 2.1$ or anything more unlikely is less than 5% under both null hypotheses. But we might run into a problem deciding which hypothesis to assume as true. It might be tempting to accept the results of the test in (b) given that x is closer to 0 than 4. However, it is possibly objectionable to decide on a test after the results have already been observed (for example, this is one of the rationales for requiring pre-analysis plans in clinical trials). 4 than 0? Yes.

One possible ex-ante way to decide (before having observed x) is to determine what makes for a more 'natural' null hypothesis. This could be in Bayesian terms: we set the null hypothesis to be the one we think is a priori more likely to be true. But this could also be decided with respect to the context using some loss criteria: we might decide that a given test is more appropriate if the cost of making a type I error is smaller. For example, μ could be the percentage increase in mortality following a new medical procedure. It is probably more costly to assume the procedure does not lead to higher mortality when it actually does, than to assume the procedure leads to higher mortality when it does not. Yes.

Question 3

Yes.

The MLE of θ where $X_1, \ldots, X_n \sim \text{Exp}(\theta)$ is $\bar{X}^{-1}$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sample mean. Letting $\Theta_0 \subset \mathbb{R}_+^2$ be the parameter space corresponding to H_0, the likelihood ratio given H_0 and H_1 and observations $x = x_1, \ldots, x_n$ and $y = y_1, \ldots, y_n$ is

$$\Lambda_{x,y}(H_0, H_1) = \frac{\sup_{(\theta_1,\theta_2)\in\mathbb{R}_+^2\setminus\Theta_0} f(x, y \mid \theta_1, \theta_2)}{\sup_{(\theta_1,\theta_2)\in\Theta_0} f(x, y \mid \theta_1, \theta_2)} = \frac{f\left(x, y \mid \bar{x}^{-1}, \bar{y}^{-1}\right)}{f\left(x, y \mid \frac{2}{\bar{x}+\bar{y}}, \frac{2}{\bar{x}+\bar{y}}\right)}$$

Yes.

The last equality holds because $\bar{x}^{-1}$ and $\bar{y}^{-1}$ maximise the likelihood function when there is no restriction for θ_1 and θ_2 to be equal, whereas constraining θ_1 and θ_2 to be equal means the likelihood function is maximised as though x and y were drawn from a single exponential distribution,

Yes.

in which case the MLE is $\frac{2n}{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i} = \frac{2}{\bar{x}+\bar{y}}$. Therefore, the likelihood ratio is

$$
\begin{aligned}
\Lambda_{x,y}(H_0, H_1) &= \frac{\prod_{i=1}^n \bar{x}^{-1} e^{-\bar{x}^{-1} x_i} \times \prod_{i=1}^n \bar{y}^{-1} e^{-\bar{y}^{-1} y_i}}{\prod_{i=1}^n \frac{2}{\bar{x}+\bar{y}} e^{-\frac{2}{\bar{x}+\bar{y}} x_i} \times \prod_{i=1}^n \frac{2}{\bar{x}+\bar{y}} e^{-\frac{2}{\bar{x}+\bar{y}} y_i}} \\
&= \frac{(\bar{x}\bar{y})^{-n} \exp(-2n)}{\left(\frac{2}{\bar{x}+\bar{y}}\right)^{2n} \exp\left(-\frac{2n\bar{x}}{\bar{x}+\bar{y}} - \frac{2n\bar{y}}{\bar{x}+\bar{y}}\right)} \\
&= \left[\frac{(\bar{x}+\bar{y})^2}{4\bar{x}\bar{y}}\right]^n \\
&= \left[\frac{\bar{x}^2 + \bar{y}^2 + 2\bar{x}\bar{y}}{4\bar{x}\bar{y}}\right]^n \\
&= \frac{1}{4^n}\left[\frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n y_i} + \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} + 2\right]^n \\
&= \frac{1}{4^n}\left[\frac{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}{\sum_{i=1}^n y_i} + \frac{\sum_{i=1}^n x_i + \sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}\right]^n \\
&= \frac{1}{4^n}\left[\frac{1}{1-T} + \frac{1}{T}\right]^n = \frac{1}{4^n}\left[\frac{1}{T(1-T)}\right]^n = \frac{1}{4^n}\left[-\frac{1}{(T-1/2)^2 - 1/4}\right]^n
\end{aligned}
$$

Yes.

Yes.

Yes.

$T - T^2 - \frac{1}{4} + \frac{1}{4} = -(T-\frac{1}{2})^2 + \frac{1}{4}$

which shows that the likelihood ratio is a monotone function of $|T - 1/2|$. The distribution of the sums of X_i and Y_i is such that $\sum_{i=1}^n X_i \sim \Gamma(n, \theta_1)$ and $\sum_{i=1}^n Y_i \sim \Gamma(n, \theta_2)$. Under the null hypothesis, $\theta_1 = \theta_2$, and following the arguments in Sheet 1, $T \sim \mathrm{Beta}(n, n)$.

Yes.

We want to find some positive k such that, conditional on $\theta_1 = \theta_2$,

$$
\Pr\left\{\frac{1}{4^n}\left[-\frac{1}{(T-1/2)^2 - 1/4}\right]^n > k\right\} = \alpha
$$

Here it is better to simplify the likelihood ratio depending on when it increases or decreases for various parameters.

Since $T \in (0,1) \implies 0 < 4(T-1/2)^2 < \frac{1}{4}$, we have

$$
\begin{aligned}
\Pr\left\{\frac{1}{4^n}\left[-\frac{1}{(T-1/2)^2 - 1/4}\right]^n > k\right\} &= \Pr\left[\frac{1}{4(T-1/2)^2 - 1} < -k^{\frac{1}{n}}\right] \\
&= \Pr\left[4(T-1/2)^2 - 1 > -k^{-\frac{1}{n}}\right] \\
&= \Pr\left[|T - 1/2| > \frac{k^{\frac{1}{n}} - 1}{4k^{\frac{1}{n}}}\right] \\
&= \Pr\left[T > \frac{3k^{\frac{1}{n}} - 1}{4k^{\frac{1}{n}}} \cup \cancel{\cap}\, T < \frac{k^{\frac{1}{n}} + 1}{4k^{\frac{1}{n}}}\right] = \alpha
\end{aligned}
$$

and we can implicitly find a k which satisfies this given that T has a $\mathrm{Beta}(n, n)$ distribution.

Yes.

this looks like two tails of a test and OR operator in the ~~end~~ end.

Question 4

Under the null hypothesis,

$$
p_i(\theta) = \binom{3}{i} \theta^i (1-\theta)^{3-i}
$$

Yes.

It is possible to simplify quantiles to $t < -c + \frac{1}{2}$ and $t > c + \frac{1}{2}$.

The hypotheses could be written more formally: some parameters have to be defined.

Rather θ determines all probabilities, not the other way round.

For each bunch, i follows a multinomial distribution with parameters $p = (p_0, p_1, p_2, p_3)$. Under the unrestricted model for p_i, we have 3 free parameters (one element of p is implied by the other three since all 4 must sum to 1), whereas in the restricted model under the null hypothesis, there is one free parameter: given any element of p, the value of the underlying parameter θ can be imputed and the 3 other probabilities are identified. Therefore, the difference in dimensionality between H_1 and H_0 is $3 - 1 = 2$. ← Yes.

Under the unrestricted model, the likelihood function is maximised at $\hat{p}_i = \frac{n_i}{n}$ where n_i is the number of bunches with i defective articles and $n = \sum_{i=0}^{3} n_i$. Therefore, Yes.

~~Note that constants are already omitted.~~

$$\ln \Lambda_x(H_0, H_1) = \frac{\ln \prod_{i=0}^{3} \hat{p}_i^{n_i}}{\ln \sup_{\theta\in[0,1]} \prod_{i=0}^{3} p_i(\theta)^{n_i}} = \sum_{i=0}^{3} n_i \ln \hat{p}_i - \sup_{\theta\in[0,1]} \sum_{i=0}^{3} n_i \ln p_i(\theta)$$

where $\hat{\theta}$ is the maximiser in the second term. The maximand is equal to

$$\begin{aligned}
&3n_0 \ln(1-\theta) + n_1 \left[\ln 3 + \ln\theta + 2\ln(1-\theta)\right] + n_2 \left[\ln 3 + 2\ln\theta + \ln(1-\theta)\right] + 3n_3 \ln\theta \\
=&(3n_0 + 2n_1 + n_2)\ln(1-\theta) + (n_1 + 2n_2 + 3n_3)\ln\theta + (n_1 + n_2)\ln 3
\end{aligned}$$

Yes.

The maximiser $\hat{\theta}$ satisfies the following first-order condition:

$$\frac{n_1 + 2n_2 + 3n_3}{\hat{\theta}} = \frac{3n_0 + 2n_1 + n_2}{1 - \hat{\theta}} \implies \hat{\theta} = \frac{n_1 + 2n_2 + 3n_3}{3n}$$

Yes.

Going back to the log-likelihood ratio, we find that it is equal to

$$\ln \Lambda_x(H_0, H_1) = \sum_{i=0}^{3} n_i \ln \hat{p}_i - \sup_{\theta\in[0,1]} \sum_{i=0}^{3} n_i \ln p_i(\theta) = \sum_{i=0}^{3} n_i \ln \frac{n_i}{n p_i(\hat{\theta})}$$

Yes.

using $\hat{p}_i = \frac{n_i}{n}$. Denoting n_i by o_i (what is observed), and $np_i(\hat{\theta})$ by e_i (what is expected given the maximum-likelihood parameter), the test statistic for Pearson's Chi-squared test is

$$T = \sum_{i=0}^{3} \frac{(o_i - e_i)^2}{e_i}$$

The above is approximately equal to $2\ln\Lambda_x(H_0, H_1)$, which by Wilks' theorem converges to a χ^2_2 distribution where the degrees of freedom is equal to the difference in dimensionality between H_1 and H_0. In our case, we have

$$\begin{aligned}
\hat{\theta} &= \frac{n_1 + 2n_2 + 3n_3}{3n} = 0.25 \\
p_0(\hat{\theta}) &= \frac{27}{64},\ p_1(\hat{\theta}) = \frac{27}{64},\ p_2(\hat{\theta}) = \frac{9}{64},\ p_3(\hat{\theta}) = \frac{1}{64} \\
e_0 &= 216,\ e_1 = 216,\ e_2 = 72,\ e_3 = 8 \\
T &= \frac{9}{216} + \frac{144}{216} + \frac{225}{216} + \frac{36}{216} = \frac{414}{216} \approx 1.9167
\end{aligned}$$

Yes.

The critical value for a test at the 5% level is 5.991, so we do not reject the null hypothesis here.

The final test value should be larger and sufficient to reject H_0.

Question 5

not official term

The Neyman–Pearson lemma states that the likelihood-ratio test of size α between $H_0\colon f = f_0$ and $H_1\colon f = f_1$ has the weakly greatest power among all tests with size less than or equal to α. The likelihood ratio test rejects the null hypothesis when $\Lambda_x(H_0, H_1) > k$, where $\Pr[\Lambda_x(H_0,H_1) > k \mid H_0] = k$. We define $C \subseteq \mathcal{X}$ such that $\Lambda_x(H_0,H_1) > k$ for all $x \in C$. Again, we denote the size of the test as α, and the probability of a type II error as β, and we assume that there exists C such that $\Pr[x \in C \mid H_0] = \alpha$. Suppose we have another test which rejects the null hypothesis when $x \in \mathcal{D} \subseteq \mathcal{X}$, and that it has size $\alpha' \le \alpha$ and the probability of a type II error is β'. We need to show that $\beta' \ge \beta$. We have

α

Yes.

$$
\begin{aligned}
\beta - \beta' &= \sum_{x\in C^{\complement}} f_1(x) - \sum_{x\in\mathcal{D}^{\complement}} f_1(x) \\
&= \sum_{x\in C^{\complement}\cap\mathcal{D}} f_1(x) + \sum_{x\in C^{\complement}\cap\mathcal{D}^{\complement}} f_1(x) - \sum_{x\in\mathcal{D}^{\complement}\cap C} f_1(x) - \sum_{x\in\mathcal{D}^{\complement}\cap C^{\complement}} f_1(x) \\
&= \sum_{x\in C^{\complement}\cap\mathcal{D}} f_0(x)\,\underbrace{\Lambda_x(H_0,H_1)}_{\le k} - \sum_{x\in\mathcal{D}^{\complement}\cap C} f_0(x)\,\underbrace{\Lambda_x(H_0,H_1)}_{>k} \\
&\le k\left[\sum_{x\in C^{\complement}\cap\mathcal{D}} f_0(x) - \sum_{x\in\mathcal{D}^{\complement}\cap C} f_0(x)\right] \\
&= k\left[\sum_{x\in C^{\complement}\cap\mathcal{D}} f_0(x) + \sum_{x\in C\cap\mathcal{D}} f_0(x) - \sum_{x\in\mathcal{D}^{\complement}\cap C} f_0(x) - \sum_{x\in C\cap\mathcal{D}} f_0(x)\right] \\
&= k\left[\sum_{x\in\mathcal{D}} f_0(x) - \sum_{x\in C} f_0(x)\right] \\
&= k(\alpha' - \alpha) \le 0
\end{aligned}
$$

which shows that $\beta \le \beta'$ and that the likelihood-ratio test has weakly greater power $1-\beta \ge 1-\beta'$.

Yes.

Question 6

Testing that sex and eye colour are independent amounts to testing that the joint probability mass function is the product of the marginal probability mass functions, or

$$
\begin{aligned}
H_0 &: \Pr(A_i = a, B_i = b) = \Pr(A_i = a)\Pr(B_i = b) = p_a p_b \\
H_1 &: \Pr(A_i = a, B_i = b) = p_{ab}
\end{aligned}
$$

where $\sum_b \sum_a p_{ab} = 1$. and $\sum_a p_a = 1$ and $\sum_b p_b = 1$

Letting n_{ab} be the number of observations in both category $a \in A$ and category $b \in B$, the log-likelihood ratio is

$$
\ln \Lambda_x(H_0,H_1) = \ln \frac{\prod_b \prod_a p_{ab}^{n_{ab}}}{\prod_b \prod_a (p_a p_b)^{n_{ab}}} = \sum_b \sum_a n_{ab} \ln p_{ab} - \sum_b \sum_a n_{ab}(\ln p_a + \ln p_b)
$$

Yes.

Although supremum of parameters has to be mentioned for a likelihood ratio.

MLE ~~estimators~~ for H_0 (at least) could be ~~shown~~ derived by Lagrangian methods. Yes.

The first term is maximised by setting $p_{ab} = \frac{n_{ab}}{n}$ where $n = \sum_b \sum_a n_{ab}$. The second term is maximised by setting $p_a = \frac{n_a}{n}$ and $p_b = \frac{n_b}{n}$, where $n_a = \sum_b n_{ab}$ and $n_b = \sum_a n_{ab}$. Using the same test statistic as before, and letting p_M and p_{Bl} be the MLE for the probability one is male or has blue eyes under H_1, we have

$$p_M = \frac{29}{59},\ p_{Bl} = \frac{28}{59}$$

$$e_{M,Bl} = \frac{29}{59} \times \frac{28}{59} \times 59,\ e_{M,Br} = \frac{29}{59} \times \frac{31}{59} \times 59,\ e_{F,Bl} = \frac{30}{59} \times \frac{28}{59} \times 59,\ e_{F,Br} = \frac{30}{59} \times \frac{31}{59} \times 59$$

$$T = \sum_b \sum_a \frac{(o_{ab} - e_{ab})^2}{e_{ab}}$$

$$= \frac{\left(19 - \frac{812}{59}\right)^2}{\frac{812}{59}} + \frac{\left(10 - \frac{899}{59}\right)^2}{\frac{899}{59}} + \frac{\left(9 - \frac{840}{59}\right)^2}{\frac{840}{59}} + \frac{\left(21 - \frac{930}{59}\right)^2}{\frac{930}{59}} \approx 7.460$$ Yes.

Under H_1, there are 3 free parameters, and under H_0, there are 2 free parameters, so the difference in dimensionality between H_1 and H_0 is 1, and a χ_1^2 test is appropriate. The 5% critical value is 3.841, so the null hypothesis is rejected. Yes.

To test that each of the cell probabilities is 1/4, we now have $e_{ab} = \frac{59}{4}$ for all a, b. The difference in dimensionality between H_1 and H_0 is now 3 since the parameters under H_0 are fully specified, and the test statistic is

$$T = \frac{(19 - 14.75)^2}{14.75} + \frac{(10 - 14.75)^2}{14.75} + \frac{(9 - 14.75)^2}{14.75} + \frac{(21 - 14.75)^2}{14.75} \approx 7.644$$ Yes.

and the 5% critical value is now 7.815, which means the null hypothesis is not rejected this time. The null hypothesis H_0' that all the cell probabilities are equal to 1/4 is nested within the original null hypothesis H_0 that sex and eye colour are independent; H_0 implies H_0'. Therefore the H_0 is a stronger statement, and there is no contradiction in our tests rejecting stronger hypothesis H_0 while failing to reject the weaker hypothesis H_0'.

Need more data to reject H_0'?

H_0' is more specific than H_0. Look at the parameter spaces for both hypotheses.

Question 7

We have a two-way contingency table with r rows and c columns, and n_{ij} indicates the entry in the i^{th} row and j^{th} column. For a test of homogeneity, there is the assumption that the row totals $\sum_{j=1}^{c} n_{ij} = n_{i+}$ for some pre-determined values of n_{i+}. Under the model, the n_{ij} are distributed like so:

$$(n_{i1}, \ldots, n_{ic}) \sim \text{Multinomial}(n_{i+}, p_{i1}, \ldots, p_{ic}) \text{ independently for } i \in \{1, \ldots, r\}$$ Yes.

In testing for homogeneity down the rows, we have

$$H_0 : p_{1j} = p_{2j} = \ldots = p_{rj} = p_j,\ j \in \{1, \ldots, c\}$$ and $\sum p_j = 1$

$$H_1 : p_{ij} \text{ are unrestricted}$$ not entirely and $\sum_j p_{ij} = 1$ for $\forall i$

Under H_0, the likelihood function is

$$\prod_{i=1}^{r} \frac{n_{i+}!}{n_{i1}! \times \ldots \times n_{ic}!} \times p_1^{n_{i1}} \times \ldots \times p_c^{n_{ic}}$$ Yes.

and the log-likelihood is

$$\ln A + \sum_{j=1}^{c} n_{+j} \ln p_j$$

for some constant A and where $n_{+j} = \sum_{i=1}^{r} n_{ij}$. To maximise this subject to the constrant that $\sum_{j=1}^{c} p_j = 1$, we have the Lagrangian

$L(p, \lambda)$

$$\mathscr{L} = \ln A + \sum_{j=1}^{c} n_{+j} \ln p_j - \lambda \left(\sum_{j=1}^{c} p_j - 1 \right)$$

Better to keep the 'j' index.

and the first-order condition implies

$$\frac{n_{+j}}{p_j} = \lambda \implies p_j = \frac{n_{+j}}{n_{+k}} p_k \implies \sum_{j=1}^{c} \frac{n_{+j}}{n_{+k}} p_k = \frac{p_k}{n_{+k}} n = 1 \implies p_k = \frac{n_{+k}}{n}$$

Yes.

Therefore, the MLE under H_0 is $\hat{p}_j = \frac{n_{+j}}{n}$ for $j = 1, \ldots, c$. Going through similar steps, the MLE under H_1 is $\hat{p}_{ij} = \frac{n_{ij}}{n_{i+}}$ for $i = 1, \ldots, r$ and $j = 1, \ldots, c$. The log-likelihood ratio is

Yes.

$$\begin{aligned}
\ln \Lambda_x(H_0, H_1) &= \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \ln \hat{p}_{ij} - \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \ln \hat{p}_j \\
&= \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \ln \frac{\hat{p}_{ij}}{\hat{p}_j} \\
&= \sum_{i=1}^{r} \sum_{j=1}^{c} n_{ij} \ln \frac{n_{ij}}{n_{i+} n_{+j} / n}
\end{aligned}$$

which is equivalent to that under the test for independence, and the same approximation can be used to justify Pearson's chi-squared test. Yes.

From the clinical trial data, we have

$$e_{iI} = \frac{63}{150} \times 50 = 21, \ e_{iN} = \frac{40}{150} \times 50 = \frac{40}{3}, \ e_{iW} = \frac{47}{150} \times 50 = \frac{47}{3}$$

for $i = 1, \ldots, r$ and where I, N, and W refer to the categories 'Improved', 'No difference', and 'Worse'. Therefore,

$$\begin{aligned}
T = {} & \frac{(18-21)^2}{21} + \frac{(20-21)^2}{21} + \frac{(25-21)^2}{21} \\
& + \frac{\left(17 - \frac{40}{3}\right)^2}{\frac{40}{3}} + \frac{\left(10 - \frac{40}{3}\right)^2}{\frac{40}{3}} + \frac{\left(13 - \frac{40}{3}\right)^2}{\frac{40}{3}} \\
& + \frac{\left(15 - \frac{47}{3}\right)^2}{\frac{47}{3}} + \frac{\left(20 - \frac{47}{3}\right)^2}{\frac{47}{3}} + \frac{\left(12 - \frac{47}{3}\right)^2}{\frac{47}{3}} \approx 5.173
\end{aligned}$$

Yes.

The difference in dimensionality between H_1 and H_0 is 6, and the 5% critical value for a χ^2_6 distribution is 9.49, so the null hypothesis is not rejected.

Should be 4

H_0 is not rejected indeed.

Question 8

The likelihood ratio is

$$\Lambda_x(H_0, H_1) = \frac{\prod_{i=1}^n \theta_1 e^{-\theta_1 x_i}}{\prod_{i=1}^n \theta_0 e^{-\theta_0 x_i}} = \left(\frac{\theta_1}{\theta_0}\right)^n \exp\left[(\theta_0 - \theta_1)\sum_{i=1}^n x_i\right]$$

Yes.

The likelihood ratio is monotonically decreasing in the statistic $T(X) = \sum_{i=1}^n X_i \sim \Gamma(n, \theta)$. Therefore, a likelihood ratio test of size α is equivalent to one with $T(X)$ as the test statistic, which rejects the null hypothesis when $T(X) < k$. We require that

Yes.

$$\Pr[T(X) < k \mid H_0] = \int_0^k \frac{\theta_0^n}{\Gamma(n)} y^{n-1} e^{-\theta_0 y}\, \mathrm{d}y = \alpha$$

Assuming we have such k, the power function can be expressed as

$$W(\theta) = \Pr[T(X) < k \mid \theta] = \int_0^{k(\theta_0)} \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y}\, \mathrm{d}y$$

Yes.

where the dependence on θ_0 is implicit through k. For a test of H_0 against H_1 to be uniformly most powerful of size α, we must have

i. $\sup_{\theta \in \Theta_0} W(\theta) = \alpha$

ii. For any other test with size $\le \alpha$ and power function W^*, we have $W(\theta) \ge W^*(\theta)$ for all $\theta \in \Theta_1$

Yes.

When testing $H_0: \theta = \theta_0$ against $H_1: \theta > \theta_0$, we have $\Theta_0 = \theta_0$ so the size condition is satisfied since $W(\theta_0) = \alpha$ by definition. For any other test with size $\alpha^* \le \alpha$ and power function W^*, the Neyman–Pearson lemma implies that $W(\theta_1) \ge W^*(\theta_1)$ for all $\theta_1 \in \Theta_1$, so the likelihood ratio test is the uniformly most powerful test.

Yes.

Neyman–Pearson lemma works for simple hypotheses – there has to be specialisation and then returning to $H_1: \theta > \theta_0$.

When testing $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$, we use the same test and critical region as the one we used when testing $H_0': \theta = \theta_0$ against $H_1': \theta = \theta_1$, with $\theta_1 > \theta_0$. We have to check that the size condition is satisfied. Fixing the value of k as the one defined before $k(\theta_0)$, we have

$$\begin{aligned}
\frac{\partial}{\partial\theta} W(\theta) &= \int_0^k \left[\frac{n\theta^{n-1}}{\Gamma(n)} y^{n-1} e^{-\theta y} - \frac{\theta^n}{\Gamma(n)} y^n e^{-\theta y}\right] \mathrm{d}y \\
&= \frac{n}{\theta}\int_0^k \frac{\theta^n}{\Gamma(n)} y^{n-1} e^{-\theta y}\, \mathrm{d}y - \frac{n}{\theta}\int_0^k \frac{\theta^{n+1}}{\Gamma(n+1)} y^n e^{-\theta y}\, \mathrm{d}y \\
&= \frac{n}{\theta}\left[\Pr\left(\sum_{i=1}^n X_i < k\right) - \Pr\left(\sum_{i=1}^{n+1} X_i < k\right)\right] > 0
\end{aligned}$$

Yes.

both under H_0

Yes.

which means $W(\theta)$ is increasing in θ and $\sup_{\theta \le \theta_0} W(\theta) = W(\theta_0) = \alpha$. We now have to satisfy the second condition. Suppose there is some other test of $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$ with critical region C^* and power function W^*, such that $\sup_{\theta \le \theta_0} W^*(\theta) \le \alpha$. We need to show that $W(\theta_1) \ge W^*(\theta_1)$ for all $\theta_1 > \theta_0$. Since $\sup_{\theta \le \theta_0} W^*(\theta) \le \alpha$, it must be that $W^*(\theta_0) \le \alpha$. This means if C^* is set as the critical region for a test of $H_0': \theta = \theta_0$ against $H_1': \theta = \theta_1$, it has size $\le \alpha$. Then, by the Neyman–Pearson lemma, $W^*(\theta_1) \le W(\theta_1)$, and this applies for all $\theta_1 > \theta_0$, which is what is needed. Therefore the simple test of $H_0: \theta = \theta_0$ against $H_1: \theta = \theta_1$ is also the uniformly most powerful test for the composite test of $H_0: \theta \le \theta_0$ against $H_1: \theta > \theta_0$.

~~That~~ has to be

Yes.

any θ from the H_0 region has to be taken here

Question 9

With $X \sim N(0,1)$ and $Y \sim \chi^2_n$, we have

$$f_{X,Y}(x,y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}}$$

Yes.

We have $T = \frac{X}{\sqrt{Y/n}}$, and applying the change of variables $g(X,Y) = (U,V)$ where $U = T, V = Y$, we have $(X,Y) = g^{-1}(U,V) = (U\sqrt{V/n}, V)$, and

$\sqrt{\frac{V}{n}}$?

$$\det(J) = \det\begin{pmatrix} \sqrt{U/n} & \frac{U}{2\sqrt{nV}} \\ 0 & 1 \end{pmatrix} = \frac{\sqrt{U}}{\sqrt{n}}$$

$\frac{\sqrt{V}}{\sqrt{n}}$?

And we get

$$\begin{aligned}
f_T(t) &= \int_0^\infty f_{T,Y}(t,y)\,\mathrm{d}y \\
&= \int_0^\infty f_{X,Y}(g^{-1}(x,y))\det(J)\,\mathrm{d}y \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2(y/n)}{2}} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} \frac{\sqrt{y}}{\sqrt{n}}\,\mathrm{d}y \\
&= \frac{1}{2^{\frac{n+1}{2}} n^{\frac{1}{2}} \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^\infty \exp\left[-\frac{\left(\frac{t^2}{n}+1\right)}{2} y\right] y^{\frac{n+1}{2}-1}\,\mathrm{d}y \\
&= \frac{1}{2^{\frac{n+1}{2}} n^{\frac{1}{2}} \pi^{\frac{1}{2}} \Gamma\left(\frac{n}{2}\right)} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\left[\frac{\left(\frac{t^2}{n}+1\right)}{2}\right]^{\frac{n+1}{2}}} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{(n\pi)^{\frac{1}{2}}} \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}
\end{aligned}$$

u, v

dv

Yes.

Yes.

Yes.

The variable use could be more consistent.

as needed.

Question 10

under which hypothesis? Correct for H_1 here.

The MLE of μ and σ^2 are $\bar{X}$ and $n^{-1}S_{XX}$, so the likelihood ratio is

$$\Lambda_x(H_0,H_1) = \frac{\sup_{(\mu,\sigma^2)\in\mathbb{R}\times\mathbb{R}_+} f(x\mid\mu,\sigma^2)}{\sup_{\sigma^2\in\mathbb{R}_+} f(x\mid\mu_0,\sigma^2)} = \frac{f(x\mid\bar{x}, n^{-1}s_{xx})}{\sup_{\sigma^2\in\mathbb{R}_+} f(x\mid\mu_0,\sigma^2)}$$

This notation is not particularly usual - rather estimates are plugged into a general likelihood function, but parameter values could be different from estimates.

To find the supremum in the denominator, we maximise the following

$$-\frac{n}{2}\ln(2\pi) - n\ln\sigma - \frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2\sigma^2}$$

Yes.

Better to leave σ^2 here as a parameter.

Derivatives can also be taken with respect to σ^2 instead of σ.

We get the following from the first-order condition

$$-\frac{n}{\sigma}+\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{\sigma^3}=0 \implies \sigma^2 = n^{-1}\sum_{i=1}^n (x_i-\mu_0)^2$$

Yes.

Therefore,

$$\Lambda_x(H_0,H_1)=\frac{\prod_{i=1}^n (2\pi n^{-1}s_{xx})^{-\frac{1}{2}}\exp\left[-\frac{(x_i-\bar{x})^2}{2n^{-1}s_{xx}}\right]}{\prod_{i=1}^n \left[2\pi n^{-1}\sum_{i=1}^n (x_i-\mu_0)^2\right]^{-\frac{1}{2}}\exp\left[-\frac{(x_i-\mu_0)^2}{2n^{-1}\sum_{i=1}^n (x_i-\mu_0)^2}\right]}$$
$$=\left[\frac{\sum_{i=1}(x_i-\mu_0)^2}{s_{xx}}\right]^{\frac{n}{2}}\exp\left[\frac{\sum_{i=1}^n (x_i-\mu_0)^2}{2n^{-1}\sum_{i=1}^n (x_i-\mu_0)^2}-\frac{s_{xx}}{2n^{-1}s_{xx}}\right]$$
$$=\left[\frac{\sum_{i=1}(x_i-\bar{x}+\bar{x}-\mu_0)^2}{s_{xx}}\right]^{\frac{n}{2}}$$

Yes.

$$=\left[\frac{s_{xx}+n(\bar{x}-\mu_0)^2}{s_{xx}}\right]^{\frac{n}{2}}$$

What happens to the interaction terms here?

$$=\left[1+\frac{n(\bar{x}-\mu_0)^2}{s_{xx}}\right]^{\frac{n}{2}}=\left[1+\frac{n(\bar{x}-\mu_0)^2}{s_{xx}/(n-1)}\times\frac{1}{n-1}\right]^{\frac{n}{2}}=\left(1+\frac{T^2}{n-1}\right)^{\frac{n}{2}}$$

Yes.

We have that

$$T=\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sqrt{S_{XX}/(n-1)}}=\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\frac{1}{\sqrt{\frac{S_{XX}}{\sigma^2}/(n-1)}}$$

Under the null hypothesis, $\frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma}\sim N(0,1)$, while $\frac{S_{XX}}{\sigma^2}\sim\chi^2_{n-1}$ and both are independent of one another. Therefore, like the previous question, $T\sim t_{n-1}$ and the size α likelihood ratio test rejects the null hypothesis when $|T|>t_{n-1}(\alpha/2)$.

Yes.

Mean yields some information, which can be used under some models.

Question 11

For $X_1,\dots,X_n$, we have as before $\frac{S_{XX}}{\sigma^2}\sim\chi^2_{n-1}$ and S_{XX} is independent of $\bar{X}$ (so the values of $\bar{X}$ yield no information, and it's as though we observed nothing). We want to compare the following value between the two statisticians:

$$\Pr\left(\frac{\frac{1}{n-1}S_{XX}}{\sigma^2}>1.5\right)=\Pr\left[\frac{S_{XX}}{\sigma^2}>1.5(n-1)\right]$$

Yes.

where we replace X with Y for statistician B. and adjust n accordingly

For statistician A, we want to find the probability a χ^2_9 variable exceeds 13.5, and for statistician B, we want to find the probability a χ^2_{16} variable exceeds 24. The probabilities are approximately 0.1426 and 0.08950, so $S_{XX}/9$ is more likely to have exceeded the true value by more than 50%.

Yes.

Question 12

We can redefine the test function $\varphi(x)$ such that

$$\varphi(x) = \begin{cases} 1 & \text{if } \Lambda_x(H_0, H_1) > k \\ \gamma & \text{if } \Lambda_x(H_0, H_1) = k \\ 0 & \text{if } \Lambda_x(H_0, H_1) < k \end{cases}$$

which means φ rejects the null hypothesis if the likelihood ratio exceeds the critical value, does not reject if the likelihood ratio is below the critical value, and rejects with probability γ when $\Lambda_x = k$ where γ is constructed to yield

$$\mathbb{E}[\varphi(x) \mid H_0] = \alpha$$

which gives us the test of the required size. Yes.

1. Can look for a test $C = \{X : X > k\}$ instead of the likelihood ratio test as the ratio is an increasing function in X.

a) For the test of $\mu = 0$ against $\mu = 4$:

$$\Lambda_X(H_0, H_1) = \exp(4X - 8)$$

Test: $C = \{X : X > k\}$ is suitable due to increasing function in X.

$P(X > k) = 0.05$ under H_0

$\Rightarrow P(X \leq k) = 0.95$

$\Rightarrow k = \Phi^{-1}(0.95)$

because we assume a random variable distributed as $N(0, 1)$.

$\Rightarrow$ The same numbers.

$k = 1.645$ for $\alpha = 0.05$ and

$k = 2.326$ for $\alpha = 0.01$.

b) $\Lambda_X(H_0, H_1) = \exp(8 - 4X)$

from the second case

Decreasing function in x.

Test $C = \{x : x < k\}$.

$P(X < k) = \alpha$ under H_0

$P(X-4 < k-4) = \cancel{\ } \Phi$

should correspond to $N(0,1)$ case

$$k - 4 = \Phi^{-1}(\alpha)$$

Then $k = 2.355$ and

$k = 1.674$ for $\alpha = 0.01$

In both cases reject H_0 only at ~~$\alpha = 0.01$~~ level

$\alpha = 0.05$

3. To simplify a likelihood test can take a derivative of the ratio function with respect to

$$\left|T - \frac{1}{2}\right|.$$

Alternatively, omit the power n and simplify the function.

6. Larger degrees of freedom for a test will require more data.
Also $\Theta \setminus \theta_0$ parameter space can be broad and not precise enough for a meaningful test.

7. Degrees of freedom and likelihood ratios have to be the same to obtain analogous Pearson's chi-squared test.
Degrees of freedom for independence test:

$$|\Theta_1| = r \cdot c - 1$$

$$|\theta_0| = (\cancel{r-1})(\cancel{c-1}) \quad r-1+c-1$$

$$|\theta_1| - |\theta_0| = \cancel{r \cdot c - 1 - (r-1)(c-1)} =$$

$$= \cancel{r c - 1 - r c + r + c - 1}$$

$$r \cdot c - 1 - r + 1 - c + 1 =$$

$$= r \cdot c - r - c + 1$$

Degrees of freedom for the trial test:

$$|\theta_1| = r \cdot c - r$$

$$|\theta_0| = c - 1$$

$$|\theta_1| - |\theta_0| = r \cdot c - r - c + 1$$

If $\sum_{i=1}^{n} X_i \sim \text{Gamma}(n, \theta_0)$

then $\theta_0 \sum_{i=1}^{n} X_i \sim \text{Gamma}(n, 1)$

11. Can use non-central χ^2 distribution to exploit $\overline{X}$ and $\overline{Y}$ values.

Non-central χ^2 arises from $N(\mu_i, \sigma^2)$ distributed variables.

Estimate general μ as

$$\hat{\mu} = \frac{10 \cdot \overline{X} + 17 \cdot \overline{Y}}{27} \approx 5.7$$

$$S_{XX} \sim \sum_{i=1}^{10} N(-0.2, \sigma^2)^2 = \\ = \sigma^2 \cdot \chi^2_{10}\left(\sum_{i=1}^{10} (0.2)^2\right)$$

as non-centrality parameter is

$$\sum_{i=1}^{n} \mu_i^2$$

$$S_{YY} \sim \sum_{i=1}^{17} N(0.1, \sigma^2)^2 = \sigma^2 \chi^2_{17}\left(\sum_{i=1}^{17} (0.1)^2\right)$$

Non-centrality parameter is essential to battle with inflated significance levels

Here, it yields to a larger power of detecting deviations and also to increased difference between both samples.