Imperial College London

Problem Sheet 4 Solutions

MATH50011
Statistical Modelling 1

Week 5

Lecture 9 (Hypothesis Testing)

1. A clinical trial is conducted to establish whether treatment A has a different effect on systolic blood pressure than treatment B. The treatment effect will be measured by comparing the mean of a group receiving A and a group receiving B. Write a few sentences suitable for a non-statistician (better yet, a non-mathematician!) explaining what is meant by the terms *type I error* and *type II error* in this context.

   Solution. Here, a "different effect" most likely coincides with testing $H_0 : \mu_A = \mu_B$ against $H_1 : \mu_A \neq \mu_B$ where μ_A and μ_B are the true mean SBP for each group.

   A type I error would mean that we reject H_0 when *we should not have*. In this context we can tell a non-statistician that this corresponds to deciding erroneously to claim treatment B has an effect on SBP different than treatment A.

   A type II error means that we do not reject H_0 when *we should have*. This would mean that we act as if there were no difference between treatment A and B even though there is a true difference.

   Note that in applications we can never be certain that we have or have not made either of these errors (or a correct decision), since they are defined with respect to the unknown parameters.

2. In the results of a study, you read the following sentence:

   > "Since $p < 0.05$, there is a less than 5% chance that the null hypothesis is true."

   Write an explanation of why this statement is false. Give a correct interpretation of a p-value suitable for a non-statistician.

   Solution. This statement is wrong because a p-value is not a probability statement about the null hypothesis. Rather, the p-value is a statement about observing the data when the null hypothesis is true. A correct interpretation would be a statement such as, "If the null hypothesis is true, the probability of observing a test statistic at least as extreme as this in repeated sampling is less than 5%."

3. The mean level of prothrombin in the general population is known to be 22.0 mg/100 ml of plasma. A sample of 30 patients showing vitamin K deficiency has a mean prothrombin level of 19.5 mg/100ml and standard deviation 4 mg/100ml of plasma.

(a) Is the mean prothrombin level in patients with vitamin K deficiency different from that in the general population? Set up a null and alternative hypothesis that addresses this question.

Solution. Here, we have $H_0 : \mu = 22.0$ mg/100 ml of plasma vs $H_1 : \mu \neq 22.0$ mg/100 ml of plasma.

(b) Find the critical values for the sample mean to test the hypotheses in (a) using $\alpha = .05$. Use these critical values to test the hypotheses in (a); state your statistical and scientific conclusions. Clearly state the test statistic and distribution you are using.

Solution. We do not assume that the individual prothrombin levels are normally distributed in the problem statement. However, we can still use a *normal approximation* based on the central limit theorem:

$$\bar{X} \sim N(\mu, \sigma^2/n) = N(22.0, \sigma^2/30)$$

This means that, approximately,

$$\frac{\bar{X} - 22.0}{\sigma/\sqrt{30}} \sim N(0,1)$$

Note that σ (the standard deviation of the population) is unknown, but we have the sample standard deviation, that is $S = 4$. Thus, we have

$$\frac{\bar{X} - 22.0}{4\sqrt{30}} \sim t_{n-1} = t_{29}$$

For a two-sided test, as in (a), we have that the critical values based on the student's t-distribution with 29 degrees of freedom at level $\alpha = 0.05$ is $\pm t_{29,0.975} \approx \pm 2.045$. The test is equivalent to rejecting H_0 if 22.0 is not contained in the 95% confidence interval with upper/lower confidence limits given by

$$\bar{x} \pm 2.045 \times 4/\sqrt{30}.$$

For the given data, we have $\bar{x} = 19.5$ mg/100ml. This leads to the 95% confidence interval

$$19.5 \pm 2.045 \times 4/\sqrt{30} \implies 95\% \text{ CI: } 18 \text{ mg/100ml to } 21 \text{ mg/100ml}$$

which does not contain 22.0 mg/100ml. Hence, we reject H_0 at the $\alpha = 0.05$ level. We conclude based on this evidence that mean prothrombin level in patients with vitamin K deficiency differs from that in the general population.

(c) Compute the p-value and use that to test the hypotheses in (a) using $\alpha = .05$. State both your statistical and scientific conclusions.

Solution. The two-sided test has a p-value of

$$p = 2P\left(Z > \left|\frac{\bar{x} - 22.0}{4/\sqrt{30}}\right|\right) \approx 0.0018$$

where $Z \sim t_{29}$. We find that $p = 0.0018$, and reject H_0 at the $\alpha = 0.05$ level. We conclude based on this evidence that mean prothrombin level in patients with vitamin K deficiency differs from that in the general population.

4. Suppose that, in the general population, birth weights are approximately normally distributed with a mean weight of 3200g and a standard deviation of 400g. A sample of 25 babies born to teenage mothers has an average birth weight of 2980g.

   (a) You would like to use this sample to determine if the average birth weight of babies born to teenage mothers is different from the general population. Set up a null and alternative hypothesis that addresses this question and carry out a hypothesis test using $\alpha = .05$. Be sure to state both your statistical and scientific conclusions.

   Solution. Here, $H_0 : \mu = 3200$g and $H_1 : \mu \neq 3200$g. We have that the two-sided test has a p-value of

   $$p = 2P\left(Z > \left|\frac{2980 - 3200}{400/\sqrt{25}}\right|\right) = 2P\left(Z > 2.75\right) = 0.005959526$$

   where $Z \sim N(0, 1)$. We reject the null hypothesis at the $\alpha = 0.05$ level. This means that based on these data, we find evidence that birth weight of babies born to teenage mothers is different from the general population.

   (b) Give a 95% confidence interval for the mean weight of babies born to teenage mothers.

   Solution. A corresponding two-sided 95% confidence interval is given by

   $$2980 \pm z_{0.975} 400/\sqrt{25} \implies 95\% \text{ CI: } 2823.2\text{g to } 3136.8\text{g}$$

   (c) Describe how you could use the confidence interval from (b) to test the hypotheses in (a).

   Solution. We note that since the value defining the null hypothesis, 3200g, is not contained in the 95% confidence interval for μ that we reject H_0 at the $\alpha = 0.05$ level.

5. In this exercise, we consider a typical *sample size calculation*. These types of methods are common during the planning of a study.

   Let $X_1, \dots, X_n$ be a random sample from a $N(\theta, \sigma^2)$ population with σ^2 known. We consider testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$. We will reject H_0 if $(\bar{X} - \theta_0)/(\sigma/\sqrt{n}) > c$.

   (a) Find the power function $\beta(\theta)$ of the test. Express your answer in terms of the standard normal cdf $\Phi(z)$;

Solution. The power function of this test is

$$\begin{aligned}\beta(\theta) &= P_\theta\left(\frac{\bar{X}-\theta_0}{\sigma/\sqrt{n}} > c\right)\\ &= P_\theta\left(\frac{\bar{X}-\theta}{\sigma/\sqrt{n}} > c + \frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right)\\ &= P\left(Z > c + \frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right)\\ &= 1-\Phi\left(c + \frac{\theta_0-\theta}{\sigma/\sqrt{n}}\right)\end{aligned}$$

where $Z \sim N(0,1)$ has a distribution that is free of θ.

(b) Define the value c_α of c such that the test has level α;

Solution. As θ increases, we see that $\beta(\theta) \to 1$ monotonically. Hence, it suffices to find c_α such that $\beta(\theta_0) = \alpha$. In particular, we have that

$$\beta(\theta_0) = 1-\Phi\left(c + \frac{\theta_0-\theta_0}{\sigma/\sqrt{n}}\right) = 1-\Phi(c) \equiv \alpha$$

so that $c_\alpha = \Phi^{-1}(1-\alpha)$.

(c) Find values c and n such that the test has level α and $\beta(\theta) \geq b$ for all $\theta \geq \theta_0 + \sigma$.

Solution. We take $c_\alpha = \Phi^{-1}(1-\alpha) \equiv z_{1-\alpha}$ as before, since this does not depend on n. Since $\beta(\theta)$ is increasing, the requirement $\beta(\theta) \geq b$ for all $\theta \geq \theta_0 + \sigma$ will be met if n is chosen so that $\beta(\theta_0+\sigma) = b$. We have

$$\beta(\theta_0+\sigma) = 1-\Phi\left(c_\alpha + \frac{\theta_0-\theta_0-\sigma}{\sigma/\sqrt{n}}\right) = 1-\Phi(c_\alpha - \sqrt{n}) \equiv b$$

Solving for n yields

$$n = (c_\alpha - \Phi^{-1}(1-b))^2 = (z_{1-\alpha} - z_{1-b})^2$$

which can be interpreted as squaring the difference between the $1-\alpha$ and $1-b$ quantiles of the $N(0,1)$ distribution.

6. Binomial data gathered from more than one population are often presented in a *contingency table*. For the case of two populations, the table might look like this:

	Population		
	1	2	Total
Successes	S_1	S_2	$S = S_1 + S_2$
Failures	F_1	F_2	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

where Population 1 is Binomial(n_1, p_1), with S_1 successes and F_1 failures, Population 2 is Binomial(n_2, p_2), with S_2 successes and F_2 failures, and S_1 and S_2 are independent.

We consider testing the hypothesis that $H_0 : p_1 = p_2$ against $H_1 : p_1 \neq p_2$.

(a) Consider the statistic

$$W = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}\,(1-\hat{p})}$$

where $\hat{p}_k = S_k/n_k$ for $k = 1, 2$ and $\hat{p} = S/n$.

Show that $W \to_d \chi_1^2$ as $n_1, n_2 \to \infty$ with $n_1/n_2 \to \gamma \in (0, \infty)$. Explain how an approximate level α test can be constructed based on W.

Solution. Note $\hat{p}(1-\hat{p}) \to_p p(1-p)$ where $p = p_1 = p_2$ is the common value under H_0. By the central limit theorem we have that both

$$\frac{\sqrt{n_1}(\hat{p}_1 - p)}{\sqrt{p(1-p)}} \to_d N(0,1) \qquad \textit{and} \qquad \frac{\sqrt{n_2}(\hat{p}_2 - p)}{\sqrt{p(1-p)}} \to_d N(0,1).$$

Now, considering the statistic of interest, we will write $W = T^2$ with

$$\begin{aligned}
T &= \frac{\hat{p}_1 - p}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}\,(1-\hat{p})}} - \frac{\hat{p}_2 - p}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{p}\,(1-\hat{p})}} \\
&= \frac{\sqrt{n_1}(\hat{p}_1 - p)}{\sqrt{\left(1 + \frac{n_1}{n_2}\right)\hat{p}\,(1-\hat{p})}} - \frac{\sqrt{n_2}(\hat{p}_2 - p)}{\sqrt{\left(\frac{n_2}{n_1} + 1\right)\hat{p}\,(1-\hat{p})}} \\
&= \frac{\sqrt{(1+\gamma)p(1-p)}}{\sqrt{\left(1 + \frac{n_1}{n_2}\right)\hat{p}\,(1-\hat{p})}} \frac{\sqrt{n_1}(\hat{p}_1 - p)}{\sqrt{(1+\gamma)p(1-p)}} - \frac{\sqrt{(1+\gamma^{-1})p(1-p)}}{\sqrt{\left(1 + \frac{n_2}{n_1}\right)\hat{p}\,(1-\hat{p})}} \frac{\sqrt{n_2}(\hat{p}_2 - p)}{\sqrt{(1+\gamma^{-1})p(1-p)}} \\
&\to_d 1 \cdot \frac{1}{\sqrt{1+\gamma}} Z_1 + 1 \cdot \frac{1}{\sqrt{1+\gamma^{-1}}} Z_2 = N(0,1)
\end{aligned}$$

where Z_1 and Z_2 are independent $N(0,1)$ random variables. This convergence follows by Slutsky's lemma where $n_1/n_2 \to \gamma$, so that the variance in the last equality is indeed

$$\frac{1}{1+\gamma} + \frac{1}{1+\gamma^{-1}} = 1.$$

Since $T \to_d N(0,1)$ we have $T^2 \to_d \chi_1^2$ as $n_1, n_2 \to \infty$, as claimed.

(b) We may alternatively measure the departure from H_0 in terms of the difference between the observed frequencies S_1, S_2, F_1, F_2 and the *expected frequencies*:

Expected Frequencies

	1	2	Total
Successes	$n_1 S/n$	$n_2 S/n$	$S = S_1 + S_2$
Failures	$n_1 F/n$	$n_2 F/n$	$F = F_1 + F_2$
Total	n_1	n_2	$n = n_1 + n_2$

Consider the statistic

$$W^* = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

where the sum is taken across all 4 cells of the tables.

Show that $W^* = W$ so that W^* is also asymptotically chi squared.

(This is the most common form of the *chi squared test of independence*.)

Solution. Substituting $\hat{p}_i$s for S_is and F_is gives

$$\begin{aligned} W^* &= \frac{n_1^2(\hat{p}_1 - \hat{p})^2}{n_1\hat{p}} + \frac{n_n^2(\hat{p}_2 - \hat{p})^2}{n^2\hat{p}} \\ &\quad \frac{n_1^2[(1-\hat{p}_1) - (1-\hat{p})]^2}{n_1(1-\hat{p})} + \frac{n_2^2[(1-\hat{p}_2) - (1-\hat{p})]^2}{n_2\hat{p}} \\ &= \frac{n_1(\hat{p}_1 - \hat{p})^2}{\hat{p}(1-\hat{p})} + \frac{n_2(\hat{p}_2 - \hat{p})^2}{\hat{p}(1-\hat{p})} \end{aligned}$$

Now, write $\hat{p} = (n_1\hat{p}_1 + n_2\hat{p}_2)/(n_1 + n_2)$. Substitute this into the numerator, and some algebra yields

$$n_1(\hat{p}_1 - \hat{p})^2 + n_2(\hat{p}_2 - \hat{p})^2 = \frac{(\hat{p}_1 - \hat{p}_2)^2}{\frac{1}{n_1} + \frac{1}{n_2}}$$

so that $W^* = W$, as desired.

7. A famous medical experiment was conducted by Joseph Lister in the late 1800s. In his experiment, Lister tested whether carbolic acid (a disinfectant) could reduce the risk of mortality following surgery. Data based on 75 amputations with and without the use of carbolic acid are presented in the following table:

		Carbolic acid used?	
		Yes	No
Patient lived?	Yes	34	19
	No	6	16

Use these data and the test you derived in the previous exercise to test whether the use of carbolic acid is associated with patient mortality.

You may use the `R` function `prop.test()` to check your by-hand solution.

Solution. Here $\hat{p}_1 = 34/40$, $\hat{p}_2 = 19/35$, and $\hat{p} = 53/75$ with $n_1 = 40$ and $n_2 = 35$.

Computing, e.g., W we find

$$w = \frac{(19/35 - 53/75)^2}{\left(\frac{1}{40} + \frac{1}{35}\right) 53/75(22/75)} \approx 8.49516$$

Noting that $P(\chi_1^2 > 8.5) \approx 0.00356$, we would likely decide to reject the null hypothesis that $p_1 = p_2$ in this case.

In `R`, we reproduce these calculations exactly using the command

```
prop.test(x=c(34,19),n=c(40,35),correct=F)
```

8. Suppose that $X_1, \dots, X_m$ and $Y_1, \dots, Y_n$ are random samples and X_i is independent of Y_j for all i and j. Suppose further that $X_i \sim N(\mu_x, \sigma_x^2)$ and $Y_j \sim N(\mu_y, \sigma_y^2)$ for all i and j. We consider testing $H_0 : \theta = 0$ against $H_1 : \theta \neq 0$ for $\theta = \mu_x - \mu_y$.

   (a) Assume that $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Derive the LRT for testing the hypotheses given above. Show that the LRT can be based on the statistic

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_P^2\left(\frac{1}{m} + \frac{1}{n}\right)}}$$

   where

$$S_P^2 = \frac{1}{m+n-2}\left(\sum_{i=1}^{m}(X_i - \bar{X})^2 + \sum_{j=1}^{n}(Y_i - \bar{Y})^2\right).$$

   The quantity S_P^2 is called the *pooled variance estimator*.

Solution. By definition, the likelihood ratio test statistic for this test is given by

$$\lambda(x, y) = \frac{\sup L(\mu_x, \mu_y, \sigma^2|x, y)}{\sup_{H_0} L(\mu_x, \mu_y, \sigma^2|x, y)}$$

First, we compute the MLE under H_0, that is when $\mu_x = \mu_y$. Under H_0, the X_is and Y_is are one sample of size $m + n$ from a $N(\mu, \sigma^2)$ population, where $\mu = \mu_x = \mu_y$. Hence,

$$\hat{\mu} = \frac{\sum_i X_i + \sum Y_i}{m+n} = \frac{m\bar{x} + n\bar{y}}{m+n}$$

and

$$\hat{\sigma}_0^2 = \frac{\sum_i (X_i - \hat{\mu})^2 + \sum_i (Y_i - \hat{\mu})^2}{m+n}.$$

For the unrestricted MLEs we use the likelihood

$$L(\mu_x, \mu_y, \sigma^2|x, y) = (2\pi\sigma^2)^{-(n+m)/2} \exp\{-[\sum_i (X_i - \mu_x)^2 + \sum_i (Y_i - \mu_y)^2]/2\sigma^2\}.$$

By inspection, maximizing over μ_x does not depend on μ_y and vice versa so that $\hat{\mu}_x = \bar{x}$ and $\hat{\mu}_y = \bar{y}$. Then,

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n+m}{2}\frac{1}{\sigma^2} + \frac{1}{2}\left[\sum_i (x_i - \hat{\mu}_x)^2 + \sum_i (y_i - \hat{\mu}_y)^2\right]\frac{1}{\sigma^4} \equiv 0$$

implies

$$\hat{\sigma}^2 = \left[\sum_i (x_i - \bar{x})^2 + \sum_i (y_i - \bar{y})^2\right]\frac{1}{n+m}.$$

One can verify the maximum by checking the second derivative.

Therefore, the likelihood ratio test statistic becomes

$$\lambda(x, y) = \frac{\sup L(\mu_x, \mu_y, \sigma^2|x, y)}{\sup_{H_0} L(\mu_x, \mu_y, \sigma^2|x, y)} = \frac{L(\hat{\mu}_x, \hat{\mu}_y, \hat{\sigma}^2|x, y)}{L(\hat{\mu}, \hat{\sigma}_0^2|x, y)}$$

$$= \frac{(2\pi\hat{\sigma}^2)^{-\frac{n+m}{2}} \exp\left\{\frac{-1}{2\hat{\sigma}^2}\left[\sum_i(x_i-\bar{x})^2+\sum_i(y_i-\bar{y})^2\right]\right\}}{(2\pi\hat{\sigma}_0^2)^{-\frac{n+m}{2}} \exp\left\{\frac{-1}{2\hat{\sigma}_0^2}\left[\sum_i(x_i-\hat{\mu})^2+\sum_i(y_i-\hat{\mu})^2\right]\right\}} = \left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}\right)^{\frac{n+m}{2}}.$$

It suffices to consider the ratio $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$ since the exponent is not random. Now, in the numerator of $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$ we use $\hat{\mu} = (m\bar{x}+n\bar{y})/(n+m)$ to write

$$\sum_{i=1}^{n}\left(x_i - \frac{m\bar{x}+n\bar{y}}{m+n}\right)^2 = \sum_{i=1}^{n}\left((x_i-\bar{x}) + \left(\bar{x} - \frac{m\bar{x}+n\bar{y}}{m+n}\right)\right)^2 = \sum_{i=1}^{m}(x_i-\bar{x})^2 + \frac{n^2m}{(n+m)^2}(\bar{x}-\bar{y})^2.$$

A similar calculation for the Y sum gives

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = \frac{\sum(x_i-\bar{x})^2+\sum(y_i-\bar{y})^2+\frac{nm}{n+m}(\bar{x}-\bar{y})^2}{\hat{\sigma}^2(n+m)} = 1 + \frac{nm}{(n+m)^2}\frac{(\bar{x}-\bar{y})^2}{\hat{\sigma}^2}.$$

Since $\hat{\sigma}^2 = \frac{n+m-2}{n+m}S_p^2$, we have that

$$\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2} = 1 + \frac{nm}{(n+m)(n+m-2)}\frac{(\bar{x}-\bar{y})^2}{S_p^2} = 1 + \frac{1}{n+m-2}\frac{(\bar{x}-\bar{y})^2}{S_p^2(\frac{1}{m}+\frac{1}{n})} = 1 + \frac{1}{n+m-2}t^2.$$

where here t denotes the observed value of T. Thus, large values of $\frac{\hat{\sigma}_0^2}{\hat{\sigma}^2}$ are equivalent to large values of T. Since we reject the null hypothesis H_0 for a large value of the likelihood ratio test statistic $\lambda(x,y)$ and a large value of $\lambda(x,y)$ is equivalent to a large value of T, we can base our rejection rule on T.

(b) Show that, under H_0, $T \sim t_{n+m-2}$. This yields the *two-sample t-test*.

Solution. Under H_0, $\bar{X}-\bar{Y}$ has a $N(0,(m^{-1}+n^{-1})\sigma^2)$ distribution and $(n+m-2)S_p^2/\sigma^2 = (m-1)S_X^2/\sigma^2 + (n-1)S_Y^2/\sigma^2$ has a χ^2_{n+m-2} distribution (by independence). Hence, $T \sim t_{n+m-2}$.

(c) Samples of wood were obtained from the core and periphery of a certain Byzantine church. The date of the wood was determined, giving the following data.

```
core      <- c(1294, 1279, 1274, 1264, 1263, 1254, 1251,
1251, 1248, 1240, 1232, 1220, 1218, 1210)
periphery <- c(1284, 1272, 1256, 1254, 1242,
1274, 1264, 1256, 1250)
```

Use the two-sample t-test to determine if the mean age of the core is the same as the mean age of the periphery.

You should complete this exercise by computing the relevant quantities (you may use a calculator). The command `ttest(x,y)` function in R can be used to check your answer.

Solution. You should find that

```
> t.test(x = core, y = periphery, var.equal = T)

Two Sample t-test
```

```
data:  core and periphery
t = -1.2907, df = 21, p-value =
0.2109
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-29.967604   7.015223
sample estimates:
mean of x mean of y
1249.857  1261.333
```

So that we do not reject the null hypothesis that the mean age of the core is the same as the mean age of the periphery.

9. Let $Y_1, \dots, Y_n$ be i.i.d. Geometric(p) for unknown $p \in (0, 1)$. Then the pmf for each Y_i is

$$P(Y_i = k) = p(1-p)^{k-1}, \quad k = 1, 2, \dots$$

(a) Find the maximum likelihood estimator for p.

Solution. The likelihood is

$$L(p) = \prod_{i=1}^{n} p(1-p)^{y_i - 1}.$$

The corresponding log-likelihood is

$$\log L(p) = \sum_{i=1}^{n} (\log p + (y_i - 1)\log(1-p)) = n \log p + \left(-n + \sum_{i=1}^{n} y_i\right) \log(1-p)$$

with derivative

$$\frac{\partial}{\partial p} \log L(p) = \frac{n}{p} - \left(-n + \sum_{i=1}^{n} y_i\right) \frac{1}{1-p}.$$

The maximum likelihood estimator $\hat{p}$ must satisfy $\frac{\partial}{\partial p} \log L(p)|_{p=\hat{p}} = 0$. Thus

$$0 = \frac{n}{\hat{p}} - \left(-n + \sum_{i=1}^{n} y_i\right) \frac{1}{1-\hat{p}}.$$

This leads to $\hat{p}\left(-n + \sum_{i=1}^{n} y_i\right) = n(1-\hat{p})$. Collecting terms gives $\hat{p} = \frac{n}{\sum_{i=1}^{n} y_i}$.

To check that this indeed a maximizer:

$$\left(\frac{\partial}{\partial p}\right)^2 \log L(p) = -\frac{n}{p^2} - \underbrace{\left(-n + \sum_{i=1}^{n} y_i\right)}_{\geq 0} \frac{1}{(1-p)^2} < 0.$$

(b) Construct the likelihood ratio test statistic t for testing $H_0 : p = 0.5$ against $H_0 : p \neq 0.5$.

Solution. $L(0.5) = 0.5^{\sum_{i=1}^n y_i}$, thus

$$t = \frac{\sup_{p \in (0,1)} L(p)}{L(0.5)} = \frac{L(\hat{p})}{L(0.5)} = \left(\frac{n}{\sum_{i=1}^n y_i} \right)^n \left(\prod_{i=1}^n (1 - \frac{n}{\sum_{j=1}^n y_j})^{y_i - 1} \right) 0.5^{-\sum_{i=1}^n y_i}.$$

(c) Under H_0, state the asymptotic distribution of t as $n \to \infty$. You do not need to verify regularity conditions.

Solution. As $n \to \infty$, we have $2 \log t \xrightarrow{d} \chi_1^2$

(d) Describe how to use t to construct an asymptotic level α test.

Solution. We reject H_0 if $2 \log t > c$ where c is such that $P(X > c) = \alpha$ for $X \sim \chi_1^2$.

10. Let $X_1, \dots, X_n$ be a random sample from a parametric model with marginal density function $f_\theta(x)$. Assume below that regularity conditions hold.

(a) Let $\hat{\theta}_n$ be the MLE of θ_0. Show that if $\theta = \theta_0$, the *Wald statistic* defined as $W = (\hat{\theta}_n - \theta_0)^2 / \mathrm{SE}(\hat{\theta}_n)^2$ is asymptotically χ_1^2.

Explain how this can be used to construct an approximate level α test of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

Solution. The $\mathrm{SE}(\hat{\theta}_n)^2 = I_1(\hat{\theta}_n)^{-1}/n$ so we have

$$W = \frac{[\sqrt{n}(\hat{\theta}_n - \theta_0)]^2}{I_1(\hat{\theta}_n)^{-1}} \to_d N(0,1)^2 = \chi_1^2.$$

Here we have used asymptotic normality of the MLE, Slutsky's lemma and continuous mapping to arrive at the asymptotic chi-squared distribution.

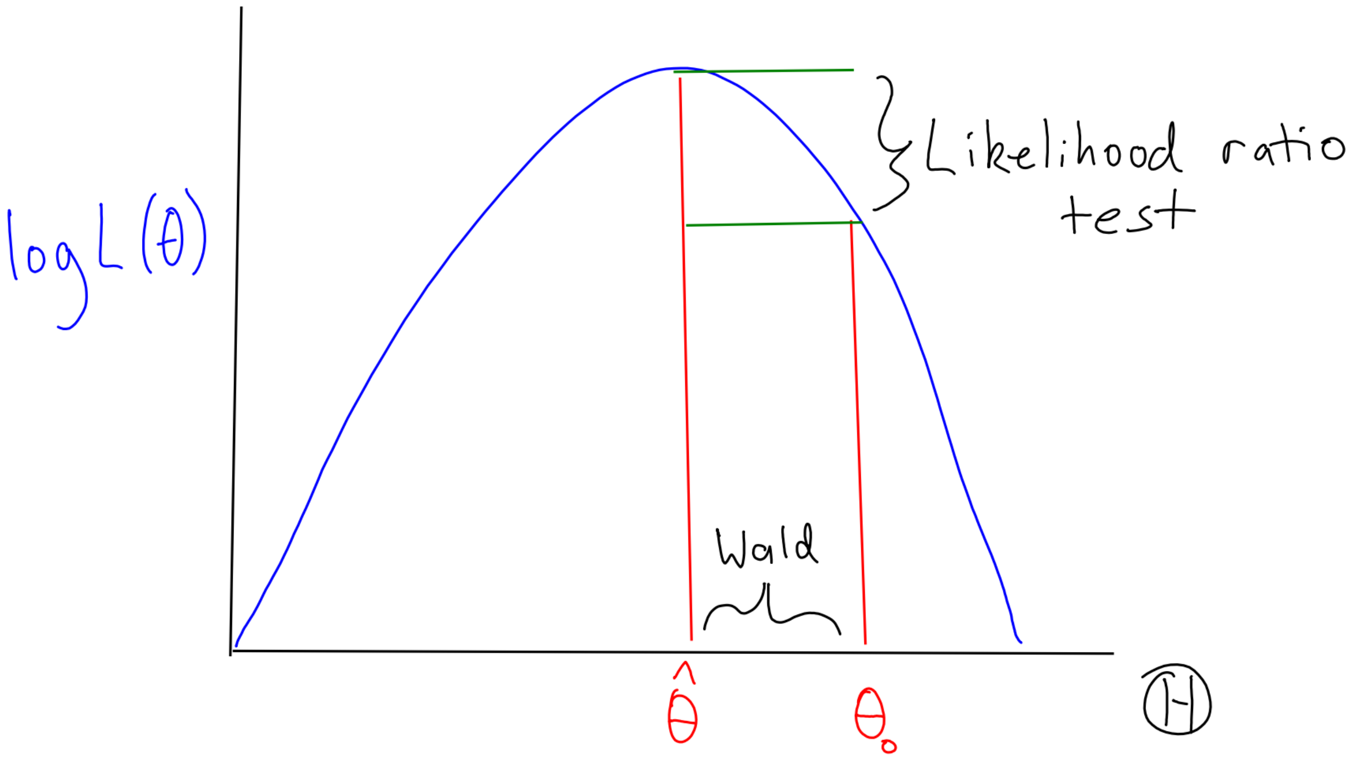
(b) How would you expect the results of the *Wald test* in part (a) to compare to the results of a likelihood ratio test for large sample sizes? Explain.

Solution. Both test statistics converge in distribution to χ_1^2 in this case, so in large samples we would reach the same decisions for either test.

(c) Sketch a graph of a log-likelihood function and label the MLE $\hat{\theta}_n$ and θ_0 on the horizontal axis. Indicate on your graph the quantities used to calculate the Wald test statistic and likelihood ratio test statistic.

Solution. The differences in the log-likelihood are compared for the LRT and the differences in the parameter values themselves are compared for the Wald test.

See sketch below:



R lab: The Bootstrap

This exercise introduces concepts through use of the R software package.

Let T_n be an asymptotically normal estimator of θ based on a random sample $Y_1, \dots, Y_n$. We now consider a flexible method called the bootstrap that allows us to approximate the sampling distribution of T_n using observations $y_1, \dots, y_n$.

The bootstrap sampling distribution can be used to construct confidence intervals for θ by either:

i. Computing $\mathrm{SE}(T_n)$ with respect to the bootstrap sampling distribution and using the formula $T_n \pm c_{\alpha/2}\,\mathrm{SE}(T_n)$ from the notes;

ii. Computing the $\alpha/2$ and $1 - \alpha/2$ quantiles of the bootstrap sampling distribution.

This procedure is widely applicable, but is most useful for estimators where it is difficult to obtain a closed-form expression for $\mathrm{SE}(T_n)$.

In R, the code below shows how we usually compute $\bar{y}$ and estimate its standard error based on 4 data points: 2, 4, 9, and 12.

```
y <- c(2,4,9,12)
ybar <- mean(y)
se.ybar <- sqrt(var(y)/4)
```

Running the above code, we find that the standard error is about 2.29.

The bootstrap sampling distribution of $\bar{Y}$ is obtained by resampling the data points with replacement and computing $\bar{Y}$ based on the resampled data. There are 4 data points, so there are $4^4 = 256$ equally likely resamples. We can use R to obtain all 256 values in the bootstrap sampling distribution as follows:

```
# All 4^4 = 256 possible resamples with replacement
y.star <- expand.grid(y,y,y,y)

# All 256 sample means based on resampling w/replacement
ybar.star <- apply(y.star, 1, mean)

# The standard error based on this is
se.ybar.star <- sqrt(var(ybar.star))
```

From the above, we find that the bootstrap standard error is about 1.98.

It is a good idea to also visualise the bootstrap distribution of $\bar{Y}$. This can be achieve with `hist(ybar.star)`. For large sample sizes, we would expect the bootstrap sampling distribution to look approximately normal. The normal approximation for $n = 4$ seems to be less than ideal.

The number of bootstrap resamples n^n grows too quickly to be reasonable for the average statistician. Instead, we usually approximate the bootstrap sampling distribution by drawing a large number of random samples as follows.

```
set.seed(50011)

ybar.boot <- numeric(length = 10000)
for(i in 1:10000){
y.boot <- sample(y, size = 4, replace = TRUE)
ybar.boot[i] <- mean(y.boot)
}
se.ybar.boot <- sqrt(var(ybar.boot))
```

From the above, we find that the bootstrap standard error is about 1.99. This is nearly the same as the value obtained by enumerating all 256 samples.

11. Using the code examples above:

    (a) Construct three approximate 95% confidence intervals for the mean μ based on the formula $T_n \pm c_{\alpha/2}\,\mathrm{SE}(T_n)$ where the standard error is based on `se.ybar`, `se.ybar.star`, and `se.ybar.boot`.

    (b) Construct two additional approximate 95% confidence intervals for the mean with limits define by the 2.5% and 97.5% percentiles of `ybar.star` and `ybar.boot`. (Hint: use the `quantile()` function.)

    (c) Compare the similarities/differences in the confidence intervals you constructed in parts (a) and (b).

    (d) Replace the data `y` in your code with a random sample of $n = 30$ standard exponential random variables: `y <- rexp(n=30)`. Based on your previous code, construct an approximate 95% confidence interval for the mean based on a normal approximation. Construct two different bootstrap 95% confidence intervals for the mean based on 10000 resamples. (Note: you are not being asked to enumerate all 30^{30} resamples.)

Solution.

(a) We ran the following additional code in R to generate the 95% confidence intervals:

```
> ybar +c(-1,1)* 1.96*se.ybar
[1]  2.267995 11.232005
```

```
> ybar +c(-1,1)* 1.96*se.ybar.star
[1]  2.860867 10.639133
> ybar +c(-1,1)* 1.96*se.ybar.boot
[1]  2.848146 10.651854
```

So the three intervals are, to two decimals,

$$(2.27, 11.23), \qquad (2.86, 10.64), \qquad \text{and} \qquad (2.85, 10.65).$$

(b) We use the following code to generate the 95% confidence intervals based on the percentiles:

```
> quantile(ybar.star, c(0.025,0.975))
2.5% 97.5%
3.0  10.5
> quantile(ybar.boot, c(0.025,0.975))
2.5% 97.5%
3.0  10.5
```

So the two intervals are both $(3.0, 10.5)$ using this method.

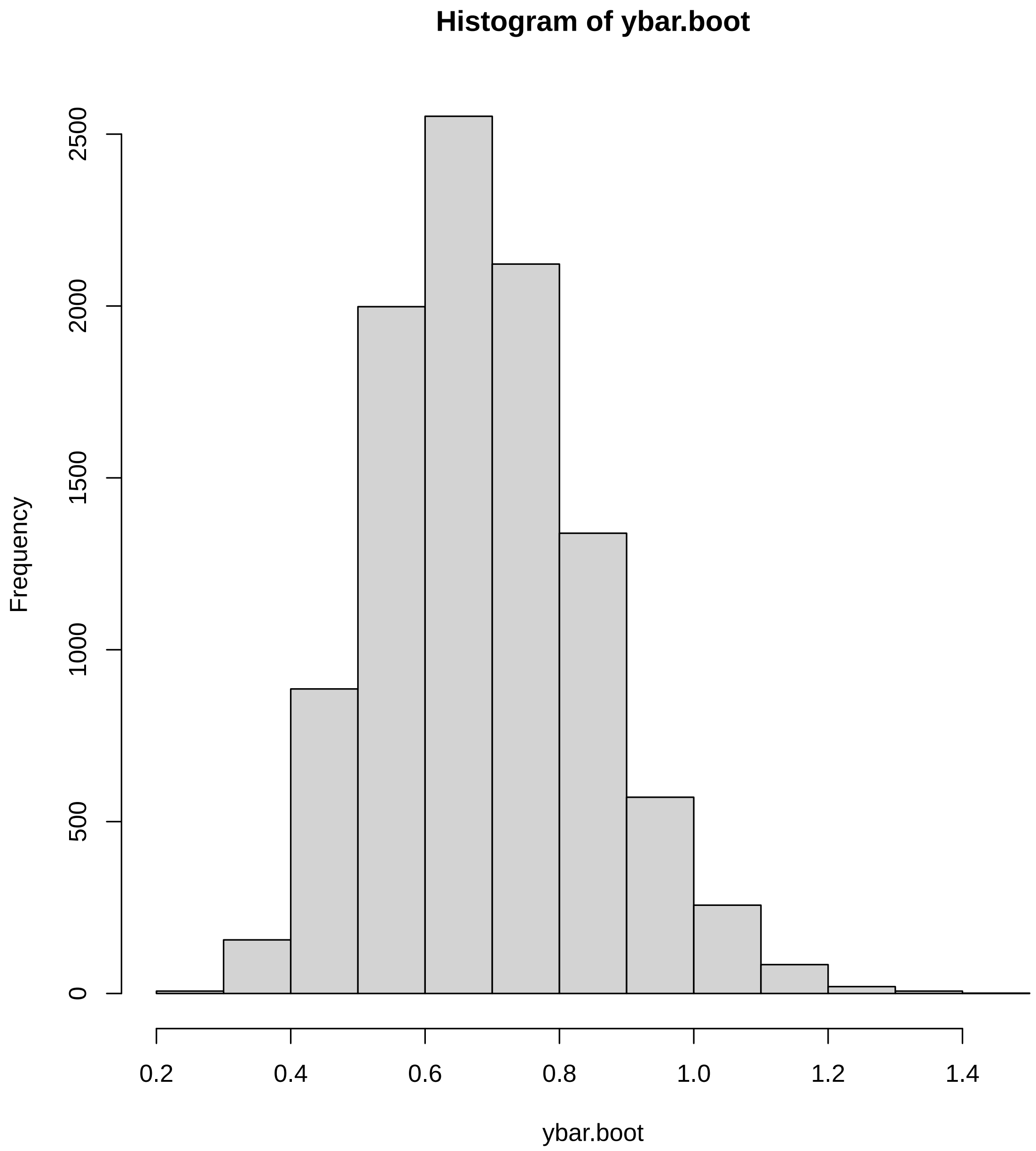
(c) The interval based on the usual standard error estimate $s/\sqrt{n}$ is wider than the bootstrap confidence intervals. The results of taking 10000 resamples do not differ greatly from enumerating the full sampling distribution. The percentile-based bootstrap confidence intervals result in the narrowest 95% confidence intervals in this example.

(d) When we replace the original data with the $n = 30$ random samples from the exponential distribution, we run the following code:

```
y <- rexp(n=30)
ybar <- mean(y)
se.ybar <- sqrt(var(y)/30)

# The bootstrap using Monte Carlo sampling
set.seed(50011)

ybar.boot <- numeric(length = 10000)
for(i in 1:10000){
y.boot <- sample(y, size = 30, replace = TRUE)
ybar.boot[i] <- mean(y.boot)
}
se.ybar.boot <- sqrt(var(ybar.boot))

# Confidence intervals using se.ybar, se.ybar.boot
# c_{\alpha/2} is approx 1.96
ybar +c(-1,1)* 1.96*se.ybar
ybar +c(-1,1)* 1.96*se.ybar.boot

# Confidence intervals using quantiles of ybar.boot
quantile(ybar.boot, c(0.025,0.975))
```

Note that your results may differ slightly depending on whether you reset your random seed.

We obtain a 95% confidence interval of (0.3770655, 0.9975993) using the normal approximation, (0.3793849 0.9952799) using the bootstrap standard error, and (0.418570 1.031513) using the percentile method. Here, the first two intervals are fairly similar. The percentile-based bootstrap interval is shifted upward relative to the other two intervals. This may be because the percentiles are not exactly symmetric for our bootstrap distribution. The histogram of the resamples contained in `ybar.boot` supports this.


Histogram of ybar.boot
Frequency
ybar.boot