STAT 510: Homework 10

David Dalpiaz

Due: Monday, April 25, 11:59 PM

Exercise 1 (Normal Means MLE)

Consider $X_1, \ldots, X_k \sim N(\theta_i, \sigma_i^2)$. That is, each observation is drawn independently from a normal distribution with potentially different means and variances. Assume the variances are known.

Define $\theta = (\theta_1, \ldots, \theta_k)$.

- Find the MLE for θ, $\hat{\theta}$.
- Find $\mathbb{E}\left[\hat{\theta}\right]$.
- Find $\mathbb{V}\left[\hat{\theta}\right]$. Also note what this result simplifies to when $\sigma_1 = \ldots = \sigma_k = 1$.

Solution

First, we write the likelihood and the log-likelihood.

$$\mathcal{L}(\theta) = \prod_{i=1}^{n} f\left(x_i; \theta_i, \sigma_i^2\right) \propto \exp\left[-\frac{\sum_{i=1}^{n}(x_i - \theta_i)^2}{2\sigma_i^2}\right]$$

$$\log \mathcal{L}(\theta) \propto -\frac{\sum_{i=1}^{n}(x_i - \theta_i)^2}{2\sigma_i^2}$$

Next, we maximize to find the MLE.

$$\frac{\partial}{\partial \theta_i} \log \mathcal{L}(\theta) \propto \frac{(x_i - \theta_i)}{\sigma_i^2}$$

$$\hat{\theta}_i = x_i$$

$$\hat{\theta} = (X_1, X_2, \ldots, X_n)$$

Next, we note that this estimator is unbiased.

$$\mathbb{E}\left[\hat{\theta}\right] = (\theta_1, \ldots, \theta_k)$$

With heterogeneity, the variance is a diagonal matrix, with the individual variances on the diagonal and zeros on the off diagonals, which is due to independence.

$$\mathbb{V}\left[\hat{\theta}\right] = \begin{bmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_k^2 \end{bmatrix}$$

With homogeneity, the variance matrix becomes the identity matrix.

$$\mathbb{V}\left[\hat{\theta}\right] = I_k$$

Exercise 2 (Estimating a Variance with One Observation)

Consider a single observation, $X \sim N(0, \sigma^2)$.

- Find an unbiased estimator of σ^2.
- Find the MLE of σ.

Solution

Note that

$$\mathbb{E}\left[X^2\right] = \mu^2 + \sigma^2 = \sigma^2.$$

Thus X^2 is an unbiased estimator of σ^2.

Next, we find the MLE for σ.

$$\mathcal{L}(\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{x^2}{2\sigma^2}\right]$$

$$\log \mathcal{L}(\sigma) = -\log\sigma - \log(\sqrt{2\pi}) - \frac{x^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\sigma} \log \mathcal{L}(\sigma) = -\frac{1}{\sigma} + \frac{x^2}{\sigma^3}$$

We equate this quantity to 0 and solve, while noting that σ is restricted to only take non-negative values.

$$\hat{\sigma} = |x|$$

This problem is made somewhat simple by assuming a known mean of 0. Suppose we did not do that. The instructor has spent some time thinking about a couple related problems:

1. Can we create a (finite length) interval for μ given a single observations?
2. Can we create a (finite length) interval for σ given a single observations?

The instructor gave a somewhat non-technical talk on these ideas: The Smallest Data Analysis: What can we learn from a single observation?

Referenced in that talk are a series of papers that take this subject much more seriously.

- Wall, Boen, and Tweedie
- Letters to the Editor about Wall, Boen, and Tweedie
- Rodriguez
- Portnoy

Notably, the last paper by Portnoy is the most recent and probably contains the best review of the subject. Interestingly, Portnoy was a founding member of the Department of Statistics at the University of Illinois. However, that is not the only way he is related to this homework! You might also notice that in the introduction to his paper, he notes that he learned of this problem during his time at Stanford in a class taught by Charles Stein. Yes, that is the Stein of James-Stein estimation fame!

Exercise 3 (Inverse Guassian MLE)

Let $X_1, \ldots, X_n$ be a random sample from the inverse Gaussian distribution

$$f(x; \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left(-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right), \quad x > 0.$$

Find the MLE of μ and λ.

Solution

First we write the likelihood and log-likelihood.

$$\mathcal{L}(\mu, \lambda) \propto \lambda^{n/2} \cdot \exp\left[-\frac{\lambda}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\mu^2 x_i}\right]$$

$$\log \mathcal{L}(\mu, \lambda) \propto \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\mu^2 x_i} = \frac{n}{2}\log\lambda - \frac{\lambda}{2}\sum_{i=1}^{n}\frac{(x_i/\mu-1)^2}{x_i} =$$

We'll need to take two partial derivatives.

$$\frac{\partial}{\partial\mu}\log\mathcal{L}(\mu,\lambda) \propto \frac{\lambda}{\mu^2}\sum_{i=1}^{n}(x_i/\mu - 1)$$

Setting this derivative equal to zero and solving yields the sample mean as the MLE of μ.

$$\hat{\mu} = \bar{x}$$

$$\frac{\partial}{\partial\lambda}\log\mathcal{L}(\mu,\lambda) \propto \frac{n}{2\lambda} = \sum_{i=1}^{n}\frac{(x_i/\mu-1)^2}{2x_i}$$

After substituting the MLE of μ for μ, setting equal to zero, and some re-arrangement, we arrive at

$$\frac{n}{\lambda} = \frac{1}{\bar{x}^2}\sum_{i=1}^{n}\frac{(x_i-\bar{x})^2}{x_i}.$$

If we bother to wrestle with a lot of algebra, we arrive at the MLE for λ.

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n}\left(\frac{1}{x_i} - \frac{1}{\bar{x}}\right)}.$$

Exercise 4 (A Regression MLE)

Consider $Y_1, \ldots, Y_n$ such that

$$Y_i = \beta x_i + \epsilon_i$$

where

- the x_i are fixed, known constants

- $\epsilon_i \sim N(0, \sigma^2)$
- σ^2 is unknown.

Find the MLE of β as well as its mean and variance.

First, note we can re-write this probability model in a more useful manner.

$$Y_i \mid X_i = x_i \sim N\left(\beta x_i, \sigma^2\right)$$

We should be careful to note that we are assuming that the x_i are fixed, known constants. Only the Y_i are random.

Now, we can write the (conditional) likelihood and log-likelihood.

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} f\left(y_i; \beta, \sigma^2\right) \propto \prod_{i=1}^{n} \exp\left[-\frac{1}{2}\frac{(y_i - \beta x_i)^2}{\sigma^2}\right] = \exp\left[-\frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2}{2\sigma^2}\right]$$

$$\log \mathcal{L}(\beta) \propto -\frac{\sum_{i=1}^{n}(y_i - \beta x_i)^2}{2\sigma^2}$$

We optimize this log-likelihood as we usually do. Thankfully, we only need one equation.

$$\frac{\partial}{\partial \beta} \log \mathcal{L}(\beta) \propto \frac{\sum_{i=1}^{n} x_i(y_i - \beta x_i)}{\sigma^2} = 0$$

Finally, we arrive at the MLE.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

Lastly, we find the mean and variance of this estimator.

$$\mathbb{E}\left[\hat{\beta}\right] = \frac{\sum_{i=1}^{n} x_i \cdot \mathbb{E}\left[Y_i\right]}{\sum_{i=1}^{n} x_i^2} = \beta \cdot \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i^2} = \beta$$

$$\mathbb{V}\left[\hat{\beta}\right] = \frac{\sum_{i=1}^{n} x_i^2 \cdot \mathbb{V}\left[Y_i\right]}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \sigma^2 \cdot \frac{\sum_{i=1}^{n} x_i^2}{\left(\sum_{i=1}^{n} x_i^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$$

Exercise 5 (Beta-Geometric Model)

Assume:

- Likelihood: $X_1, \ldots, X_n \sim \text{Geometric}(p)$
- Prior: $p \sim \text{Beta}(\alpha, \beta)$

Find the posterior mean of $p \mid X_1, \ldots, X_n$, that is, the Bayes estimator of p under squared error loss.

Solution

First, we note the likelihood.

$$\mathcal{L}(p) = (1-p)^{\sum x_i - n} p^n$$

Next, we note the prior density.

$$f(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} p^{\alpha-1} \cdot (1-p)^{\beta-1}$$

Then, recall that the posterior density is proportional to the prior times the likelihood.

$$f(p \mid x_1, \ldots x_n) \propto f(p)\mathcal{L}(p) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cdot\Gamma(\beta)} p^{\alpha-1} \cdot (1-p)^{\beta-1} \cdot (1-p)^{\sum x_i - n} p^n \propto (1-p)^{\beta + \sum x_i - n - 1} p^{\alpha + n - 1}$$

Here, we find that the beta distribution is a conjugate prior for a geometric likelihood.

$$p \mid x_1, \ldots x_n \sim \text{Beta}(\alpha + n, \beta + \sum_{i=1}^{n} x_i - n)$$

Finally, we find the mean of this posterior distribution, which in this case, is the Bayes estimator.

$$\hat{p} = \frac{\alpha + n}{\alpha + \beta + \sum_{i=1}^{n} x_i}$$

Note that this answer supposes we are using a geometric distribution where the sample space is the number of trials. Sometimes, number of failures is used instead. This results in a slightly different "looking" answer.

Exercise 6 (A Simple LRT)

Suppose $X_1, \ldots, X_n \sim N(\mu, \sigma^2 = 2)$. Derive the likelihood ratio test for

$$H_0 : \mu = 10 \quad \text{versus} \quad H_1 : \mu \neq 10.$$

Use the data stored below in `norm_data` to carry out the test by calculating an approximate p-value using the large sample properties of the likelihood ratio test statistic.

```
set.seed(42)
norm_data = rnorm(n = 100, mean = 10.4, sd = sqrt(2))
```

Solution

```
calc_like = function(data, mu) {
  prod(dnorm(data, mean = mu, sd = sqrt(2)))
}
```

```
like_alt =  calc_like(data = norm_data, mu = mean(norm_data))
like_null = calc_like(data = norm_data, mu = 10)
```

```
lr = like_alt / like_null
lrts = 2 * log (lr)
```

```
pchisq(lrts, df = 1, lower.tail = FALSE)
```

```
## [1] 0.001612836
```

Exercise 7 (A LRT for Two Proportions)

Suppose $X_1, \ldots, X_{n_x} \sim \text{Bernoulli}(p_x)$ and $Y_1, \ldots, Y_{n_y} \sim \text{Bernoulli}(p_y)$. Derive the likelihood ratio test for

$$H_0 : p_x = p_y \quad \text{versus} \quad H_1 : p_x \neq p_y.$$

Assuming $n_x = 20$, $n_y = 30$, and $p_x = p_y = 0.3$, repeatedly simulate from this setup and for each simulation:

- Calculate the likelihood ratio test statistic.
- Calculate the value of the usual "textbook" test statistic where $\hat{p}$ is the pooled estimate of the proportion.

$$z = \frac{\hat{p}_x - \hat{p}_y}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_x} + \frac{1}{n_y}\right)}}$$

Using the results of these simulations:

- Plot a histogram of the calculated likelihood ratio test statistics and overlay the approximate distribution of the test statistic under the null hypothesis.
- Create a scatter plot of the likelihood ratio versus the textbook test statistics. What do you notice?

Solution

Under the null, the likelihood is given by

$$\mathcal{L}(p) = \prod_{i=1}^{n_x} p^{x_i} \cdot (1-p)^{1-x_i} \prod_{i=1}^{n_y} p^{y_i} \cdot (1-p)^{1-y_i}.$$

It can be shown (or intuited) that the MLE for p is given by

$$\hat{p} = \frac{\sum_{i=1}^{n_x} x_i + \sum_{i=1}^{n_y} y_i}{n_x + n_y}.$$

Under the alternative, the likelihood is given by

$$\mathcal{L}(p_x, p_y) = \prod_{i=1}^{n_x} p_x^{x_i} \cdot (1-p_x)^{1-x_i} \prod_{i=1}^{n_y} p_y^{y_i} \cdot (1-p_y)^{1-y_i}.$$

It can be shown (or intuited) that the MLEs for the two parameters are

$$\hat{p}_x = \frac{\sum_{i=1}^{n_x} x_i}{n_x}$$

and

$$\hat{p}_y = \frac{\sum_{i=1}^{n_y} y_i}{n_y}.$$

The likelihood ratio test statistic is given by

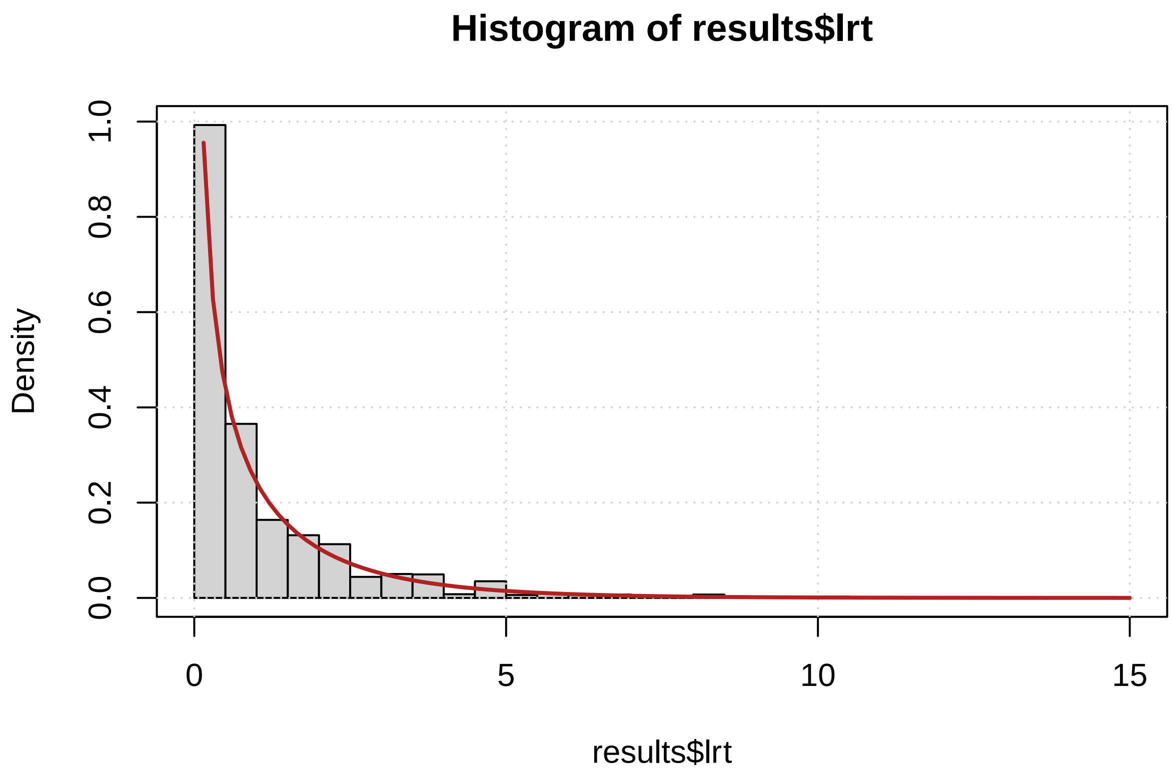
$$\lambda = 2 \log \frac{\mathcal{L}(\hat{p}_x, \hat{p}_y)}{\mathcal{L}(\hat{p})}.$$

```
# function to perform single simulation
gen_and_calc_test_stats = function(n_x, n_y, p_x, p_y) {

  # generate data
  x_data = rbinom(n = n_x, size = 1, prob = p_x)
  y_data = rbinom(n = n_y, size = 1, prob = p_y)

  # calculate estimates
  p_hat_x = mean(x_data)
  p_hat_y = mean(y_data)
  p_hat   = mean(c(x_data, y_data))

  # calculate likelihood
  like_x = prod(dbinom(x = x_data, size = 1, prob = p_hat_x))
  like_y = prod(dbinom(x = y_data, size = 1, prob = p_hat_y))
  like_alt  = like_x * like_y
  like_null = prod(dbinom(x = c(x_data, y_data), size = 1, prob = p_hat))

  # calculate likelihood ratio test statistic
  lr = like_alt / like_null
  lrts = 2 * log (lr)

  # calculate "textbook" statistic
  z = (p_hat_x - p_hat_y) / sqrt(p_hat * (1 - p_hat) * (1 / n_x + 1 / n_y))

  # return
  c(lrts, z)
}
```

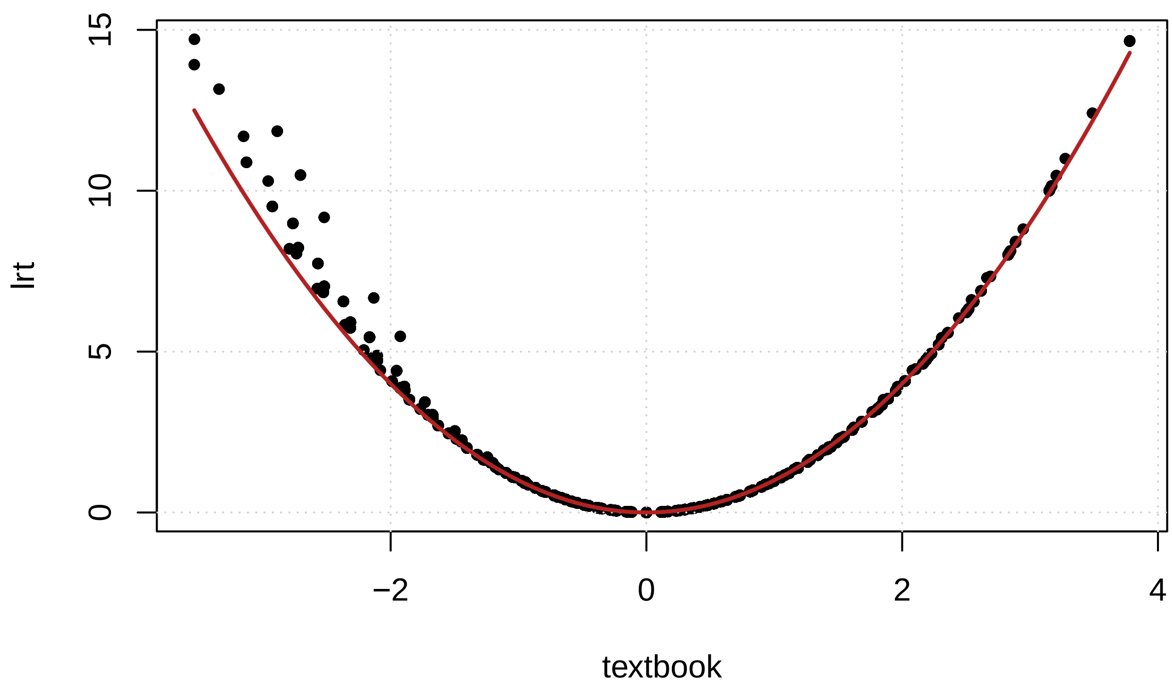
```
# perform all simulations
set.seed(42)
test_stats = replicate(
  n = 10000,
  gen_and_calc_test_stats(n_x = 20, n_y = 30, p_x = 0.3, p_y = 0.3)
)
```

```
# wrangle results
results = as.data.frame(t(test_stats))
colnames(results) = c("lrt", "textbook")
```

```
hist(results$lrt, probability = TRUE, breaks = 50)
box()
grid()
curve(dchisq(x, df = 1), add = TRUE, col = "firebrick", lwd = 2)
```




```
plot(lrt ~ textbook, data = results, pch = 20)
grid()
curve(x ^ 2, col = "firebrick", add = TRUE, lwd = 2)
```



Exercise 8 (An ANOVA Adjacent LRT)

Suppose

- $X_1, \ldots, X_{n_x} \sim N(\mu_x, \sigma_x^2)$.
- $Y_1, \ldots, Y_{n_y} \sim N(\mu_y, \sigma_y^2)$.
- $Z_1, \ldots, Z_{n_z} \sim N(\mu_z, \sigma_z^2)$.

Derive the likelihood ratio test for $H_0 : \sigma_x^2 = \sigma_y^2 = \sigma_z^2$ versus an alternative that allows for at least one unequal variance.

Use the data stored below in the vectors `x`, `y`, and `z` to carry out the test by calculating an approximate p-value using the large sample properties of the likelihood ratio test statistic. (Note that this data is **not** tidy, but is instead stored in a format that is easy to understand.) Does the result match your expectation?

```
set.seed(42)
x = rnorm(n = 50, mean = -5, sd = 1)
y = rnorm(n = 60, mean =  0, sd = 1)
z = rnorm(n = 70, mean =  5, sd = 1)
```

Solution

Here, the alternative model allows for all three variances to be different. Hence, the likelihood is given by

$$\mathcal{L}(\theta_1) = \prod_{i=1}^{n_x} \frac{1}{\sigma_x\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x_i-\mu_x)^2}{\sigma_x^2}\right] \cdot \prod_{i=1}^{n_y} \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(y_i-\mu_y)^2}{\sigma_y^2}\right] \cdot \prod_{i=1}^{n_z} \frac{1}{\sigma_z\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(z_i-\mu_z)^2}{\sigma_z^2}\right].$$

After going through the tedious optimization mechanics (which are very similar to previous homework) we can obtain the maximum likelihood estimates of the parameters.

$$\hat{\mu}_x = \bar{x}, \quad \hat{\mu}_y = \bar{y}, \quad \hat{\mu}_z = \bar{z}$$

$$\hat{\sigma}_x^2 = \frac{1}{n_x}\sum_{i=1}^{n_x}(x_i-\bar{x})^2, \quad \hat{\sigma}_y^2 = \frac{1}{n_y}\sum_{i=1}^{n_y}(y_i-\bar{y})^2, \quad \hat{\sigma}_z^2 = \frac{1}{n_z}\sum_{i=1}^{n_z}(z_i-\bar{z})^2$$

$$\hat{\theta}_1 = (\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z, \hat{\sigma}_x^2, \hat{\sigma}_y^2, \hat{\sigma}_z^2)$$

Under the null, we assume a common variance.

$$\mathcal{L}(\theta_0) = \prod_{i=1}^{n_x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(x_i-\mu_x)^2}{\sigma^2}\right] \cdot \prod_{i=1}^{n_y} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(y_i-\mu_y)^2}{\sigma^2}\right] \cdot \prod_{i=1}^{n_z} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\frac{(z_i-\mu_z)^2}{\sigma^2}\right].$$

The estimation of the means remains the same. The MLE for σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{n_x+n_y+n_z}\left[\sum_{i=1}^{n_x}(x_i-\bar{x})^2 + \sum_{i=1}^{n_y}(y_i-\bar{y})^2 + \sum_{i=1}^{n_z}(z_i-\bar{z})^2\right]$$

$$\hat{\theta}_0 = (\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_z, \hat{\sigma}^2)$$

Plugging everything in, we arrive at the likelihood ratio test statistic.

$$\lambda = 2 \log \frac{\mathcal{L}\left(\hat{\theta}_1\right)}{\mathcal{L}\left(\hat{\theta}_0\right)}.$$

Note that there is a nice satisfying analytic simplification here, but, for our purposes, we can simply calculate as-is in R. (Although, doing some analytic simplification would probably be necessary with bigger sample sizes due to numeric precision issues. Here thought, we can get away with being lazy.)

```
sd_mle = function(x) {
  sqrt(var(x) * (length(x) - 1) / length(x))
}
```

```
lr_alt =
  prod(dnorm(x, mean = mean(x), sd = sd_mle(x))) *
  prod(dnorm(y, mean = mean(y), sd = sd_mle(y))) *
  prod(dnorm(z, mean = mean(z), sd = sd_mle(z)))
```

```
pooled_var =
  (sum((x - mean(x)) ^ 2) +
   sum((y - mean(y)) ^ 2) +
   sum((z - mean(z)) ^ 2)) /
    (length(x) + length(y) + length(z))
```

```
lr_null =
  prod(dnorm(x, mean = mean(x), sd = sqrt(pooled_var))) *
  prod(dnorm(y, mean = mean(y), sd = sqrt(pooled_var))) *
  prod(dnorm(z, mean = mean(z), sd = sqrt(pooled_var)))
```

```
lrts = 2 * log(lr_alt / lr_null)
```

```
pchisq(lrts, df = 2, lower.tail = FALSE)
```

```
## [1] 0.1395509
```

As excepted, because we can see that this data was generated with equal variances, we fail to reject for any reasonable α.

Exercise 9 (Free Points)

It's been a long semester! Draw a smiley face for a free point!

Solution

:)

Exercise 10 (Free Points)

It's been a long semester! Draw a smiley face for a free point!

Solution

:)

Exercise 11 (Free Points)

It's been a long semester! Draw a smiley face for a free point!

Solution

:)