Statistical Methods

Ugurcan Lacin

November 1, 2017

Exercises

4.84

Gamma1(alpha = 4, beta = 1), Gamma2(alpha = 40, beta = 1), and Gamma3(alpha = 80, beta = 1).

a) *What do you observe about the shapes of these three density functions? Which are less skewed and more symmetric?*

Gamma1 is skewed to the right, which is positive skewed. Gamma2 is also positive skewed.Gamma3 looks more symmetric. As seen, When alpha value is increased, the shape is getting symmetric.

b) *What differences do you observe about the location of the centers of these density functions?*

When alpha value is increased, the center of the density functions aim to increase as well.

c) *Give an explanation for what you observed in part b.*

Means are also getting increased, when alpha value is increased.

4.117

Beta1(alpha = 9, beta = 7), Beta2(alpha = 10, beta = 7), and Beta3(alpha = 12, beta = 7)

a) *Are these densities symmetric? Skewed left? Skewed right?*

The lower alpha looks symmetric, but increasing alpha gives left skewed.

b) *What do you observe as the value of alpha gets closer to 12?*

Graphs loses symmetricity when the value of alpha gets closer to 12.

c) *Graph some more beta densities with alpha > 1, beta > 1, and alpha > beta. What do you conjecture about the shape of beta densities with alpha > beta and both alpha > 1 and beta > 1?*

The shape of Beta is always skewed left with alpha > beta and both alpha > 1 and beta > 1

4.118

Beta1(alpha = .3, beta = 4), Beta2(alpha = .3, beta = 7), and Beta3(alpha = .3, beta = 12)

a) *Are these densities symmetric? Skewed left? Skewed right?*

All of them are skewed right. Higher beta is more skewness to the right.

b) *What do you observe as the value of beta gets closer to 12?*

Spread is getting decreased, and highest point is growing.

c) *Which of these beta distributions gives the highest probability of observing a value larger than 0.2?*

Beta1(alpha = .3, beta = 4) has the highest probability.

d) *Graph some more beta densities with alpha $<$ 1 and beta $>$ 1. What do you conjecture about the shape of beta densities with alpha $<$ 1 and beta $>$ 1?*

the close alpha to 1, the skewness decreases.

10.11

a) *In the scenario to be simulated, what is the only kind of error that can be made?*

The only error that can be made is an error Type II. Since H0 is not true, the error beta would be accept H0 when Ha is true.

b) *Click the button "Clear Summary." Conduct at least 200 simulations. What proportion of the simulations resulted in type II errors (hover the pointer over the box about "Error" in the lower right portion of the display)? How is the proportion of type II errors related to the proportion of times that H0 is rejected?*

H0 is rejected in only 8% of the cases and accepted in 92%. So, 92% is high value for beta.

> A type I error is made if H 0 is rejected when H 0 is true. The probability of a type I error is denoted by alpha. The value of alpha is called the level of the test. A type II error is made if H 0 is accepted when H a is true. The probability of a type II error is denoted by beta.

c) *Change n,the number of trials used for each simulated test, to 30 and leave all other settings unchanged. Simulate at least 200 tests. Repeat for n = 50 and n = 100. Click the button "Show Summary." How do the values of ??(.6), the probability of a type II error when p = .6, change as the sample size increases?*

When sample size is increased, the value of error rate beta decreases.

d) *Leave the window with the summary information open and continue with Exercise 10.12.*

I did.

10.12

a) *Which of the two tests ?? = .05 or ?? = .10 gives the smaller simulated values for ??, using samples of size 15?*

for alpha 0.05 the beta is higher than with alpha 0.1 at a sample size of n = 15.

b) *Which gives the smaller simulated values for ?? for each of the other sample sizes?*

Overall beta is smaller for all the sample sizes for alpha 0.1 compared to when alpha is 0.05.

11.6

a) *What is the intercept of the line with 0 slope? What is the value of SSE for the line with 0 slope?*

The intercept of the line with 0 slope is 43.362 and the SSE is 1002.839

b) *Do you think that a line with negative slope will ???t the data well? If the line is dragged to produce a negative slope, does SSE increase or decrease?*

A line with negative slope would not fit the data well, since the price is growing over the years and not decreasing. If the line is dragged to produce a negative slope, the SSE increases.

c) *Drag the line to obtain a line that visually ???ts the data well. What is the equation of the line that you obtained? What is the value of SSE? What happens to SSE if the slope (and intercept) of the line is changed from the one that you visually ???t?*

The equation of the obtained line is Y = 21.802 + 4.791 * X The SSE is 18.393. If the slope of the line is changed, the SSE will probably increase.

d) *Is the line that you visually ???t the least-squares line? Click on the button "Find Best Model" to obtain the line with smallest SSE. How do the slope and intercept of the least-squares line compare to the slope and intercept of the line that you visually ???t in part (c)? How do the SSEs compare?*

The line that I visually fit is nearly the least-squares line. The SSE of the best model is just 0.107 lower.

e) *Refer to part (a). What is the y-coordinate of the point around which the blue line pivots?*

It pivots around the intercept of a.

f) *Click on the button "Display/Hide Error Squares."What do you observe about the size of the yellow squares that appear on the graph? What is the sum of the areas of the yellow squares?*

Overall the size of the yellow squares is quite small. The sum of the areas of the yellow squares is the SSE, so 18.286

11.50

a) *Drag the blue line to obtain an equation that visually ???ts the data well. What do you notice about the value of* r^2 *as the ???t of the line improves?*

the value of r^2 is growing with improving fit of the line.

b) *Click the button "Find Best Model" to obtain the least-squares line. What is the value of* r^2 *What is the value of the correlation coef???cient?*

The value of r^2 for the best fit is 0.982 The value of the correlation coefficient is $sqrt(r^2) = \sqrt{0}.982 = 0.99$

c) *Why is the value of* r^2 *so much larger than the value of* r^2 *that you obtained in Exercise 11.49(b) that used the data from Example 11.1?*

Because a linear fit is not perfect choice for the reflect data.

16.4

a) *Click the button "Next Trial" to observe the result of taking a sample of size n = 1 from a Bernoulli population with p = .1. Did you observe a success or a failure? Does the posterior look different than the prior?*

if posterior looks different than the prior, there is a failure.

b) *Click the button "Next Trial" once again to observe the result of taking a sample of total size n = 2 from a Bernoulli population with p = .1. How many successes and failures have you observed so far? Does the posterior look different than the posterior you obtained in part (a)?*

I observed two failures and the posterior looks different than the one in a. The area under the curve for p > 0.5 is going towards 0.

c) *If you observed a success on either of the ???rst two trials, click the "Reset" button and start over. Next, click the button "Next Trial" until you observe the ???rst success. What happens to the shape of the posterior upon observation of the ???rst success?*

As long as there is no success the graph is running towards infinity for p going towards 0. But the moment we observe the first success, the graph origins at (0,0).

d) *In this demonstration,we assumed that the true value of the Bernoulli parameter is p = .1. The mean of the beta prior with ?? = 1,??= 3 is .25. Click the button "Next Trial" until you obtain a posterior that has mean close to .1. How many trials are necessary?*

6 trials were necessary to obtain a mean of 0.1.

Imputation techniques

Which type of missing mechanism do you prefer to get a good imputation?

I prefer *Missingness that depends on unobserved predictors*. Because, in that case, there is a dependency to a situation, such as, education level, character type, etc. This kind of information can be extracted from existing data. To find good imputation for missing values, at least, there is a high chance to find it as reasonable.

Say something about simple random imputation and regression imputation of a single variable.

Simple random imputation is the simplest approach to impute missing values. All missing values are changed with values which are produced random according to existing values for that variable

Regression imputation is a better system. Other variables are used to predict missing values. Comparing to simple random imputation, regression will have produced values.

Explain shortly what Multiple Imputation is.

Multiple imputation is a system to estimate missing values by creating several imputed values and analyzing them to find best fit for missing values.