Problem Set 3 Corrections

October 13, 2020

1 1) Let X, Y, and Z be independent RVs with variances X , Y , Z . Let:

$$U = Z - XV = Z - Y \tag{1}$$

Answer the following questions: (a) Find Cov(U, V) and U,V

(b)If U=Z+X and V=Z+Y, do the values of Cov(U,V)and U,V computed inpart a) change? Explain.

(c) How does U,V change if ^2_Z is much larger than ^2_X or ^2_Y?

(d) How does U,V change if ^2_Z is much smaller than ^2_X or ^2_Y?

(e) How do the answers for parts c) and d) relate to variable standardization?

I had no issues with this problem.

2 2) Let X and Y be jointly distributed RVs with correlation define the standardized random variables Xbar and Ybar as:

2.1 Answer the following questions:

(a) Show that Cov(Xbar,Ybar)= X,Y.

(b) Principal component analysis (PCA) is normally defined by the eigenvalue decomposition of the sample covariance matrix for multivariate data. Look at the R PCA function prcomp(). What is the impact of setting center=T and scale=T when calling prcomp()? When might this be desirable?

Part a is correct, for b) I should have mentioned the result is that the eigenvalue decomposition is performed on the sample correlation matrix ratherthan the sample covariance matrix

3 3) The number of offspring of an organism is a discrete random variable with mean and variance ^2 . Each of its offspring reproduces in the same manner.

(a) Find the expected number of offspring in the third generation.

(b) Find the variance of the number of offspring in the third generation.

(c) Validate your answers to a) and b) via simulation with the number of offspring repre- sented by a Poisson RV with = 2. Create 1000 separate populations that each include 3 generations and use a histogram to visualize the empirical distribution of the number of offspring in the third generation. Estimate the expected number of 3rd generation offspring using the average across all 1000 simulations and estimate the variance of the number using the R var() function (we will learn the basis for these estimates in Chapter 8). Compare these estimates with the values computed according to the results in part a) and b).

Used wrong variance formula in b thus leading to wrong variance calculation. Also my estimates were wrong in my graphs.

4 5) Let X1,X2,... be a sequence of independent random variables with E[Xi]= and Var(Xi)= ^2i. Show that if

Used correct theorm and got same answer.

5 7) Suppose that X1, ..., X20 are independent random variables with density functionsf(x)=3x^2,0 x 1. LetS=X1+...+X20.

(a) Use the central limit theorem to approximate P (S 14).

(b) If you are instead asked to approximate P(S 15), what simplification can be made to the calculation?

(c) Validate the approximation by plotting the CLT-based density (compute this using dnorm()) and true density of S. Use the inverse CDF method to simulate from the true density and plot using a kernel density estimate (R code plot(kernel()), we'll learn the details of kernel density estimation later in the course).

didnt take part a all the way to final, but did 3/4 correct. Didnt show full work for part c, but my intuition was correct.

6 8) (Based on Rice 5.21) We wish to evaluate the integral... (couldn't copy this problem)

(a) Show that E(I(f)) = I(f) Correct, but didn't show full steps. Missed the second last step.