Statistical Computing - Homework 1

Wentao Huang, 1303160224, Statistics 1602

2019/03/09

Contents

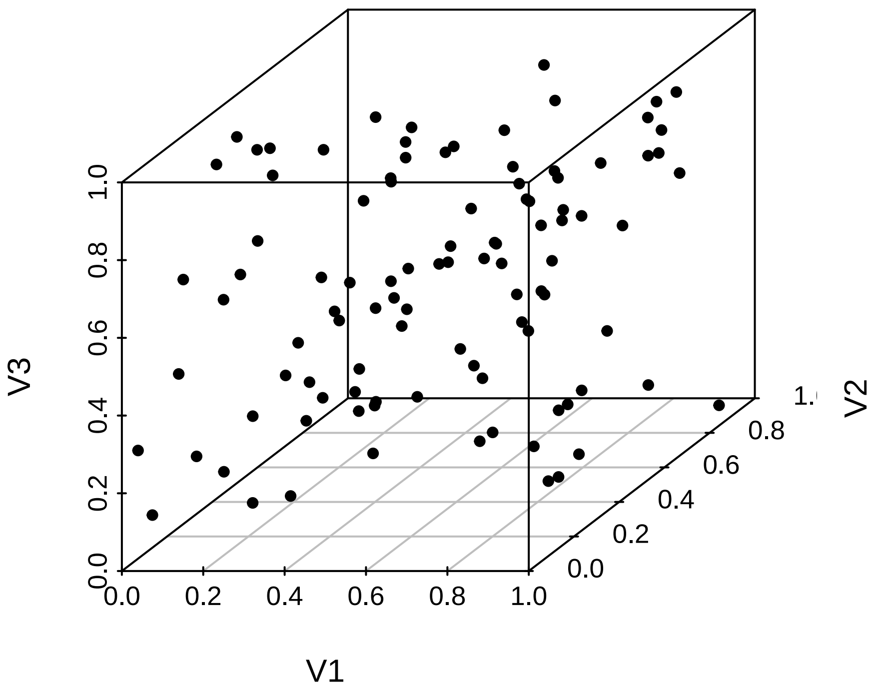
Problem

Simulate a sequence of uniform random numbers with R function `runif()`. Make 3-tuples with every 3 consecutive numbers from the sequence. Visualize the tuples in the cube $[0,1]^3$ and test the uniformity of these tuples with Pearson's χ^2 test.

Solution:

```
# generate 3-tuples dataframe
set.seed(666)
dat1 = as.data.frame(array(runif(300),c(100,3)))

# plot in 3d
## install.packages("scatterplot3d")
library(scatterplot3d)
scatterplot3d(dat1,xlim=c(0,1),ylim=c(0,1),zlim=c(0,1),pch=20)
```




```
# Pearson's chisq-test
chisq.test(dat1)
```

```
## Warning in chisq.test(dat1): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  dat1
## X-squared = 36.87, df = 198, p-value = 1
```

Problem 2

Prove that $F(X) \sim U(0,1)$ for a continuous random variable X with CDF $F(\cdot)$. And explain why this is not true if X follows a discrete distribution.

Proof:

Since the random variable X is continuous, it's PDF $f(x)$ must exist. We denote it's CDF as:

$$F(X) = \int_{-\infty}^{X} f(t)dt$$

which is monotone increasing and continuous. Hence the CDF's inverse function F^{-1} must exist. If $F(X) \sim U(0,1)$, then $F^{-1}(F(X)) = X \sim F(\cdot)$.

In the discrete case, there is no inverse function, therefore this is not true if X follows a discrete distribution.

Q.E.D.

Problem 3

A fault-tolerant memory bank is built with 5 memory units. These units have independent random failure times, each with a distribution F. The memory bank is operable as long as 4 or more units are still working, so it fails when the second unit fails. Let $G(\cdot|\alpha,\beta)$ be the CDF of the $Beta(\alpha,\beta)$.

a)

Express the failure time of the memory bank as a function of a single variable $U \sim U(0,1)$ via $F(\cdot)$ and $G(\cdot|\alpha,\beta)$ and/or their inverses.

Solution:

b)

Suppose that F is the exponential distribution with mean $500,000$ (hours). Sample $10,000$ memory bank lifetimes using the expression from part a). Estimate the mean memory bank lifetime and give a 99% confidence interval for it.

Solution:

c)

If we didn't use a redundant system, we might instead use 4 memory units and the bank would fail when the first of those failed. Find the expected lifetime of such a system without redundancy and give a 99% confidence interval. The redundant system takes 1.25 times as many memory units. Does it last at least 1.25 times as long on average?

Solution:

d)

Compare the speed of sampling by inversion with that of generating 5 exponential variables, sorting them and picking out the second smallest.

Solution:

Problem 4

For acceptance-rejection sampling,

a)

Show that the ratio $f(x)/g(x)$ of $N(0, 1)$ to Cauchy densities is maximized at $x = \pm 1$.

Proof:

b)

Can we use acceptance-rejection with proposals from $N(0, 1)$ to sample from Cauchy distribution? Explain why or why not.

Solution: