STAT3023 Statiscal Inference

Lab Week 4

Tutor: Wen Dai

SID: 470408326

School of Mathematics and Statistics

The University of Sydney

Semester 1, 2021

1. (a) Generate 100 iid Unif(0,1) (use `runif`) random variables and store them in `u`. Apply the function $-\log(1-u)$ to each element, and store the results in `x`.

Solution

Here we sample 100 times from the uniform distribution (unif(0, 1)) with a random variable U and hence, the random variable $1 - e^{-X} = U \sim$Unif(0, 1)

```
set.seed(2021)
u = runif(100)
x = -1 * log(1 - u)
```

(b) Plot the histogram of `x` and overlay it with the density curve of exponential(1) (use `dexp(x, rate=1)`). Why do we have good agreement here? (Hint: $-\log(1-u)$ is the inverse function of the c.d.f. of exponential(1).)

Solution

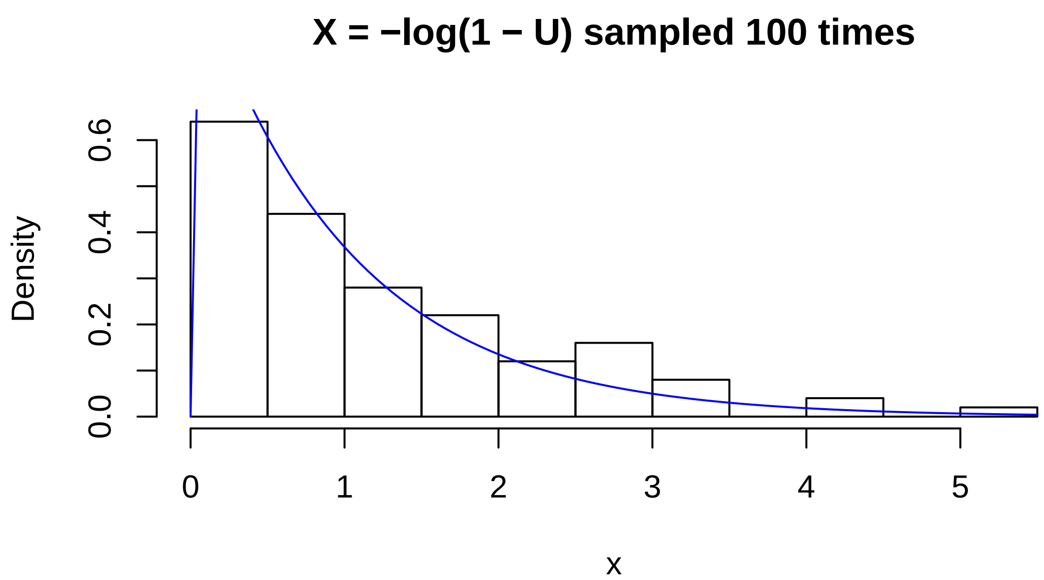
We see this to be the case because:

$$\begin{aligned} P_X(x) &= P(X \le x) \\ &= P(-\log(1-U) \le x) \\ &= P\left(1 - U \ge e^{-x}\right) \\ &= P\left(U \le 1 - e^{-x}\right) \\ &= P_U\left(1 - e^{-x}\right) \end{aligned}$$

Hence we can say that $f_X(x) = \frac{\mathrm{d}}{\mathrm{d}x}(1 - e^{-x}) = e^{-x}$ for $x > 0$ because the values of $U \in [0, 1]$ and $X = -\log(1-U)$

That is, $X \sim exp(1)$

```
hist(x,
     main = "X = -log(1 - U) sampled 100 times",
     probability = TRUE)
curve(dexp(x, rate = 1), add = TRUE, col = "blue")
```

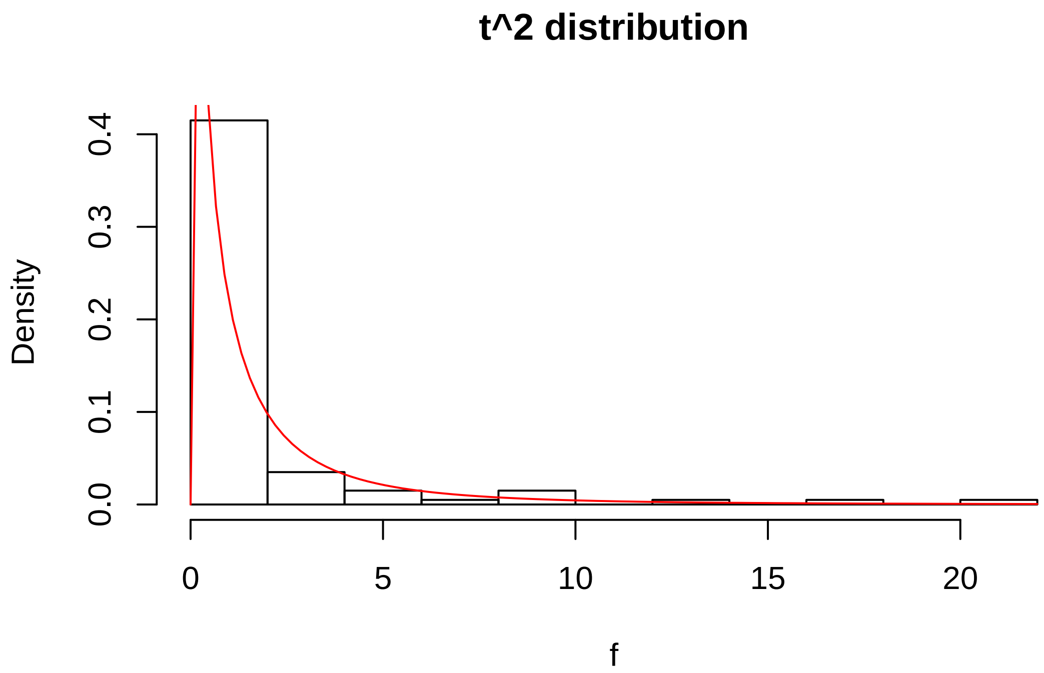


2. Transformation of random variables

(a) Generate 100 random variables from a t distribution with 5 degrees of freedom (use `rt(100, df=5)`). Store them in `t`. Make another vector `f` by `f = t^2`. Overlay the histogram of `f` with the density curve of a $F_{1,5}$ distribution (use `df(x, df1=1, df2=5)`). Comment on the plot.

Solution

```
set.seed(2021)
t = rt(n = 100, df = 5)
f = t^2
hist(x = f, main = "t^2 distribution", probability = TRUE)
curve(df(x, df1=1, df2=5), add = TRUE, col = "red")
```



We see that the the density of the $F_{1,5}$ seems to fit a t^2 distribution with 5 degrees of freedom. In general, a logical question to ask is:

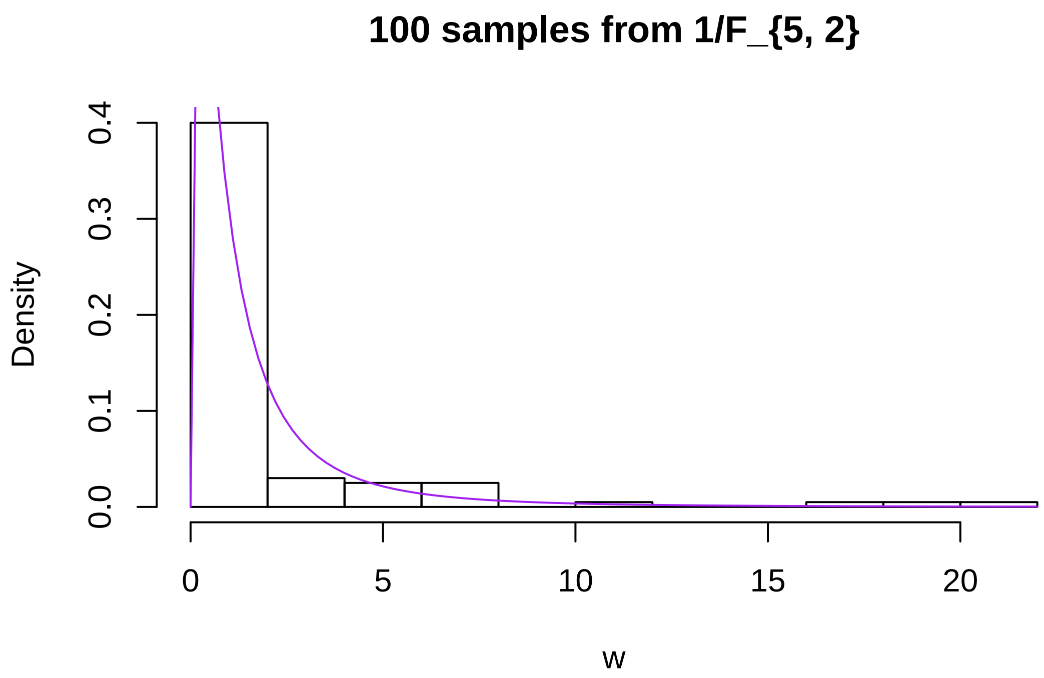
$$\text{Is } X \sim F_{1,p} \iff X \sim t_p^2 \text{ true?}$$

(b) Generate 100 random variables from a $F_{5,2}$ distribution (use `rf(100, df1=5, df2=2)`). Store them in `y`. Make another vector `w = 1/y`. Overlay the histogram of `w` with the density curve of a $F_{2,5}$ distribution. Comment on the plot.

Solution

```
# 100 samples from the F_{5, 2} distribution
set.seed(2021)
y = rf(100, df1=5, df2=2)
w = 1/y
```

```r
hist(x = w, main = "100 samples from 1/F_{5, 2}", probability = TRUE)
curve(df(x, df1=2, df2=5), add = TRUE, col = "purple")
```



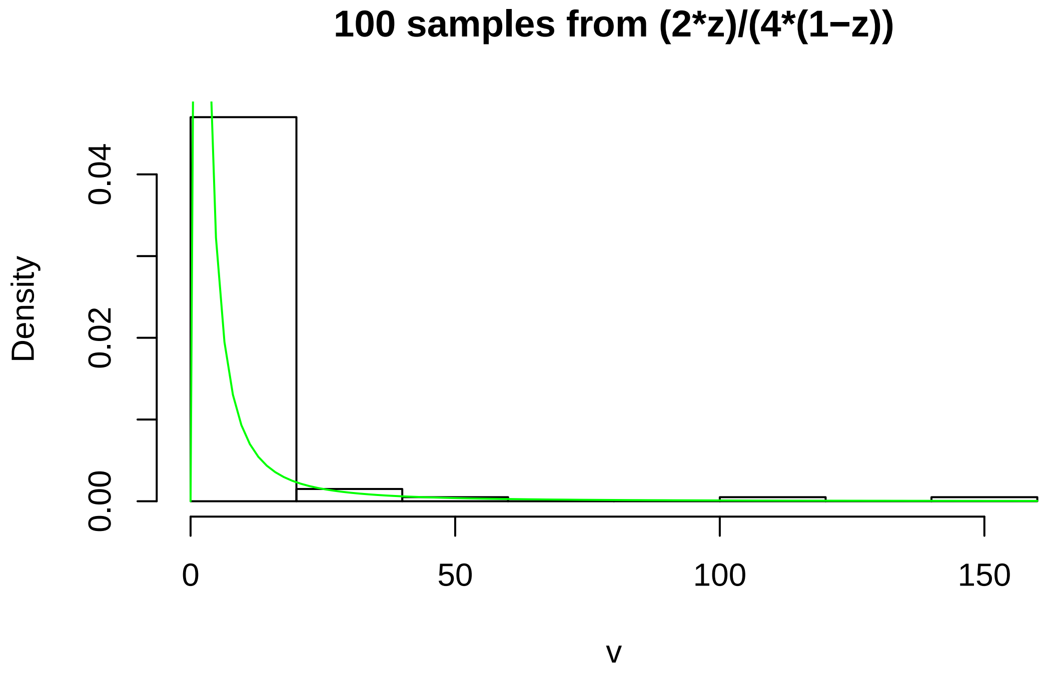
It seems like when we sample from the distribution $\frac{1}{Y}$ where $Y \sim F_{5,2}$ the density curve which fits this is $F_{2,5}$. It's like taking the reciprocal of the F distribution switches the two parameters!

A natural question which this prompts is does this hold true in general- that is if $Y \sim F_{a,b}$ then is $W = \frac{1}{Y} \sim F_{b,a}$?

(c) Generate 100 random variables from a beta(2, 1) distribution (use `rbeta(100, shape1=2, shape2=1)`). Store them in `z`. Make another vector `v = 2*z/(4*(1-z))`. Overlay the histogram of `v` with the density curve of a $F_{4,2}$ distribution. Comment on the plot.

Solution

```r
set.seed(2021)
z = rbeta(100, shape1=2,shape2=1)
v = (2*z)/(4*(1-z))
hist(x = v, main = "100 samples from (2*z)/(4*(1-z))", probability = TRUE)
curve(df(x, df1=4, df2=2), add = TRUE, col = "green")
```



Hence we see that the $F_{4,2}$ density seems to fit $W = \frac{2X}{4(1-X)}$ where $X \sim$ beta(2, 1) distribtion quite well.