# STAT3023 Statiscal Inference

Lab Week 3

Tutor: Wen Dai

SID: 470408326

School of Mathematics and Statistics

The University of Sydney

Semester 1, 2021

1. (a) Generate 100 realisations of the sample variance of 10 independent $N(0, 1)$ random variables and store them in `s2`

**Solution**

To do this, we have to recall that sample variance is given by:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{2}(X_i - \overline{X})^2$$

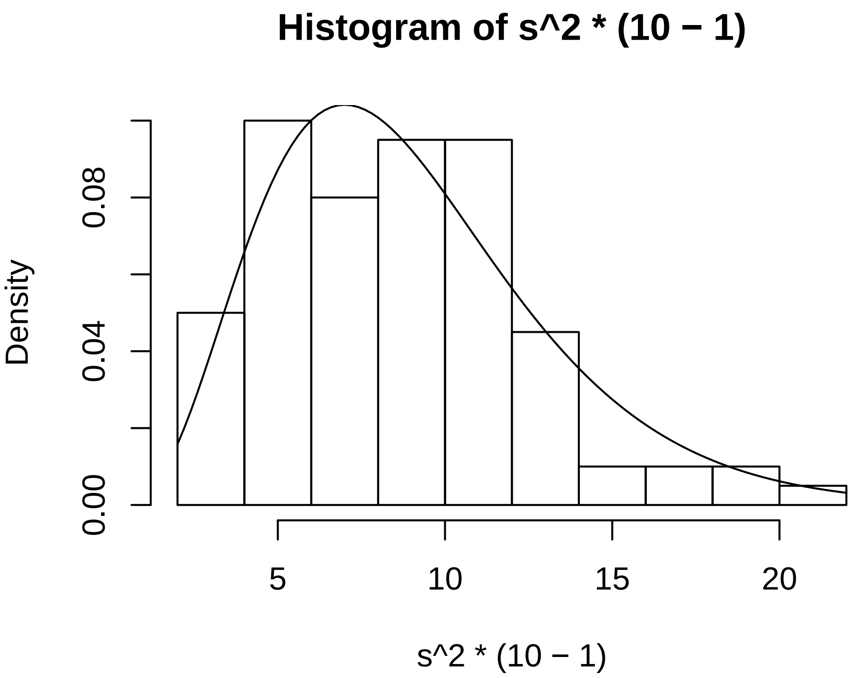
We can use the R functin `var` to obtain the sample variance from a vector

```r
set.seed(3023)
num_sim = 100
num_rvs = 10
experiment = rnorm(n = num_sim * num_rvs, mean = 0, sd = 1)
data_mat = matrix(experiment, nrow = num_sim)
s2 = apply(data_mat, MARGIN = 1, FUN = var)
```

(b) Plot the histogram of `(10 - 1) * s2` and overlay it with the density function of the $\chi^2_9$ distribution (use `dchisq`)

**Solution**

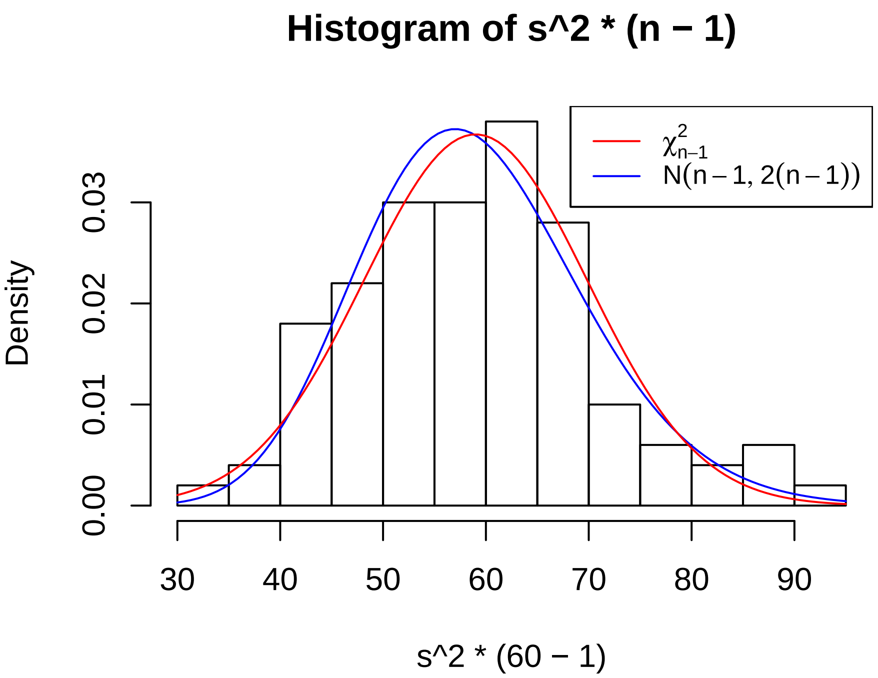
```r
hist(s2 * (num_rvs - 1),
     main = sprintf("Histogram of s^2 * (%d - 1)", num_rvs),
     probability = TRUE,
     xlab = sprintf("s^2 * (%d - 1)", num_rvs))
curve(dchisq(x, df = num_rvs - 1), add = TRUE)
```




(c) Repeat (a) and (b) with $n = 60$ independent $N(0, 1)$ random variables. Overlay the histogram with both the density curve of $\chi^2_{n-1}$ and the density curve of $N(n-1, 2(n-1))$ (in two different colours). Comment on the fit.

**Solution**

```r
set.seed(3023)
num_sim = 100
num_rvs = 60
experiment = rnorm(n = num_sim * num_rvs, mean = 0, sd = 1)
data_mat = matrix(experiment, nrow = num_sim)
s2 = apply(data_mat, MARGIN = 1, FUN = var)

hist(s2 * (num_rvs - 1),
     main = sprintf("Histogram of s^2 * (n - 1)"),
     probability = TRUE,
     xlab = sprintf("s^2 * (%d - 1)", num_rvs),
     breaks = 15)
curve(dchisq(x, df = num_rvs - 1),
      add = TRUE,
      col = "blue")
curve(dnorm(x, mean = num_rvs - 1, sd = sqrt(2*(num_rvs - 1))),
      add = TRUE,
      col = "red")
legend(x = "topright",
       legend = c(expression(chi[n-1]^2), expression(N(n-1, 2(n-1)))),
       col = c("red", "blue"),
       lty = c(1,1),
       cex = 0.8)
```




(d) For $n = 60$ compute $P\left[(n-1)S^2 > 68\right]$ using both the exact distribution $\left(\chi^2_{n-1}\right)$ and the normal approximation. Compare the results.

**Solution**

Recall that

$$\frac{(n-1)\,S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}{\sigma^2} \sim \chi^2_{n-1}$$

And in this case, $\sigma = 1$ and $\mu = 0$. Hence we have:

$$(n-1)S^2 = \sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \sim \chi^2_{n-1}$$

Thus to compute the $P\left[(n-1)\,S^2 > 68\right]$ we can use the $\chi^2_{n-1}$ distribution or approximate it with $X \sim N(n-1, 2(n-1))$

```
pchisq(68, df = num_rvs - 1, lower.tail = FALSE)

[1] 0.1975349

pnorm(68, mean = num_rvs - 1, sd = sqrt(2*(num_rvs - 1)), lower.tail = FALSE)

[1] 0.2036888
```

2. (a) When two random variables $(X, Y)$ follow a bivariate normal distribution, the covariance matrix $\Sigma$ is defined as

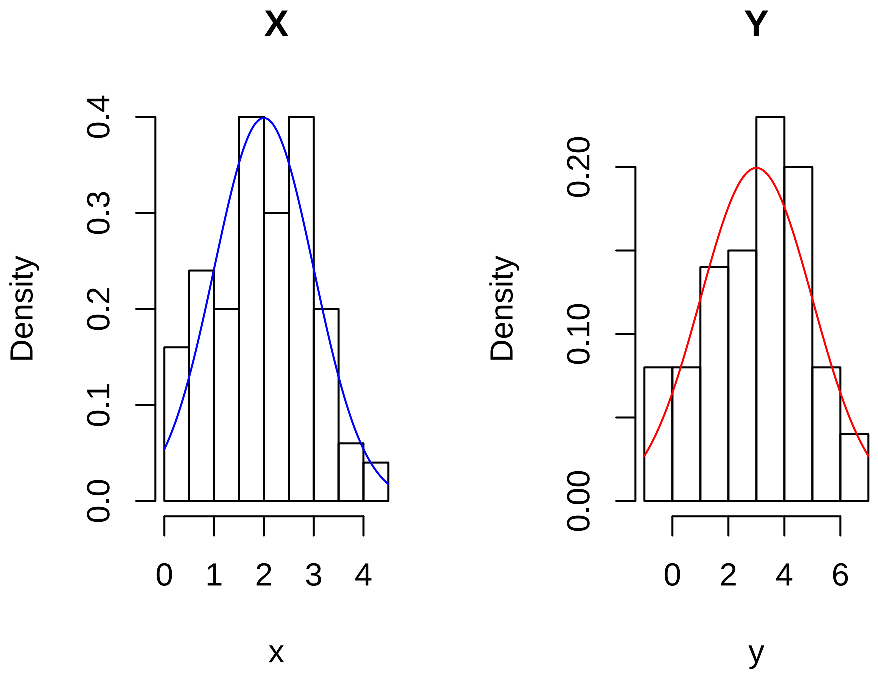
$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Where $\rho$ is the correlation, $\sigma_1, \sigma_2$ are the variances of $X$ and $Y$ respectively. Use `mvrnorm` from the `MASS` library to generate 100 samples from a bivariate normal distribution with $\boldsymbol{\mu} = (\mu_1, \mu_2)$ with $\mu_1 = 2, \mu_2 = 3$ and $\Sigma = \begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$. Call the first column `x` and the second column `y`

```
library(MASS)
set.seed(3023)
cov_matrix = matrix(c(1, 1, 1, 4), nrow = 2, ncol = 2)
cov_matrix

     [,1] [,2]
[1,]    1    1
[2,]    1    4

xy = mvrnorm(100, mu = c(2, 3), Sigma = cov_matrix)
x = xy[ , 1]
y = xy[ , 2]
```

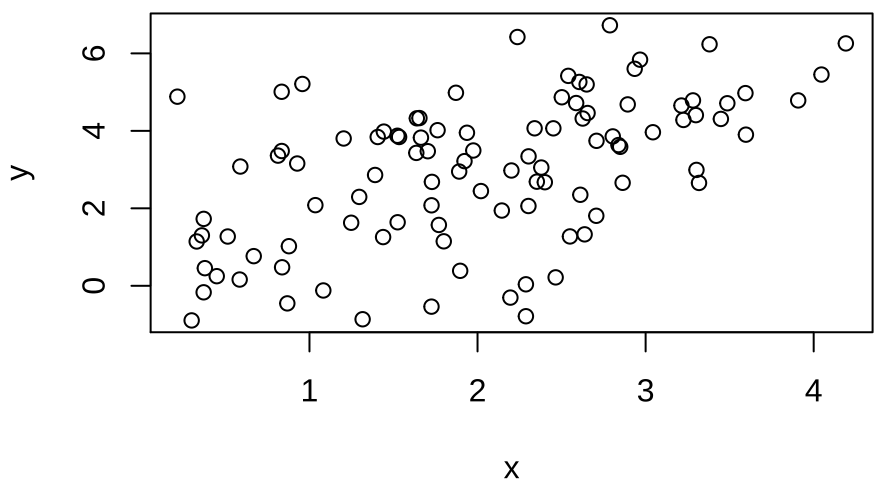
(b) Plot the histogram of `x` and overlay it with the corresponding marginal normal density. Repeat for `y`. (Recall the marginal distribution of $X$ is $N(\mu_1, \sigma_1^2)$)

```
par(mfrow = c(1,2))
hist(x, probability = TRUE, main = "X")
curve(dnorm(x, mean = 2, sd = 1), add = TRUE, col = "blue")


hist(y, probability = TRUE, main = "Y")
curve(dnorm(x, mean = 3, sd = 2), add = TRUE, col = "red")
```




(c) Produce a scatter plot of `x` and `y` (use `plot`). Compute the sample correlation coefficient (use `cor`) and compare with the population correlation $\rho$. (First work out $\rho$ in the $\Sigma$ given.)

```
par(mfrow = c(1,1))
plot(x, y)
```




```
# Get sample correlation (theoretical correlation is 0.5)
cor(x, y)

[1] 0.5283183
```

3. (a) Generate 100 realizations of the minimum of 10 independent exponential(1) random variables. Note the rate parameter in `rexp` is defined as the reciprocal the expectation (check the density function in the help file ?`rexp`).

**Solution**

We recall that if $X \sim exp(\lambda)$ then $\mathbb{E}[X] = \frac{1}{\lambda}$. Hence the `rate` parameter in R is $\lambda$.

```
n = 10
num_sim = 100
lambda = 1
mat_sim = matrix(rexp(n = num_sim * n, rate = lambda),
                 nrow = num_sim,
                 ncol = n)
simulate_mins = apply(mat_sim, MARGIN = 1, FUN = min)
```

(b) Plot the histogram and overlay it with the density of exponential(1/10) (`rate`=10) distribution. Comment on the fit.

```
par(mfrow = c(1,1))
hist(x = simulate_mins, probability = TRUE,
     xlab = "minimums",
     ylab = "probability")
curve(dexp(x, rate = n), add = TRUE)
```