SABS PROBABILITY PRACTICAL SOLUTIONS

DAVID STEINSALTZ

(1) Blood groups are distributed in the UK as follows:

O	A	B	AB
48%	39%	10%	3%

(a) If two people are picked at random from the population, what is the chance that their blood is of the same type? Of different types?

$$\begin{aligned}\mathbb{P}\{\text{same type}\} &= \mathbb{P}\{\text{both A}\} + \mathbb{P}\{\text{both B}\} + \mathbb{P}\{\text{both O}\} + \mathbb{P}\{\text{both AB}\}\\ &= 0.48^2 + 0.39^2 + 0.10^2 + 0.03^2\\ &= 0.39.\end{aligned}$$

Also

$$\mathbb{P}\{\text{different type}\} = 1 - \mathbb{P}\{\text{same type}\} = 0.61.$$

(b) If four people are picked at random, let p_k be the probability that there are exactly k different blood types among them. Find all values of p_k.

$$\begin{aligned}p_1 &= \mathbb{P}\{\text{all A}\} + \mathbb{P}\{\text{all B}\} + \mathbb{P}\{\text{all O}\} + \mathbb{P}\{\text{all AB}\}\\ &= 0.48^4 + 0.39^4 + 0.10^4 + 0.03^4\\ &= 0.076.\\ p_4 &= 4! \cdot 0.48 \cdot 0.39 \cdot 0.10 \cdot 0.03\\ &= 0.013.\\ p_3 &= 12\Big(0.48 \cdot 0.39 \cdot 0.10\,(0.48 + 0.39 + 0.10)\\ &\qquad 0.48 \cdot 0.10 \cdot 0.03\,(0.48 + 0.10 + 0.03)\\ &\qquad 0.48 \cdot 0.39 \cdot 0.03\,(0.48 + 0.39 + 0.03)\\ &\qquad 0.03 \cdot 0.39 \cdot 0.10\,(0.03 + 0.39 + 0.10)\Big)\\ &= 0.296.\\ p_2 &= 1 - p_1 - p_3 - p_4\\ &= 0.615.\end{aligned}$$

(2) In a genetic experiment, the offspring of a particular cross have 25% chance of being yellow, and 75% chance of being green.

(a) If there are 10 offspring, calculate the probability that at least 8 are green.

The probability of any one being green is 0.75. So the number of greens has binomial distribution with parameters $(10, 0.75)$. The probability of at least 8 is

$$\sum_{i=8}^{10} \binom{10}{i} (0.75)^i (0.25)^{10-i} = 0.526.$$

(b) Suppose they also have a 25% chance of being short and 75% chance of being tall. Calculate the expected number that are tall and yellow. What assumptions do you need to make?
Assuming the two characters are independent, the probability of being tall and yellow is $0.75 \cdot 0.25 = 0.1875$. So the expected number out of ten trials is 1.875.

(3) Continuing an example from the lecture, suppose there is a disease that occurs in three forms: Mild, Severe, and Lethal. There is a gene that is known to occur in two *alleles* (variants), denoted A_1 and A_2, where the A_1 allele provides some protection against the symptoms of the disease, but does not prevent the disease. 75% of the general population has A_1, and 75% of those with the disease has A_1. Of those people with A_1 who have the disease, 90% have the Mild form, and the rest have the Severe form. The A_2 sufferers are evenly split between the Severe and Lethal forms.

Suppose you observe a patient with the Severe form. Calculate the probability that the patient is of type A_1. Do this with Bayes' Rule and with natural frequencies.

$$\begin{aligned}\mathbb{P}\{\text{ Severe } \,|\, A_1\} &= 0.1,\\ \mathbb{P}\{\text{ Severe } \,|\, A_2\} &= 0.5,\\ \mathbb{P}(A_1) &= 0.75.\end{aligned}$$

By Bayes Rule,

$$\begin{aligned}\mathbb{P}\{A_1 \,|\, \text{ Severe }\} &= \frac{\mathbb{P}\{\text{ Severe } \,|\, A_1\}\mathbb{P}(A_1)}{\mathbb{P}\{\text{ Severe } \,|\, A_1\}\mathbb{P}(A_1) + \mathbb{P}\{\text{ Severe } \,|\, A_2\}\mathbb{P}(A_2)}\\ &= \frac{0.1 \times 0.75}{0.1 \times 0.75 + 0.5 \times 0.25}\\ &= 0.375.\end{aligned}$$

Using Natural Frequencies, we consider 1000 randomly chosen patients. 750 are type A_1, and of them 75 have Severe disease. 250 are type A_2, and of them $250 \times 0.5 = 125$ have Severe disease. So there are a total of 200 with Severe disease, of whom 75 are of type A_1, yielding a proportion $75/200 = 0.375$.

(4) I roll a fair die, and then flip a number of fair coins equal to the number that came up on the die.

(a) Calculate the probability that exactly four heads come up on the coins.
Let X be the outcome of the die roll, and Y the number of heads in the coin flips. Then

$$\begin{aligned}\mathbb{P}\{Y = 4\} &= \sum_{i=1}^{6} \mathbb{P}\{X = i\}\mathbb{P}\{Y = 4 \,|\, X = i\}\\ &= \sum_{i=4}^{6} \frac{1}{6}\binom{i}{4}\left(\frac{1}{2}\right)^4\left(\frac{1}{2}\right)^{i-4}\\ &= \frac{1}{6}\left(\frac{1}{16} + \frac{5}{32} + \frac{15}{64}\right)\\ &= \frac{29}{384}\\ &= 0.0755.\end{aligned}$$

(b) Given that three heads come up, calculate the probability that the die roll was 5.

$$\begin{aligned}\mathbb{P}\{X = 5 \,|\, Y = 4\} &= \frac{\mathbb{P}\{X = 5 \,\&\, Y = 4\}}{\mathbb{P}\{Y = 4\}} \\ &= \frac{5/32 \cdot 1/6}{29/384} \\ &= \frac{10}{29} \\ &= 0.345.\end{aligned}$$

(5) Suppose we have a sequence of independent trials, each with probability p of success. Let X be the number of the trial on which you have the first success.

(a) Calculate the probability mass function $\mathbb{P}\{X = k\}$.
The event $\{X = k\}$ occurs exactly when there are $k-1$ failures followed by a success. The probability is $\mathbb{P}\{X = k\} = (1-p)^{k-1}p$.

(b) Calculate the expectation and variance of X.
Starting from the formula for the sum of a geometric series

$$\frac{1}{p} = \sum_{k=0}^{\infty}(1-p)^k$$

we take two derivatives to get

$$\frac{1}{p^2} = \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$\frac{2}{p^3} = \sum_{k=1}^{\infty} k(k-1)(1-p)^{k-2} = \sum_{k=1}^{\infty}(k+1)k(1-p)^{k-1}.$$

And subtracting these we get

$$\frac{2}{p^3} - \frac{1}{p^2} = \sum_{k=1}^{\infty} k^2(1-p)^{k-1}.$$

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=1}^{\infty} k\mathbb{P}\{X = k\} \\ &= \sum_{k=1}^{\infty} kp(1-p)^{k-1} \\ &= p \cdot \frac{1}{p^2} \\ &= \frac{1}{p}.\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=1}^{\infty} k^2 \mathbb{P}\{X=k\} \\ &= p\left(\frac{2}{p^3} - \frac{1}{p^2}\right) \\ &= \frac{2}{p^2} - \frac{1}{p} \\ \mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\ &= \frac{1}{p^2} - \frac{1}{p} \\ &= \frac{1-p}{p^2}\end{aligned}$$

(6) Stroke patients with aphasic deficits are each given a number of straightforward tasks in a psychometric test. The number of errors made by 123 patients are shown in the table below. Calculate the mean and variance of the number of errors per patient and comment on these values. Fit a Poisson distribution and comment on how well it fits the observed data.

Number of errors	0	1	2	3	4	5 or more
Number of patients	5	30	56	15	10	7

$\bar{x} = \frac{5\times0+30\times1+56\times2+15\times3+10\times4+7\times5}{123} = 2.13$

$s^2 = \frac{5(0-\bar{x})^2+30(1-\bar{x})^2+56(2-\bar{x})^2+15(3-\bar{x})^2+10(4-\bar{x})^2+7(5-\bar{x})^2}{123-1} = 1.36$

The mean and variance are not very close which suggests a Poisson distribution may not be a good fit.

x	0	1	2	3	4	5 or more
$P(X=x)$	0.119	0.253	0.270	0.191	0.102	0.065
Expected	14.6	31.1	33.2	23.5	12.5	8.01
Observed	5	30	56	15	10	7

The fitted values are not especially good. The observed values are more peaked around the value of 2 than the fitted Poisson.

(7) Let X have uniform distribution on the interval $[a,b]$, defined as the continuous distribution whose density is constant on that interval and 0 outside it.

(a) What are the density and the cdf of this distribution?

The density is

$$f(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{if } x > b. \end{cases}$$

The cdf is

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b. \end{cases}$$

(b) What are the expectation and variance?

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} f(x)x dx \\
&= \frac{1}{b-a}\int_a^b x dx \\
&= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^2-a^2}{2}\right] \\
&= \frac{b+a}{2}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[X^2] &= \int_{-\infty}^{\infty} f(x)x^2 dx \\
&= \frac{1}{b-a}\int_a^b x^2 dx \\
&= \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b \\
&= \frac{1}{b-a}\left[\frac{b^3-a^3}{2}\right] \\
&= \frac{b^2+ab+b^2}{3} \\
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \frac{b^2+ab+b^2}{3} - \frac{(b+a)^2}{4} \\
&= \frac{b^2-2ab+b^2}{12} \\
&= \frac{(b-a)^2}{12}
\end{aligned}$$

(8) In a certain country the heights of adult males have mean 170cm and standard deviation 10cm, and the heights of adult females have mean 160cm and standard deviation 8cm; for each sex the distribution of heights approximates closely to a normal probability model. On the hypothesis that height is not a factor in selecting a mate, calculate the probability that

(a) a husband and wife selected at random are both taller than 164cm;

Let X =wife height, Y =husband height, and let the cdfs be F_X and F_Y. This is just the product $(1-F_X(164))(1-F_Y(164))$. We could compute it directly with the `R` command `(1-pnorm(164,170,10))*(1-pnorm(164,160,8))`, giving the result 0.224. Alternatively, we can convert this to statements about a standard normal random variable Z. For X we take $Z = (X-160)/8$. This is standard normal (mean 0, variance 1), so

$$\mathbb{P}\{X > 164\} = \mathbb{P}\{Z > 0.5\}.$$

Similarly for Y.

(b) in a randomly selected husband and wife the wife is taller than the husband;
The difference is a normal random variable. We know that the expectation of $X - Y$ is the difference in expectations -10, and the variance is the sum of the variances $8^2 + 10^2 = 164$, since we assumed X and Y to be independent. Thus we can compute $\mathbb{P}\{X - Y > 0\}$ with `1-pnorm(0,-10,sqrt(164))`, yielding 0.217.

(c) the average height of a random couple is greater than 168cm. [You may use the fact that the sum of two independent normal random variables is also normal.]
$(X + Y)/2$ is normal with mean 165 and variance $(8^2 + 10^2)/4 = 41$. We than compute `1-pnorm(168,165,sqrt(41))`, yielding 0.320.

(9) 100 students each perform an experiment to estimate a parameter μ, and each one independently computes a 99% confidence interval for μ. What is the probability that there will be at least 3 students whose confidence intervals do not include μ? (Hint: Use the binomial distribution or the Poisson distribution.)
The probability of a confidence interval not including μ is 0.01 and there are 100 of them, so the expected number X of "bad confidence intervals" is 1. Assuming they are independent, we may model the number X as Poisson with parameter 1. The probability mass function is

$$\mathbb{P}\{X = k\} = e^{-1} \cdot \frac{1^k}{k!}.$$

Thus

$$\begin{aligned}\mathbb{P}\{X \geq 3\} &= 1 - \mathbb{P}\{X \leq 2\} \\ &= 1 - e^{-1}\left(1 + 1 + \frac{1}{2}\right) \\ &= 0.080.\end{aligned}$$

We could also compute this with the `R` command `1-ppois(2,1)`. The more exact solution would use the binomial distribution with parameters $(100, 0.01)$. In `R` this would be `1-pbinom(2,100,.01)`, yielding 0.079.

(10) (a) Using `R` compute the following:

(i) $\mathbb{P}\{X = 112\}$ where X is binomial with $n = 200$, $p = 0.6$.

$$P\{X = 112\} = \binom{200}{112} 0.6^{112} 0.4^{88} = .0293.$$

```
> dbinom(112,200,.6)
[1] 0.02933229
```

(ii) $\mathbb{P}\{X \geq 4\}$ where X is Poisson with parameter 8.

$$\mathbb{P}\{X \geq 4\} = 1 - e^{-8} \sum_{k=0}^{3} \frac{8^k}{k!} = 0.958.$$

```
> 1-ppois(3,8)
[1] 0.9576199
```

(iii) $\mathbb{P}\{1 < X < 2\}$ where X is Exponential with parameter 2. The density is $2e^{-2x}$ for $x \geq 0$. So

$$\mathbb{P}\{1 < X < 2\} = 2\int_1^2 e^{-2y} dy = e^{-2} - e^{-4} = 0.117.$$

```
> pexp(2,2)-pexp(1,2)
[1] 0.1170196
```

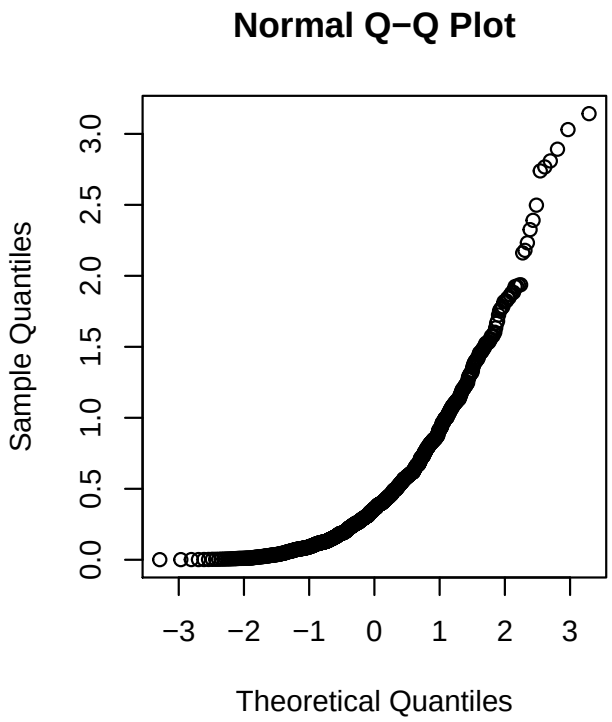
(iv) $\mathbb{P}\{X < 2\}$ where X is normal with mean 3 and variance 7. $Z = (X-3)/\sqrt{7}$ is standard normal. So $\mathbb{P}\{X < 2\} = \mathbb{P}\{Z < (2-3)/\sqrt{7}\} = \mathbb{P}\{Z < -0.378\}$.

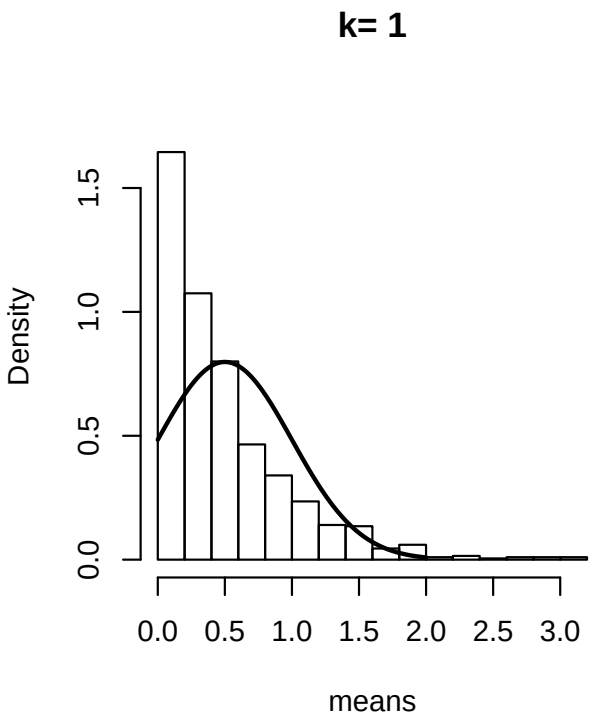
```
> pnorm(2,3,sqrt(7))
[1] 0.3527285
```

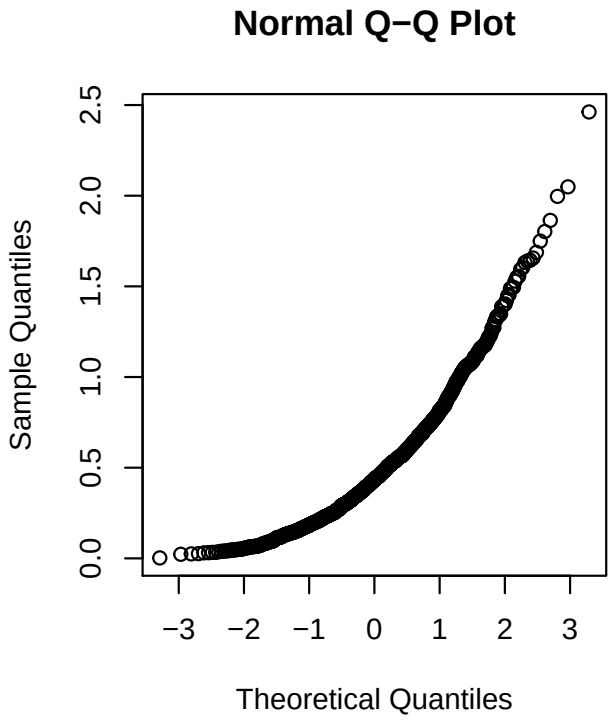
```
par(mfrow=c(2,2)) # Make a 2x2 grid of plots
for (k in c(1,2,10,100)){
x=seq(0,2,.01)
sigma=.5/sqrt(k)
y=dnorm(x,.5,sigma)
maxdens=1.8*k^.35 # Somewhat arbitrary limit to make the
    plot
#tall enough to include the whole plot
samples=array(rexp(k*1000,2),c(1000,k))
means=apply(samples,1,mean)
qqnorm(means)
hist(means,freq=FALSE,breaks=15,ylim=c(0,maxdens))
lines(x,y,lwd=2)
wait=readline() # Wait for <return>
}
```

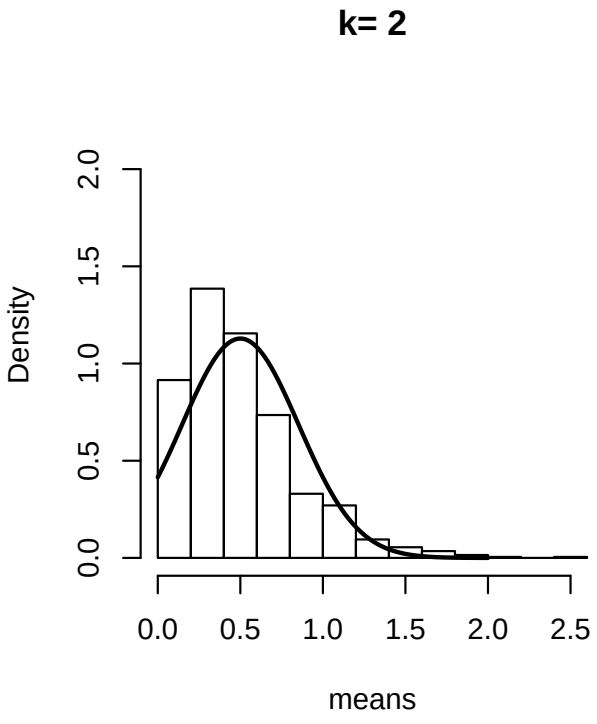
.

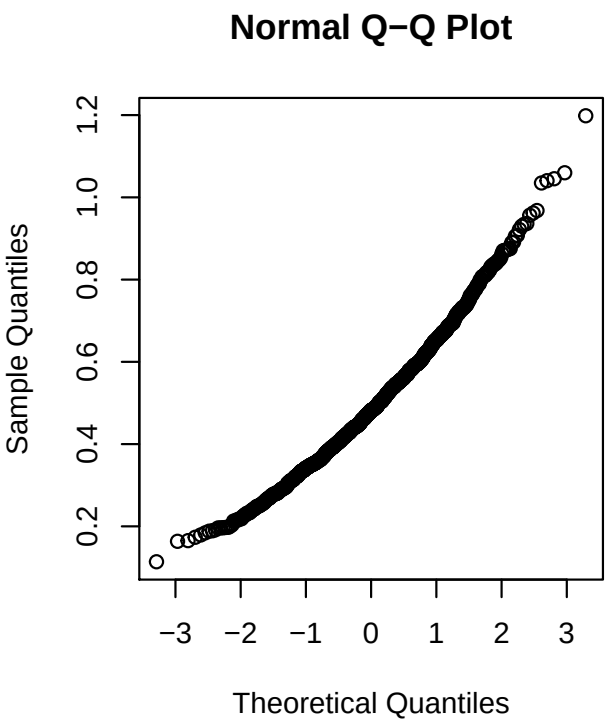
(b)

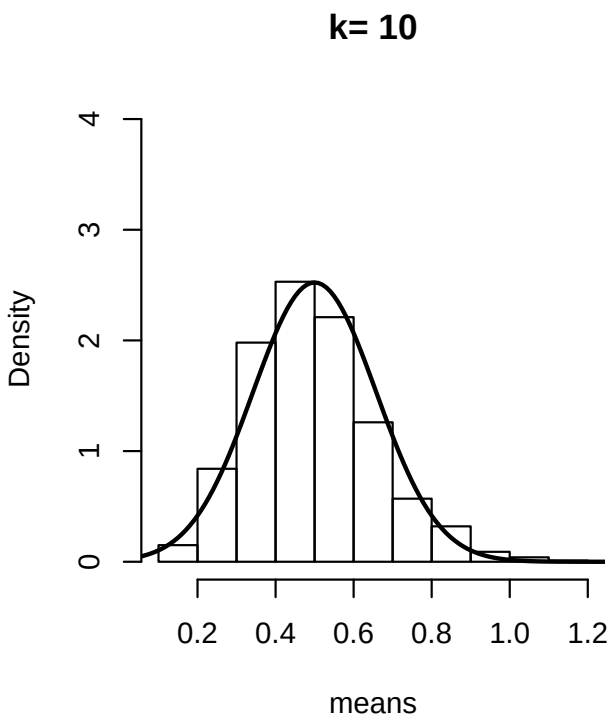

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0.0
0.5
1.0
1.5
2.0
2.5
3.0
−3
−2
−1
0
1
2
3

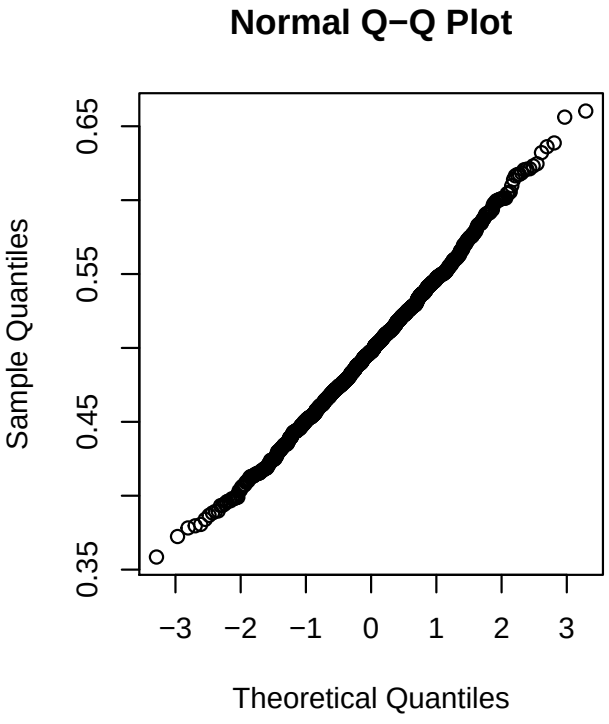

k= 1
Density
means
0.0
0.5
1.0
1.5
2.0
2.5
3.0

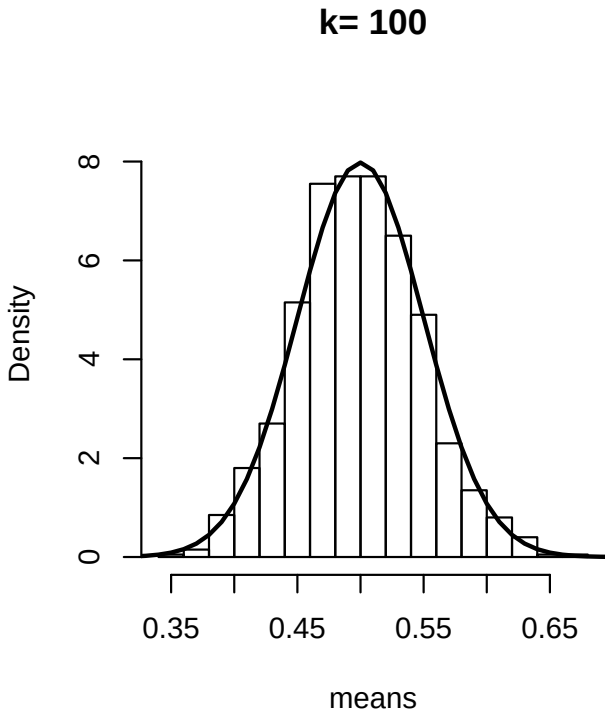

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0.0
0.5
1.0
1.5
2.0
2.5
−3
−2
−1
0
1
2
3


k= 2
Density
means
0.0
0.5
1.0
1.5
2.0
2.5


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0.2
0.4
0.6
0.8
1.0
1.2
−3
−2
−1
0
1
2
3


k= 10
Density
means
0
1
2
3
4
0.2
0.4
0.6
0.8
1.0
1.2


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
0.35
0.45
0.55
0.65
−3
−2
−1
0
1
2
3


k= 100
Density
means
0
2
4
6
8
0.35
0.45
0.55
0.65