Monte Carlo simulations

The ASTA team

Contents

1 Basic principle **1**
1.1 Many, many simulations . . . 2

2 Example 1 **2**

3 (Pseudo) random number generation **2**

4 Example 1 - cont'd **3**

5 Example 2 **3**

6 More complicated simulations **6**

7 Variance is important **6**

8 Example 3: Bivariate reliability problem **6**
8.1 Model . . . 6
8.2 Simulations . . . 7
8.3 Results . . . 7
8.4 Theoretical answer . . . 8

9 Sensitivity **8**
9.1 Illustrating $g(B, C) < 0$ . . . 8
9.2 `sapply`, `lapply`, `for` loops . . . 10

10 Example 4: Wholesales offer **10**
10.1 No patent . . . 10
10.2 With patent . . . 11
10.3 Simulating sales . . . 11
10.4 Wholesale offer . . . 12

1 Basic principle

Monte Carlo (MC) estimation of the probability of event A:

- Run an experiment N times, and count how many times A occurred, N_A, say

$$P(A) \approx \frac{N_A}{N}$$

(Frequentist)

1.1 Many, many simulations

Letting the number of experiments approach infinity, $\approx$ becomes $=$:

$$P(A) = \lim_{N \to \infty} \frac{N_A}{N}$$

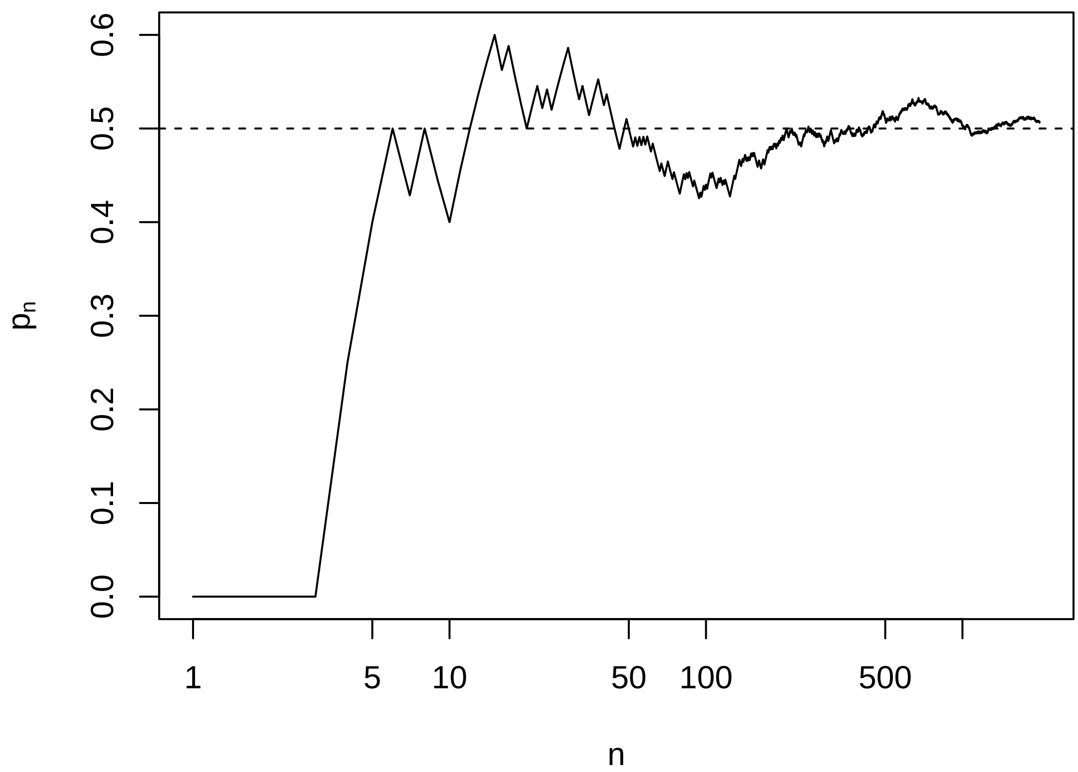
2 Example 1

Recall (lecture 1.2):

- John Kerrich, a South African mathematician, was visiting Copenhagen when World War II broke out. Two days before he was scheduled to fly to England, the Germans invaded Denmark. Kerrich spent the rest of the war interned at a camp in Hald Ege near Viborg, Jutland. To pass the time he carried out a series of experiments in probability theory. In one, he tossed a coin 10,000 times.
- The first 25 observations were (0 = tail, 1 = head):

$$0, 0, 0, 1, 1, 1, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 1, 0, \dots$$

- Plot of the empirical probability $\hat{p}$ of getting a head against the number of tosses n:



(The horizontal axis is on a log scale).

3 (Pseudo) random number generation

```
sample(1:10, 2)
```

```
## [1] 6 9
```

```r
sample(1:10, 2)
```

```
## [1]  2 10
```

```r
set.seed(48)
sample(1:10, 2)
```

```
## [1] 5 3
```

```r
set.seed(48)
sample(1:10, 2)
```

```
## [1] 5 3
```

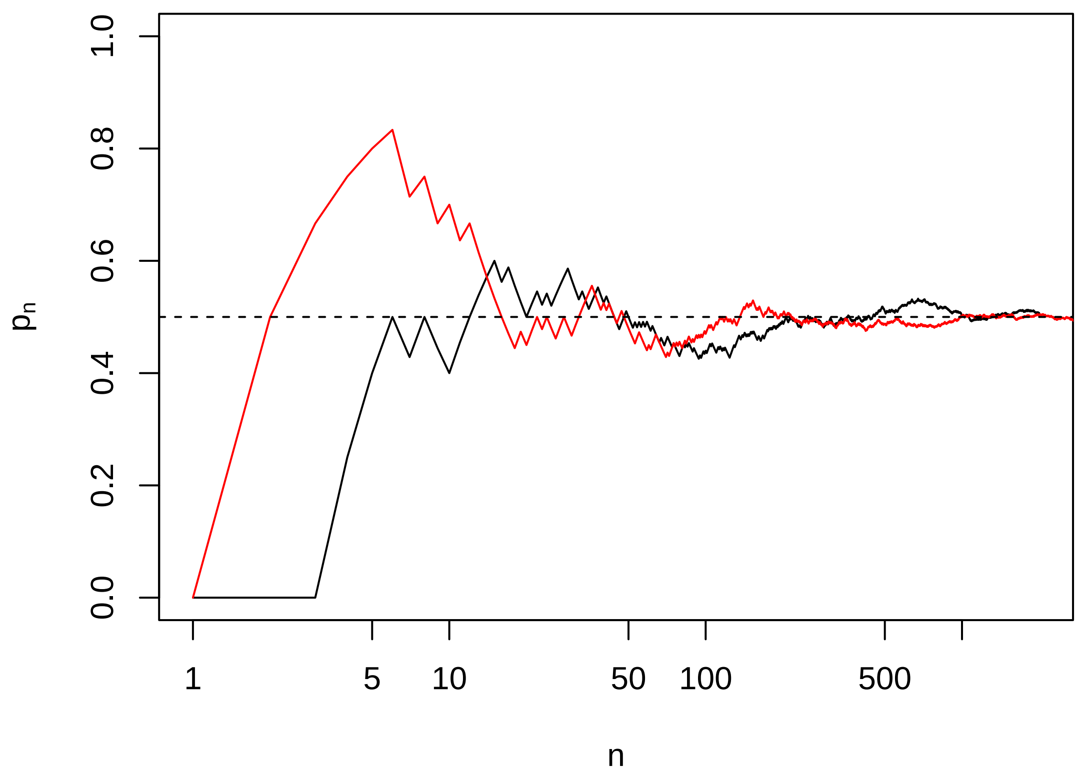
4 Example 1 - cont'd

Let the computer toss the coin (MC simulation):

```r
set.seed(48)
x2 <- sample(c(0, 1), 10000, replace = TRUE, prob = c(0.5, 0.5))
head(x2)
```

```
## [1] 0 1 1 1 1 1
```

- Black: John Kerrich
- Red: MC



5 Example 2

Recall (lection 2.1):

- A company produces cylindrical components for automobiles. It is important that the mean component diameter is $\mu = 5$mm. The standard deviation is $\sigma = 0.1$mm.
- An engineer takes a random sample of $n = 100$ components. These have an average diameter of $\bar{x} = 5.027$. Is it reasonable to think $\mu = 5$?
- If the population of components has the correct mean, then

$$\bar{X} \approx \texttt{norm}\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = \texttt{norm}\left(5, \frac{0.1}{\sqrt{100}}\right) = \texttt{norm}(5, 0.01).$$

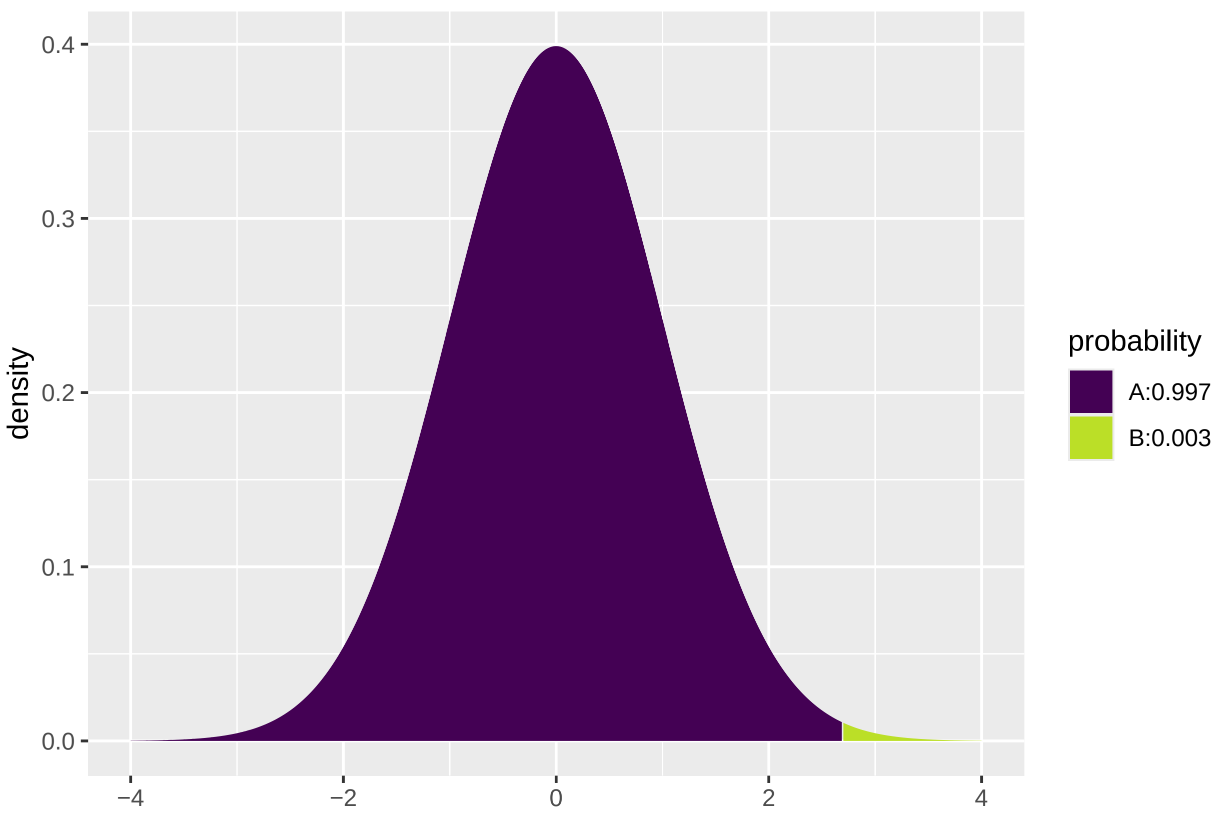
- For the actual sample this gives the observed z-score

$$z_{obs} = \frac{\bar{x} - 5}{0.01} = 2.7$$

which should come from an approximate standard normal distribution.

- The probability of getting a higher z-score is:

```
prob <- 1 - pdist("norm", mean = 0, sd = 1, q = 2.7, xlim = c(-4, 4))
```




```
prob
```

```
## [1] 0.0035
```

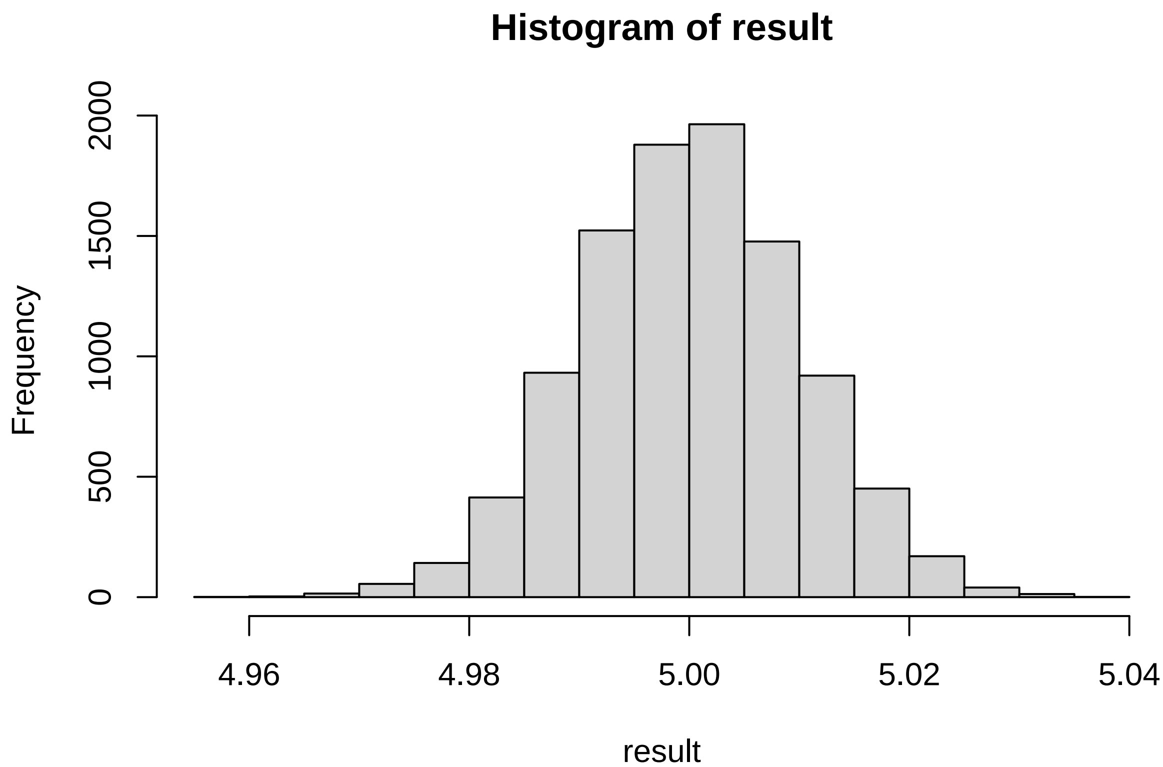
- Thus, it is highly unlikely that a random sample has such a high z-score. A better explanation might be that the produced components have a population mean larger than 5mm.

Use MC simulation:

```r
set.seed(48)
N <- 10000
result <- rnorm(n = N, mean = 5, sd = 0.1/sqrt(100))
sprintf("%0.5f", head(result))
```

```
## [1] "5.00200" "4.97220" "4.99304" "5.02075" "5.00790" "5.00490"
```

```r
hist(result)
```



```r
sum(result >= 5.027)
```

```
## [1] 34
```

```r
sum(result >= 5.027)/N
```

```
## [1] 0.0034
```

```r
prob1 <- mean(result >= 5.027)
prob1
```

```
## [1] 0.0034
```

```r
prob1
```

```
## [1] 0.0034
```

```r
prob
```

```
## [1] 0.0035
```

```
prob1 - prob
```

```
## [1] -6.7e-05
```

```
(prob1 - prob)/prob
```

```
## [1] -0.019
```

6 More complicated simulations

- `sample()`
- `rnorm()`
- `replicate()`:

```
set.seed(48)
x2 <- sample(c(0, 1), 10000, replace = TRUE, prob = c(0.5, 0.5))
mean(x2)
```

```
## [1] 0.49
```

```
set.seed(48)
x2 <- replicate(N, {
  sample(c(0, 1), 1, prob = c(0.5, 0.5))
})
mean(x2)
```

```
## [1] 0.49
```

7 Variance is important

- Average of 45M and 55M is 50M
- Average of 1M and 99M is 50M

Averages are not always sufficient. Think of budgets:

- On average the profit will be 50M
  - But… What is the risk that it will become e.g. below 20M?
  - …or 0 even?
- What is the probability that it is e.g. below 0 (deficit)?

8 Example 3: Bivariate reliability problem

Limit state function:

- defines failure criterion
- function of all random variables
- failure when $g(\dots) \leq 0$

8.1 Model

$$g(B, C) = B - C \leq 0$$

- $B \sim \mathcal{N}(100, 10)$: budget of a given project (in \$1,000)
- $C \sim \mathcal{N}(50, 10)$: expected costs (in \$1,000)

Assess

$$P(g(B,C) < 0)$$

8.2 Simulations

```
set.seed(48)
N <- 100000
Bdata <- rnorm(N, mean = 100, sd = 10)
Cdata <- rnorm(N, mean = 50, sd = 10)
gdata <- Bdata - Cdata
```

```
sum(gdata < 0)
```

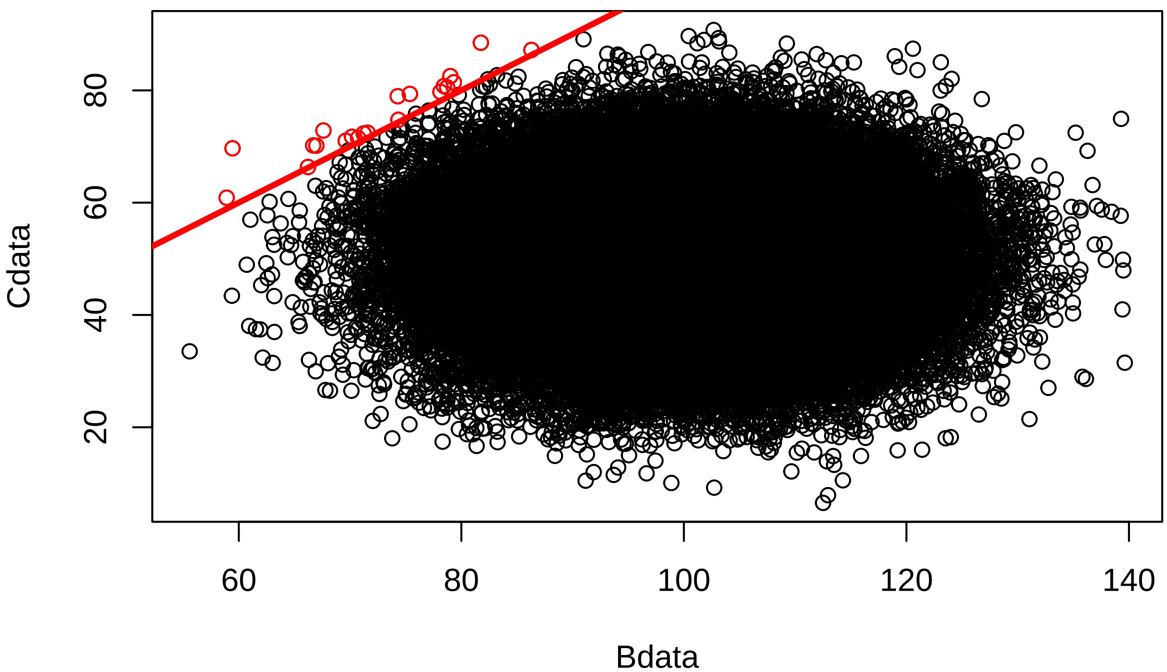
```
## [1] 21
```

```
prob1 <- mean(gdata < 0)
prob1
```

```
## [1] 0.00021
```

8.3 Results

```
plot(Bdata, Cdata)
abline(a = 0, b = 1, col = "red", lwd = 3)
points(Bdata[gdata < 0], Cdata[gdata < 0], col = "red")
```



8.4 Theoretical answer

For the particular case, we can exploit that the sum of two normal distributions is also a normal distribution, and due to independence the variance is the sum of the two variances:

```
g_mean <- 100 - 50
g_var <- 10^2 + 10^2
prob <- pdist("norm", q = 0,
              mean = g_mean,
              sd = sqrt(g_var), plot = FALSE)
prob
```

```
## [1] 2e-04
```

```
prob - prob1
```

```
## [1] -6.5e-06
```

```
(prob - prob1) / prob
```

```
## [1] -0.032
```

9 Sensitivity

- $B \sim \mathcal{N}(100, 10)$: budget of a given project (in \$1,000)
- $C \sim \mathcal{N}(50, 10)$: expected costs (in \$1,000)
- $g(B, C) = B - C$

How does

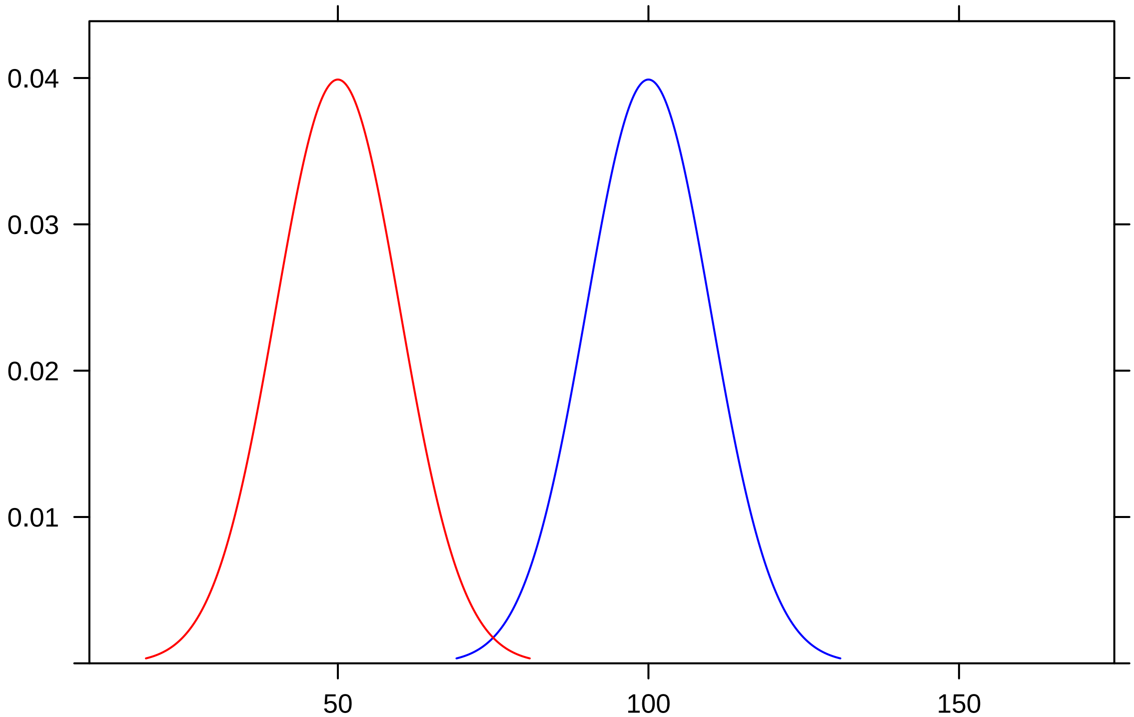
$$P(g(B, C) < 0)$$

depend on the variance/standard deviations?

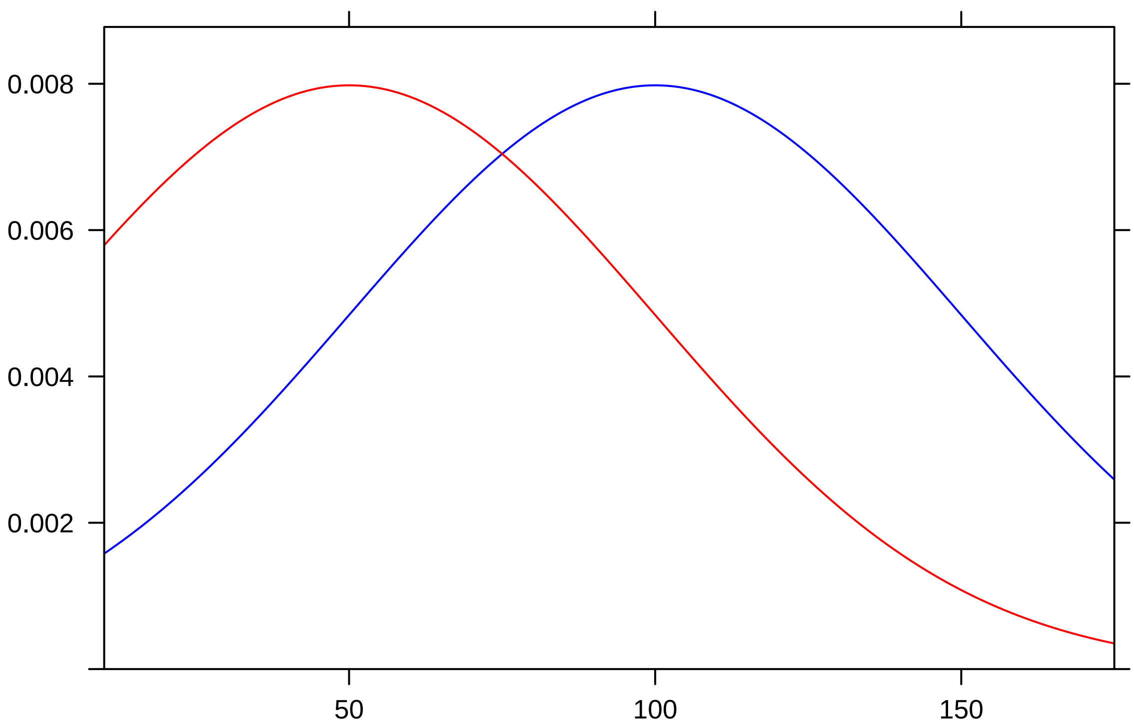
9.1 Illustrating $g(B, C) < 0$

Area where red is larger than blue illustrates $P(g(B, C) < 0)$:

```
plotDist("norm", mean = 100, sd = 10,
         xlim = c(10, 175), col = "blue")
plotDist("norm", mean =  50, sd = 10,
         add = TRUE, col = "red")
```




```r
plotDist("norm", mean = 100, sd = 50,
         xlim = c(10, 175), col = "blue")
plotDist("norm", mean =  50, sd = 50,
         add = TRUE, col = "red")
```

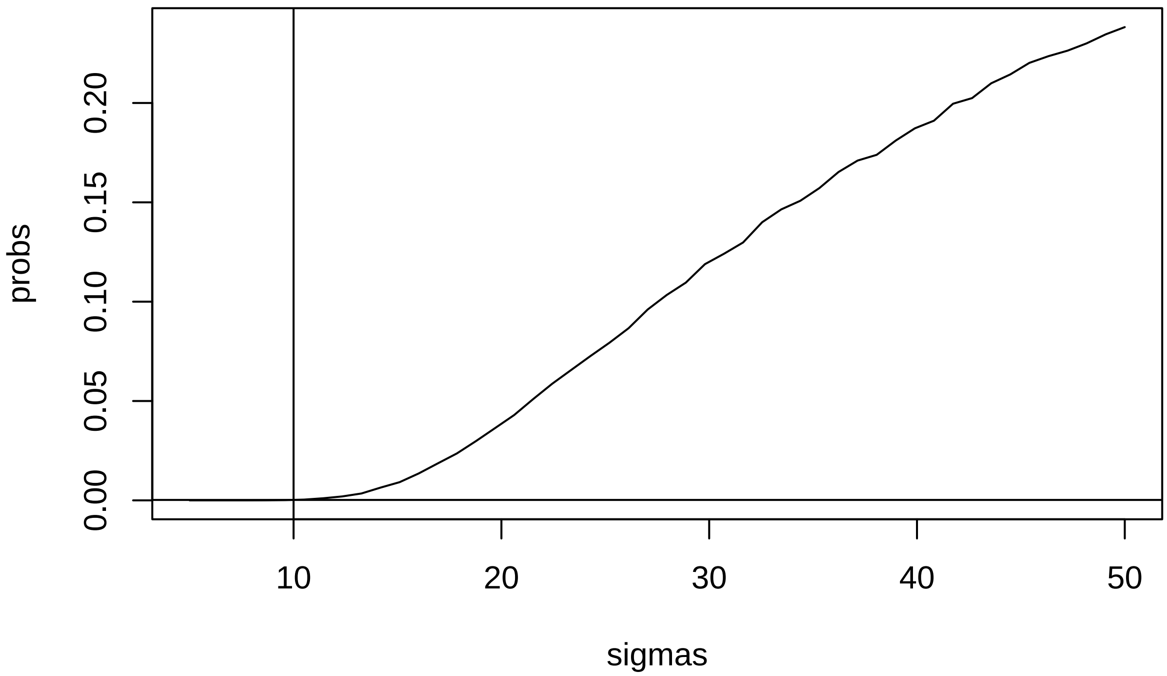


9.2 sapply, lapply, for loops

Using `sapply` (other possibilities: `lapply`, `for` loops):

```
set.seed(48)
N <- 100000
sigmas <- seq(5, 50, length.out = 50)
probs <- sapply(sigmas, function(sigma) {
  Bdata <- rnorm(N, mean = 100, sd = sigma)
  Cdata <- rnorm(N, mean = 50, sd = sigma)
  gdata <- Bdata - Cdata
  mean(gdata < 0)
})
```

```
plot(sigmas, probs, type = "l")
abline(h = prob1)
abline(v = 10)
```



10 Example 4: Wholesales offer

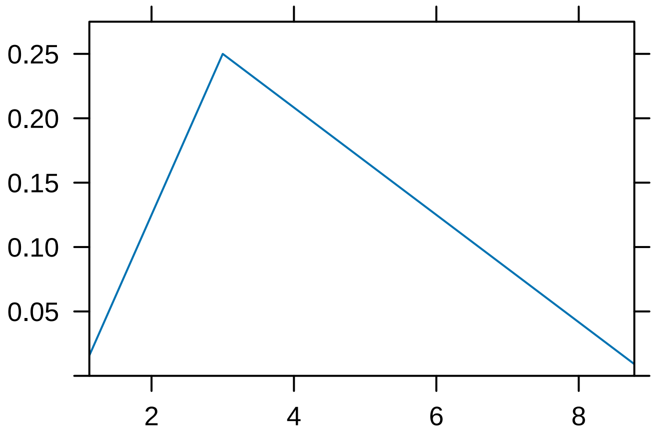
- Newly established business with a innovative product
- Decided to apply for a patent
  - 40% chance to get patent

10.1 No patent

The expected sales for next year is \$1-\$9 mio. with \$3 mio. being the most likely value.

Triangular distribution:

```r
plotDist("tri", params = list(min = 1, max = 9, mode = 3))
```

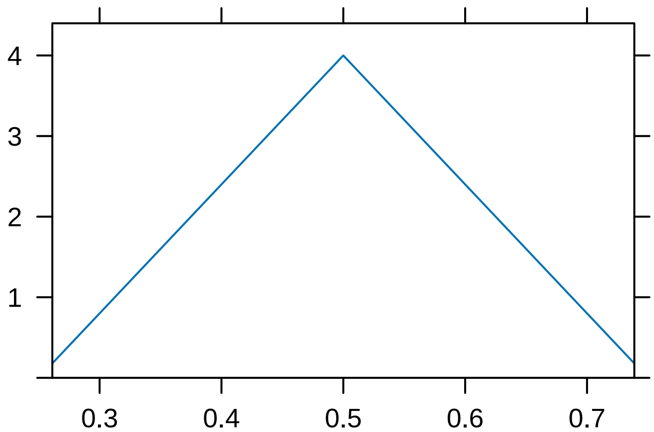


10.2 With patent

Sales goes up 25%-75% with 50% percent being the most likely value

Triangular distribution:

```r
plotDist("tri", params = list(min = 0.25, max = 0.75, mode = 0.5))
```



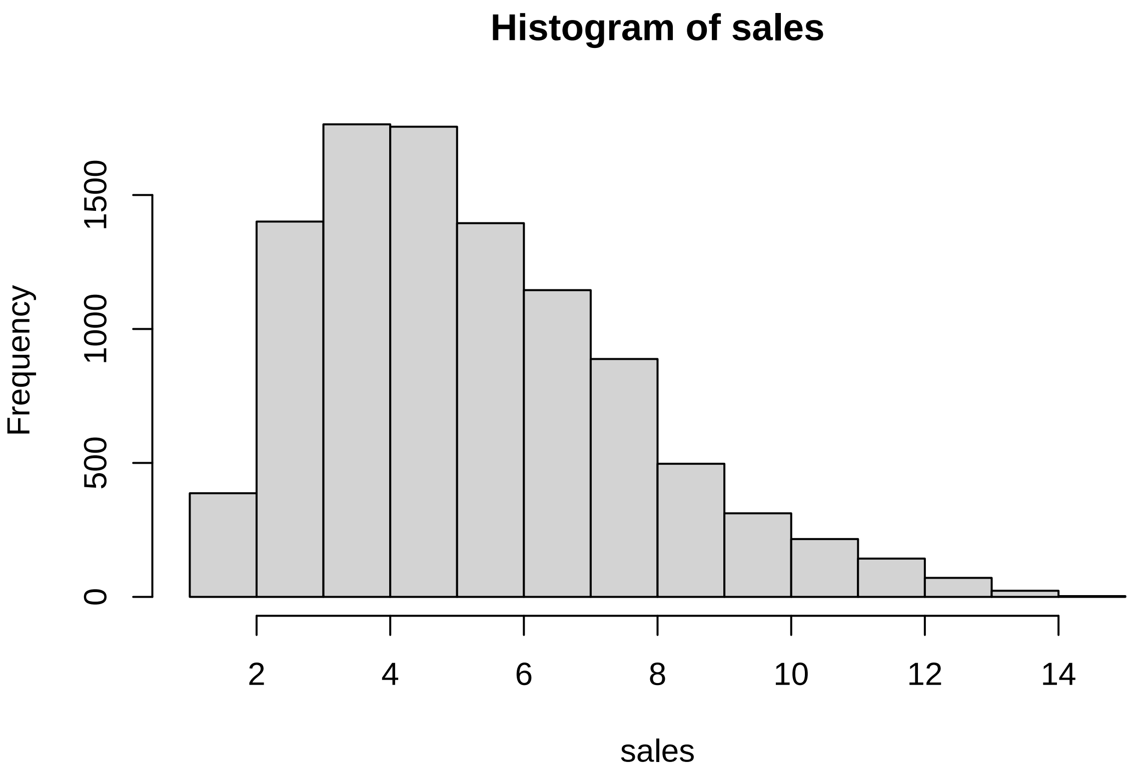
10.3 Simulating sales

```r
set.seed(48)
N <- 10000
got_patent <- sample(x = c(0, 1), prob = c(0.6, 0.4), size = N, replace = TRUE)
table(got_patent)
```

```
## got_patent
##    0    1
## 5935 4065
```

```r
sales_base <- rtri(N, min = 1, max = 9, mode = 3)
patent_markup <- rtri(N, min = 0.25, max = 0.75, mode = 0.5)
sales <- sales_base + got_patent*patent_markup*sales_base
```

```r
hist(sales)
```




```r
sales_avg <- mean(sales)
sales_avg
```

```
## [1] 5.2
```

```r
quantile(sales, probs = c(0.025, 0.975))
```

```
##  2.5% 97.5%
##   1.8  10.9
```

10.4 Wholesale offer

Wholesale offer to buy your entire production for the next year for $6 mio.

Take the deal?

Many, many aspects:

- Willing to take a risk?
- Future plans

Initially: How likely is it that we can sell the production for more ourselves?

```r
prob <- mean(sales > 6)
prob
```

```
## [1] 0.33
```

Does not answer the question on its own, but it helps (e.g. in assessing the risk).