

2 Further simulation methods

In this section we look at two topics:

1. Use of the `sample` command in R to perform simulations that help solve combinatorial problems.
2. The inverse CDF method for simulating continuous random variables.

2.1 The sample command

The `sample` command samples elements repeatedly from a given set, and sampling can either be with or without replacement.

```

1 sample(x, n) # Create a random sample of n elements from x without replacement
2 sample(x, n, replace = TRUE) # Create a random sample with replacement
3 sample(x, n, prob = p) # Create a random sample from x using probabilities stored in p

```

Example 2.1: Suppose we randomly permute the letters of the word MISSISSIPPI. What is the expected number of times we have consecutive S's in the jumbled up word? What is the probability the word contains no consecutive S's? (This problem is relevant to the design of clinical trials.)

Solution is possible analytically, but simulation is easy:

```

1 mississippi = function(num_sims) {
2   total_pairs = 0
3   freq = 0
4   letters = c("M","I","S","S","I","S","S","I","P","P","I")
5   len = length(letters)
6   for (i in 1:num_sims) {
7     word = sample(letters, len)
8     #Where are the S's?
9     s = (word == "S")
10    #Where are the pairs?
11    pairs = s[2:11] & s[1:10]
12    if (sum(pairs)>0) {
13      freq = freq + 1;
14    }
15    total_pairs = total_pairs + sum(pairs)
16  }
17  cat("Expectation = ",total_pairs/num_sims, "\n")
18  cat("Probability = ",freq/num_sims, "\n")
19 }

```

Make some notes about how this function works.

Example 2.2: The Ehrenfest model of diffusion. Suppose we have two urns, urn A initially containing n orange balls and urn B containing n grey balls. At every time-step, a ball is chosen uniformly at random from A and a ball is chosen uniformly at random from B . The two balls are then swapped (i.e. the ball from A is put in B and the ball from B is put in A). This repeats for k time steps. Suppose $n = 50$ and let X be the number of grey balls in B after $k = 40$ steps. Use sampling to investigate the distribution of X .

```

1 #This function returns a single value of X
2 ehrenfest = function(n, k) {
3   A = rep(0, n)
4   B = rep(1, n)
5   for (i in 1:k) {
6     ballA = sample(1:n, 1)
7     ballB = sample(1:n, 1)
8
9     if ((A[ballA]==1)&&(B[ballB]==0)) {
10      A[ballA] = 0
11      B[ballB] = 1
12    }
13    else if ((A[ballA]==0)&&(B[ballB]==1)) {
14      A[ballA] = 1
15      B[ballB] = 0
16    }
17  }
18  sum(B)
19 }

```

```

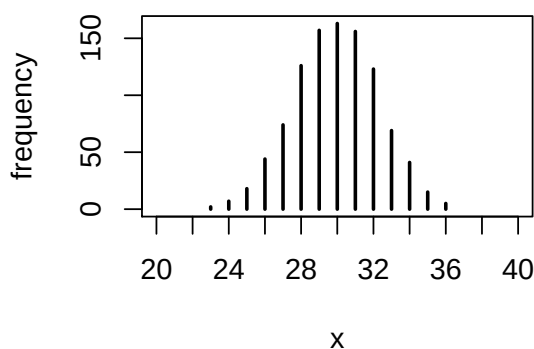
1 # Sample X many times to see the distribution
2 n = 50
3 k = 40
4 num_iterations = 1000
5 x = vector(mode = 'numeric', length = num_iterations)
6 for (j in 1:num_iterations) {
7   x[j] = ehrenfest(n,k)
8 }

```

```

1 plot(table(x), xlim = c(20, 40), xaxt = 'n', xlab = 'x', ylab = 'frequency')
2 axis(1, at = seq(20, 40, 2))

```



Make some notes about how this function works.

2.2 Inverse CDF method of simulating univariate random variables

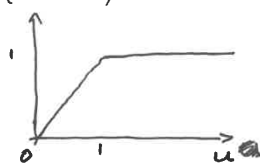
Let $U \sim U(0, 1)$. Suppose $F(x)$ is a well-defined CDF which is invertible. Then the random variable $X = F^{-1}(U)$ has CDF $F(x)$.

Proof:

$$X = F^{-1}(U)$$

$$\begin{aligned} \text{so } \Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(U \leq F(x)) \quad \because F \text{ is an increasing function.} \end{aligned}$$

$$\Pr(U \leq u) = F(x) \quad \text{Since } 0 \leq F(x) \leq 1 \text{ (see graph).}$$



This result enables us to simulate the random variable X very readily provided we can simulate uniform random variables. In R we would make use of the command:

```
x = inv.cdf(runif(1))
```

Here, the function `inv.cdf` is a function that we have written ourselves to return the inverse of the CDF of a particular distribution of interest; see practical 1 for more details.

Example 2.3: Suppose we obtain two observations from a $U(0, 1)$ distribution: 0.1 and 0.85. Use these values to obtain two observations from $X \sim \text{Exp}(2)$.

$$\text{If } X \sim \text{Exp}(2) \text{ then } F_X(x) = 1 - e^{-2x}$$

$$\begin{aligned} \text{If } y = 1 - e^{-2x} \text{ then } e^{-2x} &= 1 - y \\ -2x &= \log(1 - y) \\ x &= -\frac{1}{2} \log(1 - y) \end{aligned}$$

$$\text{Plug in } y = 0.1 \text{ \& } y = 0.85$$

$$\text{Gives realized values } 0.0527, 0.949.$$

Example 2.4: Patients arriving at a doctor's surgery have appointment times 15 minutes apart. The random variable L measures how late a patient is for his/her appointment, with negative values denoting early arrives. The PDF of L is

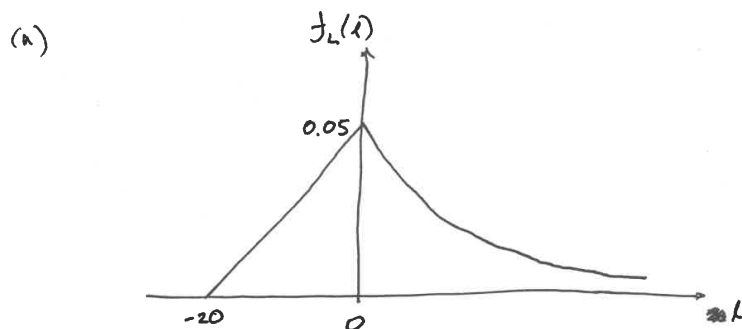
$$f_L(l) = \begin{cases} 0.0025(l+20), & -20 \leq l \leq 0, \\ 0.05e^{-l/10}, & l > 0. \end{cases}$$

(a) Sketch $f_L(l)$.

(b) Show that the CDF of L is

$$F_L(l) = \begin{cases} 0, & l < -20, \\ 0.00125(l+20)^2, & -20 \leq l \leq 0, \\ 1 - \frac{e^{-l/10}}{2}, & l > 0. \end{cases}$$

(c) Suppose that 0.205 and 0.713 are two realized values of $U \sim U(0,1)$. Show that the corresponding realized values of L are -7.19 and 5.55 under the inverse CDF simulation method.



(b) L lies between -20 & $+\infty$. So $F_L(l) = 0$ for $l \leq -20$.

$$\begin{aligned} F_L(l) &= \int_{-20}^l f_L(s) ds \\ &= \int_{-20}^l 0.0025(s+20) ds \\ &= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^l \\ &= 0.00125(l+20)^2 \quad \text{for } l \leq 0 \\ F_L(0) &= 0.5 \end{aligned}$$

$$\begin{aligned}
F_L(l) &= \int_{-20}^0 f_L(s) ds + \int_0^l f_L(s) ds \quad \text{for } l > 0. \\
&= 0.5 + \int_0^l 0.05 e^{-s/10} ds \\
&= 0.5 + \left[0.05 \times -10 e^{-s/10} \right]_0^l \\
&= 0.5 + 0.5 - 0.5 e^{-l/10} \\
&= 1 - \frac{e^{-l/10}}{2}
\end{aligned}$$

(c) $u = 0.205 < 0.5$ so realized value lies between -20 & 0 .

$$\text{and } 0.205 = 0.00125(l+20)^2$$

$$\Rightarrow (l+20)^2 = 164$$

$$\Rightarrow l+20 = \pm 12.8$$

$$\Rightarrow l = -20 + 12.8 \quad (\text{so that } -20 \leq l \leq 0) \\ = -7.2$$

Sim $u = 0.713 > 0$ so realized value > 0

$$\begin{aligned}
1 - \frac{e^{-l/10}}{2} &= 0.713 \Rightarrow -l/10 = \log(2 - 2 \times 0.713) \\
&\Rightarrow l = 5.55.
\end{aligned}$$