Lecturas Importantes

- An Introduction to Statistical Learning with Applications in R
  `http://www-bcf.usc.edu/~gareth/ISL/`.

Preguntas

1. Aquí está un programa recursivo que imprime todas las maneras posibles que una cantidad x (en centavos) se puede componer usando monedas australianas (que vienen en 5, 10, 20, 50, 100, y 200 centavos). Para evitar la repetición, cada posible descomposición se ordena:

```
> cambio <- function(x, y.vec = c()) {
+   # encuentra las posibles formas posibles de tener x usando monedas Australianas
+   # x es dado en centavos y asumimos que es divisible por 5
+   # y.vec son monedas ya utilizada (asi la cantidad total es x + sum(y.vec))
+   if (x == 0) {
+     cat(y.vec, "\n")
+     } else {
+       monedas <- c(200, 100, 50, 20, 10, 5)
+       new.x <- x -monedas
+       new.x <- new.x[new.x >= 0]
+       for (z in new.x) {
+         y.tmp <- c(y.vec, x - z)
+         if (identical(y.tmp, sort(y.tmp))) {
+           cambio(z, y.tmp)
+         }
+       }
+ }
+   return(invisible(NULL))
+ }
```

   Vuelve a escribir este programa para que en lugar de escribir la salida en la pantalla, la devuelve como una lista, donde cada elemento es un vector que da una posible descomposición de x.

2. A Jessica le encantan los juegos y sobretodo calcular la probabilidad de ganar uno. Un dia, su amigo César le propuso dibujar un círculo y luego colocar n reflectores distribuidos regularmente a lo largo de la circunferencia. Luego Jessica elegirá un número X, tal que si un haz de luz choca inicialmente en el reflector 1, chocará cada X reflectores. Si el haz de luz llega a golpear en todos los reflectores, Jessica ganará. Diseña un algoritmo para ayudar a Jessica a calcular la probabilidad que tiene de ganar el juego dado N si $1 \leq N \leq 10^5$.

3. Jessica y Vilma decidieron realizar el siguiente juego:

   (a) El primer jugador elige un número n, el cual será el tamaño de una pila de monedas.

   (b) Luego de elegir n, inicia el primer jugador.

   (c) Se alternan los movimientos de ambos jugadores hasta que el último que no pueda sacar más monedas de la pila pierda. En cada turno, ellos pueden sacar una cantidad de monedas entre 1 y 3 $(1, 2, 3)$.

   Decidieron jugar 6 rondas alternando el que juegue primero. Vilma sabe cómo es la estrategia óptima y el valor de n óptimo para ganarle a Jessica, mientras que Jessica solamente usa la estrategia de sacar 4-`(Jugada anterior de Vilma)` (Si le toca jugar primero, entonces saca en su primer turno n

mod 4 monedas) hasta que la pila se reduzca a un tamaño menor que 4, en cuyo caso sacará todas las monedas.

Si el vencedor es aquel que gane 4 rondas o más. ¿ Qué probabilidad tiene Vilma de ganar asumiendo que Jessica no sabe qué valores de n le beneficiarían para ganar de manera segura?.

4. Una reserva natural boscosa tiene 13 plataformas de observación de aves dispersas en un gran bloque de tierra. Los naturalistas afirman que en cualquier momento, hay un 75 por ciento de posibilidades de ver aves en cada plataforma. Supongamos que caminas a través de la reserva y visitas cada plataforma. Si se asume que todas las condiciones relevantes son satisfechas, sea X una variable aleatoria binomial que representa el número total de plataformas en las que ves aves.

   - Visualice la función de masa de probabilidad de la distribución binomial de interés.
   - ¿Cuál es la probabilidad de ver aves en todos los sitios?.
   - ¿Cuál es la probabilidad de ver aves en más de 9 plataformas?.
   - ¿Cuál es la probabilidad de ver aves entre 8 y 11 plataformas (inclusive)? Confirma tu respuesta, usando funciones de R ya definidas.
   - Simula realizaciones de X que representan 10 visitas diferentes a la reserva. Almacene el vector resultante como un objeto. Calcula la media y la desviación estándar de la distribución de interés.

5. Cada sabado, al mismo tiempo, una puesto individual al lado de un camino cuenta el número de coches que pasan por una ventana en 120 minutos. Con base en el conocimiento previo, el número promedio de coches que pasan durante ese tiempo es exactamente 107. Sea X la variable aleatoria de Poisson del número de coches que pasan por su posición en cada sesión del dia sábado.

   - ¿ Cuál es la probabilidad de que más de 100 autos pasen un dia sábado?.
   - Determina la probabilidad de que no pasen coches.
   - Grafica la función de masa de Poisson correspondiente a los valores en $60 \leq x \leq 150$.
   - Simula 260 resultados de esta distribución (alrededor de cinco años de sesiones semanales de supervisión de dias sábado). Grafica los resultados simulados usando `hist`.Utiliza `xlim` para establecer los límites horizontales de 60 a 150. Compara tu histograma con la forma de la función de masa de probabilidad del ítem anterior.

6. Un tutor sabe que el tiempo que se tarda en completar una cierta pregunta de estadística por parte de los estudiantes de primer año X, se distribuye normalmente con una media de 17 minutos y una desviación estándar de 4.5 minutos.

   - ¿ Cuál es la probabilidad de que un estudiante seleccionado al azar tarda más de 20 minutos en completar la pregunta?.
   - ¿ Cuál es la probabilidad de que un estudiante tome entre 5 y 10 minutos para terminar la pregunta? Encuentra el tiempo que marca el 10 por ciento más lento de los estudiantes en completar la pregunta.
   - Grafica la distribución normal de interés entre $\pm 4\sigma$ y sombrea el área de probabilidad del ítem anterior (el 10 por ciento más lento de los estudiantes en completar la pregunta).

7. Responde las siguientes preguntas:

   (a) Usa `lapply()` y una función anónima para encontrar el coeficiente de variación (desviación estándar dividida por la media), para todas las columnas en el conjunto de datos `mtcars`

   (b) ¿ Qué hace la siguiente función estadística?. ¿ Cuál sería un mejor nombre para la función?.

```
> bc <- function(lambda) {
+   if (lambda == 0) {
+     function(x) log(x)
+   } else {
+     function(x) (x ^ lambda - 1) / lambda
+   }
+ }
```

(c) ¿ Porqué son las siguientes dos invocaciones de `lapply()` equivalentes?

```
> trims <- c(0, 0.1, 0.2, 0.5)
> x <- rcauchy(100)
>
> lapply(trims, function(trim) mean(x, trim = trim))
> lapply(trims, mean, x = x)
```

(d) La función a continuación escala un vector para que esté en el rango $[0, 1]$. ¿ Cómo se aplicaría a cada columna de un data frame?. ¿ Cómo se aplicaría a cada columna numérica en un data frame?.

```
> scale01 <- function(x) {
+   rng <- range(x, na.rm = TRUE)
+   (x - rng[1]) / (rng[2] - rng[1])
+ }
```

8. Escribe una función `camino(N)`, en el cual comenzamos en 0 y generamos N pasos de longitud unitaria tomados aleatoriamente a la izquierda o a la derecha. Seleccionamos la dirección de cada paso generando un número aleatorio x uniformemente distribuido entre 0 y 1, moviéndose en dirección negativa si $x \leq 0.5$ y en sentido positivo si $x > 0.5$. La función debe devolver el desplazamiento del camino desde su punto de partida. A continuación, repetimos el camino aleatorio muchas veces (100 en la función anterior), generando un componente del desplazamiento de vectores para cada repetición que llamamos `multicamino`. Calcula la media y la desviación estándar del desplazamiento para este conjunto de caminos. aleatorios.

9. Lanzamos una moneda 20 veces y dejamos que X sea la longitud de la secuencia más larga de caras. Deseamos estimar la función de probabilidad p de X. Es decir, para $x = 1, 2, \ldots, 20$ deseamos estimar:

$$p(x) = \mathbb{P}(X = x).$$

Escribe una función `fsimul`, que simula X.

10. Explica los siguientes resultados:

(a)
```
> S = seq(from = 1, to = 6, length = 200)
> par(cex.main = 1.5, cex.axis = 1.2, cex.lab = 1.5)
> plot(1:200, S, type = "n", xlab = "$n$", ylab = "titulo....")
> for (s in 1:100) lines(1:200, cumsum(sample(1:6, 200,
+     replace = TRUE))/1:200, col = rgb(0, 0, 0, alpha = 0.3))
```

(b)
```
> iter = 2000
> avg0 <- avg1 <- avg2 <- avg3 <- rep(0, iter)
> for (i in 1:iter) {
+     S = rexp(90)
+     avg0[i] = S[1]
+     avg1[i] = mean(S[1:3])
+     avg2[i] = mean(S[1:30])
```

```
+     avg3[i] = mean(S[1:90])
+ }
> SR = stack(list(`n=1` = avg0, `n=3` = avg1, `n=30` = avg2,
+     `n=90` = avg3))
> names(SR) = c("averages", "n")
> ggplot(SR, aes(x = averages, y = ..density..)) + facet_grid(n ~
+     .) + geom_histogram() + scale_x_continuous(limits = c(0,
+     3))
```