Quiz Simulación estadística

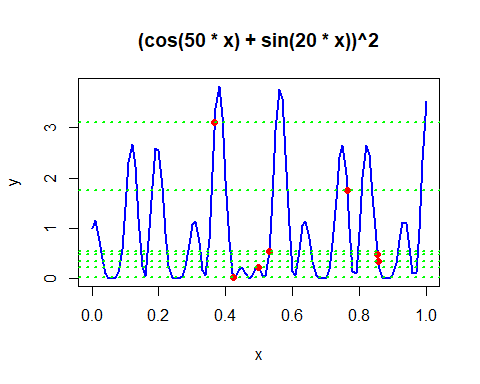
María Alejandra Vahos Blanco

## QUUIZ

1. sUtilizar el método de integración Monte Carlo para evaluar la integral

#### Método Monte-Carlo

En el ejercicio:



El método de Monte Carlo para calcular integrales consiste en obtener el promedio de áreas que se calculan multiplicando la base, es decir, , por la altura, que estará dada por un número aleatorio evaluado en la función.

# Establecer la semilla para reproducibilidad  
set.seed(1317)  
m <- 5000 # Cantidad de números aleatorios que generaremos  
aleatorios <- runif(m) # Generamos los números aleatorios  
  
# escribimos la función que deseamos integrar  
funcion <- (cos(50 \* aleatorios) + sin(20 \* aleatorios))^2  
#aplicamos el método montecarlo   
theta <- (1-0)\*funcion/m  
media <- sum(theta)  
  
#calculamos el valor real de la integral para compararlo  
integrand <- function(aleatorios) (cos(50 \*aleatorios) + sin(20 \* aleatorios))^2  
real <- integrate(integrand, lower = 0, upper = 1)$value;real

[1] 0.9652009

media

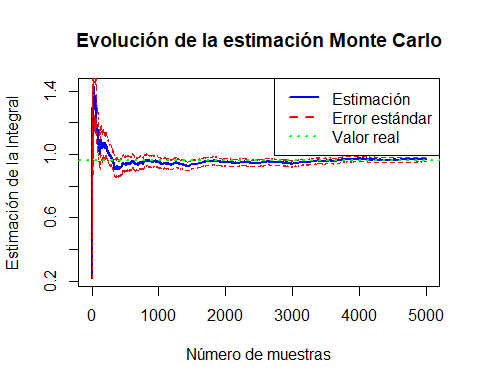
[1] 0.9699554

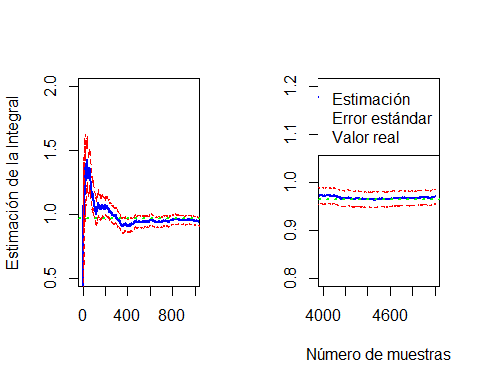
El valor de la integral que nos dio mediante las 5000 simulaciones es muy cercano al valor al realizar la integral, por lo que se puede decir que la estimacion por el método montecarlo es muy precisa.

Realizar una grafica que muestre la evolución de las estimaciones, junto con el error estándar de cada una de ellas.

= Desviación estándar de las simulaciones

= Error estándar





La evolucion de las estimaciones es muy rapida, relativamente converge rápido a nuestro valor real, con 200 simulaciones ya se observa un valor cercano al de nuestra integral.

2. Para el desarrollo de este ejercicio tenga en cuenta lo siguiente: Si es una muestra aleatoria de una distribución normal, y es la varianza muestral, entonces

Un intervalo de confianza de una sola vía está dado por donde es el -cuantil de la distribución .

Utilizar el método Monte Carlo para estimar el nivel de confianza en la construcción de un intervalo de confianza de una sola vía para la varianza, tomando como referencia una muestra aleatoria de tamaño de una distribución normal . Para el procedimiento realice replicas y tome .

Nota: Recuerde que el objetivo final es establecer la proporción de veces que la varianza queda contenida en los intervalos construidos.

set.seed(1713)  
contador <- 0  
for (i in 1:1000) {  
 numeros <- rnorm(n = 20, mean = 0, sd = 2) #creamos los datos que se distribuyen normal   
 s2 <- var(numeros)# sacamos la varianza d elos datos  
 limite <- (20-1) \* s2 / qchisq(0.05, df = 20-1) #escribimos el limite del intervalo  
 if (4 < limite) {  
 contador <- contador + 1 #guardamos las veces que 4 se encuentra en el intervalo  
 }  
}  
  
# Calcular y mostrar la proporción  
proporcion <- contador / 1000  
cat("La proporción de veces que la varianza queda contenida es:", proporcion, "\n")

La proporción de veces que la varianza queda contenida es: 0.957

Con un , podemos decir que la proporción de varianzas simuladas que contienen el parámetro es aproximadamente 95.7%. Esto es coherente con un nivel de confianza del 95% y refleja la probabilidad de que el intervalo de confianza construido para la varianza muestral contenga el valor verdadero de la varianza poblacional en aproximadamente el 95% de las muestras.

3. Para la simulación del punto 2, se partió del supuesto que la variable se distribuye normal, ¿que sucede si la población muestreada no es normal? Por ejemplo, suponer que la población muestreada es y que tiene una varianza de 4. Repetir la simulación, reemplazando las N (0, 4) muestras con . muestras.

set.seed(1317)  
contador2 <- 0  
for (i in 1:1000) {  
 numeros <- rchisq(n = 20, df = 2) # generamos los datos chi cuadrado con 2 grados de libertad  
 s2 <- var(numeros) #generamos la varianza de los datos generados   
 limite <- (20-1)\*s2/qchisq(0.05, df = 20-1) # Calculamos el límite del intervalo   
 if (4 <= limite) {  
 contador2 <- contador2 + 1 # Guardamos el numero de veces que 4 esta contenido en el intervalo   
 }  
}  
proporcion2 <- contador2 / 1000  
cat("La proporción de veces que la varianza queda contenida es:", proporcion2, "\n")

La proporción de veces que la varianza queda contenida es: 0.775

Si la población muestreada no se comporta de manera normal y se usa el mismo intervalo de confianza que para una que sí se distribuye normalmente, la proporción de valores contenidos por el intervalo no será equivalente a nuestro nivel de confianza deseado. Aunque una distribución tiene una varianza de 4, la proporción de intervalos que contienen el valor verdadero de la varianza es solo de 77.5%, ya que el intervalo no está considerando que nuestra población proviene de una distribución , que es asimétrica, en lugar de una normal.

1. Consultar sobre el método denominado Muestreador de Gibbs (Gibbs Sampler), explicar en qué consite, en qué casos se utiliza, cu ́al es el algoritmo general que lo define, presentar un ejemplo de aplicación.

* El muestreador de Gibbs es utilizado para generar muestras de distribuciones complejas, especialmente en estadística bayeciana, donde se quiere realizar estimaciones de los parámetros y es dificil o casi imposible muestrear de dicha distribución.
* En problemas bayesiones es común que la distribución posterior sea compleja, y no tenga una forma análitica sencilla, el muestreador de gibbs se puede ver como una caminata aleatoria, donde se realizan las estimaciones condicionando a la función de los parametros,

Es un método basado en simulación que permite explorar la distribución posterior de los parámetros, incluso cuando la distribución posterior no tiene una forma cerrada conocida, por lo que no se pueden extraer muestras independientes de manera fácil para hacer los métodos de inferencia posterior. Se aplica generarmente en estádistica bayesiana, pero sirve para explorar cualquier distribución probabilistica.

Se por lo general con distribuciones previas semi conjugadas, es decir, la distribucion posterior es desconocida, pero las distribuciones condicionales completas si lo son, tambien se puede implementar con distribuciones condicionales completas desconocidas, pero seria un proceso un poco mas extenso y se puede implementar con el muestreador gibbs pero tambien se le debe agregar paso especificos con algoritmos un poco mas especializados como el metropolis.

#### Algoritmo

1.Iniciación: se selecciona un conjunto inicial de valores para las variables

1. paso iterativo:
2. Para cada variable , se fija el valor actual de todas las demás variables y se muestrea un nuevo valor para utilizando la distribucion condicional
3. Repetición: La distribución estacionaria es el punto al que llega la cadena de Markov cuando deja de depender de los valores iniciales, se suelen quemar los primeros datos, ya que estos dependen de los valores iniciales que nosotros ponemos.

## Ejemplo:

La empresa tiene un presupuesto limitado que debe asignar entre los tres departamentos clave:

1. **Marketing**: Necesita recursos para campañas publicitarias y de promoción.
2. **Desarrollo de Producto**: Requiere financiamiento para investigación, desarrollo y mejora de productos.
3. **Atención al Cliente**: Gestiona el soporte a los clientes y la satisfacción post-venta.

La asignación debe reflejar las prioridades estratégicas de la empresa, donde:

* **Marketing** es visto como un área crucial para el crecimiento, por lo que se espera que reciba una mayor proporción del presupuesto.
* **Desarrollo de Producto** también es importante, pero quizás no tanto como Marketing.
* **Atención al Cliente** es una prioridad más baja este año, por lo que su asignación es la menor.

Para capturar estas creencias iniciales, los parámetros de la **distribución de Dirichlet** se han configurado como:

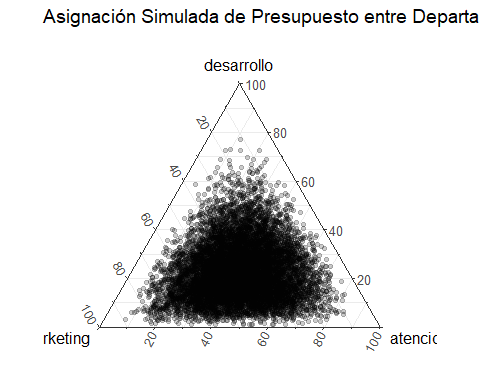
* para Marketing (creencia de que recibirá una proporción mayor del presupuesto),
* para Desarrollo de Producto (recibirá una proporción media),
* para Atención al Cliente (probabilidad de recibir una menor proporción del presupuesto).

### Objetivo de la simulación:

El objetivo de la simulación es generar múltiples escenarios posibles de cómo podrían distribuirse los recursos entre estos departamentos, respetando las creencias iniciales de la empresa. Estos escenarios te ayudarán a explorar la variabilidad en la asignación de presupuesto y a tomar una decisión más informada sobre cómo distribuirlo de forma efectiva.

# Parámetros de la distribución de Dirichlet  
alpha <- c(4, 3, 1) # α1 (Marketing), α2 (Desarrollo de Producto), α3 (Atención al Cliente)  
n\_samples <- 10000 # Número de muestras  
  
# Inicializamos las proporciones para cada departamento  
marketing <- numeric(n\_samples)  
desarrollo <- numeric(n\_samples)  
atencion <- numeric(n\_samples)  
  
# Condiciones iniciales (empezamos con valores arbitrarios para las proporciones)  
marketing[1] <- 0.3  
desarrollo[1] <- 0.4  
atencion[1] <- 1 - marketing[1] - desarrollo[1]  
  
# Muestrador de Gibbs  
for (i in 2:n\_samples) {  
   
 # Muestreamos la proporción para Marketing dado Desarrollo y Atención al Cliente  
 marketing[i] <- rbeta(1, alpha[1], alpha[2] + alpha[3]) \* (1 - desarrollo[i-1])  
   
 # Muestreamos la proporción para Desarrollo de Producto dado Marketing y Atención  
 desarrollo[i] <- rbeta(1, alpha[2], alpha[1] + alpha[3]) \* (1 - marketing[i])  
   
 # Calculamos la proporción de Atención al Cliente basado en la restricción total (suma 1)  
 atencion[i] <- 1 - marketing[i] - desarrollo[i]  
}  
  
# Ahora tenemos las muestras para las proporciones de presupuesto en cada departamento  
asignacion\_presupuesto <- data.frame(marketing = marketing, desarrollo = desarrollo, atencion = atencion)  
  
# Veamos un resumen de los resultados  
summary(asignacion\_presupuesto)

marketing desarrollo atencion   
 Min. :0.03889 Min. :0.01106 Min. :0.05086   
 1st Qu.:0.28068 1st Qu.:0.14380 1st Qu.:0.28645   
 Median :0.37709 Median :0.21660 Median :0.37658   
 Mean :0.38379 Mean :0.23193 Mean :0.38428   
 3rd Qu.:0.48025 3rd Qu.:0.30276 3rd Qu.:0.47537   
 Max. :0.89195 Max. :0.76910 Max. :0.84271



* La **variabilidad** en Marketing es considerable, con un rango de 0.04075 a 0.88126. Esto podría indicar que, dependiendo de las condiciones, la empresa podría estar distribuyendo el presupuesto de manera muy desigual, lo que es un aspecto a considerar para mantener una asignación equilibrada.
* En el caso de Desarrollo de Producto, el **mínimo** es notablemente bajo, lo que sugiere que podría haber años donde este departamento no reciba suficiente apoyo. Esto es preocupante si el desarrollo de productos es crucial para el crecimiento de la empresa

1. Generar una distribuci ́on normal bivariada con vector de medias (μ1, μ2), varianzas y correlación usando el muestreadsor de Gibbs.

La distribución normal bivariada denotada como , tiene una matriz de covarianza dada por:

### Distribuciones condicionales:

## Ejemplo:

Supongamos que

set.seed(1317)  
  
# Parámetros de la distribución normal bivariada  
mu1 <- 0 # Media de X  
mu2 <- 0 # Media de Y  
sigma1 <- 1 # Desviación estándar de X  
sigma2 <- 1 # Desviación estándar de Y  
rho <- 0.71 # Correlación  
  
# Inicialización de vectores de muestras  
X <- numeric(5000)  
Y <- numeric(5000)  
  
# Valores iniciales  
X[1] <- 0  
Y[1] <- 0  
  
# Iteración del Muestreador de Gibbs  
for (i in 2:5000) {  
 # Muestreo de X dado Y[i-1]  
 X[i] <- rnorm(1, mean = mu1 + rho \* (sigma1 / sigma2) \* (Y[i-1] - mu2),   
 sd = sigma1 \* sqrt(1 - rho^2))  
   
 # Muestreo de Y dado X[i]  
 Y[i] <- rnorm(1, mean = mu2 + rho \* (sigma2 / sigma1) \* (X[i] - mu1),   
 sd = sigma2 \* sqrt(1 - rho^2))  
}

Media de X: -0.002811098

Media de Y: 0.003420154

Varianza de X: 0.9866312

Varianza de Y: 0.9942671

Correlación entre X e Y: 0.7075389

