

# GUÍA DE TRABAJO ESPECIAL: IMPORTANCE SAMPLING

## Introducción

Los métodos tradicionales de integración por Monte Carlo no siempre son los más eficientes. Por ejemplo, si queremos encontrar la probabilidad de que una variable  $X$  con distribución normal sea mayor que 3. Una forma de encontrar este valor es resolviendo la integral:

$$P(X > 3) = \int_3^{\infty} f_X(t) dt = \frac{1}{\sqrt{2\pi}} \int_3^{\infty} e^{-t^2/2} dt$$

Si definimos la función  $h(t) = I_{[3, \infty)}(t)$  como la función característica del intervalo  $[3, \infty)$ , podemos reescribir la integral original como:

$$\int_3^{\infty} f_X(t) dt = \int_{-\infty}^{\infty} f_X(t) h(t) dt = E[h(X)]. \quad (1)$$

El muestreo de importancia o *Importance Sampling* es una forma de hacer que las simulaciones de Monte Carlo converjan más rápido. Se elige una distribución diferente para muestrear puntos que pueda generar más puntos *importantes*. Con el ejercicio anterior, se quiere generar más números alrededor de 3 para tener una mejor estimación. La teoría se puede resumir en este resultado, teniendo en cuenta que  $f_X$  representa a una variable  $X$  y  $g_Y$  a una variable  $Y$ .

$$E[h(X)] = \int_{-\infty}^{\infty} h(t) f_X(t) dt = \int_{-\infty}^{\infty} h(t) f_X(t) \left( \frac{g_Y(t)}{g_Y(t)} \right) dt \quad (2)$$

$$= \int_{-\infty}^{\infty} \frac{h(t) f_X(t)}{g_Y(t)} g_Y(t) dt \quad (3)$$

$$= E \left[ \frac{h(Y) f_X(Y)}{g_Y(Y)} \right] \quad (4)$$

La función  $f_X$  es denominada la *distribución objetivo*,  $g_X$  es la *distribución de importancia* y el cociente de ambas es la *ponderación de importancia*. Esto nos permite realizar un muestreo con cualquier distribución con función de densidad  $g_Y$ , siempre que multipliquemos  $h(X)$  por la ponderación de importancia.

## Ejercicio 1.

Dada una  $X$  normal estándar, estimar  $P(X > 3)$  de dos formas distintas.

- Escriba una función en `python` para estimar  $P(X > 3)$  usando la forma dada en 1, que tome un valor  $n$  como entrada. Graficar la diferencia entre el valor obtenido y el valor real (aproximado a 7 cifras decimales), para 50 valores de  $n$  equidistantes (pre-redondeo) entre 1 y 500.000.
- Escriba otra función que estime  $P(X > 3)$  como se exhibe en 2, que reciba  $n$  y una función de importancia `imp_fun`. Elija 3 funciones de importancia distintas, entre ellas  $Y \sim \mathcal{N}(4, 1)$ . Grafique todas las diferencias con el valor real al igual que en *a)* y escriba sus conclusiones.

## Ejercicio 2. Llamadas a Soporte Técnico

Una línea de llamadas para soporte técnico recibe un promedio de 2 llamadas por minuto. ¿Cuál es la probabilidad de que tengan que esperar al menos 10 minutos para recibir 9 llamadas?

- a) Implementar un estimador para resolver este problema con Monte Carlo tradicional y calcular estimaciones usando 10000, 20000, ..., 500000 puntos de muestreo.
- b) Implementar un estimador para resolver este problema usando importance sampling. Calcular estimaciones usando 10000, 20000, ..., 500000 puntos de muestreo.
- c) En un mismo gráfico, muestre los errores de aproximación comparando con la verdadera solución del problema:  $1 - \text{scipy.stats.gamma}(a=9, \text{scale}=0.5).cdf(10)$ . Escriba sus conclusiones.

## Reporte:

Deberá presentar un informe en formato pdf que conste de las siguientes secciones:

1. **Introducción:** En esta sección debe presentar el problema, y resumir el procedimiento mediante el cual lo va a solucionar.
2. **Algoritmo y descripción de las Variables:** En esta sección debe introducir las variables que utilizó en los algoritmos de los ejercicios propuestos, y explicar brevemente como funcionan dichos algoritmos. Adjuntar además el código desarrollado en `python` debidamente documentado (dentro del código).
3. **Resultados y Conclusiones:** Debe presentar los resultados obtenidos para los 2 ejercicios con sus correspondientes conclusiones.
4. **Conclusiones:** Esta sección debe contener conclusiones globales del método de muestro de importancia. Incluya un resumen con las conclusiones que obtuvo en la sección anterior. No olvide que un lector lee la Introducción y las Conclusiones, y si le interesa el trabajo, lee el contenido.

La bibliografía asociada se encuentra en el Capítulo 8 del libro *Simulación* (Segunda Edición ed.) de S. Ross (1999).