Problemas de Física 4
§ Probabilidad, Variables Aleatorias y Simulaciones

1. **Estimación de π por Monte Carlo**:

   La palabra *estocástico* viene del griego (*stokhos*) donde se usaba para designar al juego de lanzar dardos en un blanco. Si consideramos que los dardos se lanzan con una probabilidad uniforme en un cuadrado de lado 2, la probabilidad de que los mismos caigan en el círculo inscrito de radio $r = 1$ es igual a la proporción de las áreas:

   $$P = \frac{A_{circ}}{A_{cuad}} = \frac{\pi}{4}$$

   (a) Utilice una simulación computacional del proceso descrito para estimar el valor de π en base a probabilidades frecuenciales, con $N = 100, 500, 1000$.

   (b) Realice ahora la estimación con la cantidad de iteraciones necesaria para que se estabilicen 3 dígitos de la estimación.

   La utilización de un proceso estocástico para llevar a cabo una estimación numérica se conoce habitualmente como un *método de Monte Carlo*. Fueron introducidos por John von Neumman, para su utilización en problemas de difusión de neutrones.

2. **Una *paradoja* en probabilidad**:

   Supongamos que lanzamos un dado equilibrado dos veces, de manera independiente. La probabilidad de obtener el número 6 en cada tiro es, naturalmente $p = 1/6$. Como tiramos el dado 2 veces de manera independiente, la probabilidad de obtener el 6 en *al menos uno* de los tiros resulta $\frac{1}{6} + \frac{1}{6} = \frac{1}{3}$.

   Por otro lado, la probabilidad de *no* obtener el número 6 en un tiro es $1-p = 5/6$ y por lo tanto la probabilidad de *no obtener ningún* 6 al tirar 2 veces será $(1-p)^2 = (5/6)^2 = 25/36$.

   Ahora bien, si lanzamos un dado dos veces, tiene que pasar **una sola** de las dos cosas: o bien obtenemos *algún* 6 o bien no obtenemos *ningún* 6 (eventos excluyentes). Y sin embargo tenemos que $1/3 + 25/36 = 12/36 + 25/36 = 37/36 > 1$.

   (a) Realice experimentos computacionales para ver cuál de los supuestos anteriores es falso.

   (b) Explíquelo en base a la teoría de probabilidades.

3. **La Aguja de Buffon**:

   Una fundamentación axiomática rigurosa de la teoría de probabilidades tuvo que esperar hasta la década de 1930 con A. N. Kolmogorov para ver la luz. Los probabilistas clásicos sin embargo, desde Pascal, Newton y los Bernoulli en el Siglo XVII, confiaron siempre en los *experimentos* para corroborar la validez de sus afirmaciones.

   En el siglo XVII, el mayor cultor de este enfoque (un pionero de los métodos de Monte Carlo!) fue probablemente Georges LeClerc, Comte de Buffon, quien llevaba siempre consigo su generador de números aleatorios favorito (una moneda) y llegó a utilizarlo 2084 veces para estudiar la paradoja de San Petesburgo (cfr. infra). Entre los problemas interesantes que nos dejó

§Preparado por Luciano Perez

el Conde de Buffon, se encuentra el siguiente: *dibujamos una par de líneas paralelas a distancia 1; ahora lanzamos al azar una aguja de longitud 1 sobre el dibujo. Cuál es la probabilidad de que la aguja intereseque el menos una de las líneas?*

(a) (Opcional) Demuestre que la probabilidad de que la aguja toque alguno de los extremos es $p = 2/\pi$.

(b) Utilice el resultado del item anterior para diseñar un experimento que estime π por un método de Monte Carlo. En 1850, R. Wolf de Zurich lanzó la aguja 5000 veces y obtuvo $\pi \approx 3.1596$, así que probablemente necesite más de esos intentos para obtener una buena aproximación. Aproveche que dispone de computadora para no tener que lanzar una aguja física!

Si le gustan los problemas geométricos de probabilidad, probablemente disfrute el libro de M. G. Kendall y A. P. Moran, *Geometric Probability* (Hafner Publishing Co. 1963).

4. **Hombre Precavido**: Un estudiante de medicina leyó en un informe que el 70% de la gente muere mientras duerme en su cama. A partir de entonces comenzó a dormir en el sillón del living de su casa.

   (a) Explique si tiene sentido o no lo que hace en base a la información que leyó.

   (b) El 63% de las multas por infracciones de tránsito son levantadas a conductores de género masculino. Manejan mejor las mujeres que los hombres?

   Sobre falacias en probabilidad condicional, y de todo tipo, se puede ver el libro de Darrell Huff, *How to Lie with Statistics* (Norton & Co. 1954). El libro (completo) es una lectura de un par de horas, y se recomienda fuertemente para todo aquel que trabaja en o con estadística.

5. **Paradoja de San Petesburgo**:

   La familia Bernoulli incluyó a varios de los principales matemáticos de los siglos XVII y XVIII. Entre ellos, Nicolás Bernoulli fue el que dedicó mayores esfuerzos a la teoría de probabilidades, escribiendo su tesis doctoral sobre la aplicación de las probabilidades ... en problemas jurídicos!

   Entre los problemas interesantes que nos legó N. Bernoulli está el siguiente, que fue presentado por su primo Daniel Bernoulli ante la Academia de Ciencias de San Petesburgo, de donde deriva su nombre: *una moneda equilibrada (o sea con la misma probabilidad para cara y seca*[†]*) se arroja sucesivamente; llamemos T al número de veces que se arrojó la moneda hasta obtener la primer cara. El experimento permite diseñar una lotería, cuyo boleto costará \$$p$, y en la cual si $T = k$ nos pagan \$$2^k$. El "precio justo" del boleto se define como aquel que nos deja indiferentes entre comprar el boleto o no, es decir el que iguala la ganancia esperada con el costo del boleto, o simbólicamente $p = E(X)$ donde X es la ganancia obtenida en el juego.*

   (a) Realice experimentos computacionales para encontrar el precio justo del boleto para la lotería de San Petesburgo. Explique luego sus resultados con el cálculo teórico de $E(X)$.

   (b) Si el premio fuera de \$$k$ cuando $T = k$, cambia la situación?

[†]John von Neumman diseñó una forma de transformar cualquier moneda con cara y ceca en una moneda equilibrada, googléelo.

6. **Simulación de Variables no Uniformes**:

(a) *Binomial y Aproximación de la Normal:* Una variable aleatoria binaria X se dice *Bernoulli de parámetro* p si $P(X = 1) = p$ y $P(X = 0) = q = 1 - p$. Si se tiene una sucesión de experimentos de Bernoulli, a cada uno asociada una v.a. X_n, se puede definir una nueva variable aleatoria $S_n = X_1 + X_2 + ... + X_n$.
La variable S_n sigue entonces la *distribución binomial* y en consecuencia $P(S_n = k) = C_k^n p^k q^{n-k}$, donde hemos llamado C_k^n al número combinatorio de n elementos tomados de a k. Estas variables tienen esperanza $E(S_n) = np$ y varianza $V(S_n) = npq$ (demuéstrelo!).
La variable *normalizada* o *estandarizada*:

$$Z_n = \frac{S_n - np}{\sqrt{npq}}$$

nos permite introducir la versión más elemental del *Teorema del Límite Central* (llamado *teorema local*) ya que cuando n es grande:

$$P\{a < Z_n \leq b\} \sim \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2} dx$$

La altura de la población masculina de 20 a 65 años se distribuye $\sim N(1.70, 0.1)$. Utilice el resultado anterior para estimar:

i. La probabilidad de encontrar una persona entre $1.83mts$ y $1.88mts$.
ii. La altura cuadrática media.

(b) *Método del Acoplamiento Cuantil:* Nuestro objetivo es *simular* una variable aleatoria X con una función de distribución $F(t) = P(X \leq T)$ dada.

i. Supongamos que F es inversible y que U es una variable aleatoria unifome en el [0,1]; definimos la variable $Z = F^{-1}(U)$. Demuestre que Z tiene la distribución deseada, o sea $Z \sim F$.¶
ii. *Exponencial:* Utilice el resultado anterior para experimentar computacionalmente con la variable *exponencial de parámetro* λ, es decir aquella con densidad:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

iii. *Cauchy:* El mismo método anterior pero aplicado a la variable que tiene densidad:

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

Qué pasa cuando intenta encontrar $E(X)$ o $E(X^2)$? Demuéstrelo analíticamente.

(c) *Método de Rechazo:* El método del acoplamiento cuantil requiere encontrar la función $F^{-1}(x)$, algo que no siempre es trivial (cómo lo haría para la normal?). El *método de rechazo* permite simular una variable

¶El método funciona igual cuando F no es inversible, definiendo la función o *acoplamiento cuantil*: $F^{-1}(x) = inf\{t : F(t) \geq x\}$. Ver las *Notas de Probabilidades y Estadística* de Victor Yohai.

X con densidad $f_X(x)$ a partir de una variable uniforme U y una variable aleatoria Y con densidad $f_Y(x)$ de la que podamos disponer muestras más fácilmente (e.g. porque su cuantil sea más fácil de obtener analíticamente).

La idea es utilizar un algoritmo como el que sigue:

i. Calcular $M = \max_x \left(\frac{f_X(x)}{g_Y(x)}\right)$
ii. Generar Y de g_Y y U uniforme [0,1].
iii. Si $R > \frac{1}{M}\frac{f_X(Y)}{g_Y(Y)}$ se rechaza Y y se vuelve al paso anterior.
iv. Si no, $X = Y$

Utilice este método para experimentar computacionalmente con la distribución de Maxwell–Boltzmann.

Sobre simulación de variables aleatorias, es muy práctico el libro de J. Dagpunar, *Principles of Random Variate Generation* (Claredon Press 1988).

(d) *Distribución Empírica y Bootstrap:* En *estadística* uno intenta reconstruir la distribución de ciertas variables aleatorias a partir de *muestras* dadas y supuestos ad hoc. El herramental típico de esta ciencia es la *aproximación asintótica*, que permite obtener, bajo las hipótesis de trabajo (test, estimación, etc) la distribución que tienen las funciones de interés (e.g. los momentos de X, estimados a partir de los momentos muestrales) a medida que el tamaño de la muestra se hace más y más grande (límite asintótico).

Este método tiene ciertas deseventajas, como ser la dificultad intrínseca de calcular las distribuciones límite analíticamente, aun para funciones sencillas como la mediana muestral. Una alternativa de creciente interés es la metodología de *bootstrap*, que hace uso intensivo de la computadora para simular la distribución muestral subyacente a partir de la *distribución empírica*, es decir la distribución de frecuencias muestral.

La idea del algoritmo básico de bootstrap es la siguiente:

i. Se parte de una muestra $S = X_1, X_2, ..., X_N$.
ii. Se *re–muestrea* aleatoriamente desde S con reposición, N valores (*muestra de bootstrap*, $X_1^*, .., X_N^*$).
iii. Se computa la función de interés sobre la muestra de bootstrap, e.g. el promedio, la media, etc f_i^*.
iv. Se repite un número B de veces y se construye la distribución frecuencial de las f_i^*.
v. Con dicha distribución frecuencial se estiman las propiedades de la distribución asintótica de la función (esperanza, varianza, etc).

Use este método para estimar funciones de variables con una distribución conocida, a partir de remuestrear de una muestra fija de tamaño $N = 100$. Compare con la estimación puntual obtenida para muestras de tamaño $N = 100, 200, 500$ y con el valor obtenido analíticamente.