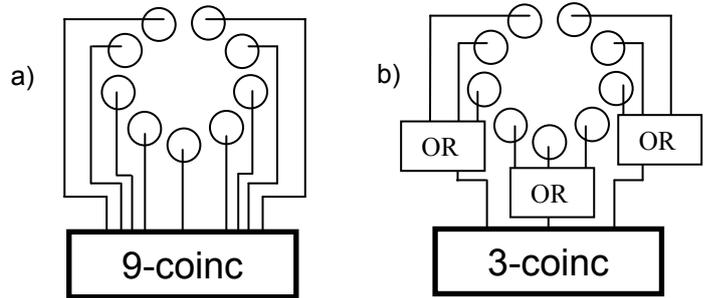


# Problemas de Métodos Estadísticos

## 1.- Conceptos generales.

1.- Calcular la eficiencia del sistema de nueve fotomultiplicadores distribuidos según las figuras a) y b) teniendo en cuenta que la eficiencia de cada uno de ellos por separado es  $\xi = 0.93$ . La detección en a) exige la coincidencia, AND, de los 9 fotomultiplicadores, mientras que en b) solo la coincidencia, AND, de los 3 grupos OR.



2.- En los haces de antiprotones de las antiguas cámaras de burbujas era frecuente encontrar contaminación de partículas más ligeras como piones y muones. Se puede estimar la contaminación pues las partículas más ligeras producen rayos delta de mayor energía. Contando los rayos delta con una energía mayor que  $E_{max}$ , la máxima energía posible de los rayos delta producidos por partículas pesadas, estimar dicha contaminación a partir del número de trazas con un rayo delta,  $N_1$ , y con dos rayos delta,  $N_2$ .

3.- Un haz de mesones, compuesto en un 90% de piones y en un 10% de kaones atraviesa un contador Cherenkov. En principio el contador proporciona señal para los piones y no para los kaones. En la práctica, la eficiencia para piones es del 95%, mientras que hay un 6% de probabilidad de que el contador proporcione una señal accidental para los kaones. Si un mesón da una señal, ¿qué probabilidad hay de que sea un pión? Y si no la da, ¿qué probabilidad hay de que sea un kaón?

4.- Los piones neutros  $\pi^0$  decaen electromagnéticamente a dos  $\gamma$ . Supongamos que estudiamos la detección de  $\pi^0$  en un contador y que la probabilidad promedio de observar una conversión de un  $\gamma$  en un par  $e^+e^-$  es  $\alpha$ . ¿Cuáles son las probabilidades de detectar o ver dos, uno o ninguno de los productos de un  $\pi^0$ ?

5.- En el concurso televisivo *Let's Make a Deal*, el concursante escoge una puerta entre tres, y su premio consiste en lo que se encuentra detrás. Una de ellas oculta un coche, y tras las otras dos hay una cabra. Sin embargo, antes de abrirla, el presentador abre una de las otras puertas y muestra que detrás de ella hay una cabra. ¿Debe el concursante mantener su elección original o escoger la otra puerta? ¿Hay alguna diferencia?

6.- El paludismo mongólico (PM) es una rara enfermedad que los médicos solo esperan encontrar en uno de cada 10000 pacientes. Siempre produce granos y apatía aguda, normalmente (60% de los casos) viene acompañada de una sed furiosa, y ocasionalmente (20% de los casos) de violentos estornudos. Dichos síntomas pueden ser debidos a otras causas: en particular, de los pacientes que no sufren PM, el 3% tiene granos, el 10% sufre apatía, el 2% sed y el 5% se quejan de violentos estornudos. Estas cuatro probabilidades son independientes. Demostrar que si vas al médico con los cuatro síntomas, la probabilidad de sufrir de PM es del 80%. ¿Cuál es la probabilidad si tienes todos los síntomas menos estornudos violentos?

PROBLEMAS TEMA 1

① Los sucesos (detecciones) en cada fotomultiplicador son independientes:  
 $P(A \text{ y } B) = P(A)P(B)$

a) La probabilidad de detectar luz en el fotomultiplicador  $i$  es:

$$P(D_i) = \xi = 0,93, \quad i = 1, \dots, 3$$

La probabilidad de detectar en 3 en coincidencia es:

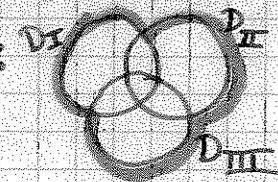
$$P(D_{1+2+3}) = \prod_{i=1}^3 P(D_i) = (0,93)^3 \approx 0,81 = 81\% \quad \checkmark$$

b) En este caso, es suficiente <sup>agrupados</sup> con que uno de los tres fotomultiplicadores haga una detección en coincidencia con los otros dos grupos.

El hecho de tener 3 en un grupo afecta <sup>la eficiencia global</sup> porque la puerta es de tipo OR y no AND. La meta es cubrir la misma sección que en el caso A con un sistema más eficiente.

La podemos interpretar el sistema como 3 <sup>detectores grandes con 3</sup> fotomultiplicadores <sup>entre grupo</sup>  $X$ , con puertas de tipo AND <sup>del grupo (I-III)</sup>

a un grupo de 3



$$P(D_I \text{ ó } D_{II} \text{ ó } D_{III}) = \text{ley que cubre las regiones que cubren este = } P(\text{triple detección})$$

$$= P(D_I) + P(D_{II}) + P(D_{III}) - P(D_I \text{ y } D_{II}) - P(D_I \text{ y } D_{III}) - P(D_{II} \text{ y } D_{III}) + P(D_I \text{ y } D_{II} \text{ y } D_{III})$$

$$= 3 \cdot \xi - 3 \cdot \xi^2 + \xi^3 \approx 0,9997 \approx 99,97\% = X \quad \checkmark$$

Alternativamente,

$$P(D_I \text{ ó } D_{II} \text{ ó } D_{III}) = \overset{1-}{1-P(D_I)} \cdot \overset{1-}{1-P(D_2)} \cdot \overset{1-}{1-P(D_3)} = (1-\xi)^3$$

$$= 1-1 + 3\xi - 3\xi^2 + \xi^3 = 3\xi - 3\xi^2 + \xi^3$$

La eficiencia efectiva del grupo es  $X$

Para que haya señal, los 3 grupos deben detectar:

$$P(\text{detecc}) = \xi^3 \approx 99,9\% \quad \checkmark$$

②  $E > E_{max} \rightarrow$  rayos  $\delta$  de partículas ligeras pesadas

$N_{\mu}$ : no total de muones;  $d$ : probabilidad de producir un rayo delta

$N_1$ : trazas con 1 rayo  $\delta$  con  $E > E_{max}$

$N_2$ : " con 2 rayos  $\delta$

$$N_1 = N_{\mu} \cdot d$$

$$\rightarrow N_2 = N_{\mu}^2 \cdot d^2$$

$$N_2 = N_{\mu} \cdot d \cdot d = N_{\mu} d^2$$

$$\frac{N_1^2}{N_2} = \frac{N_{\mu}^2 d^2}{N_{\mu} d^2} = N_{\mu} \Rightarrow N_{\mu} = \frac{N_1^2}{N_2}$$

$$N_0 = N_{\mu} \sum_{i=1}^{\infty} N_i = N_{\mu} (1 - (\frac{1}{10} - 1)) = N_{\mu} (2 - 2d) (1-d)$$

③ haz de muones  $\begin{cases} 90\% \text{ piones} = P(\pi) \\ 10\% \text{ kaones} = P(K) \end{cases} \rightarrow$  Cherenkov

$$\xi_{\pi} = 55\% = P(ds|\pi) \quad \text{detectar señal}$$

$$\xi_K = 6\% = P(ds|K)$$

$$P(\text{detectar señal}) = 90\% \cdot \xi_{\pi} + 10\% \cdot \xi_K$$

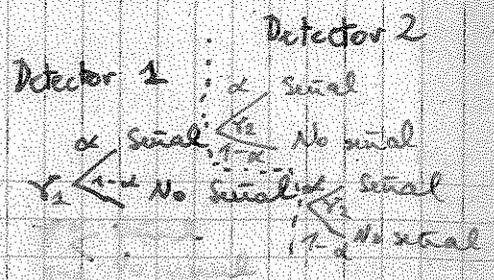
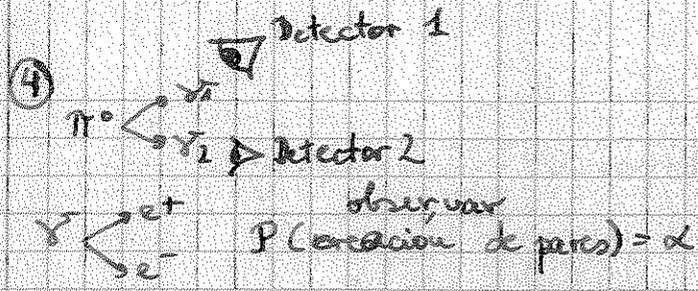
$\xrightarrow{\text{Teorema de Bayes}}$

$$P(\text{mesón detectado sea } \pi) = \frac{P(\pi) P(ds|\pi)}{P(ds)} \approx 99,3\%$$

$$P(\text{no detectar señal}) = 1 - P(\text{detect. señal}) = 90\% (1 - \xi_{\pi}) + 10\% (1 - \xi_K)$$

$$P(\text{mesón no detectado sea } K) = \frac{10\% (1 - \xi_K)}{90\% (1 - \xi_{\pi}) + 10\% (1 - \xi_K)} = 67,6\%$$

Si lo detectamos, estamos casi seguros que es un pión; si no lo detectamos, es más probable que sea un kaón.



$P(1) = \alpha \cdot (1-\alpha) \cdot 2$  (Señal 1 + No señal 2) + viceversa

$P(2) = \alpha^2 = \alpha \cdot (1-\alpha)$

$P(0) = (1-\alpha)^2 = 1 - P(1) - P(2) = 1 - 2\alpha + 2\alpha^2 - \alpha^2 = 1 - 2\alpha + \alpha^2$

∴ distribución binomial

5) El concursante debe escoger la otra puerta, pues al escoger por primera vez entre las tres, era más probable que le haya tocado cabra, y por tanto al cambiar obtendrá coche.

En la primera elección, es  $\frac{2}{3}$  probable que le haya tocado cabra. Por tanto, si cambia será  $\frac{2}{3}$  probable de obtener coche, mientras que si mantiene la elección solo es  $\frac{1}{3}$  probable que sea coche.

Se puede utilizar el tma. de Bayes para demostrarlo:

Supongamos que hay 3 puertas A, B, C. Yo selecciono la puerta A y el presentador abre la C, lo que me da una información D)

$P(A = \text{coche} | D) = \frac{P(D | A = \text{coche}) \cdot P(A = \text{coche})}{P(D)}$

$P(B = \text{coche} | D) = \frac{P(D | B = \text{coche}) \cdot P(B = \text{coche})}{P(D)}$   
idem con C

A priori,  $P(A = \text{coche}) = P(B = \text{coche}) = P(C = \text{coche}) = \frac{1}{3}$

$P(D | C = \text{coche}) = 0$  ya que D revela que en C hay cabra.

$P(D | B = \text{coche}) = 1$  ya que A está seleccionada, y el presentador no tiene opción.

$P(D | A = \text{coche}) = \frac{1}{2}$  ya que el presentador podría abrir B ó C

$P(D) = \sum_i P(i) \cdot P(D | i = \text{coche}) = \frac{1}{3} \cdot \frac{1}{2} + \frac{1}{3} \cdot 1 + \frac{1}{3} \cdot 0 = \frac{1}{2}$

$P(A = \text{coche} | D) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$ ,  $P(B = \text{coche} | D) = 1 \cdot \frac{1}{3} = \frac{1}{3}$  ∴ Por tanto

⑥ (PM) = Paludismo  
molególico

$$P(\text{PM}) = \frac{1}{10000}$$

$$P(\text{graves}) = 100\%$$

$$P(\text{apatía}) = 100\%$$

$$P(\text{sed}) = 60\%$$

$$P(\text{est.}) = 20\%$$

Normal

$$P(\text{Normal}) = 0,9999 \approx 1$$

$$P(\text{graves}) = 3\%$$

$$P(\text{apatía}) = 10\%$$

$$P(\text{sed}) = 2\%$$

$$P(\text{est.}) = 5\%$$

Utilizamos la estadística bayesiana (teorema de Bayes).  
 $\Omega = \{ \text{PM}; \text{Normal} \}$  ↗ independientes

$$P(4 \text{ síntomas} | \text{Normal}) = 3\% \cdot 10\% \cdot 2\% \cdot 5\% = 3 \cdot 10^{-6}$$

$$P(4 \text{ síntomas} | \text{con PM}) = 12\% = 100\% \cdot 100\% \cdot 60\% \cdot 20\%$$

$$P(\text{PM} | 4 \text{ síntomas}) = \frac{P(4 \text{ síntomas} | \text{con PM}) \cdot P(\text{con PM})}{P(4 \text{ síntomas} | \text{PM}) \cdot P(\text{PM}) + P(4 \text{ síntomas} | \text{Normal}) \cdot P(\text{Normal})}$$

$$= \frac{12 \cdot 10^{-6}}{12 \cdot 10^{-6} + 3 \cdot 10^{-6} \cdot 0,9999} \approx 80\%$$

$$P(3 \text{ sínt./est.} | \text{PM}) = 100\% \cdot 100\% \cdot 60\% \cdot (100\% - 20\%) = 48\%$$

$$P(3 \text{ sínt./est.} | \text{Normal}) = 3\% \cdot 10\% \cdot 2\% \cdot (100\% - 5\%) = 5,7 \cdot 10^{-5}$$

$$P(\text{PM} | 3 \text{ sínt./est.}) = \frac{P(3 \text{ sínt./est.} | \text{PM}) \cdot P(\text{PM})}{P(3 \text{ sínt./est.} | \text{PM}) \cdot P(\text{PM}) + P(3 \text{ sínt./est.} | \text{Normal}) \cdot P(\text{Normal})}$$

$$= 46\%$$

# Problemas de Métodos Estadísticos

## Curso 2011-2012

### 2.- Propiedades generales de las distribuciones de probabilidad

1.- Dada una variable aleatoria  $x$  con valor esperado  $\mu$  y varianza  $\sigma^2$ , calcular el valor esperado y la varianza de las siguientes funciones:

$$f(x) = cx, \quad c = \text{const.}$$

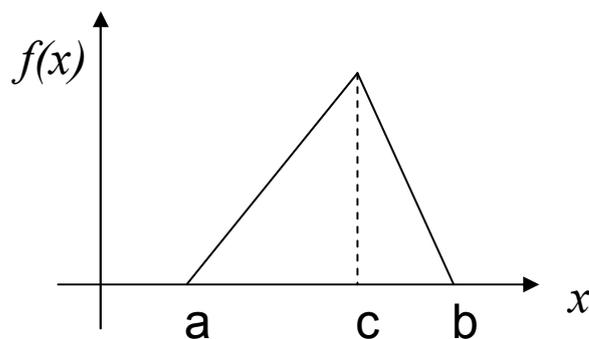
$$f(x) = \frac{x - \mu}{\sigma^2}$$

2.- Considerar la función densidad de probabilidad de una distribución triangular como la de la figura dada por:

$$f(x) = 0, \quad x < a, \quad x \geq b$$

$$f(x) = \frac{2}{(b-a)(c-a)}(x-a), \quad a \leq x < c$$

$$f(x) = \frac{2}{(b-a)(b-c)}(b-x), \quad c \leq x \leq b$$



Determinar la media, la moda y la mediana de la distribución para los casos:

- a)  $c = 0$  y  $a = -b$
- b)  $c = 0$  y  $a = -2b$ .

3.- Demostrar que el coeficiente de asimetría (*skewness*) puede expresarse como:

$$\gamma_1 = \frac{1}{\sigma^3} \left( E[x^3] - 3E[x]E[x^2] + 2E[x]^3 \right).$$

4.- Demostrar que para dos variables aleatorias  $x_1$  y  $x_2$  se cumple que:

$$V(x_1 + ax_2) = V(x_1) + a^2V(x_2) + 2a \text{cov}(x_1, x_2)$$

5.- Demostrar que el coeficiente de correlación es un número entre -1 y +1. Ayuda  $V(x_1 + ax_2) \geq 0$

6.- Mediante un programa de ordenador generar de números aleatorios es una distribución uniforme creando un histograma con 10000 números aleatorios. A continuación hacer un histograma bidimensional usando pares consecutivos de números aleatorios para las coordenadas  $x_1$  y  $x_2$ . ¿Se trata también de una distribución uniforme? Calcular el coeficiente de correlación entre coordenadas  $x_1$  y  $x_2$ . Repetir el experimento numérico anterior para los siguientes casos:

- a) La variable  $x_2$  es una gaussiana centrada en  $x_1^2$  y de sigma  $\sigma = 0.1$
- b) Mismo caso que a) pero haciendo que la variable  $x_1$  se distribuya en el intervalo  $[-1,1]$  (ayuda.- hacer el cambio  $x_1 = -1 + 2x_1'$  donde  $x_1' \in [0,1]$ ).

Comentar los resultados en función de los valores del coeficiente de correlación obtenidos.

Problemas Tema 2

①

$x$ : variable aleatoria

$E[x] = \mu$  ;  $V[x] = \sigma^2$

a)  $E[f(x)] = E[cx] = \overset{c = \text{cte}}{c} \cdot E[x] = c \cdot \mu$  ✓

$V[f(x)] = V[cx] = c^2 V(x) = c^2 \sigma^2$  ✓

b)

$E[f(x)] = E\left[\frac{x - \mu}{\sigma^2}\right] = \overset{\sigma = \text{cte}}{\frac{1}{\sigma^2}} \{E[x] - E[\mu]\} = \overset{\mu = \text{cte}}{\frac{1}{\sigma^2}} \{\mu - \mu\} = 0$  ✓

$V[f(x)] = E\left[\left(\frac{x - \mu}{\sigma^2} - E\left[\frac{x - \mu}{\sigma^2}\right]\right)^2\right] = E\left[\left(\frac{x - \mu}{\sigma^2}\right)^2\right] = \frac{1}{\sigma^4} \cdot E[(x - \mu)^2]$   
 $= \frac{1}{\sigma^4} \{\sigma^2\} = \frac{1}{\sigma^2}$  ✓

②

$$f(x) = \begin{cases} 0 & \text{si } x < a, x \geq b \\ \frac{2}{(b-a)(c-a)} (x-a) & \text{si } a \leq x \leq c \\ \frac{2}{(b-a)(b-c)} (b-x) & \text{si } c \leq x \leq b \end{cases}$$

• Moda: máximo de  $f(x) = c$

• Mediana:  $F(x_{\text{mediana}}) = 1/2$

• Media:  $E(x) = \mu$

a)  $c = 0, a = -b$



Simétrica respecto a eje vertical

$x_{\text{moda}} = 0 = c$   
 $x_{\text{mediana}} = 0$   
 $\mu = 0$  ✓

ya que  $f(x) = \begin{cases} 0 & \text{si } x < -b, x \geq b \\ \frac{1+x}{b} & \text{si } -b \leq x \leq 0 \\ \frac{1}{2} - \frac{x}{b} & \text{si } 0 \leq x \leq b \end{cases}$

...suponemos normalizado

$$\int_{-\infty}^{\infty} f(x) dx = 1 = \int_{-\infty}^0 f(x) dx + \int_0^{\infty} f(x) dx = 2 \int_{-\infty}^0 f(x) dx = 2F(0)$$

$$\hookrightarrow F(0) = \frac{1}{2} \Rightarrow x_{\text{median}} = 0$$

$$E[X] = \int_{-\infty}^{+\infty} x \cdot f(x) dx = \int_{-\infty}^0 x \cdot f(x) dx + \int_0^{+\infty} x \cdot f(x) dx$$

↙ c de variable  $x \leftrightarrow -x$   
e intercambia límites

$$= \int_{-\infty}^0 x \cdot f(x) dx - \int_0^{\infty} x \cdot f(x) dx = 0$$

b) la moda = c = 0 //

$$f(x) = \begin{cases} 0 & \text{si } x < -2b, x > b \\ \frac{x+2b}{3b^2} & \text{si } -2b \leq x < 0 \\ \frac{2}{3b^2}(b-x) & \text{si } x \geq b \end{cases}$$

$$F(x) = \int_{-\infty}^x f(x) dx$$

• Si  $x < -2b$ :  $F = 0$

• Si  $x < 0$ :

Como  $|a| > |b|$ , la mediana estará para  $x < 0$ .  
 $\int_{-\infty}^0 f(x) dx > \int_0^{\infty} f(x) dx$

$$F(x) = \int_{-2b}^x \frac{x'+2b}{3b^2} dx' = \frac{1}{3b^2} \left[ \frac{x'^2}{2} + 2bx' \right]_{-2b}^x$$

$$= \frac{1}{3b^2} \left[ \frac{x^2}{2} + 2bx - \frac{4b^2}{2} + 4b^2 \right] = \frac{1}{3b^2} \left( \frac{x^2}{2} + 2bx + 2b^2 \right)$$

$$F(x_{\text{median}}) = \frac{1}{2} = \frac{1}{3b^2} \left( \frac{x_{\text{me}}^2}{2} + 2bx_{\text{me}} + 2b^2 \right) = \frac{1}{3b^2} \left( \frac{x_{\text{me}}}{\sqrt{2}} + \sqrt{2}b \right)^2$$

$$\hookrightarrow \frac{3b^2}{2} = \left( \frac{x_{\text{me}}}{\sqrt{2}} + \sqrt{2}b \right)^2$$

$$\hookrightarrow x_{\text{me}} + 2b = \sqrt{3}b \sqrt{\frac{3}{2}}$$

$$x_{\text{me}} = -2b + \sqrt{3}b \approx -0,27b < 0. \quad \checkmark$$

Media:  $\mu = \int_{-2b}^0 x' \left( \frac{x'+2b}{3b^2} \right) dx' + \int_0^b x' \frac{2}{3b^2} (b-x') dx'$

$$= \frac{x'^3}{3} + \frac{bx'^2}{b^2} \Big|_{-2b}^0 + \frac{2}{3b^2} \left( bx'^2 - \frac{x'^3}{3} \right) \Big|_0^b$$

$$= \frac{8b^3}{3} - 4b^3 + \frac{b^3}{3} - \frac{2b^3}{3} = \frac{2b^3 - 3b^3}{3b^2} = -\frac{b}{3} \checkmark$$

3

$$\gamma_1 \equiv \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{E[(x-\mu)^3]}{\sigma^3} = \frac{1}{\sigma^3} \{E[x^3] - 3\mu E[x^2] + 3\mu^2 E[x] - \mu^3\}$$

$$\overset{\mu = E[x]}{=} \frac{1}{\sigma^3} (E[x^3] - 3E[x] \cdot E[x^2] + 3E[x]^3 - E[x]^3)$$

$$= \frac{1}{\sigma^3} (E[x^3] - 3E[x] \cdot E[x^2] + 2E[x]^3) \quad \checkmark$$

f.e.d.

4

$$V(x_1 + ax_2) = E[(x_1 + ax_2 - E[x_1 + ax_2])^2]$$

$$= E[(x_1 + ax_2 - E[x_1] - aE[x_2])^2] =$$

$$= E[(x_1 - E[x_1]) + a(x_2 - E[x_2])]^2 \quad \begin{matrix} \rightarrow \text{Desarrollo el cuadrado y} \\ \rightarrow \text{aplico linealidad de E.} \end{matrix}$$

$$= E[(x_1 - E[x_1])^2] + a^2 E[(x_2 - E[x_2])^2] + 2a E[(x_1 - E[x_1]) \cdot (x_2 - E[x_2])]$$

$$= V(x_1) + a^2 V(x_2) + 2a \text{cov}(x_1, x_2)$$

$$= \sigma_1^2 + a^2 \sigma_2^2 + 2a \text{cov}(x_1, x_2) \quad \checkmark$$

5

$$f(x_i, x_j) = \frac{\text{cov}(x_i, x_j)}{\sigma_i \cdot \sigma_j} \quad (\text{coeficiente de correlación})$$

Analizemos la expresi3n del ejercicio n=4 como funci3n del par3metro a, tomando  $V(x_1)$  y  $\text{cov}(x_1, x_2)$  fijos.

$V(x_1 + ax_2) \equiv f(a) = a^2 V(x_2) + a \cdot 2\text{cov}(x_1, x_2) + V(x_1) \geq 0 \quad \forall a$

f es un polinomio de 2º orden en a, sin cortar el eje de abscisas.



El hecho de que  $f(a) \geq 0$  implica que no tiene raices reales, es decir, que el discriminante de la ecuaci3n de 2º grado es  $\leq 0$ .

$$\hookrightarrow a^2 V(x_2) + 2a \text{cov}(x_1, x_2) + V(x_1) = 0$$

$$\hookrightarrow a_{\pm} = \frac{-2\text{cov}(x_1, x_2) \pm \sqrt{D}}{2V(x_2)} \notin \mathbb{R} \Rightarrow D \leq 0 \quad \checkmark$$

$$D = 4(\text{cov}(x_1, x_2))^2 - 4 \cdot V(x_2) \cdot V(x_1) \leq 0$$

$$\hookrightarrow (\text{cov}(x_1, x_2))^2 \leq V(x_1) \cdot V(x_2) \Rightarrow \left( \frac{\text{cov}(x_1, x_2)}{\sigma_1 \sigma_2} \right)^2 \leq 1 \Rightarrow f^2 \leq 1$$

L.O.E.F. [1,1]

⑥ - Aparte por ordenador

↳ Se lo envié por correo electrónico

ok

# -Métodos estadísticos- Problema 6 del Tema 2: Propiedades generales de las distribuciones de probabilidad

Fernando Hueso González\*

*Valencia, 24 de octubre de 2011*

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

**Enunciado del problema** Mediante un programa de ordenador, generar una distribución uniforme de 10.000 números aleatorios y crear un histograma. A continuación, hacer un histograma bidimensional usando pares consecutivos de números aleatorios para las coordenadas  $x_1$  y  $x_2$ . ¿Se trata también de una distribución uniforme? Calcular el coeficiente de correlación entre las coordenadas  $x_1$  y  $x_2$ . Repetir el experimento numérico anterior para los siguientes casos:

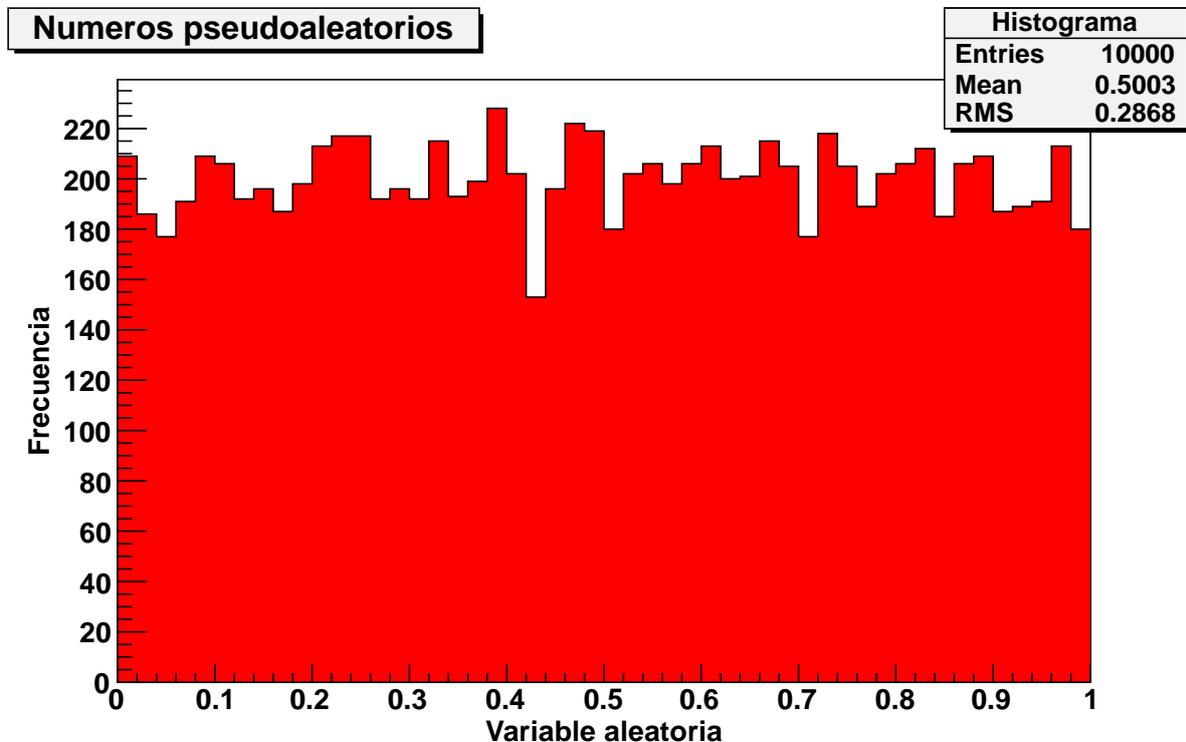
- a) La variable  $x_2$  es una gaussiana centrada en  $x_1^2$  y de sigma  $\sigma = 0.1$ .
- b) Mismo caso que a), pero haciendo que la variable  $x_1$  se distribuya en el intervalo  $[-1,1]$ . (Ayuda: realizar el cambio  $x_1 = -1 + 2 * x'_1$  donde  $x'_1 \in [0, 1]$ .)

Comentar los resultados en función de los valores del coeficiente de correlación obtenidos.

---

\*ferhue#alumni.uv.es

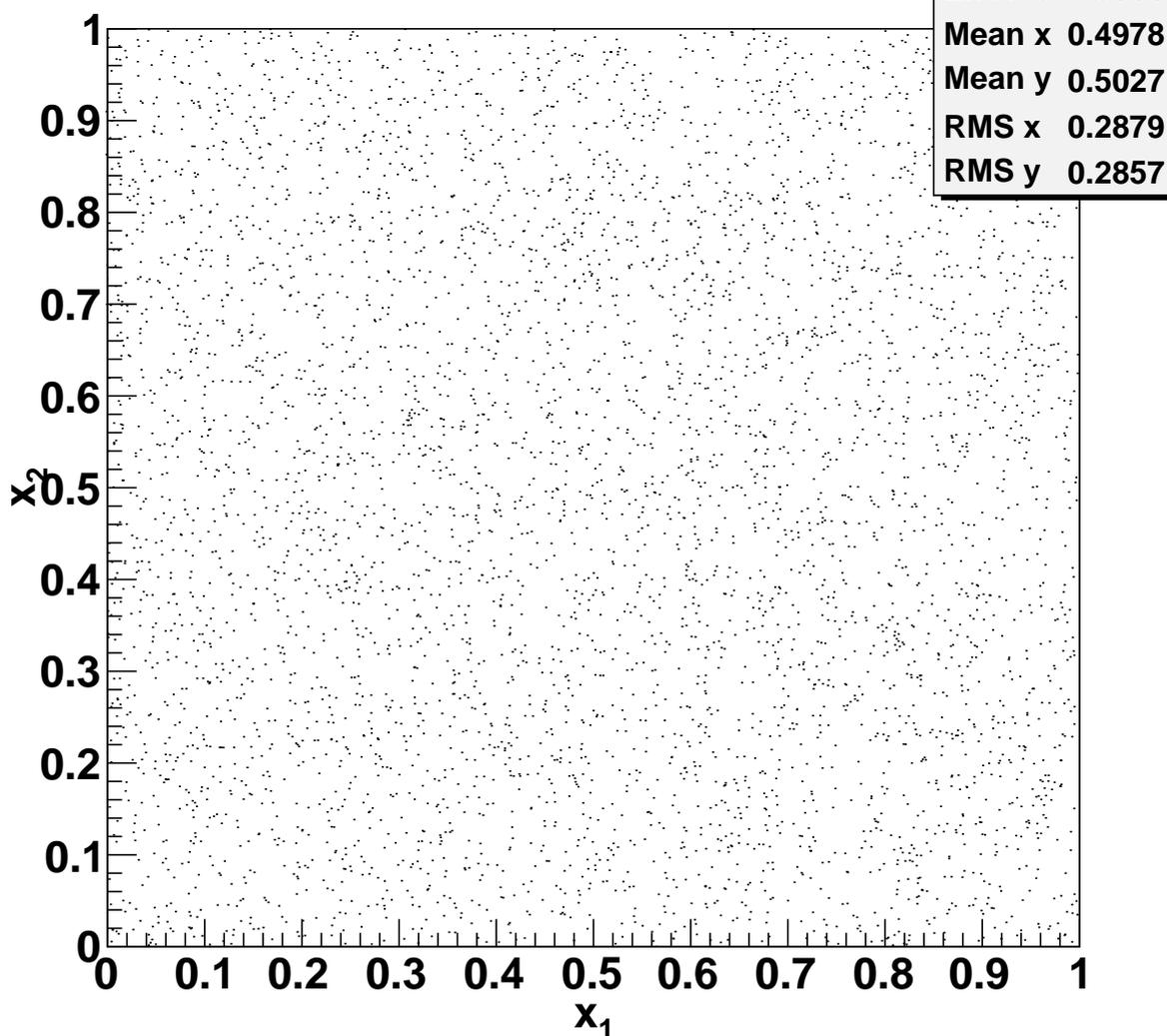
**Resolución** Mediante programación en ROOT, genero 10.000 números aleatorios usando la función preestablecida TRandom1->Rndm(), que genera un número aleatorio entre 0 y 1. Se obtiene el siguiente histograma con 50 bines:



Se observan fluctuaciones estadísticas respecto a la frecuencia con que aparece cada valor de la variable (intervalo), pero la distribución es bastante uniforme. En la mayoría de bines, esta desviación es del orden lo esperado, que sería la raíz del número de cuentas en el bin (teóricamente, el 68% deberían estar en esa zona). Si usamos funciones más rápidas pero menos fiables de generadores de números aleatorios, obtendremos un histograma con mayores desviaciones respecto al valor medio en un mayor número de bines.

Si agrupamos los números aleatorios en pares consecutivos para las coordenadas  $x_1$  y  $x_2$  y realizamos un histograma 2D (50x50 bins), se obtiene la siguiente nube de puntos:

## Numeros pseudoaleatorios



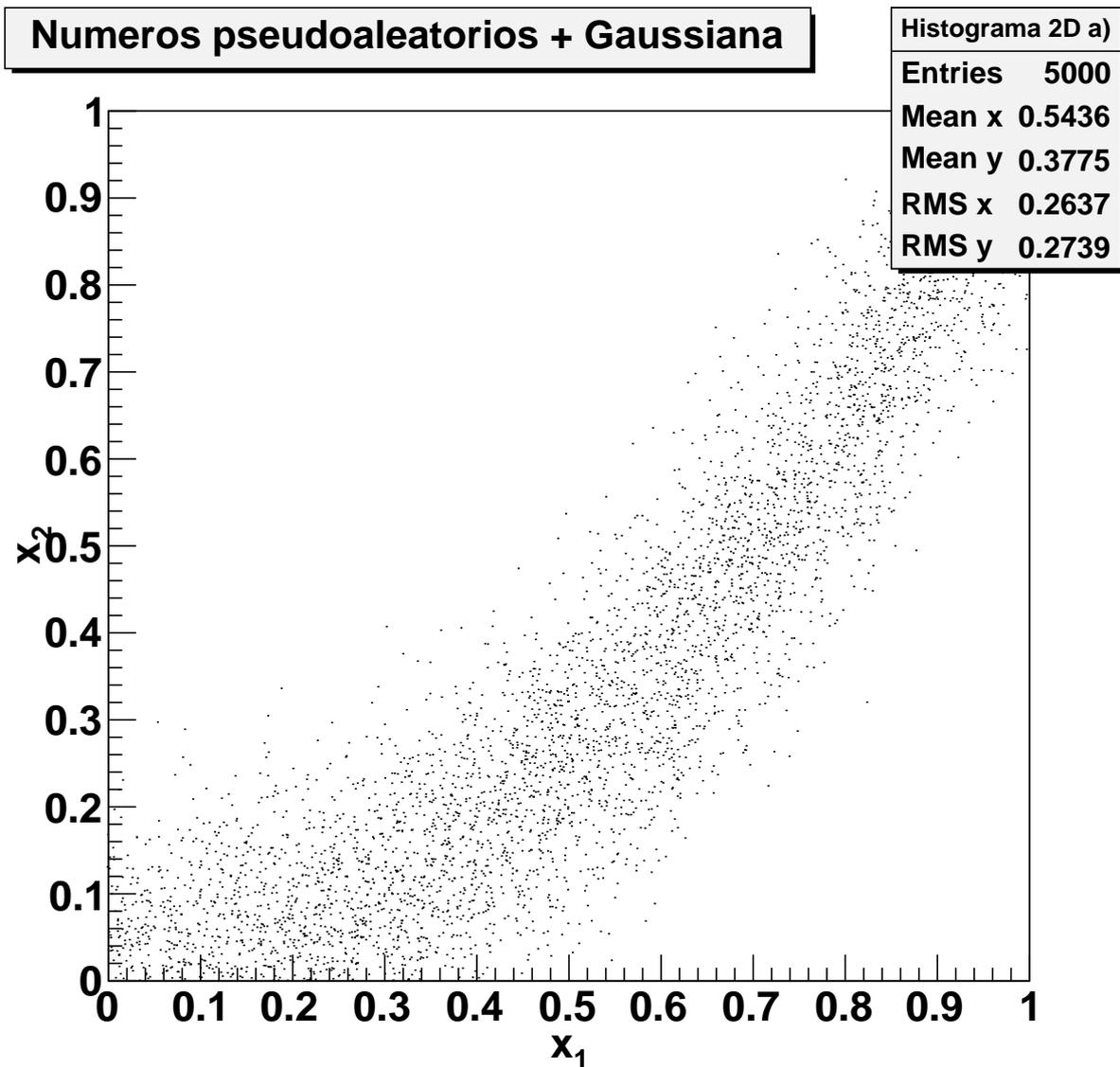
Podemos comprobar que la distribución es uniforme y que no hay ninguna tendencia o correlación clara entre las variables aleatorias, dado que la nube de puntos ocupa todo el espacio de soluciones posibles con una densidad de puntos media bastante uniforme. Se puede comprobar que si se utilizan generadores de números aleatorios menos fiables, se produciría un agrupamiento de puntos en ciertas islas.

El coeficiente de correlación lineal que se obtiene es

$$r = -0.015$$

Este valor tan cercano al cero indica la no correlación entre las variables, que en este caso, dado que sabemos que se han generado pseudoaleatoriamente, corresponde a variables independientes y por tanto descorrelacionadas.

Si repetimos el análisis para el caso a), donde sustituimos los valores de  $x_2$  por una función pseudoaleatoria de tipo gaussiana centrada en  $x_1^2$  y de sigma  $\sigma = 0.1$  (función predeterminada de ROOT: TRandom1->Gaus(mean, sigma)), se obtiene la siguiente nube de puntos:



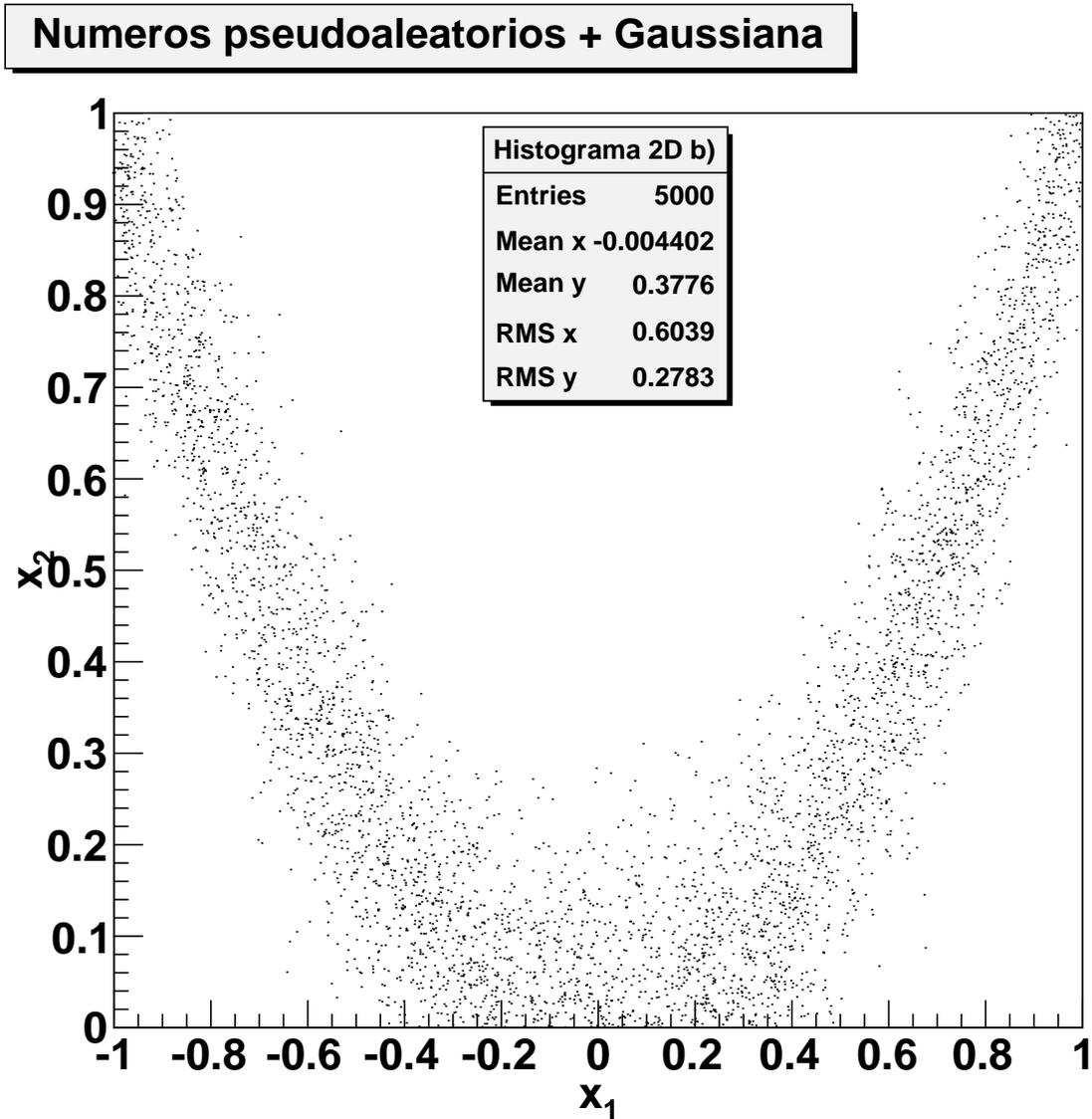
Se aprecia un claro agrupamiento de los puntos, dado que en este caso hemos forzado a que las variables estén correlacionadas (no son independientes). Se intuye la forma de parábola  $x_2 = x_1^2$  con una cierta desviación dado que la sigma de la gaussiana es de 0.1. Este hecho contrasta con el caso anterior, donde la distribución de la nube era prácticamente uniforme al no estar las variables relacionadas.

El coeficiente de correlación que se obtiene es

$$r = 0.920$$

Este valor cercano a uno indica una moderada correlación lineal entre las variables. El hecho de que no se alcance la unidad se debe a que la relación es cuadrática en lugar de lineal. Aun teniendo una correlación total entre las variables (sigma de la gaussiana nula), el coeficiente no sería la unidad, pues la relación entre las variables es no lineal.

Si repetimos el análisis de a) para el caso b), donde transformamos el intervalo  $[0,1]$  de las  $x_1$  en el  $[-1,1]$  y relacionamos  $x_2$  con el nuevo  $x_1$  de manera análoga al caso a), se obtiene la siguiente nube de puntos (100x50bines):



Se aprecia un agrupamiento de los puntos similar al caso a), dado que sólo hemos cambiado el intervalo de  $x_1$  haciéndolo simétrico y el doble de grande. Se intuye de nuevo la forma de parábola  $x_2 = x_1^2$ , esta vez simétrica, con similar dispersión de puntos debido a la sigma.

No obstante, el coeficiente de correlación que se obtiene es

$$r = -0.014$$

Este valor cercano a cero podría hacer pensar que las variables no están correlacionadas, pero no es así. La razón es que si sólo tenemos en cuenta el intervalo  $[0,1]$ , obtenderíamos un cierto valor  $R_+$  cercano a la unidad, como en el caso a), mientras que si analizamos el intervalo  $[-1,0]$ , obtendremos un valor  $R_-$  de módulo aproximadamente igual a  $R_+$  pero de signo opuesto:  $R_- = -R_+$ . Al tomar el intervalo completo  $[-1,1]$ , el cambio de pendiente en  $x=0$  hace que no exista correlación lineal entre las variables en ese intervalo (entre 0 y 1, a primer orden,

la parábola se acerca a una recta). Al tomar el intervalo completo, simétrico respecto al eje de ordenadas, ya no podemos ajustarlo a una recta y obtenemos un valor de correlación lineal cercano a cero.

Por tanto, un coeficiente de correlación nulo no siempre indica que las variables están descorrelacionadas, pues hay que tener en cuenta que para funciones que no puedan aproximarse a una recta en todo el intervalo, la correlación lineal tiende a cero.

En resumen, hemos verificado que si las variables son independientes, la correlación es nula (primer caso); si son dependientes y asimilables en primer orden a una recta, la correlación se acerca a la unidad (caso a); y que si la correlación lineal es nula, no se cumple en general que las variables sean independientes (caso b), pues incluso puede darse este efecto si las variables están totalmente correlacionadas por una expresión no lineal.

# Problemas de Métodos Estadísticos

## 3.- Propagación de errores

1.- Demostrar que si la variable  $x$  tiene como función densidad de probabilidad:

$$f(x) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}x^2\right)$$

(la pdf de la distribución normal estándar) la pdf de la variable  $y = x^2$  viene dada por:

$$g(y) = (2\pi)^{-\frac{1}{2}} y^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y\right)$$

(la distribución de  $\chi^2$  con un solo grado de libertad).

2.- Medimos los valores medios de  $x$  e  $y$  con varianzas  $\sigma_x^2$  y  $\sigma_y^2$ . La covarianza es nula. Encontrar las varianzas y covarianzas de  $r$  y  $\theta$  dadas por:

$$r^2 = x^2 + y^2 \quad \text{y} \quad \tan \theta = \frac{y}{x}$$

3.- Medimos los valores de la masa  $m$  de un objeto y su velocidad  $v$  con varianzas  $\sigma_m^2$  y  $\sigma_v^2$  y suponemos que las medidas son independientes. Considera el momento  $p = mv$  y la energía cinética  $E = \frac{1}{2}mv^2$  del objeto y calcular la matriz de covarianzas de de las variable  $(p, E)$ .

4.- Si las variables  $1/p$ ,  $\lambda$ ,  $\phi$  se han medido con errores  $\Delta\left(\frac{1}{p}\right)$ ,  $\Delta\lambda$ ,  $\Delta\phi$ , respectivamente, y sin correlación entre ellas, cuales son los errores asociados con las cantidades derivadas:

$$p_x = p \cos \lambda \cos \phi, \quad p_y = p \cos \lambda \sin \phi, \quad p_z = p \sin \lambda$$

Problemas Tema 3

①  $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$



2 ramas simétricas,  $x > 0, x < 0$

$y = x^2 \Rightarrow g(y) = \frac{f(x)}{\sum \frac{|dy|}{|dx|}} = \frac{2 \cdot f(x)}{2|x|} = \frac{1}{\sqrt{2\pi}} \frac{e^{-\frac{1}{2}x^2}}{\sqrt{y}}$

Como para un valor  $p = f(x)$  hay dos soluciones:

$x_{\pm} = \pm \sqrt{(-2) \ln(\sqrt{2\pi} p)}$ , pero  $y(x_{\pm})$  es de acuerdo (las 2 sol. coinciden)

Entonces al intervalo de  $y \in (b_1, b_2)$  hay que asociarle los intervalos en la  $x \in (\sqrt{b_1}, \sqrt{b_2}) \cup (-\sqrt{b_1}, -\sqrt{b_2})$  para conservar la probabilidad. De ahí el factor 2 que aparece en  $g(y) = \sum \frac{f(x)}{\frac{|dy|}{|dx|}}$  al ser las ramas simétricas ( $f(x) = f(-x)$ ).

②

$x \rightarrow \bar{x}, \sigma_x$   
 $y \rightarrow \bar{y}, \sigma_y$   
 $\rightarrow \vec{r} = (\bar{x}, \bar{y})$

$\cos \theta(x, y) = 0 \rightarrow V_{xy} = 0, V_{xx} = \sigma_x^2, V_{yy} = \sigma_y^2$

$r^2 = x^2 + y^2 \rightarrow r = \sqrt{x^2 + y^2}$   
 $\tan \theta = \frac{y}{x} \rightarrow \theta = \arctan\left(\frac{y}{x}\right)$   
 $\rightarrow$  Matriz de covarianza  $V_{\vec{r}}(r, \theta)$

$V[\vec{r}] = V_{rr} = V(r(x, y)) = \sum_{i,j=1}^2 \left. \frac{\partial r}{\partial x_i} \right|_{\vec{r}} \cdot \left. \left( \frac{\partial r}{\partial x_j} \right) \right|_{\vec{r}} \cdot V_{ij}(x, y) \rightarrow$  las no diagonales  $V_{xy} = 0$

$= \left( \frac{\partial r}{\partial x} \right)_{\vec{r}}^2 \cdot \sigma_x^2 + \left( \frac{\partial r}{\partial y} \right)_{\vec{r}}^2 \cdot \sigma_y^2 = \left( \frac{1 \cdot 2x}{2r} \right)_{\vec{r}}^2 \cdot \sigma_x^2 + \left( \frac{y}{r} \right)_{\vec{r}}^2 \cdot \sigma_y^2 =$

$= \frac{x^2 \sigma_x^2 + y^2 \sigma_y^2}{x^2 + y^2}$

$V[\theta] = \left( \frac{\partial \theta}{\partial x} \right)_{\vec{r}}^2 \sigma_x^2 + \left( \frac{\partial \theta}{\partial y} \right)_{\vec{r}}^2 \sigma_y^2 = \left( \frac{1 \cdot (-\frac{y}{x^2}) \cdot \sigma_x}{1 + \left(\frac{y}{x}\right)^2} \right)_{\vec{r}}^2 + \left( \frac{\frac{1}{x} \sigma_y}{1 + \left(\frac{y}{x}\right)^2} \right)_{\vec{r}}^2$

$= \left( \frac{-y \sigma_x}{x^2 + y^2} \right)_{\vec{r}}^2 + \left( \frac{x \sigma_y}{x^2 + y^2} \right)_{\vec{r}}^2 = \frac{1}{(x^2 + y^2)^2} \cdot (y^2 \sigma_x^2 + x^2 \sigma_y^2)$

$$\text{cov}(r, \theta) = V_{r\theta} = \text{Var} = \sum_{i,j=1}^2 \frac{\partial r}{\partial x_i} \bigg|_{\bar{\mu}} \frac{\partial \theta}{\partial x_j} \bigg|_{\bar{\mu}} V_{ij}(x) \quad V_{ij} = 0 \text{ si } i \neq j$$

$$= \frac{\partial r}{\partial x} \bigg|_{\bar{\mu}} \cdot \frac{\partial \theta}{\partial x} \bigg|_{\bar{\mu}} V(x) + \frac{\partial r}{\partial y} \cdot \frac{\partial \theta}{\partial y} V(y)$$

$$= \left(\frac{x}{r}\right) \cdot \frac{-y}{x^2+y^2} \cdot \sigma_x^2 \bigg|_{\bar{\mu}} + \frac{y}{r} \cdot \frac{x}{x^2+y^2} \sigma_y^2 \bigg|_{\bar{\mu}}$$

$$= \frac{\bar{x}\bar{y}}{(\bar{x}^2 + \bar{y}^2)^{3/2}} (\sigma_y^2 - \sigma_x^2)$$

③

$m \rightarrow \sigma_m, \bar{m}$  independientes:  $\text{cov}(m, v) = 0$   
 $v \rightarrow \sigma_v, \bar{v}$

$$p = mv$$

$$E = \frac{1}{2} mv^2$$

$V_{ij} = ?$  ... términos no diagonales = 0

$$V_{pp} = \sigma_p^2 = \left(\frac{\partial p}{\partial m}\right)_{\bar{\mu}}^2 \sigma_m^2 + \left(\frac{\partial p}{\partial v}\right)_{\bar{\mu}}^2 \sigma_v^2 = \bar{v}^2 \sigma_m^2 + \bar{m}^2 \sigma_v^2$$

$$V_{EE} = \sigma_E^2 = \frac{1}{4} \bar{v}^4 \cdot \sigma_m^2 + \bar{m}^2 \bar{v}^2 \cdot \sigma_v^2$$

$$V_{pE} = V_{Ep} = \frac{\partial p}{\partial m} \bigg|_{\bar{\mu}} \frac{\partial E}{\partial m} \bigg|_{\bar{\mu}} \sigma_m^2 + \frac{\partial p}{\partial v} \bigg|_{\bar{\mu}} \frac{\partial E}{\partial v} \bigg|_{\bar{\mu}} \sigma_v^2 = \bar{v} \cdot \frac{1}{2} \bar{v}^2 \sigma_m^2$$

$$+ \bar{m} \bar{m} \bar{v} \sigma_v^2 = \frac{1}{2} \bar{v}^3 \sigma_m^2 + \bar{m}^2 \bar{v} \sigma_v^2$$

$$V = \begin{pmatrix} V_{pp} & V_{pE} \\ V_{Ep} & V_{EE} \end{pmatrix}$$

todo se pide  $\Delta p_x, \Delta p_y, \Delta p_z$

$$\textcircled{4} \quad \frac{1}{p} \rightarrow \Delta\left(\frac{1}{p}\right)$$

$$\lambda \rightarrow \Delta\lambda$$

$$\varphi \rightarrow \Delta\varphi$$

$\text{cov}(x_i, x_j) = 0$  (sin correlación)

$$p_x = p \cos \lambda \cos \varphi = \frac{1}{\left(\frac{1}{p}\right)} \cdot \cos \lambda \cos \varphi$$

$$\Delta^2(p_x) = \cos^2 \lambda \cos^2 \varphi \frac{1}{\left(\frac{1}{p}\right)^2} \Delta^2\left(\frac{1}{p}\right) + p^2 \sin^2 \lambda \cos^2 \varphi \Delta^2(\lambda)$$

$$+ p^2 \cos^2 \lambda \sin^2 \varphi \Delta^2(\varphi) = \frac{1}{\left(\frac{1}{p}\right)^2} \cdot \cos^2 \lambda \cos^2 \varphi \cdot \left( \Delta^2\left(\frac{1}{p}\right) + \tan^2 \lambda \Delta^2(\lambda) + \tan^2 \varphi \Delta^2(\varphi) \right)$$

$$= p_x^2 \left( \Delta^2\left(\frac{1}{p}\right) + \tan^2 \lambda \Delta^2(\lambda) + \tan^2 \varphi \Delta^2(\varphi) \right) \equiv V_{xx}$$

$$p_y = p \cos \lambda \sin \varphi$$

$$\Delta^2(p_y) = p^4 \cos^2 \lambda \sin^2 \varphi \Delta^2\left(\frac{1}{p}\right) + p^2 \sin^2 \lambda \sin^2 \varphi \Delta^2(\lambda) + p^2 \cos^2 \lambda \cos^2 \varphi \Delta^2(\varphi)$$

$$= p_y^2 \left( \Delta^2\left(\frac{1}{p}\right) + \tan^2 \lambda \Delta^2(\lambda) + \frac{1}{\tan^2 \varphi} \Delta^2(\varphi) \right) \equiv V_{yy}$$

$$p_z = p \sin \lambda$$

$$\Delta^2(p_z) = p^4 \sin^2 \lambda \Delta^2\left(\frac{1}{p}\right) + p^2 \cos^2 \lambda \Delta^2(\lambda) =$$

$$= p_z^2 \cdot \left( \Delta^2\left(\frac{1}{p}\right) + \frac{1}{\tan^2 \lambda} \Delta^2(\lambda) \right) \equiv V_{zz}$$

Por completitud:

$$V_{xy} \equiv \text{cov}(p_x, p_y) = \left( \frac{\partial p_x}{\partial \left(\frac{1}{p}\right)} \right) \left( \frac{\partial p_y}{\partial \left(\frac{1}{p}\right)} \right) \Delta^2\left(\frac{1}{p}\right) + \frac{\partial p_x}{\partial \lambda} \frac{\partial p_y}{\partial \lambda} \Delta^2(\lambda)$$

$$+ \frac{\partial p_x}{\partial \varphi} \frac{\partial p_y}{\partial \varphi} \Delta^2(\varphi) = p^2 \cos \lambda \cos \varphi \cdot p^2 \cos \lambda \sin \varphi \Delta^2\left(\frac{1}{p}\right) +$$

$$+ p(-\sin \lambda) \cos \varphi \cdot p(-\sin \lambda) \sin \varphi \Delta^2(\lambda) + p \cos \lambda (-\sin \varphi) \cdot p \cos \lambda \cos \varphi \Delta^2(\varphi)$$

$$= p^4 \cos^2 \lambda \sin \varphi \cos \varphi \Delta^2\left(\frac{1}{p}\right) + p^2 \sin^2 \lambda \sin \varphi \cos \varphi \Delta^2(\lambda) - p^2 \cos^2 \lambda \sin \varphi \cos \varphi \Delta^2(\varphi)$$

$$= p^2 \cos^2 \lambda \sin \varphi \cos \varphi \cdot \left( p^2 \Delta^2\left(\frac{1}{p}\right) + \tan^2 \lambda \Delta^2(\lambda) - \Delta^2(\varphi) \right) = p_x \cdot p_y$$

$$V_{xz} = V_{zx} = \frac{\partial p_x}{\partial (\frac{1}{p})} \cdot \frac{\partial p_z}{\partial \varphi} \cdot \Delta^2(\frac{1}{p}) + \frac{\partial p_x}{\partial \lambda} \cdot \frac{\partial p_z}{\partial \lambda} \cdot \Delta^2(\lambda) + \frac{\partial p_x}{\partial \varphi} \cdot \frac{\partial p_z}{\partial p} \cdot \Delta^2(\varphi)$$

$$= p^2 \cos \lambda \cos \varphi p^2 \sin \lambda \Delta^2(\frac{1}{p}) + p (-\sin \lambda) \cos \varphi \cdot p \cos \lambda \Delta^2(\lambda)$$

$$= p^2 \sin \lambda \cos \lambda \cos \varphi (p^2 \Delta^2(\frac{1}{p}) - \Delta^2(\lambda))$$

$$= p_x p_z$$

$$V_{yz} = V_{zy} = \frac{\partial p_y}{\partial (\frac{1}{p})} \cdot \frac{\partial p_z}{\partial (\frac{1}{p})} \cdot \Delta^2(\frac{1}{p}) + \frac{\partial p_y}{\partial \lambda} \cdot \frac{\partial p_z}{\partial \lambda} \cdot \Delta^2(\lambda) + \frac{\partial p_y}{\partial \varphi} \cdot \frac{\partial p_z}{\partial p} \cdot \Delta^2(\varphi)$$

$$= p^2 \cos \lambda \sin \varphi p^2 \sin \lambda \Delta^2(\frac{1}{p}) + (-p \sin \lambda \sin \varphi) \cdot p \cos \lambda \Delta^2(\lambda)$$

$$= p^2 \sin \lambda \cos \lambda \sin \varphi (p^2 \Delta^2(\frac{1}{p}) - \Delta^2(\lambda))$$

$$= p_y p_z$$

Matrice de covarianza  $V = \begin{pmatrix} V_{xx} & V_{xy} & V_{xz} \\ V_{yx} & V_{yy} & V_{yz} \\ V_{zx} & V_{zy} & V_{zz} \end{pmatrix}$

# Problemas de Métodos Estadísticos

## 4.- Distribuciones de probabilidad

1.- Un estudiante se dispone a hacer auto-stop en un cruce donde en promedio pasa un coche por minuto. Si la probabilidad de que un conductor pare y lo lleve en coche es del 1 % ¿Cuál es la probabilidad de que el estudiante siga en el cruce después de que hayan pasado 60 coches? ¿Y al cabo de una hora?

2.- Las distribuciones angulares suelen caracterizarse por la asimetría forward-backward  $\frac{F}{N}$ , donde  $F$  es en número de sucesos con  $\cos\theta > 0$ ,  $B$  es el número de sucesos con  $\cos\theta < 0$ , y  $N = F + B$  es el número total de sucesos. Suponemos que los sucesos son independientes y que el promedio de sucesos por unidad de tiempo es constante, tanto para los sucesos forward como backward. Claramente solo dos de las tres variables,  $F$ ,  $B$ ,  $N$ , son independientes. Podemos ver esta situación de dos maneras.

a) El número de sucesos  $N$  se distribuye según una distribución de Poisson de media  $\mu$ , los cuales se subdividen en  $F$  y  $B = N - F$  siguiendo una distribución binomial,  $B(F; N; p_F)$ , *i.e.* las variables independientes son  $N$  y  $F$ .

b) Los sucesos  $F$  y  $B$  se distribuyen según Poisson (con parámetros  $\mu_F$  y  $\mu_B$ ), y el total es la suma de ambos, *i.e.*, las variables independientes son  $F$  y  $B$ .

Demostrar que ambos planteamientos conducen a la misma pdf.

3- Se diseña un experimento ara medir el coeficiente de asimetría de Goldhaber:

$$\gamma \equiv \frac{F - B}{F + B} \equiv \frac{2F}{N} - 1$$

Demostrar que:

a) Si  $F$  y  $B$  se consideran variables independientes de Poisson, la varianza de  $\gamma$  es aproximadamente.

$$V(\gamma) \cong \frac{4FB}{(F + B)^3}$$

b) Si consideramos  $N$  fijo, con  $F$  y  $N$  ligados en una ley binomial con constante  $N$ ,  $p$  y  $q$ , la varianza viene dad por la expresión exacta.

$$V(\gamma) = \frac{4pq}{N}$$

c) Demostrar que ambas son equivalentes.

4.- Demostrar que la media y la varianza de una distribución uniforme  $f(x) = 1/(b-a)$  vienen dadas por  $E[x] = (a+b)/2$  y  $V(x) = (b-a)^2/12$ .

5.- Para la función densidad de probabilidad binormal con variables  $x$  e  $y$  y coeficiente de correlación  $\rho$ , realizar un cambio de variable  $u$  y  $v$ , de manera que la matriz de covarianza sea diagonal y demostrar que:

$$\sigma_u^2 = \frac{\sigma_x^2 \cos^2 \theta - \sigma_y^2 \sin^2 \theta}{\cos^2 \theta - \sin^2 \theta} \quad \sigma_v^2 = \frac{\sigma_y^2 \cos^2 \theta - \sigma_x^2 \sin^2 \theta}{\cos^2 \theta - \sin^2 \theta}$$

donde:  $\tan 2\theta = \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 - \sigma_y^2}$

6.- Demostrar que FWHM (full width high maximun) de una distribución norma  $N(\mu, \sigma^2)$  viene dada por  $2\sigma\sqrt{2\ln 2}$

## Problemas con programación

### Tablero de Galton

En un tablero de Galton numeramos las filas de clavos horizontales como  $j = 1, 2, \dots, n$  de manera que la fila  $j$ -ésima tiene  $j$  clavos. Llamamos  $N_{exp}$  al número total de bolitas que caen en el clavo de la fila 1. Cada bola es desviada hacia la derecha con una probabilidad  $p$  y hacia la izquierda con una probabilidad  $(1-p)$ . Después de caer a través de  $n$  filas de clavos, cada bola estará en una de las  $n+1$  posiciones que denotaremos por  $k = 0$  (a la izquierda),  $1, 2, \dots, k = n$  (a la derecha). Después de realizar  $N_{exp}$  experimentos (i.e. bolitas) encontraremos  $N(k)$  bolas para cada valor de  $k$ .

Realizar un programa que simule el tablero de Galton de manera que en primer lugar pregunte al usuario los valores de  $N_{exp}$ ,  $n$  y  $p$ . Para cada experimento, se fija el valor de  $k$  a cero y se generan  $n$  números aleatorios  $r_j$  uniformemente distribuidos. Cada vez que  $r_j < p$ , lo que corresponde a una desviación a la derecha, el número  $k$  se incrementa en una unidad. Por cada experimento, el valor de  $k$  obtenido se almacena en un histograma. Finalmente se muestra el histograma.

Estudiar como se obtiene la distribución de Poisson eligiendo sucesivamente pares de números  $(n,p)$  como por ejemplo  $(n,p) = (1,0.3), (10, 0.03), (100, 0.003), \dots$  de tal manera que el producto  $np$  permanezca constante. Comprobar que en el límite  $n \rightarrow \infty, p \rightarrow 0$ , se obtiene la distribución de Poisson:

$$N(k) = N \frac{\lambda^k}{k!} \exp^{-\lambda}, \quad \lambda = np$$

y comparar con las distribuciones obtenidas.

### Distribución binormal y correlación.

Generar 100000 sucesos de valores  $(x,y)$  distribuidos según una distribución binormal no correlacionada con:

$$\mu_{x_1} = 3, \quad \sigma_{x_1} = 2, \quad \mu_{x_2} = 3, \quad \sigma_{x_2} = 0.5,$$

- Representar los puntos en un histograma de dos dimensiones y calcular la matriz de covarianza y el coeficiente de correlación.
- Considerar la función  $f(x,y) = 3x + 5y$ . Calcular la varianza de  $f(x,y)$  directamente usando propagación de errores y también a partir de los datos.
- Rotar ahora los sucesos respecto al centro de la distribución, no del origen, un ángulo  $\theta = 30^\circ$  y repetir los pasos anteriores.
- A partir de los datos correlacionados estimar el ángulo  $\theta$  y los valores de las  $\sigma_u, \sigma_v$ .
- Repetir todos los apartados anteriores para un ángulo  $\theta = -30^\circ$ .

Problemas Tema 4

①  $p = 1\%$

$\lambda = \frac{N}{T} = 1 \text{ min}^{-1}$

a)  $P(60 \text{ coches no paren}) = \overset{\text{binomial}}{B}(0, 60, 1\%) = \binom{60}{0} \cdot (1\%)^0 \cdot (1 - 1\%)^{60}$   
 $= (0,99)^{60} = 54,7\%$

b)  $P(\text{tras 1 hora})$  ?  $\rightarrow$  Suponemos válido aplicar Poisson

$P_r(t) = \frac{(2t)^r \cdot e^{-2t}}{r!}$

$\rightarrow$  probabilidad de que paren  $r$  coches en el tiempo  $t = 60 \text{ min}$

$P_r(1 \text{ hora}) = \frac{(60)^r \cdot e^{-60}}{r!}$

$P(r \text{ no paren}) = (1-p)^r \cdot P_r(1 \text{ hora}) = \frac{(60 \cdot 0,99)^r \cdot e^{-60}}{r!}$

En total, de que no pare ninguno:

$P_{\text{tot}} = \sum_{r=0}^{\infty} P(r \text{ no paren}) = \sum_{r=0}^{\infty} \frac{(60 \cdot 0,99)^r \cdot e^{-60}}{r!} = e^{-60} \cdot e^{60 \cdot 0,99}$   
 $= e^{-60 \cdot 0,01} = e^{-0,6} = 54,9\%$

- ②  $F$ : nº sucesos forward  
 $B$ : " backward  
 $N = F + B$ : nº total de sucesos  
 $A = \frac{F}{N}$ : asimetría

sucesos independientes

$\lambda_F \text{ cte} = \frac{N \lambda_B}{T}$

a)  $P_f(F; N, p_f) = \binom{N}{F} p_f^F (1-p_f)^{N-F}$   $\rightarrow$  binomial  
 probabilidad de obtener  $F$  sucesos tipo forward

$P_n(N; \mu) = \frac{\mu^N}{N!} \cdot e^{-\mu}$   $\rightarrow$  poissoniana  
 prob. obtener  $N$  sucesos totales

$N, F$ : independientes

$B = N - F$

¿Cuál es la probabilidad  $P_b(B; N, F)$  de obtener  $B$  sucesos tipo backward?

F	N
0	B
1	B+1
2	B+2
⋮	⋮

→ Múltiples combinaciones:

$$P_b = \sum_{N=B}^{\infty} P_n(N, \mu) \cdot P_f(N-B; N, p_f)$$

o bien:

↗ equivalente con un c. de variable

$$P_b(B) = \sum_{F=0}^{\infty} P_n(F+B, \mu) \cdot P_f(F; F+B, p_f)$$

$$= \sum_{F=0}^{\infty} \frac{\mu^{F+B} e^{-\mu}}{(F+B)!} \cdot \binom{F+B}{F} p_f^F (1-p_f)^{F+B-F}$$

$$= \sum_{F=0}^{\infty} \frac{\mu^B \mu^F e^{-\mu}}{F!} \cdot \frac{(F+B)!}{(F+B)!} \cdot \frac{1}{(F+B-F)!} p_f^F \cdot (1-p_f)^B$$

$$= \frac{(\mu p_f)^B}{B!} \left( \sum_{F=0}^{\infty} \frac{(\mu p_f)^F \cdot e^{-\mu p_f}}{F!} \right) \cdot e^{\mu p_f} \cdot e^{-\mu}$$

$$= \frac{(\mu p_f)^B}{B!} \cdot e^{-\mu(1-p_f)} = \text{Poissoniana}(B; \mu p_f)$$

centrada en  $\mu_B = \mu \cdot p_f$

↳ coincide con apartada b)

b)  $P_f(F; \mu_f) = \frac{\mu_f^F e^{-\mu_f}}{F!}$  ↗ poissoniana

$$P_b(B; \mu_b) = \frac{\mu_b^B e^{-\mu_b}}{B!}$$

B	F
0	N
1	N-1
2	⋮
⋮	⋮
N	0

$$P_n(N; F, B) = \sum_{B=0}^N P_b(B; \mu_b) \cdot P_f(N-B; \mu_f)$$

$$= \sum_{B=0}^N \frac{\mu_b^B}{B!} e^{-\mu_b} \cdot \frac{\mu_f^{N-B}}{(N-B)!} e^{-\mu_f} = e^{-(\mu_b + \mu_f)} \sum_{B=0}^N \frac{\mu_b^B \cdot \mu_f^{N-B}}{B!(N-B)!} \frac{N!}{N!}$$

$$= \frac{e^{-(\mu_b + \mu_f)}}{N!} \sum_{B=0}^N \binom{N}{B} \mu_b^B \mu_f^{N-B} = \frac{e^{-(\mu_b + \mu_f)}}{N!} \cdot (\mu_b + \mu_f)^N =$$

$$= \frac{(\mu_B + \mu_F)^N}{N!} \cdot e^{-(\mu_B + \mu_F)} = \text{Poissoniana } (N; \mu_B + \mu_F)$$

con  $\mu = \mu_B + \mu_F$

↳ coincide con apartado a)

3)  $Y = \frac{F-B}{F+B} = \frac{2F}{N} - 1$

No se puede calcular  $E[Y]$ ,  $V[Y]$  multiplicando con la pdf explícita.

a)  $F, B$ : independientes  $\rightarrow$  propagación de errores cuadrática  
 Taylor 1º orden, supones órdenes superiores despreciables  
 sigue ley

$V(Y) \approx \left(\frac{\partial Y}{\partial F}\right)^2 \cdot V(F) + \left(\frac{\partial Y}{\partial B}\right)^2 \cdot V(B)$  (no hay más términos al ser  $F$  y  $B$  indep.!!)

$$= \left(\frac{2(F+B) - (F-B) \cdot 1}{(F+B)^2}\right)^2 V(F) + \left(\frac{(-1) \cdot (F+B) - (F-B) \cdot 1}{(F+B)^2}\right)^2 V(B)$$

$$= \left(\frac{2B}{(F+B)^2}\right)^2 \cdot V(F) + \left(\frac{2F}{(F+B)^2}\right)^2 \cdot V(B)$$

Como  $F$  y  $B$  son poissonianas:  $V(F) = F$ ;  $V(B) = B$   
 la varianza coincide con el valor esperado

$$= \frac{4}{(F+B)^4} \cdot (B^2 F + F^2 B) = \frac{4BF}{(F+B)^4} (B+F) = \frac{4FB}{(F+B)^3}$$

b)  $N$  fijo

$$P_f(F; N, p, q) = \binom{N}{F} \cdot p^F q^{N-F} \quad \text{con } p = p_F$$

$q = 1 - p = 1 - p_F = p_B$

$$V[Y] = E[(Y - E[Y])^2]$$

$$E[Y] = E\left[\frac{2F}{N} - 1\right] = \frac{2}{N} \cdot E[F] - 1 \Rightarrow E[F] = \frac{N}{2} (E[Y] + 1)$$

$$= \sum_{F=0}^N \frac{F \cdot N!}{(N-F)! F!} \cdot p^F q^{N-F} = \sum_{F=1}^N \frac{N(N-1)!}{(N-1-(F-2))! (F-1)!} \cdot p \cdot p^{F-2} \cdot q^{N-2-(F-1)}$$

$$= p \sum_{M=0}^R \binom{R}{M} p^M q^{R-M} = p \cdot N (p+q)^{N-1} = NP \quad \text{(media de la binomial)}$$

El 3 sigue más abajo!

④  $f(x) = \frac{1}{b-a}$

$$E[X] = \int_a^b f(x) \cdot x \cdot dx = \frac{1}{b-a} \int_a^b x \cdot dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{(b-a)(b+a)}{2(b-a)}$$

$$= \frac{(b+a)}{2}$$

$$V[X] = \int_a^b f(x) dx \cdot (E[X] - x)^2 = \frac{1}{b-a} \int_a^b dx \cdot \left( \frac{b+a}{2} - x \right)^2$$

$$= \frac{1}{b-a} \cdot \left( -\frac{1}{3} \right) \left( \frac{b+a}{2} - x \right)^3 \Big|_a^b = -\frac{1}{3} \cdot \frac{1}{b-a} \left[ \left( \frac{a-b}{2} \right)^3 - \left( \frac{b-a}{2} \right)^3 \right]$$

$$= \frac{1}{24} \cdot \frac{1}{a-b} \cdot (2 \cdot (a-b)^3) = \frac{1}{12} (a-b)^2 = \frac{(b-a)^2}{12}$$

③ - continuación

$$E[Y] = \frac{2}{N} \cdot Np - 1 = 2p - 1$$

$p=0 \rightarrow Y(F=0) = E[Y]_{p=0} = -1$   
 $p=\frac{1}{2} \rightarrow Y(F=\frac{N}{2}) = E[Y]_{p=1/2} = 0$   
 $p=1 \rightarrow Y(F=N) = E[Y]_{p=1} = +1$

$$V(Y) = E\left[ \left( \frac{2F}{N} - 1 - (2p - 1) \right)^2 \right] = E\left[ \left( \frac{2F}{N} - 2p \right)^2 \right]$$

$$= E\left[ \frac{4F^2}{N^2} - \frac{8FP}{N} + 4p^2 \right] = \frac{4}{N^2} \cdot E[F^2] - \frac{8p}{N} E[FP] + 4p^2$$

$$E[FP] = NP$$

$$E[F^2] = \sum_{F=0}^N F^2 \frac{N! \cdot p^F q^{N-F}}{F!(N-F)!} = \sum_{F=1}^N F \cdot N! \frac{p^F q^{N-F}}{(F-1)!(N-F)!}$$

$$= \sum_{F=2}^N \frac{N!}{(F-2)!} \frac{p^F q^{N-F}}{(N-F)!} + \sum_{F=1}^N \frac{N!}{(N-F)(F-1)!} \frac{p^F q^{N-F}}{(N-F)!} = N \cdot (N-1) \sum_{F=2}^N \frac{(N-1)!}{(F-2)!} \cdot p^2 \frac{p^{F-2} q^{N-F}}{(N-F)!}$$

$$+ Np \sum_{F=1}^N \frac{(N-1)!}{(F-1)!} \frac{p^{F-1} q^{N-F}}{(N-F)!} = N \cdot (N-1) \cdot p^2 + Np = Np(Np - p + 1)$$

$$\begin{aligned}
 V(\bar{x}) &= \frac{4}{N^2} \cdot Np(Np - p + 1) - \frac{8p}{N} \cdot Np + 4p^2 \\
 &= \frac{4p}{N} (Np + q) - 4p^2 = \frac{4p}{N} (Np + q - Np) \\
 &= \frac{4pq}{N}
 \end{aligned}$$

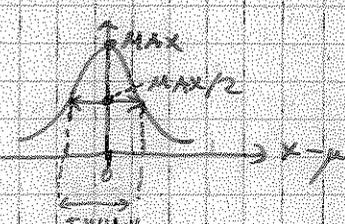
c)  $N = F + B$  ya que F, B son en realidad  $E[F], E[B]$   
(resultado de varias medidas)

$\frac{F}{N} = p$   $\rightarrow V(\bar{x}) \approx \frac{4 \overbrace{Np}^F \cdot \overbrace{Nq}^B}{N^3} = \frac{4pq}{N}$   $\rightarrow$  coincide con b) ✓

$\frac{B}{N} = q$

⑤  $\rightarrow$  eliminado

⑥ FWHM  $\rightarrow N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$

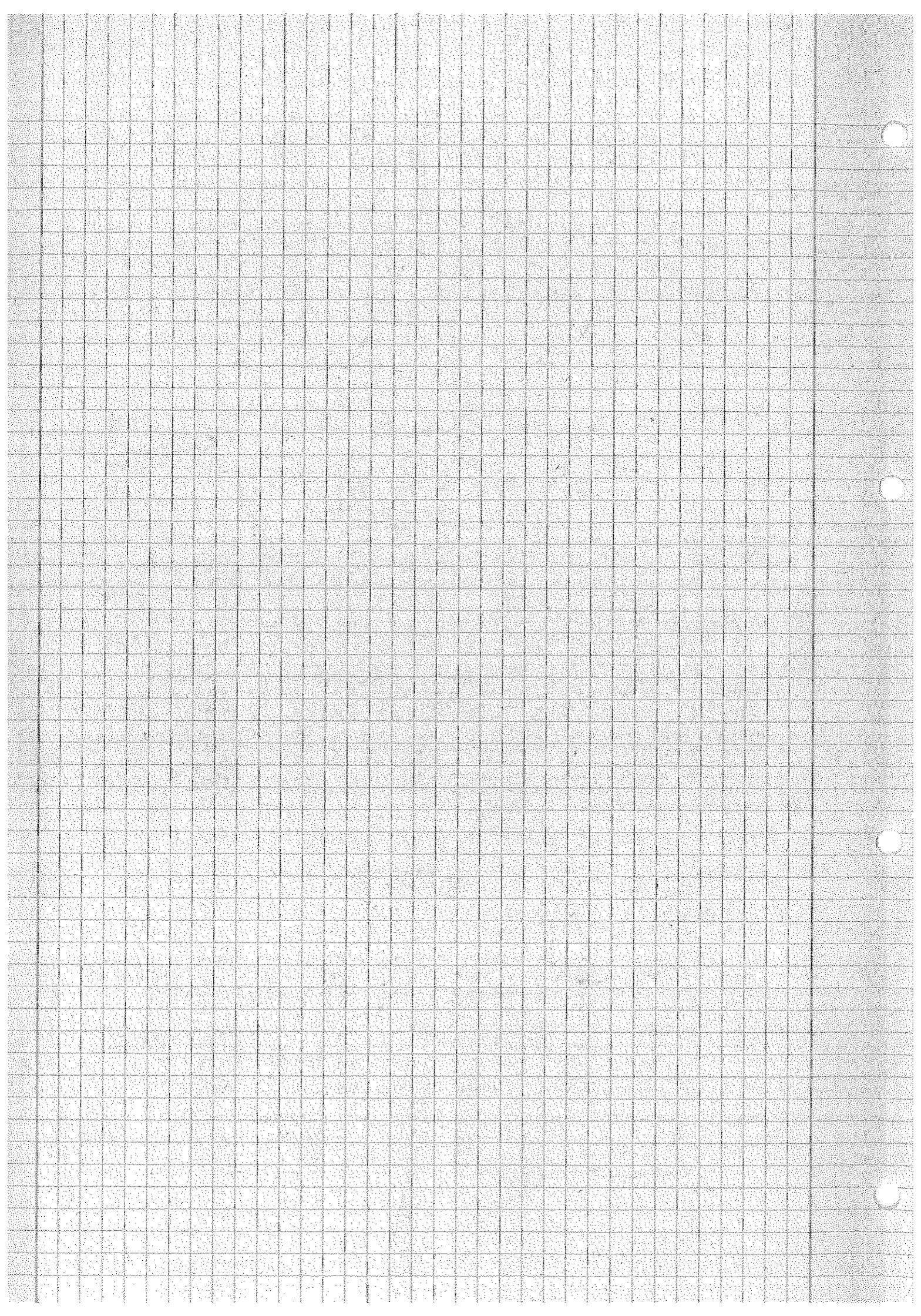


$$N(x - \mu = 0; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\frac{1}{2} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \left( \frac{x_{\pm} - \mu}{\sigma} \right)^2}$$

$$\begin{aligned}
 \ln\left(\frac{1}{2}\right) &= -\frac{1}{2} \left( \frac{x_{\pm} - \mu}{\sigma} \right)^2 \rightarrow \frac{x_{\pm} - \mu}{\sigma} = \pm \sqrt{2 \ln 2} \\
 &\quad -\ln 2 \\
 x_{\pm} &= \mu \pm \sigma \sqrt{2 \ln 2}
 \end{aligned}$$

$$\begin{aligned}
 \Rightarrow \text{FWHM} &= x_+ - x_- = \mu + \sigma \sqrt{2 \ln 2} - (\mu - \sigma \sqrt{2 \ln 2}) \\
 &= 2\sigma \sqrt{2 \ln 2}
 \end{aligned}$$



# -Métodos estadísticos-

## Problemas con programación del Tema 4: Distribuciones de probabilidad

Fernando Hueso González\*

*Valencia, 21 de noviembre de 2011*

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

### Tablero de Galton

**Enunciado** En un tablero de Galton numeramos las filas de clavos horizontales como  $j = 1, 2, \dots, n$  de manera que la fila  $j$ -ésima tiene  $j$  clavos. Llamamos  $N_{exp}$  al número total de bolitas que caen en el clavo de la fila 1. Cada bola es desviada hacia la derecha con una probabilidad  $p$  y hacia la izquierda con una probabilidad  $(1 - p)$ . Después de caer a través de  $n$  filas de clavos, cada bola estará en una de las  $n + 1$  posiciones que denotaremos por  $k = 0$  (a la izquierda), 1, 2, ...  $k = n$  (a la derecha). Después de realizar  $N_{exp}$  experimentos (i.e. bolitas) encontraremos  $N(k)$  bolas para cada valor de  $k$ .

Realizar un programa que simule el tablero de Galton de manera que en primer lugar pregunte al usuario los valores de  $N_{exp}$ ,  $n$  y  $p$ . Para cada experimento, se fija el valor de  $k$  a cero y se generan  $n$  números aleatorios  $r_j$  uniformemente distribuidos. Cada vez que  $r_j < p$ , lo que corresponde a una desviación a la derecha, el número  $k$  se incrementa en una unidad. Por cada experimento, el valor de  $k$  obtenido se almacena en un histograma. Finalmente se muestra el histograma.

Estudiar como se obtiene la distribución de Poisson eligiendo sucesivamente pares de números  $(n, p)$ , como por ejemplo  $(n, p) = (1, 0.3), (10, 0.03), (100, 0.003), \dots$  de tal manera que el producto  $np$  permanezca constante. Comprobar que en el límite  $n \rightarrow \infty, p \rightarrow 0$ , se obtiene la distribución de Poisson:

$$N(k) = N \frac{\lambda^k}{k!} \exp(-\lambda) \quad , \quad \lambda = np$$

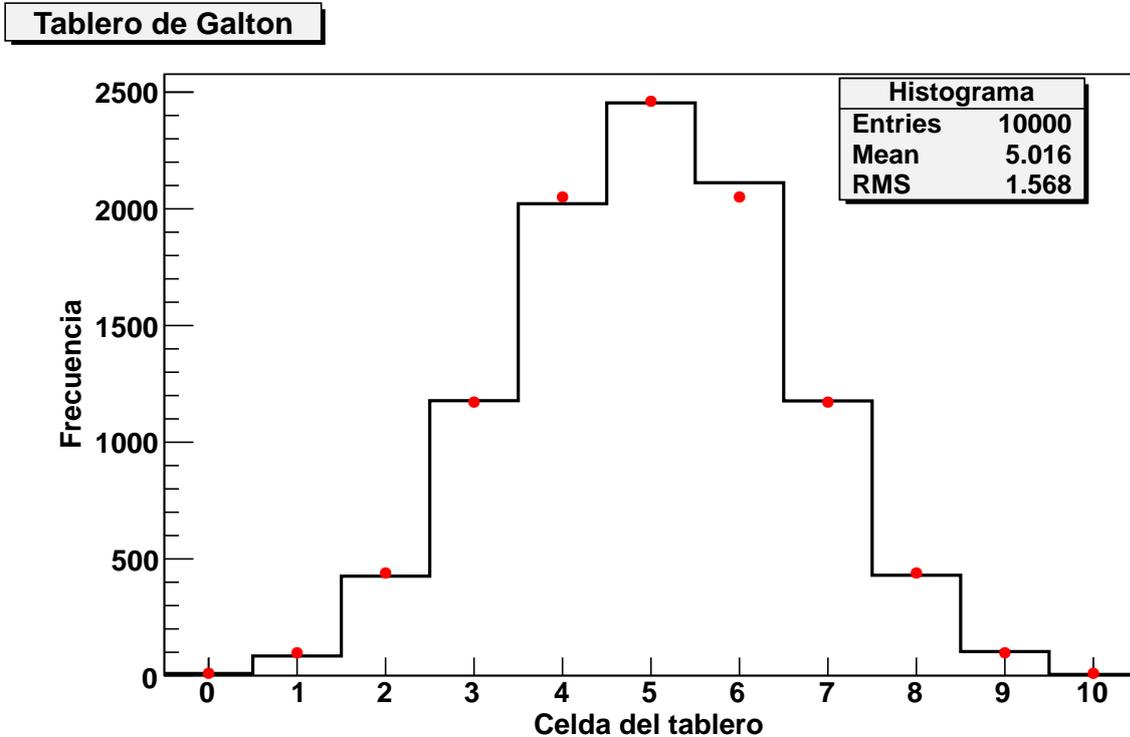
y comparar con las distribuciones obtenidas.

**Resolución** Programo un SCRIPT en ROOT que en cada fila  $j$  genera un número aleatorio  $r_j$  entre 0 y 1 (función TRandom1), que al comparar con el valor de  $p$  provocará que la bola caiga a la derecha ( $r_j < p$ ) o a la izquierda ( $r_j > p$ ), es decir que el índice  $k$  de celda aumente en una unidad.

---

\*ferhue#alumni.uv.es

Realizamos una simulación para 10.000 repeticiones del experimento (bolitas), con 10 filas (11 celdas) y una probabilidad  $p = 0.5$ , e histograma la frecuencia con la que caen las bolas en cada celda.



Las barras negras corresponden a los valores obtenidos a partir de la generación pseudoaleatorio de números. Los puntos rojos corresponden al valor esperado según la distribución de probabilidad binomial  $B(k, n, p)$ , que es la que describe este experimento, isomorfo al triángulo de Pascal cambiando clavos por números (de  $n$  filas/intentos, siendo  $p$  la probabilidad de caer a la derecha/de éxito, cuál es la probabilidad  $B$  de que la bola caiga en la celda  $k$ /de tener  $k$  éxitos).

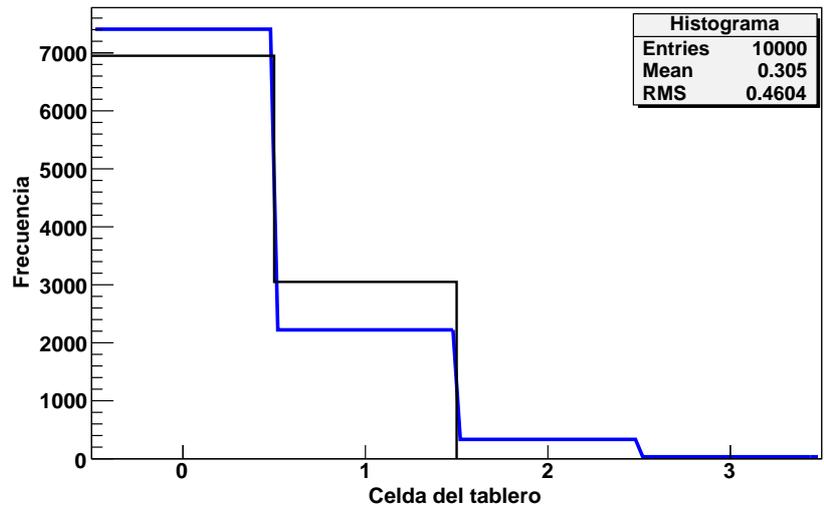
Dado que es equiprobable que caiga a derecha o izquierda, es lógico que la celda central contenga el mayor número de bolitas, y que la distribución sea aproximadamente simétrica (salvo las fluctuaciones estadísticas esperables) respecto a dicha celda.

Asimismo, se comprueba que la distribución binomial es compatible con los valores aleatorios generados, lo que indica que el generador es fiable, y los valores fluctúan respecto al de la pdf del orden de la raíz de número de cuentas del bin, como se puede comprobar si se hace zoom. Aparte, la distribución binomial tiende a la gaussiana conforme se aumenta  $N_{exp}$ .

Para estudiar cómo la distribución binomial tiende a la de Poisson o de sucesos raros, iremos disminuyendo la probabilidad  $p$  en la misma medida que aumentamos el número de filas  $n$ , de forma que el producto  $np$  permanezca constante. Realizamos la simulación y representamos el correspondiente histograma para 4 casos con  $np = 0.3 = cte$ ,  $p$  decreciente y  $n$  creciente:  $(n, p) = (1, 0.3); (2, 0.15); (5, 0.06); (10, 0.03)$ . No continúa el proceso para más casos pues no se aprecia diferencia visual en el histograma correspondiente.

En negro representamos el histograma a partir de los números pseudoaleatorios generados, en azul el histograma de la distribución poissoniana con media  $\lambda = np$  de acuerdo con el enunciado.

**Tablero de Galton**



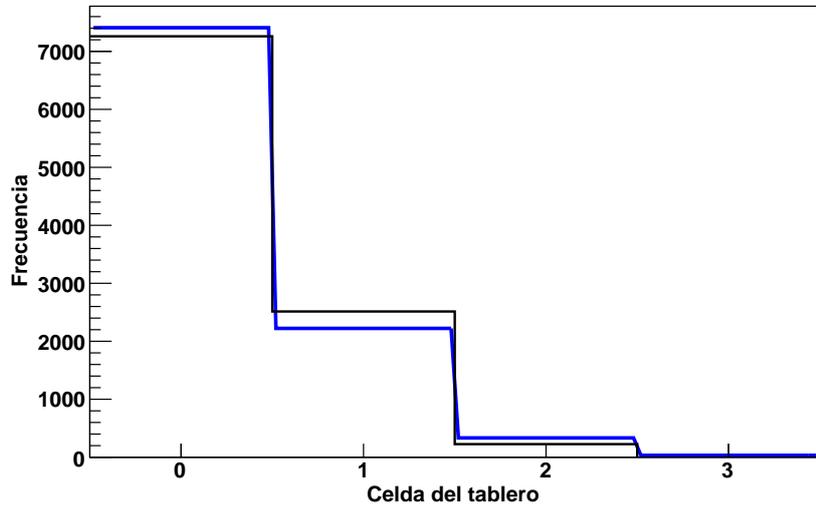
$$(n, p) = (1, 0.3)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 0.3$

Por tanto, se observa una distribución típica de sucesos raros, con una acumulación en la celda 0 (0 éxitos), que sin embargo se diferencia considerablemente de la distribución de Poisson, más allá de las fluctuaciones estadísticas esperables (raíz de cuentas del bin<sup>1</sup>). Esto se debe a que la distribución binomial (fórmula teórica que sí describe bien la curva negra, como hemos visto en el anterior histograma en los puntos rojos) no tiende suficientemente para este valor de  $p$  a la distribución de Poisson.

Si dividimos  $p$  entre 2, obtenemos el siguiente histograma:

**Tablero de Galton**



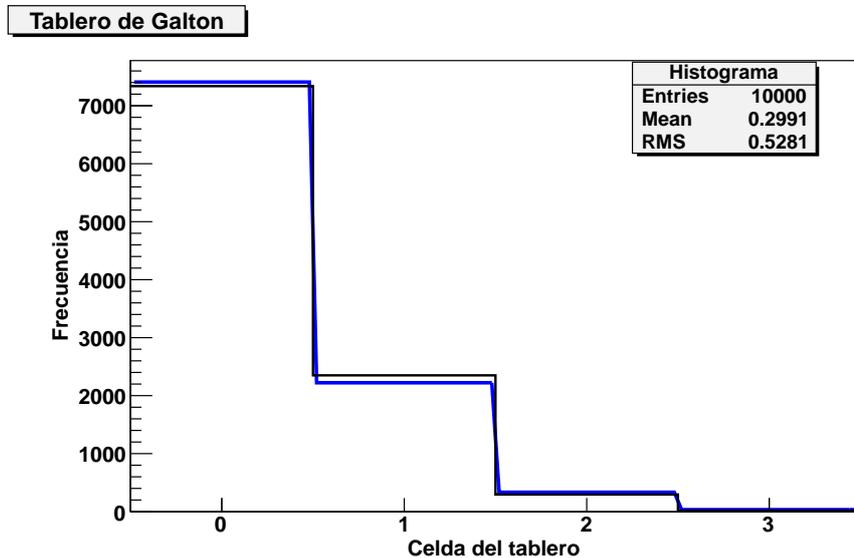
$$(n, p) = (2, 0.15)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 0.3$

Se comprueba que la distribución se asemeja más a la de Poisson, con menores fluctuaciones,

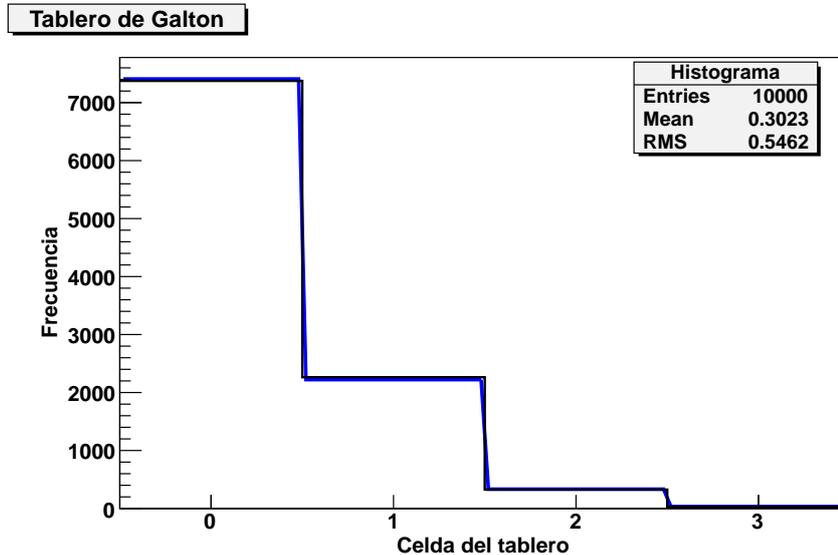
<sup>1</sup>La fórmula exacta para el error del número de cuentas  $r_k$  en el bin  $k$  es  $\sigma(r_k) = \sqrt{r_k(1 - \frac{r_k}{N_{exp}})} \sim \sqrt{r_k}$ . Ésta es una buena aproximación en la mayoría de bins ( $r_k \ll N_{exp}$ ) salvo en  $k=0$ , donde la desviación es algo menor dado el factor alto de  $r_k/N_{exp}$  al tratarse de sucesos raros.

pero todavía algo superiores a la raíz del número de cuentas en el bin. Si iteramos el proceso, obtenemos los siguientes histogramas:



$$(n, p) = (5, 0.06)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 0.3$



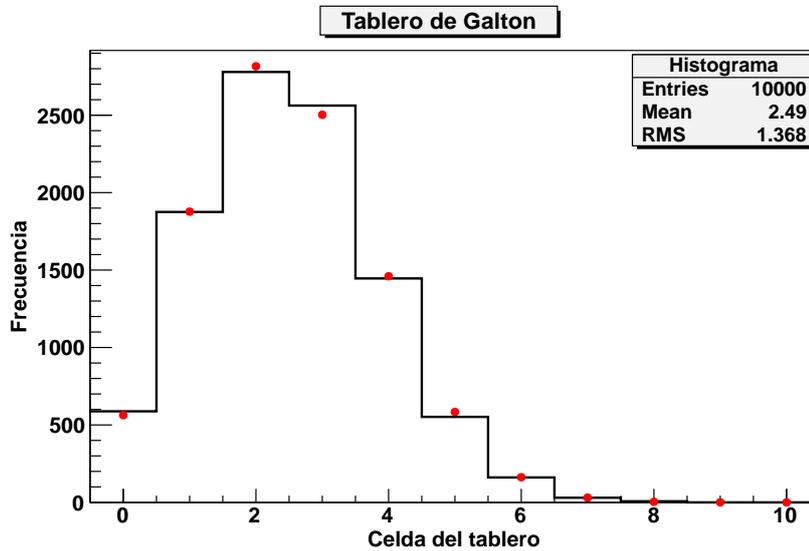
$$(n, p) = (10, 0.03)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 0.3$

Por tanto, se observa como la distribución binomial tiende a la poissoniana a medida que  $p$  decrece y  $n$  aumenta en el mismo factor, manteniendo  $np$  constante. Las fluctuaciones son del orden de la raíz del número de cuentas en el bin, como se puede comprobar visualmente haciendo zoom. Si quisiésemos hacer un estudio con mayor profundidad, una manera de comprobar objetivamente cómo la distribución binomial tiende a la poissoniana sería calculando los parámetros

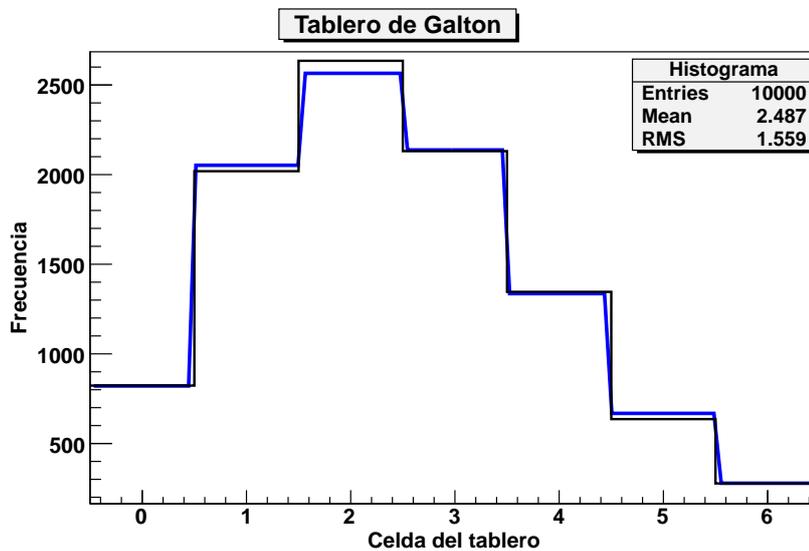
típicos de dicha distribución: varianza (tenderá al valor medio  $np$ ), skewness (tenderá a  $1/\sqrt{np}$ ), kurtosis (tenderá a  $1/np$ ).

Si jugamos con otros valores, obtenemos las siguientes distribuciones, donde se aprecia mejor cómo la distribución tiende a la poissoniana a medida que aumentamos  $n$  y reducimos  $p$  en factores equivalentes:



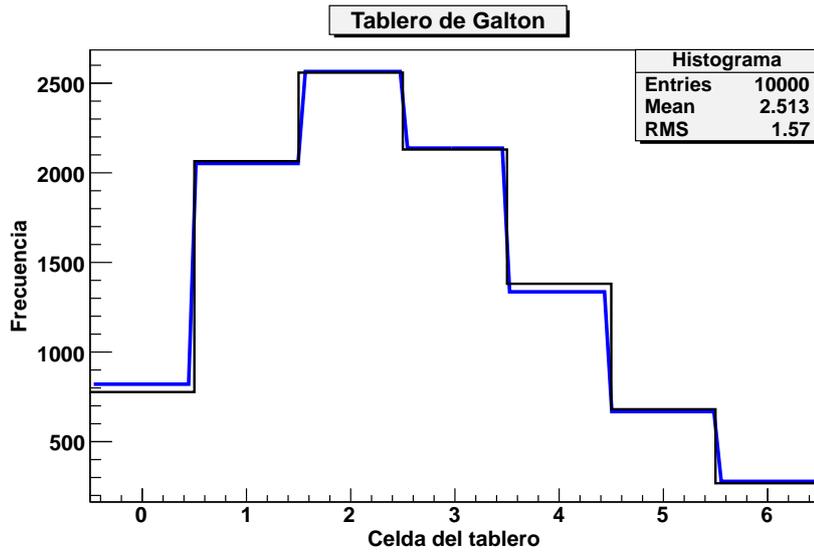
$$(n, p) = (10, 0.25)$$

Negro=Números pseudoaleatorios; Rojo=Binomial



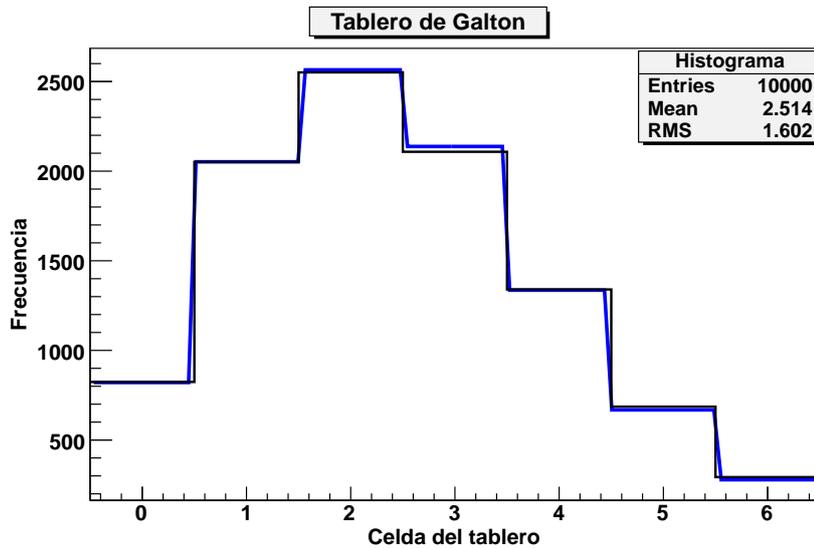
$$(n, p) = (1000, 0.0025)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 2.5$



$$(n, p) = (10000, 0.00025)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 2.5$



$$(n, p) = (100000, 0.000025)$$

Negro=Números pseudoaleatorios; Azul=Poisson con  $\lambda = 2.5$

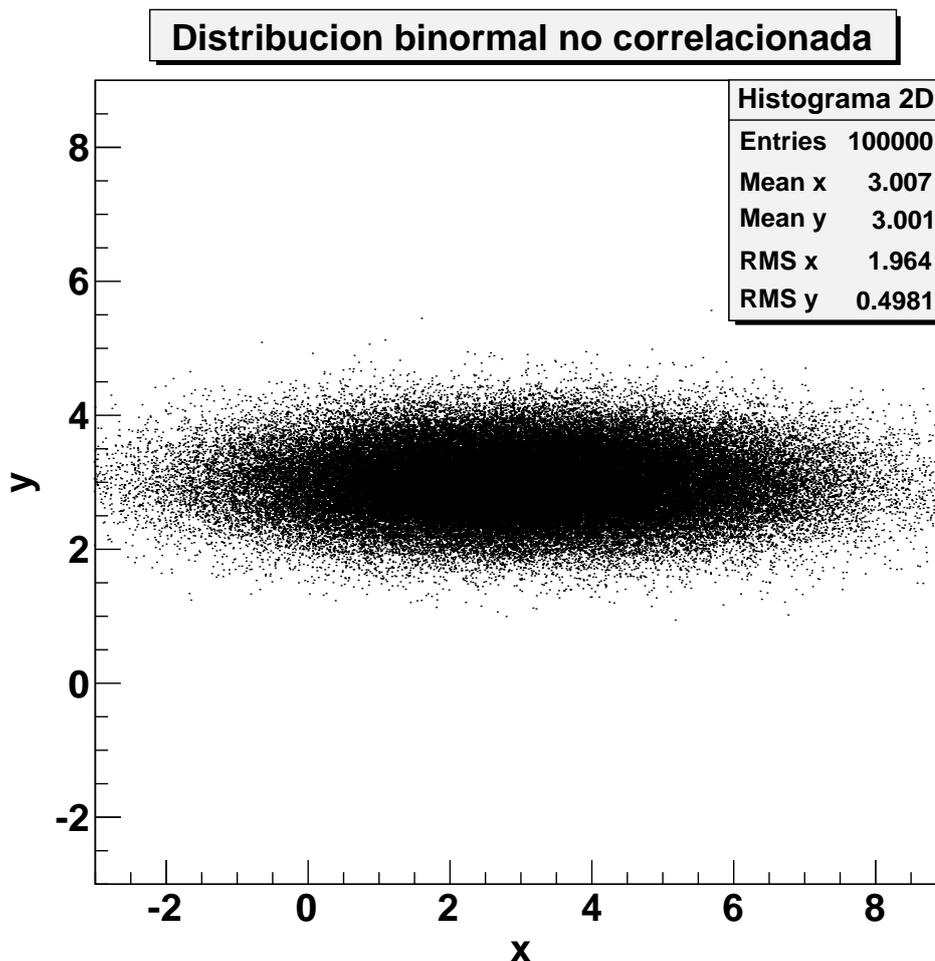
Se comprueba visualmente que a medida que aumentamos  $n$  y reducimos  $p$ , manteniendo  $np$  constante, la distribución poissoniana reproduce con menores desviaciones la distribución de números pseudoaleatorios. En el límite  $n$  tendiendo a infinito,  $p$  tendiendo a 0, con  $np$  constante, la distribución binomial tiende a la poissoniana (ley de sucesos raros).

## Distribución binormal y correlación

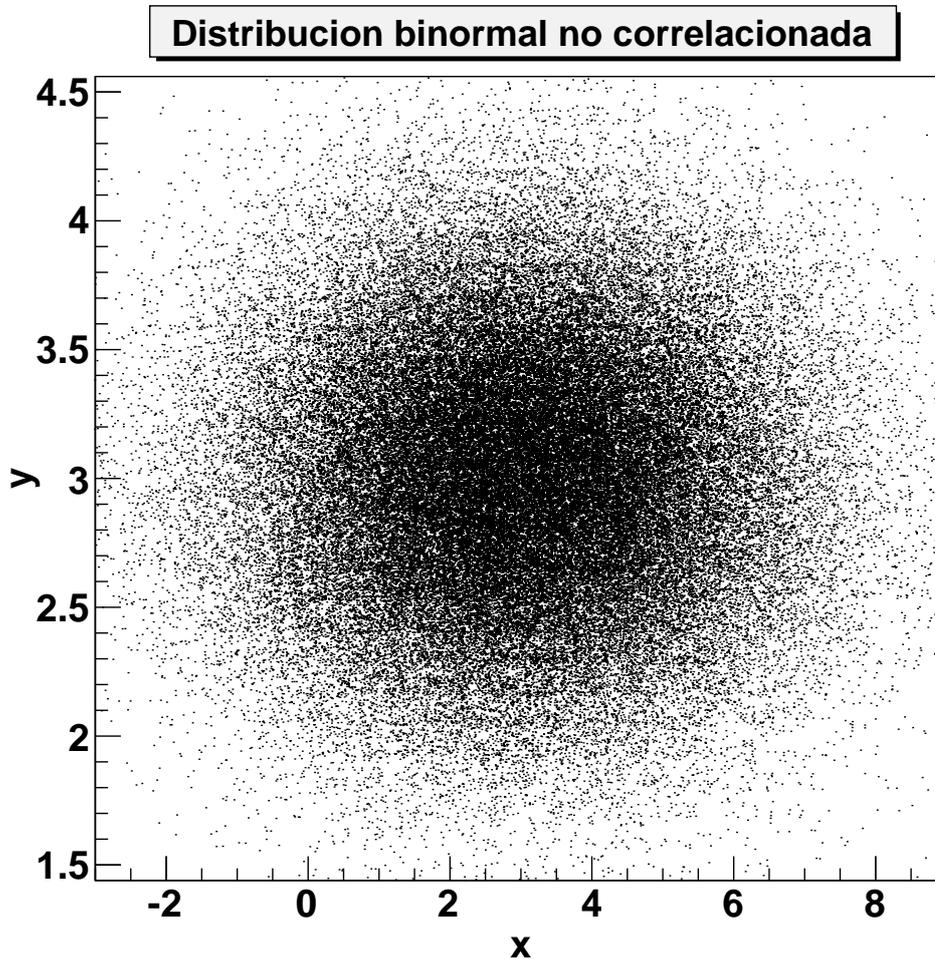
**Enunciado** Generar 100.000 sucesos de valores  $(x, y)$  distribuidos según una distribución binormal no correlacionada con:

$$\mu_{x_1} = 3, \quad \sigma_{x_1} = 2, \quad \mu_{x_2} = 3, \quad \sigma_{x_2} = 0.5$$

- Representar los puntos en un histograma de dos dimensiones y calcular la matriz de covarianza y el coeficiente de correlación.
  - Considerar la función  $f(x, y) = 3x + 5y$ . Calcular la varianza de  $f(x, y)$  directamente usando propagación de errores y también a partir de los datos.
  - Rotar ahora los sucesos respecto al centro de la distribución, no del origen, un ángulo  $\theta = 30^\circ$  y repetir los pasos anteriores.
  - A partir de los datos correlacionados estimar el ángulo  $\theta$  y los valores de las  $\sigma_u, \sigma_v$ .
  - Repetir todos los apartados anteriores para un ángulo  $\theta = -30^\circ$ .
- a)** Genero 100.000 sucesos aleatorios  $(x,y)$  con la función TRandom1-Gaus(mean,sigma) con los parámetros del enunciado. Si representamos los puntos en un histograma 2D, se obtiene la siguiente nube centrada en  $(3,3)$ :



Si en la figura anterior hacemos zoom en el eje vertical (zoom x4, el cociente entre sigmas), obtenemos una nube con simetría circular, que muestra la aleatoriedad e independencia de las variables:



La media de los datos es:

$$\bar{x} = 3.00711$$

$$\bar{y} = 3.00087$$

que son consistentes con los valores del enunciado. Se obtiene la siguiente matriz de covarianza:

$$V(x, y) = \begin{pmatrix} 3.96449 & 0.0112171 \\ 0.0112171 & 0.248034 \end{pmatrix}$$

El coeficiente de correlación es:

$$\rho = 0.0113118$$

Se comprueba que las variables  $x$  e  $y$  están descorrelacionadas (la pdf conjunta está factorizada), dado que el coeficiente de correlacion es cero y se distribuyen normalmente (en caso de pdf gaussianas, no hay posibilidad de que el coeficiente de correlación sea cero pero las variables estén relacionadas por una expresión no lineal, como sucedía en el caso de la parábola simétrica en los problemas del tema anterior).

Aparte, se comprueba que los elementos diagonales son muy cercanos a los valores verdaderos  $\sigma_x^2 = 4$ ,  $\sigma_y^2 = 0.25$  dado el alto valor de  $N$  (nos acercamos más a los valores esperados teóricos según aumenta  $N$ , por la ley de los grandes números), y los diagonales son cercanos a 0 ( $cov(x, y) = 0$  al ser variables independientes).

**b)** Si calculamos el valor esperado de la función  $f(x, y)$  y su varianza a partir de los valores del enunciado (mediante propagación de errores):

$$E[f(x, y)] = f(\mu_x, \mu_y) = 24$$

$$V(f) = (3 * \sigma_x)^2 + (5 * \sigma_y)^2 = 42.25$$

Si lo calculamos a partir de las datos pseudoaleatorios generados, obtenemos valores consistentes, con poca desviación respecto a los esperados dado el alto valor de  $N$ :

$$\bar{f} = E[f(x, y)] = f(\bar{x}, \bar{y}) = 24.0257$$

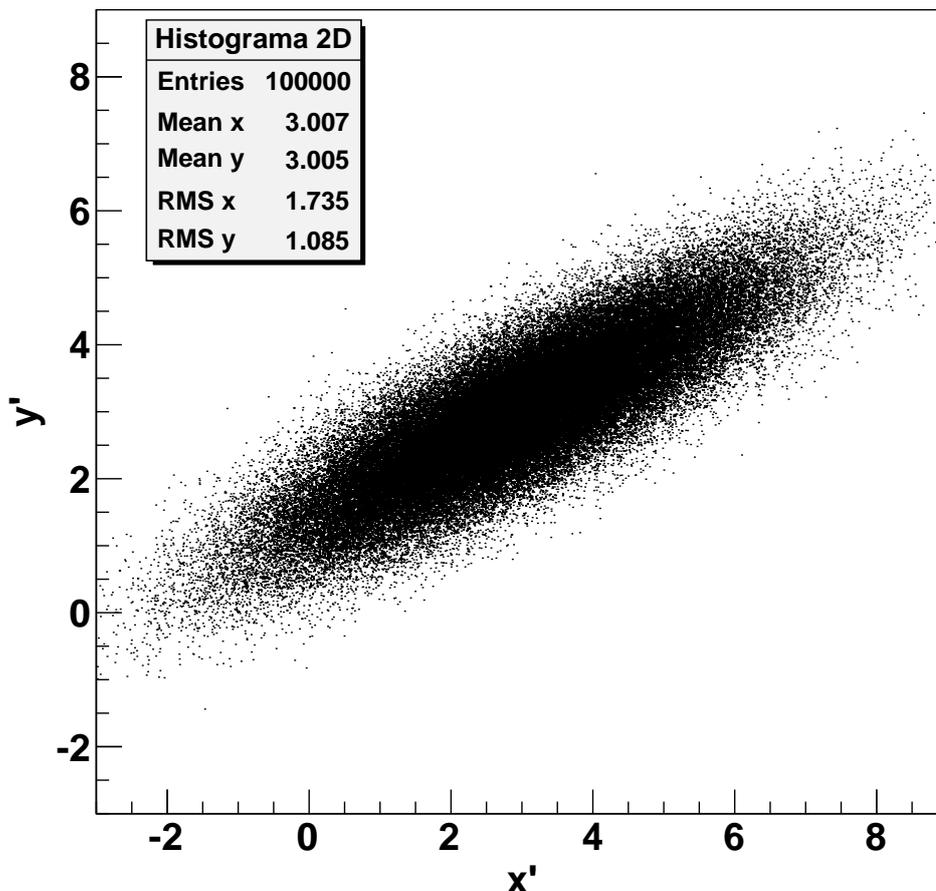
$$V(f) = E[(f(x, y) - \bar{f})^2] = 42.0481$$

**c)** Realizamos una rotación de  $\theta = +30^\circ$  sobre los sucesos  $(x, y)$  alrededor del punto  $(\mu_x, \mu_y)$ , con la que las nuevas variables aleatorias  $(x', y')$  serán:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x - \mu_x \\ y - \mu_y \end{pmatrix} + \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix} \quad (1)$$

Se obtiene el siguiente histograma:

## Distribucion binormal correlacionada



La media de los datos es:

$$\bar{x}' = 3.00572$$

$$\bar{y}' = 3.00431$$

que son consistentes con los valores del enunciado. Se obtiene la siguiente matriz de covarianza:

$$V(x', y') = \begin{pmatrix} 3.03056 & 1.61205 \\ 1.61205 & 1.18196 \end{pmatrix}$$

El coeficiente de correlación es:

$$\rho = 0.851757$$

Se comprueba que las variables  $x'$  e  $y'$  están en este caso correlacionadas, dado que el coeficiente de correlación es no nulo ( $\rho \neq 0$ ).

Aparte, se comprueba que los elementos diagonales  $V(x')$  y  $V(y')$  son muy parecidos a los valores  $\sigma_x'^2$ ,  $\sigma_y'^2$  calculados mediante propagación de errores (aplicada en la ecuación 1):

$$\begin{aligned} \sigma_x'^2 &= (\cos\theta\sigma_x)^2 + (\sin\theta\sigma_y)^2 = 3.0625 \\ \sigma_y'^2 &= (\sin\theta\sigma_x)^2 + (\cos\theta\sigma_y)^2 = 1.1875 \end{aligned} \quad (2)$$

Si calculamos el valor esperado de la función  $f'(x', y')$ , se obtiene:

$$E[f'(x', y')] = f'(\mu_x', \mu_y') = 24$$

Para calcular su varianza mediante propagación de errores, hay que desarrollar f:

$$\begin{aligned} f'(x', y') &= f'(x, y) = 3(\cos\theta(x - \mu_x) - \sin\theta(y - \mu_y)) + 5(\sin\theta(x - \mu_x) + \cos\theta(y - \mu_y)) \\ &= (x - \mu_x) * (3 * \cos\theta + 5 * \sin\theta) + (y - \mu_y) * (5 * \cos\theta - 3 * \sin\theta) \end{aligned}$$

La varianza será:

$$V(f') = ((3 * \cos\theta + 5 * \sin\theta)\sigma_x)^2 + ((5 * \cos\theta - 3 * \sin\theta) * \sigma_y)^2 = 105.964 \quad (3)$$

o bien desarrollando los cuadrados:

$$\begin{aligned} V(f') &= 3^2 * ((\cos\theta\sigma_x)^2 + (\sin\theta\sigma_y)^2) + 5^2 * ((\sin\theta\sigma_x)^2 + (\cos\theta\sigma_y)^2) \\ &\quad + 2 * 3 * 5 * \cos\theta * \sin\theta * (\sigma_x^2 - \sigma_y^2) \\ &= (3 * \sigma_x')^2 + (5 * \sigma_y')^2 + 3 * 5 * \tan(2\theta) * (\sigma_x'^2 - \sigma_y'^2) = 105.964 \end{aligned}$$

De esta última expresión, se deduce que:

$$\text{cov}(x', y') = \frac{1}{2} \tan(2\theta) * (\sigma_x'^2 - \sigma_y'^2) \quad (4)$$

Si calculamos  $f'$  y su varianza a partir de las datos pseudoaleatorios generados, obtenemos valores consistentes, con muy poca desviación respecto a los esperados por propagación de errores:

$$\bar{f}' = 24.0387$$

$$V(f') = 105.186$$

**d)** A partir de la ecuación 4, despejamos y calculamos el ángulo de rotación:

$$\theta = \frac{1}{2} \text{atan} \left( \frac{2 * \text{cov}(x', y')}{\sigma_x'^2 - \sigma_y'^2} \right) \quad (5)$$

o expresiones equivalentes:

$$\theta = \frac{1}{2} \text{asin} \left( \frac{2 * \text{cov}(x', y')}{\sigma_x'^2 - \sigma_y'^2} \right) \quad (6)$$

$$\theta = \frac{1}{2} \text{acos} \left( \frac{\sigma_x'^2 - \sigma_y'^2}{\sigma_x'^2 + \sigma_y'^2} \right) \quad (7)$$

Si sustituimos los valores obtenidos a partir de la generación pseudoaleatoria en la ecuación 5, queda un valor muy cercano a 30°:

$$\theta = 30.0857^\circ$$

Por último, podemos calcular  $\sigma_u \equiv \sigma_x$  y  $\sigma_v \equiv \sigma_y$  resolviendo el siguiente sistema inhomogéneo (derivado de la ecuación 2) por el método de Cramer o bien calculando la inversa de la matriz:

$$\begin{pmatrix} \cos^2\theta & \sin^2\theta \\ \sin^2\theta & \cos^2\theta \end{pmatrix} * \begin{pmatrix} \sigma_x'^2 \\ \sigma_y'^2 \end{pmatrix} = \begin{pmatrix} \sigma_x'^2 \\ \sigma_y'^2 \end{pmatrix} \quad (8)$$

El determinante es:

$$\Delta = \cos^4\theta - \sin^4\theta = (\cos^2\theta - \sin^2\theta) * (\cos^2\theta + \sin^2\theta) = (\cos^2\theta - \sin^2\theta) = \cos(2\theta) \quad (9)$$

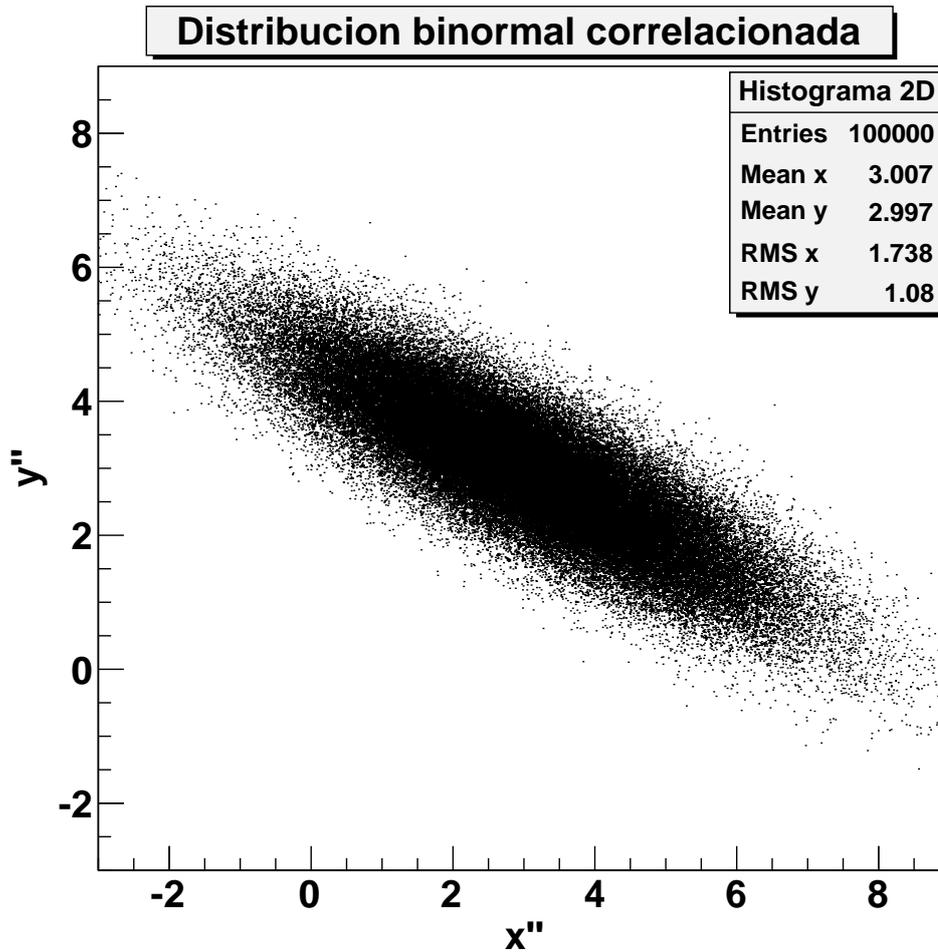
$$\sigma_x^2 = \frac{\sigma_x'^2 * \cos^2\theta * -\sin^2\theta * \sigma_y'^2}{\cos(2\theta)} \quad (10)$$

$$\sigma_y^2 = \frac{\cos^2\theta * \sigma_y'^2 - \sin^2\theta * \sigma_x'^2}{\cos(2\theta)} \quad (11)$$

Sustituyendo las varianzas  $\sigma_x'^2$  y  $\sigma_y'^2$  (a partir de los datos rotados) y el valor de  $\theta$  calculado según la ecuación 5, se obtiene valores consistentes con los verdaderos:

$$\begin{aligned}\sigma_x &= 1.9911 \\ \sigma_y &= 0.498022\end{aligned}$$

e) Si repetimos ahora todo el proceso, pero con  $\theta = -30^\circ$ , se obtiene el siguiente histograma:



La media de los datos es:

$$\begin{aligned}\bar{x}'' &= 3.00659 \\ \bar{y}'' &= 2.9972\end{aligned}$$

que son consistentes con los valores del enunciado. Se obtiene la siguiente matriz de covarianza:

$$V(x'', y'') = \begin{pmatrix} 3.04019 & -1.60649 \\ -1.60649 & 1.17233 \end{pmatrix}$$

El coeficiente de correlación es:

$$\rho = -0.850947$$

Se comprueba que las variables están también en este caso correlacionadas, dado que el coeficiente de correlación es no nulo ( $\rho \neq 0$ ). En comparación con el caso c), los valores de la diagonal de la matriz de covarianza son similares, mientras que en los no diagonales hay un cambio de signo debido al cambio de signo del ángulo de rotación (pendiente).

Aparte, se comprueba que los elementos diagonales  $V(x'')$  y  $V(y''y)$  son muy parecidos a los valores  $\sigma_x''^2$ ,  $\sigma_y''^2$  calculados mediante propagación de errores (aplicada en la ecuación 1):

$$\begin{aligned}\sigma_x''^2 &= 3.0625 \\ \sigma_y''^2 &= 1.1875\end{aligned}\tag{12}$$

Si calculamos el valor esperado de la función  $f''(x'', y'')$ , se obtiene:

$$E[f''(x'', y'')] = f''(\mu_x'', \mu_y'') = 24$$

La varianza será:

$$V(f'') = (3 * \sigma_x')^2 + (5 * \sigma_y')^2 + 3 * 5 * \tan(2\theta) * (\sigma_x'^2 - \sigma_y'^2) = 8.53607$$

Si calculamos  $f''$  y su varianza a partir de los datos pseudoaleatorios generados, obtenemos valores consistentes, con poca desviación respecto a los esperados por propagación de errores:

$$\bar{f}'' = 24.0058$$

$$V(f'') = 8.47528$$

Cabe destacar en este caso que la varianza es mucho menor que en el caso c). Para comprender esto, podemos dibujar la función  $f$  (eje  $z$ ) en función de  $x$  e  $y$ . La función viene representada por un plano, y la zona más probable de  $f$  viene seleccionada por la nube de probabilidad de la distribución binormal (con forma elíptica). En el caso de  $+30^\circ$ , la nube proyectada sobre el plano selecciona una región de  $f$  con bastante pendiente, con lo que la varianza es grande. En el caso de  $-30^\circ$ , esta región selecciona valores de  $f$  que apenas varían (varianza menor), como si se moviese por una línea de nivel en la superficie, sin cambiar la pendiente. Esta visualización intuitiva se corrobora matemáticamente a partir de la ecuación 3 ó 4.

A partir de la ecuación 5, calculamos el ángulo de rotación:

$$\theta = -29.9143^\circ$$

Por último, calculamos  $\sigma_u \equiv \sigma_x$  y  $\sigma_v \equiv \sigma_y$  resolviendo el sistema 8, y comprobamos que se obtienen valores compatibles con lo esperado:

$$\begin{aligned}\sigma_x &= 1.9911 \\ \sigma_y &= 0.498022\end{aligned}$$

# Problemas de Métodos Estadísticos

## Tema 5.- Teorema del Límite Central

### 1.- Generador de números aleatorios gaussianos.

- a) Dada una variable aleatoria  $x$  distribuida uniformemente en el intervalo  $[0,1]$ , comprobar que la variable  $s = x_1 + x_2 + \dots + x_n$  se distribuye como una variable gaussiana con:

$$\mu = \frac{N}{2} \quad \sigma^2 = \frac{N}{12}$$

Tomar  $N = 12$ , calcular los valores de  $\bar{x}$  y  $s^2$ , y comparar con los valores teóricos anteriores.

- b) Esto nos permite construir un generador de números gaussianos  $N(0,1)$  a partir de un generador uniforme construyendo la variable:

$$z = \frac{\sum_{i=1}^N x_i - \frac{N}{2}}{\sqrt{\frac{N}{12}}}$$

donde  $x_i$  es una variable uniforme entre  $[0,1]$ . Construir un histograma con 1000 entradas de dicha variable tomando  $N = 12$ . Calcular la media y la varianza de la distribución y comprobar el Teorema del Límite Central.

- c) Demostrar que con dicho generador es imposible obtener valores de  $z$  fuera del intervalo  $[-\sqrt{3N}, \sqrt{3N}]$ :

### 2.- Estudio de la distribución uniforme

Mediante un generador de números aleatorios uniforme, generar 1000 números distribuidos en el intervalo  $[0,1]$  y construir un histograma dividiendo dicho intervalo en 10 partes iguales.

- a) Calcular el valor medio y la desviación típica y comparar con lo que esperaríamos de una distribución uniforme:

$$\mu = \frac{a+b}{2} \quad \sigma^2 = \frac{(b-a)^2}{12}$$

- b) Calcular el error del valor medio según el Teorema del límite central y comprobar si el valor medio obtenido está dentro de los límites de error.
- c) Estudiar las fluctuaciones de cada *bin* y verificar que son las esperadas.

### 3.-Teorema del Límite Central

Para comprobar el Teorema del Límite Central estudiaremos la distribución de los valores medios del siguiente modo: generamos 10 números aleatorios distribuidos uniformemente en el intervalo  $[0,1]$  y calculamos su valor medio. Para esta nueva variable, repetiremos el experimento 10000 veces y calcularemos el valor medio y la desviación típica. Representaremos los valores en un histograma donde podéis elegir el tamaño del *bin*.

- a) Comprobar que los resultados son los esperados de acuerdo con el teorema del límite central.
- b) Repetir el experimento numérico tomando los valores medios cada 100 y cada 1000 números aleatorios.

### 4.- Teorema de Límite Central

Utilizando el programa que simula el tablero de Galton (ver problemas del Tema 4) visualizar como se forman distribuciones gaussianas utilizando para ello una probabilidad  $p = 0.5$  y aumentando progresivamente el número de filas del tablero.

# -Métodos estadísticos-

## Problemas Tema 5:

### Teorema del límite central

Fernando Hueso González\*

Valencia, 28 de noviembre de 2011

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

## 1. Generador de números aleatorios gaussiano

### Enunciado

- a) Dada una variable aleatoria  $x$  distribuida uniformemente en el intervalo  $[0, 1]$ , comprobar que la variable  $y = x_1 + x_2 + \dots + x_n$  se distribuye como una variable gaussiana con:

$$\mu_y = \frac{n}{2} \quad \sigma_y^2 = \frac{n}{12}$$

Tomar  $n = 12$ , calcular los valores de  $\bar{y}$  y  $s_y^2$ , y comparar con los valores teóricos anteriores.

- b) Esto nos permite construir un generador de números gaussiano  $G(0, 1)$  a partir de un generador uniforme construyendo la variable:

$$z = \frac{\sum_{i=1}^n x_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}}$$

donde  $x_i$  es una variable uniforme entre  $[0, 1]$ . Construir un histograma con  $N = 1000$  entradas de dicha variable tomando  $n = 12$ . Calcular la media y la varianza de la distribución y comprobar el Teorema del Límite Central.

- c) Demostrar que con dicho generador es imposible obtener valores de  $z$  fuera del intervalo  $[-\sqrt{3n}, \sqrt{3n}]$ .

---

\*ferhue#alumni.uv.es

## Resolución

a) La pdf uniforme entre 0 y 1 de una variable uniforme  $x$  y su valor esperado es:

$$f(x) = \begin{cases} 1 & \text{para } 0 \leq x \leq 1, \\ 0 & \text{para } x < 0 ; x > 1 \end{cases}$$

$$E[x] = \int_0^1 x dx = \frac{x^2}{2} \Big|_0^1 = \frac{1}{2}$$

Teniendo en cuenta que todas las  $x_i$  son variables uniformes, el valor esperado de  $y$  será:

$$\mu_y \equiv E[y] = E \left[ \sum_{i=1}^n x_i \right] = \sum_{i=1}^n E[x_i] = \sum_{i=1}^n \frac{1}{2} = \frac{n}{2}$$

La varianza será:

$$\sigma_y^2 \equiv V[y] = E[(y - \mu_y)^2] = E[y^2] - 2\mu_y E[y] + \mu_y^2 = E[y^2] - \mu_y^2$$

$$E[y^2] = E \left[ \sum_{i,j=1}^n x_i x_j \right] = \sum_{i,j=1}^n E[x_i x_j]$$

$$E[x_i x_j] = \begin{cases} E[x_i]E[x_j] = \frac{1}{4} & \text{para } i \neq j \\ \int_0^1 x_i^2 dx_i = \frac{1}{3} & \text{para } i = j \end{cases}$$

$$E[y^2] = \sum_{i=1}^n E[x_i^2] + \sum_{i \neq j}^n E[x_i x_j] = \frac{n}{3} + \frac{n^2-n}{4}$$

$$\sigma_y^2 = \frac{n}{3} + \frac{n^2-n}{4} - \frac{n^2}{4} = \frac{n}{3} - \frac{n}{4} = \frac{n}{12}$$

Para comprobar este desarrollo teórico, escogemos  $n = 12$  y generamos  $N = 1000$  valores de  $y$ , que resulta de sumar 12 números aleatorios generados con la función TRandom1->Rndm() de ROOT. Se obtiene el siguiente valor medio  $\bar{y}$  y varianza muestral  $s_y^2$ , que son compatibles con los valores teóricos ( $\mu$  y  $\sigma_y$ ):

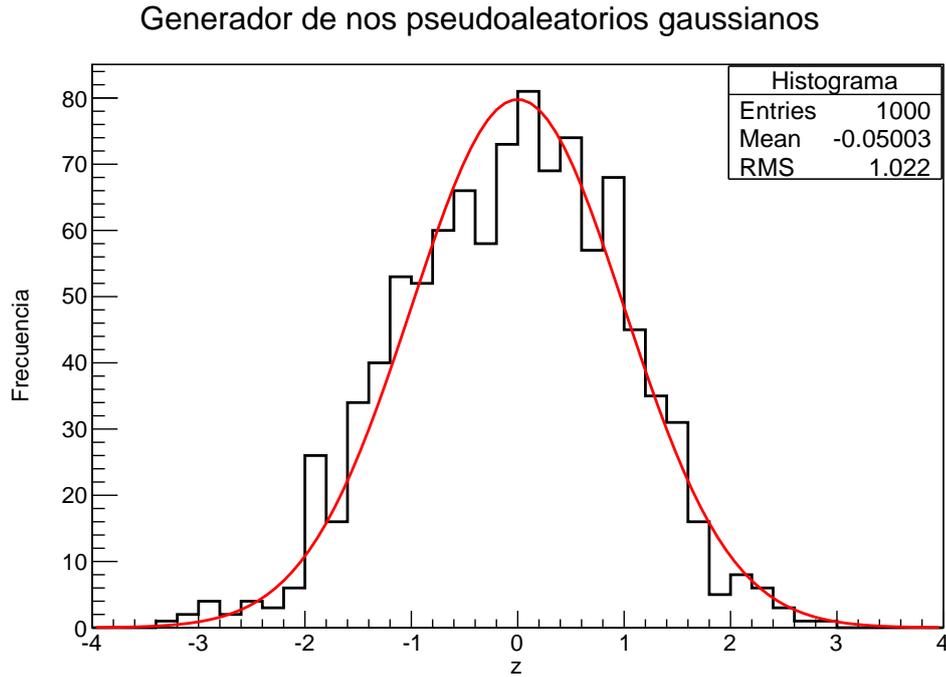
$$\begin{aligned} \mu &= 6 & \sigma_y &= 1 \\ \bar{y} &= 5,94997 & s_y &= 1,04633 \end{aligned}$$

Es fácil comprobar que aumentando  $N$  en el programa, más se acercan la media y varianza muestral a los valores teóricos, de acuerdo con la ley de los grandes números.

- b) Calculamos (a partir de los números aleatorios generados en el apartado anterior) la variable  $z$ . Se obtiene el siguiente valor medio  $\bar{z}$  y varianza muestral  $s_z^2$ , que son compatibles con los valores teóricos ( $\mu_z$  y  $\sigma_z^2$ ) de acuerdo con el teorema del límite central (la variable  $z$  tiende a una distribución normal estándar centrada en 0 y con varianza unidad):

$$\begin{aligned} \mu_z &= 0 & \sigma_z^2 &= 1 \\ \bar{z} &= -0,0500335 & s_z^2 &= 1,04633 \end{aligned}$$

Si construimos el histograma con  $n_{bin} = 40$  bins entre  $low = -4$  y  $up = 4$ , se comprueba visualmente dicho teorema:



Negro=Números pseudoaleatorios

Rojo=  $G(0, 1, x) = \frac{c}{1\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-0)^2}{1^2}}$

$$k = N * (up - low) / n_{bin} = 1000 * 8 / 40 = 200 \text{ (área de la distribución)}$$

$$RMS \text{ es equivalente a } \sqrt{s_z^2} = 1,02290$$

Vemos que las fluctuaciones en la mayoría de canales son del orden de la raíz de cuentas en el bin. Para  $N$  tendiendo a  $\infty$ , la distribución límite es la distribución normal estándar (curva roja).

- c) Teniendo en cuenta que  $\frac{\text{Max}}{\text{Min}}(x_i) = \frac{1}{0}$ , se comprueba que:

$$\frac{\text{Max}}{\text{Min}}(z) = \frac{\sum_{i=1}^n \text{Max}_{\text{Min}}(x_i) - \frac{n}{2}}{\sqrt{\frac{n}{12}}} = \frac{\sum_{i=1}^n 1 - \frac{n}{2}}{\sqrt{\frac{n}{12}}} = \begin{cases} (n - \frac{n}{2}) / \sqrt{\frac{n}{12}} \\ (0 - \frac{n}{2}) / \sqrt{\frac{n}{12}} \end{cases} = \pm \frac{\frac{n}{2}}{\sqrt{\frac{n}{12}}} = \pm \sqrt{3n}$$

Por tanto,  $z \in [-\sqrt{3n}, \sqrt{3n}]$ , como queríamos demostrar.

## 2. Estudio de la distribución uniforme

**Enunciado** Mediante un generador de números aleatorios uniforme, generar 1000 números distribuidos en el intervalo  $[0, 1]$  y construir un histograma dividiendo dicho intervalo en 10 partes iguales.

- a) Calcular el valor medio y la desviación típica y comparar con lo que esperaríamos de una distribución uniforme en el intervalo  $[a, b]$ :

$$\mu = \frac{a + b}{2} \quad \sigma^2 = \frac{(b - a)^2}{12}$$

- b) Calcular el error del valor medio según el Teorema del límite central y comprobar si el valor medio obtenido está dentro de los límites de error.
- c) Estudiar las fluctuaciones de cada bin y verificar que son las esperadas.

### Resolución

- a) Generando con la función TRandom1->Rndm() de ROOT  $N = 1000$  números pseudoaleatorios entre 0 y 1, se obtiene el siguiente valor medio  $\bar{x}$  y varianza muestral  $s_x^2$ , compatibles con los valores esperados teóricamente ( $\mu$  y  $\sigma^2$ ):

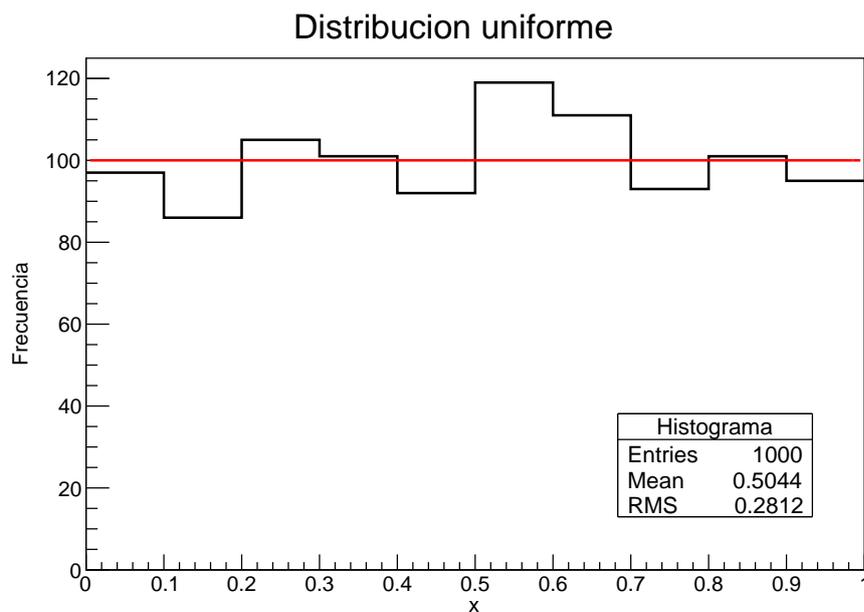
$$\begin{aligned} \mu &= 0,5 & \sigma^2 &= 0,0833333 \\ \bar{x} &= 0,504376 & s_x^2 &= 0,0791395 \end{aligned}$$

- b) Según el teorema del límite central, el error del valor medio (desviación estándar) es:

$$\sigma_{\bar{x}} = \sigma / \sqrt{N} \simeq 0,009$$

En este caso, comprobamos que  $\mu$  está a menos de 1 desviación estándar del valor medio muestral, con lo que podemos considerar que es compatible con lo esperado (comprendido dentro del margen de incertidumbre  $\mu \in [\bar{x} - \sigma_{\bar{x}}, \bar{x} + \sigma_{\bar{x}}]$ ).

- c) Representamos el histograma generado con 10 bins entre 0 y 1:



Negro=Números pseudoaleatorios  
Rojo=  $1000/10 = 100$  (frecuencia media esperada por bin)

$RMS$  es equivalente a  $\sqrt{s_x^2} = 0,281317$

A continuación analizamos las fluctuaciones en el número de cuentas  $r_k$  en cada bin  $k$ , que comparamos con la fluctuación esperada<sup>1</sup>:

k	Frecuencia $r_k$	$(r_k - 100)/\sigma(r_k)$
1	97	-0.3
2	86	-1.6
3	105	0.5
4	101	0.1
5	92	-0.9
6	119	1.9
7	111	1.1
8	93	-0.8
9	101	0.1
10	95	-0.5

Se comprueba que en la mayoría de canales (7 de 10), la fluctuación es menor a una desviación estándar (teóricamente, debería suceder esto en el 68% de los casos), y menor a dos desviaciones en su totalidad (teóricamente, 95%). Por tanto, las fluctuaciones son acordes con lo esperado.

---

<sup>1</sup>Utilizamos la fórmula exacta para el error del número de cuentas  $r_k$  en el bin  $k$ , que es  $\sigma(r_k) = \sqrt{r_k(1 - \frac{r_k}{N})}$ . Se podría aproximar por  $\sigma(r_k) \sim \sqrt{r_k}$  dado que el factor  $r_k/N \sim 1/n_{bin} = 1/10 \ll 1$ .

### 3. Teorema del Límite Central

**Enunciado** Para comprobar el Teorema del Límite Central estudiaremos la distribución de los valores medios del siguiente modo: generamos 10 números aleatorios distribuidos uniformemente en el intervalo  $[0, 1]$  y calculamos su valor medio. Para esta nueva variable, repetiremos el experimento 10000 veces y calcularemos el valor medio y la desviación típica. Representaremos los valores en un histograma donde podéis elegir el tamaño del *bin*.

- Comprobar que los resultados son los esperados de acuerdo con el teorema del límite central.
- Repetir el experimento numérico tomando los valores medios cada 100 y cada 1000 números aleatorios.

#### Resolución

- Generando con la función `TRandom1->Rndm()` de ROOT  $N = 10000$  medias  $y = \bar{x}$  de  $n = 10$  números pseudoaleatorios  $x$  entre 0 y 1 (distribución uniforme, de media  $\mu_x = 1/2$  y varianza  $\sigma_x^2 = 1/12$ ), se obtiene el siguiente valor medio  $\bar{y}$  y desviación estándar muestral  $s_y$ , compatibles con los valores esperados teóricamente ( $\mu_y = \mu_x$  y  $\sigma_y = \sigma_x/\sqrt{n}$  por el teorema del límite central):

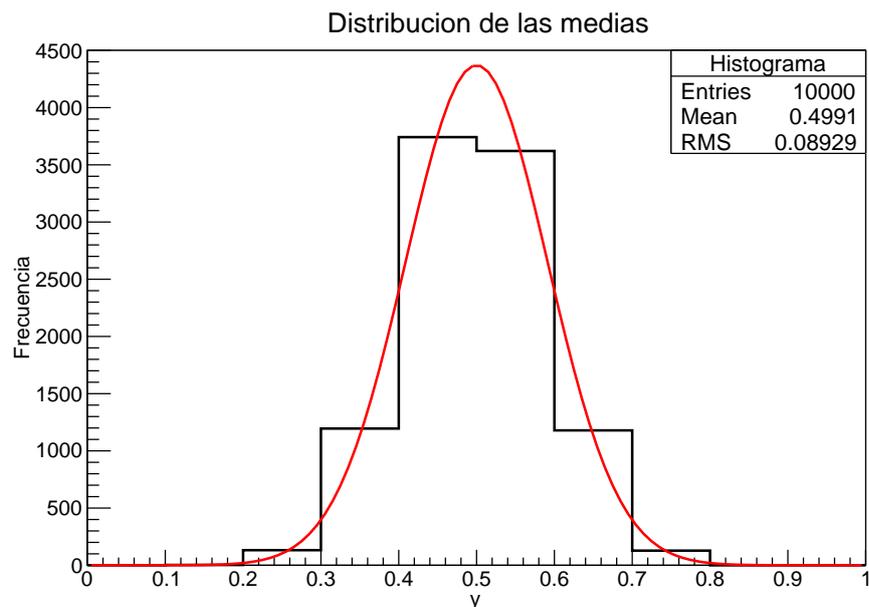
$$\begin{aligned} \mu_y &= 0,5 & \sigma_y &= 0,0912871 \\ \bar{y} &= 0,499098 & s_y &= 0,0892986 \end{aligned}$$

Según el teorema del límite central, el error del valor medio  $\bar{y}$  (desviación estándar) es:

$$\sigma_{\bar{y}} = \sigma_y/\sqrt{N} = \sigma_x/\sqrt{Nn} \simeq 0,0009$$

En este caso, comprobamos que  $\mu_y$  está a menos de 1 desviación estándar del valor medio muestral, con lo que podemos considerar que es compatible con lo esperado (comprendido dentro del margen de incertidumbre  $\mu_y \in [\bar{y} - \sigma_{\bar{y}}, \bar{y} + \sigma_{\bar{y}}]$ ).

Representamos el histograma generado con 10 bins entre 0 y 1:



Negro=Números pseudoaleatorios

$$\text{Rojo} = G(0.5, \sigma_y, y) = \frac{c}{\sqrt{2\pi \frac{1}{12n}}} e^{-\frac{1}{2} \frac{(y-0.5)^2}{1^2}}$$

$$c = N/n_{bin} = 10000/10 = 1000 \text{ (área de la distribución)}$$

Cabe señalar que el valor *RMS* de la leyenda se corresponde con el valor  $s_y$  detallado anteriormente (anchura de la distribución).

A continuación analizamos las fluctuaciones en el número de cuentas  $r_k$  en cada bin  $k$ , que comparamos con la fluctuación esperada.

k	Frecuencia $r_k$	$(r_k - G(0.5, \sqrt{\frac{1}{12*n}}, (k - 0.5)/n_{bin}))/\sigma(r_k)$
1	0	-0.2
2	3	0.1
3	132	2.6
4	1195	1.9
5	3742	-0.4
6	3621	-2.9
7	1179	1.4
8	128	2.2
9	0	-1.7
10	0	-0.2

En los bins donde  $r_k = 0$ , en la tercera columna se intercambia el denominador  $\sigma(r_k)$  por  $\sigma[G(0.5, \sqrt{\frac{1}{12*n}}, (k - 0.5)/n_{bin})]$  para evitar un infinito.

Las fluctuaciones son más significativas (varios bins a más de  $2\sigma$ , menos del 50% a menos de  $1\sigma$ ) que en el caso de la distribución uniforme, dado que elegimos pocos bins y que aproximamos el valor teórico por el valor de la gaussiana en el centro del bin. (En la distribución uniforme, el valor esperado era constante, con lo que ambos valores eran equivalentes y no era tan influyente el elegir más o menos bins). En este caso puede haber más desviaciones al haber una pendiente pronunciada, y se puede comprobar que las fluctuaciones se hacen más acordes a lo esperado si elegimos un mayor número de bins.

Con todo, podemos concluir que se cumple el teorema del límite central, y que la distribución límite a la que tienden las medias de variables aleatorias es una distribución gaussiana con los parámetros teóricos indicados anteriormente, que se han verificado satisfactoriamente con los datos pseudoaleatorios generados.

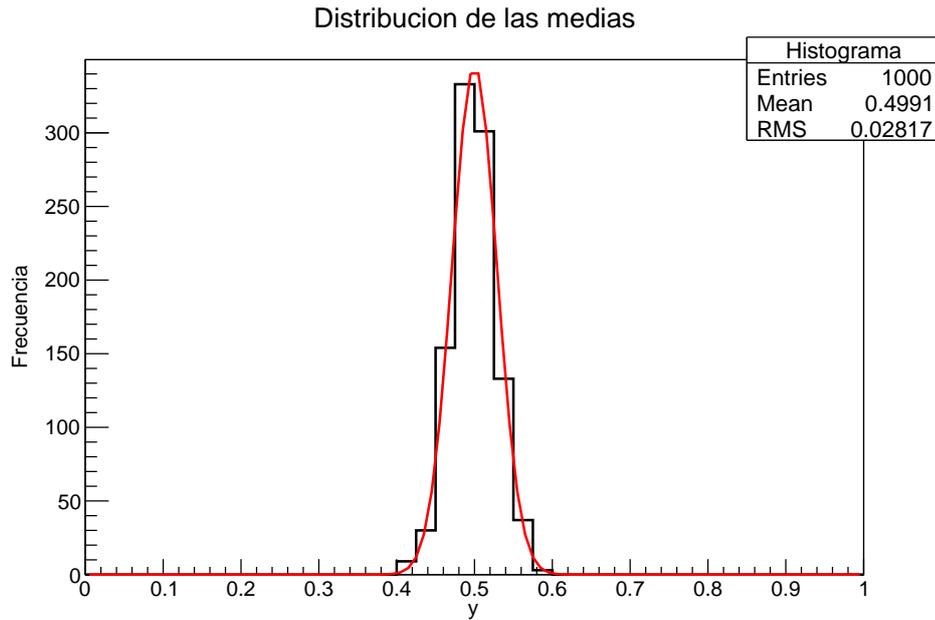
- b) Repitiendo el mismo procedimiento de a), pero con  $N = 1000$ ,  $n = 100$  (no generamos nuevos datos aleatorios, sino que reagrupamos los calculados en el apartado anterior), se obtiene el siguiente valor medio  $\bar{y}$  y desviación estándar muestral  $s_y$ , compatibles con los valores esperados (al haber reagrupado los datos de a), lógicamente la media y desviación son las mismas):

$$\begin{aligned} \mu_y &= 0,5 & \sigma_y &= 0,0288675 \\ \bar{y} &= 0,499098 & s_y &= 0,0281791 \end{aligned}$$

$$\sigma_{\bar{y}} \simeq 0,0009$$

En este caso, comprobamos que  $\mu_y$  está a menos de 1 desviación estándar del valor medio muestral, con lo que podemos considerar que es compatible con lo esperado (comprendido dentro del margen de incertidumbre  $\mu_y \in [\bar{y} - \sigma_{\bar{y}}, \bar{y} + \sigma_{\bar{y}}]$ ).

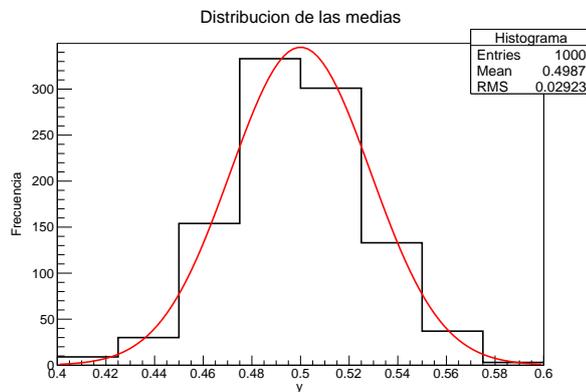
Representamos el histograma generado con 40 bins entre 0 y 1:



Negro=Números pseudoaleatorios  
 Rojo=  $G(0.5, \sigma_y, y)$

Se comprueba que la distribución es mucho más picada que en el caso a), al ser  $n$  mayor. Analizamos las fluctuaciones en el número de cuentas  $r_k$  en cada bin  $k$  en comparación con la fluctuación esperada (omitimos bins sin cuentas).

k	Frecuencia $r_k$	$(r_k - G(0.5, \sigma_y, (k-0.5)/n_{bin}))/\sigma(r_k)$
17	9	1.8
18	30	-0.6
19	154	0.5
20	333	1.2
21	301	-0.9
22	133	-1.5
23	37	0.6
24	3	-0.3



Zoom del histograma anterior

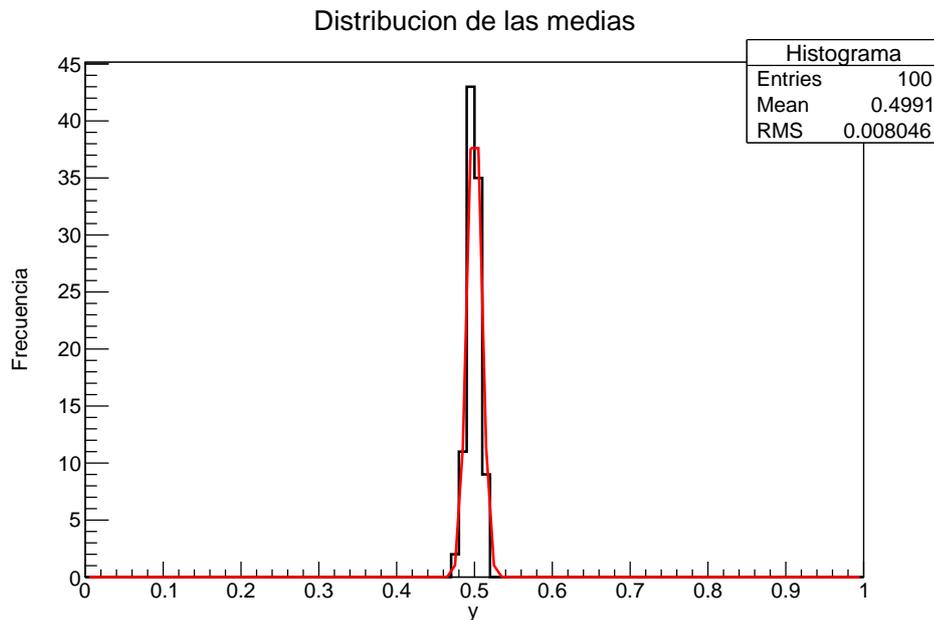
Repitiendo el mismo procedimiento de a), pero con  $N = 100$ ,  $n = 1000$ , se obtiene el siguiente valor medio  $\bar{y}$  y desviación estándar muestral  $s_y$ , compatibles con los valores esperados:

$$\begin{aligned} \mu_y &= 0,5 & \sigma_y &= 0,00912871 \\ \bar{y} &= 0,499098 & s_y &= 0,00808657 \end{aligned}$$

$$\sigma_{\bar{y}} \simeq 0,00009$$

En este caso, comprobamos que  $\mu_y$  está a menos de 1 desviación estándar del valor medio muestral, con lo que podemos considerar que es compatible con lo esperado (comprendido dentro del margen de incertidumbre  $\mu_y \in [\bar{y} - \sigma_{\bar{y}}, \bar{y} + \sigma_{\bar{y}}]$ ).

Representamos el histograma generado con 100 bins entre 0 y 1:

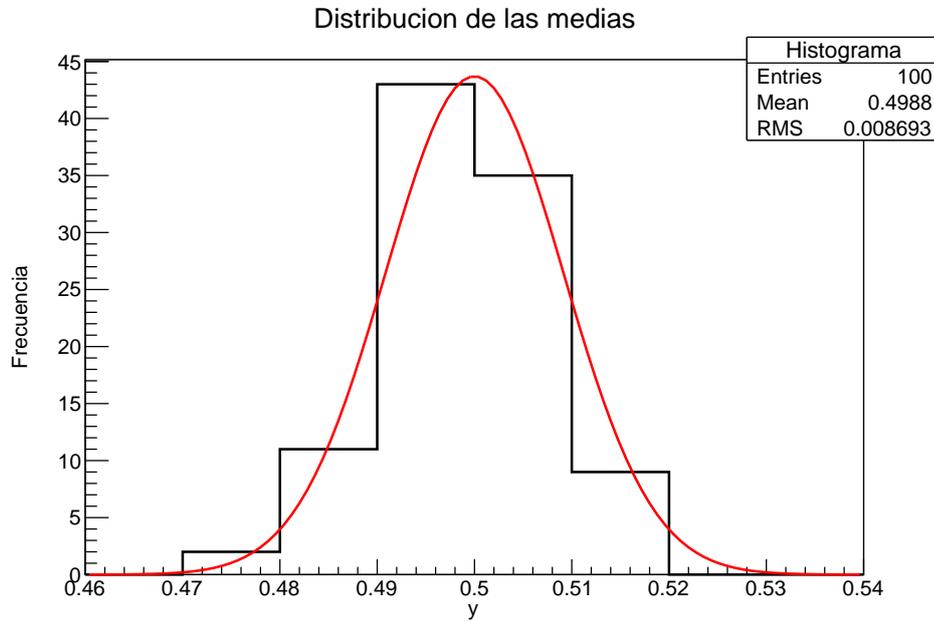


Negro=Números pseudoaleatorios  
 Rojo=  $G(0.5, \sigma_y, y)$

Se comprueba que la distribución es mucho más picada que en el caso a) y algo más que en b), al ser  $n$  mayor. Como  $N$  es pequeño (pocos experimentos/bolas), la desviación visual teoría-experimento en algunos bins es mayor. Analizamos las fluctuaciones en el número de cuentas  $r_k$  en cada bin  $k$  en comparación con la fluctuación esperada (omitimos bins sin cuentas).

k	Frecuencia $r_k$	$(r_k - G(0.5, \sigma_y, (k-0.5)/n_{bin}))/\sigma(r_k)$
48	2	0.7
49	11	-0.1
50	43	1.1
51	35	-0.5
52	9	-0.8

Las fluctuaciones en cada bin de los seleccionados son del orden de lo esperado, aunque debemos recordar que hemos seleccionado sólo algunos canales (excluidos aquellos donde  $r_k$  experimental es 0 y el esperado tiende a 0), con lo que si para comprobar rigurosamente si las las fluctuaciones respetan la estadística, deberíamos hacer el cómputo global de todos los bins (del 1 al 100).



Zoom del histograma anterior

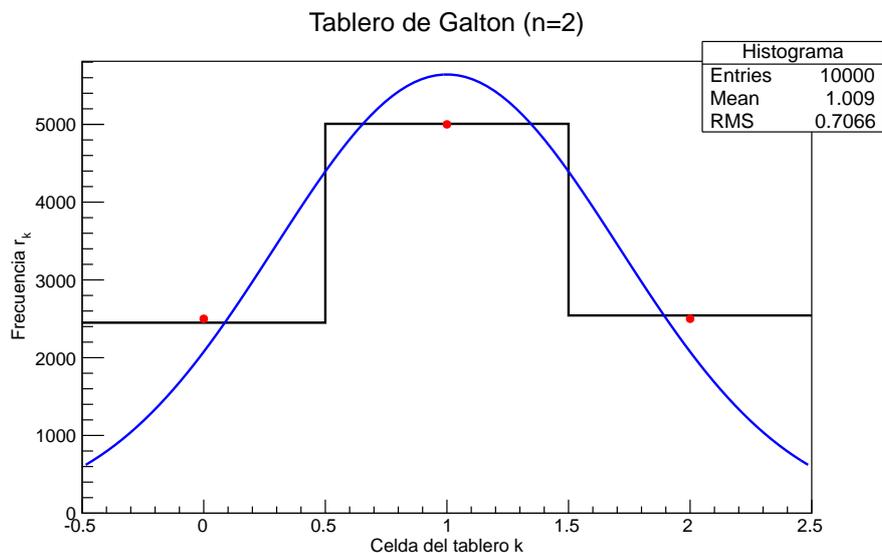
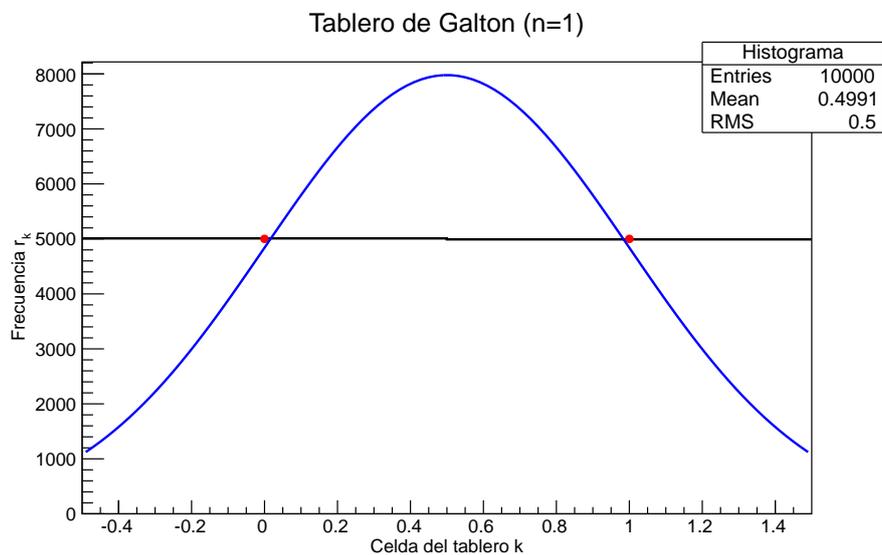
Con todo, podemos concluir que se cumple el teorema del límite central, pues hemos verificado su validez para diferentes valores de  $n$ , y en todos los casos hay acuerdo teoría-experimento en la media, el error de valor medio. Además, se comprueba que la distribución límite a la que tienden las medias de variables aleatorias es una distribución gaussiana y las fluctuaciones en cada bin son acordes con lo esperado estadísticamente.

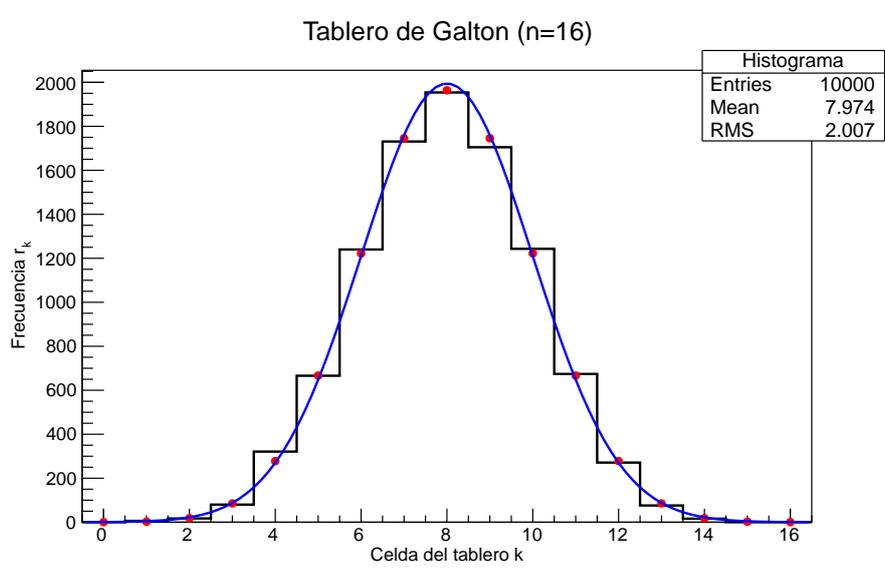
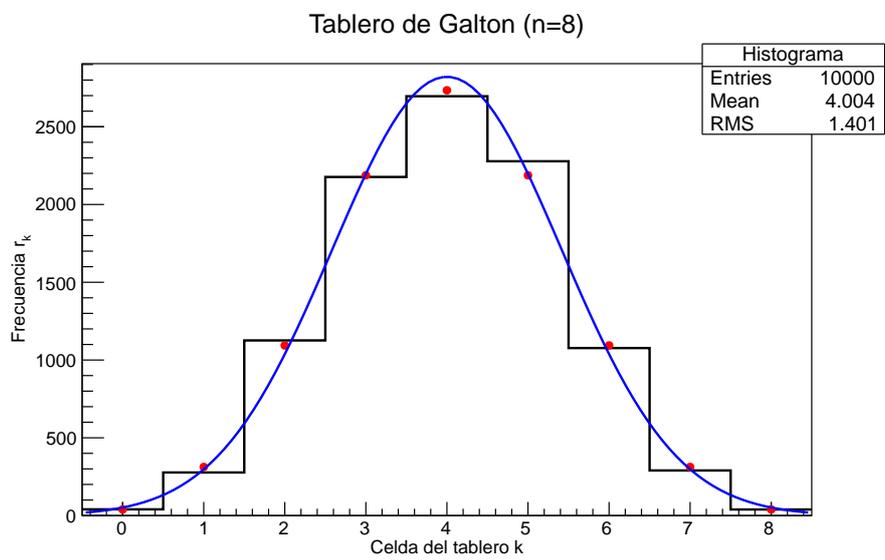
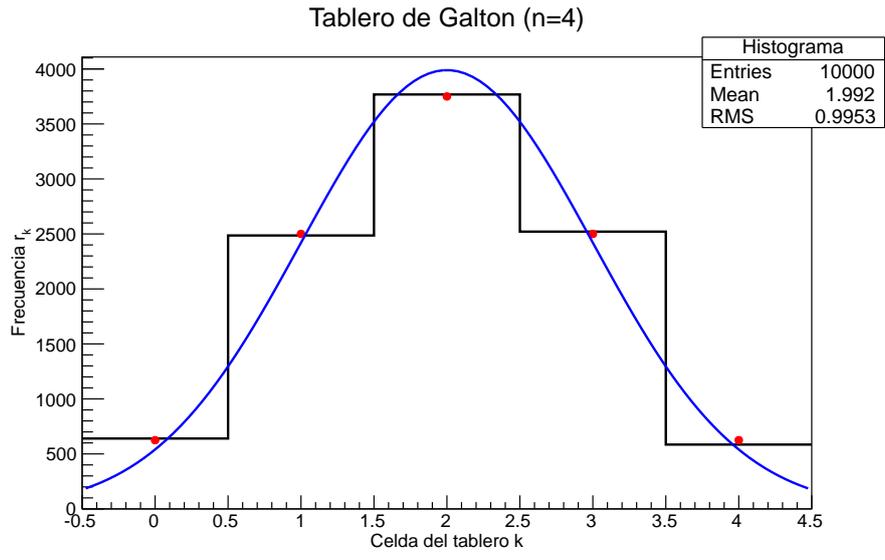
## 4. Teorema del Límite Central

**Enunciado** Utilizando el programa que simula el tablero de Galton (ver problemas del Tema 4) visualizar como se forman distribuciones gaussianas utilizando para ello una probabilidad  $p = 0.5$  y aumentando progresivamente el número de filas del tablero.

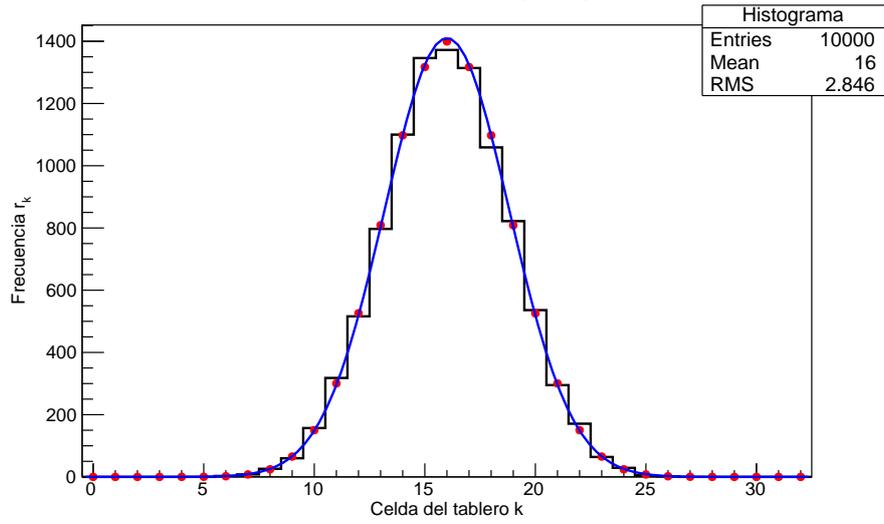
**Resolución** Dibujamos los diferentes histogramas desde  $n = 1$  filas hasta  $n = 512$ , duplicando en cada paso el valor de  $n$ . El número de bolas es  $N = 10.000$  y la probabilidad de que en una fila la bola caiga a la derecha es  $p = 0.5$ . En negro se presenta el histograma de datos pseudoaleatorios, en rojo el valor teórico según la distribución binomial, y en azul la curva gaussiana. Matemáticamente:

$$\begin{aligned} \text{Negro} &= \text{Números pseudoaleatorios} \\ \text{Rojo} &= N \binom{n}{k} p^k (1-p)^{n-k} \\ \text{Azul} &= N \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-\mu}{\sigma}\right)^2} \text{ con } \mu = np, \sigma = np(1-p) \end{aligned}$$

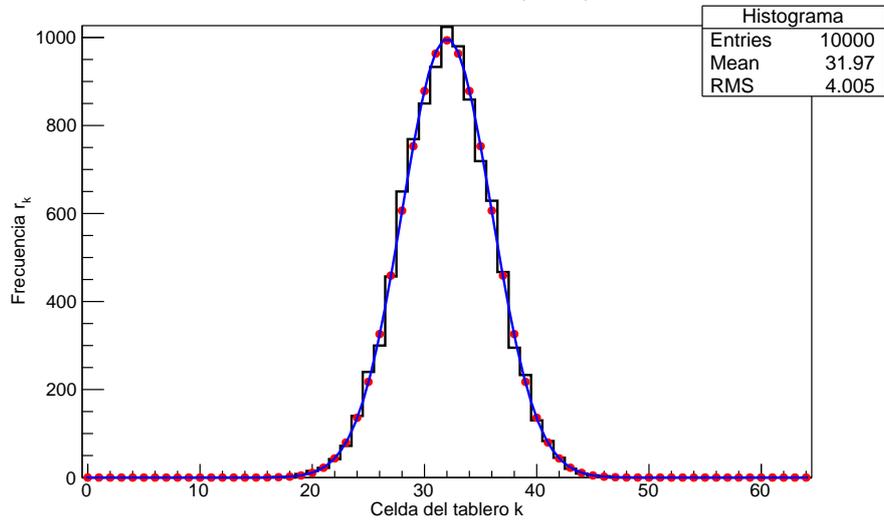




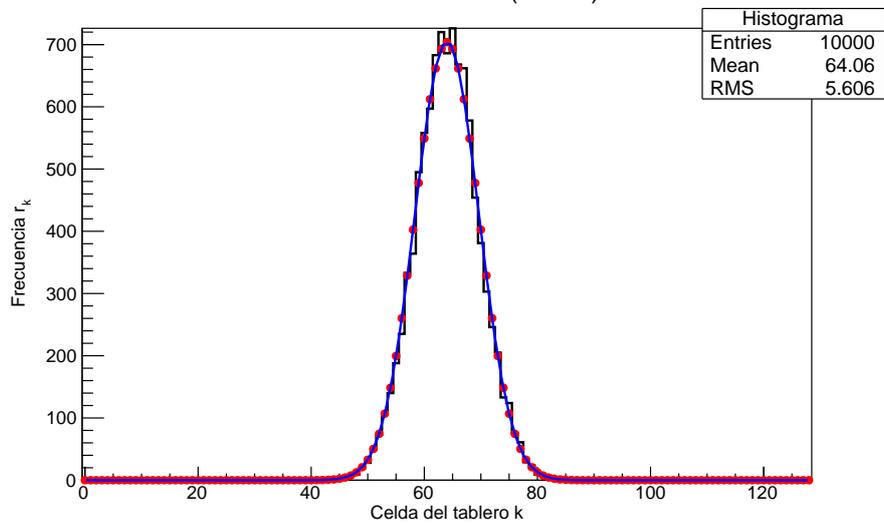
Tablero de Galton (n=32)

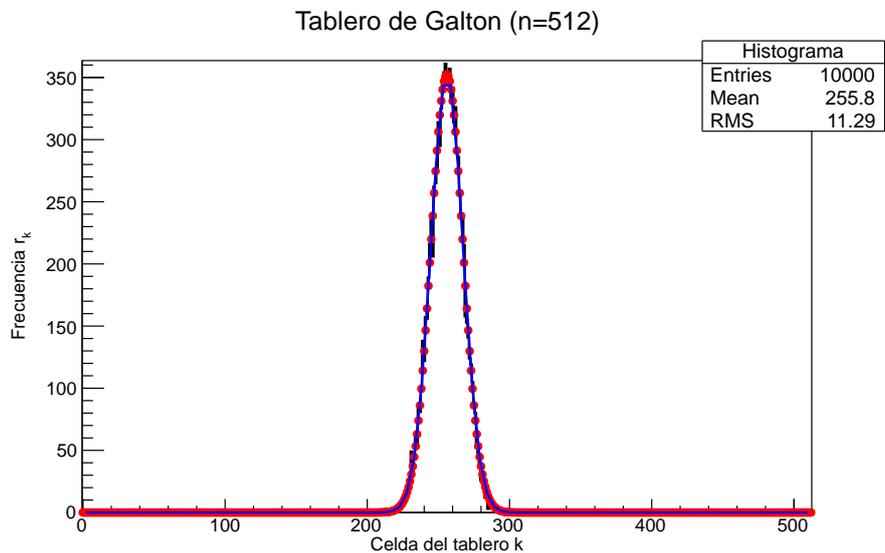
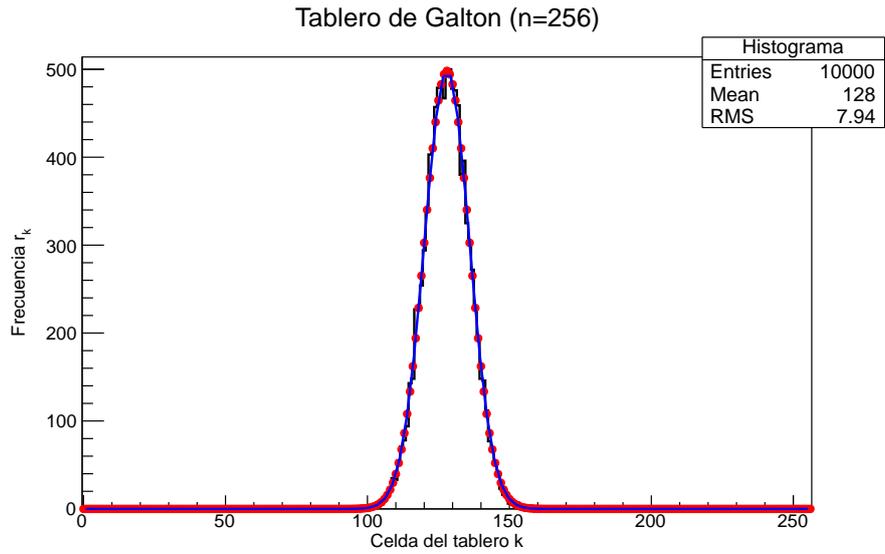


Tablero de Galton (n=64)



Tablero de Galton (n=128)





Por tanto, se comprueba cómo a mayor número de filas  $n$ , más se asemeja la distribución de datos a la distribución normal (gaussiana), la curva de color azul. En algunos canales, los datos pseudoaleatorios se desvían del valor predicho por la distribución binomial, lo cual se debe a fluctuaciones estadísticas en los canales, del orden de la raíz del número de cuentas en la mayoría de los mismos, como se ha estudiado en el problema 3. Es fácil comprobar, que si aumentamos el número de bolas en el experimento  $N$ , las desviaciones respecto a las curvas teóricas serán menores, especialmente si  $n$  es alto (el cociente  $n/N$  debe ser pequeño para que haya suficientes cuentas/bolas por bin/celda).

## Tema 6.- Métodos Monte Carlo. Generadores.

### 6.1.- La aguja del conde de Buffon

En 1777, el conde de Buffon encontró una relación entre el número  $\pi$  y la probabilidad de que una aguja de longitud  $l$ , lanzada sobre una superficie plana en la que se han trazado una serie de líneas paralelas separadas una distancia  $d$ , cruzara una de éstas líneas.

a) Demostrar analíticamente que la probabilidad viene dada por:

$$P = \frac{2l}{\pi d}$$

b) Programar el método de Buffon para la determinación del número  $\pi$ . ¿ Cuántos lanzamientos son necesarios para determinar  $\pi$  con un error de una milésima ?

### 6.2.- Propiedades de los generadores de números aleatorios

a).- Consideremos algunos ejemplos de generadores congruentes de la forma:

$$X_{n+1} = (a \cdot X_n + c) \bmod m$$

Construir la secuencia hasta completar periodo para los siguientes valores.

a)  $m = 31, c = 0, a = 7$  y  $X_0 = 1$ .

b)  $m = 31, c = 0, a = 3$  y  $X_0 = 1$ .

c)  $m = 32, c = 1, a = 5$  y  $X_0 = 1$ .

b).- Verificar que los casos (b) y (c) del ejercicio anterior satisfacen las condiciones de máximo periodo que vienen dadas por:

- Caso  $c = 0$ 
  - (i)  $m$  es un número primo.
  - (ii) Para todos los divisores primos,  $p$ , de  $(m - 1)$ , se cumple que  $a^{(m-1)/p} \bmod m \neq 1$
- Caso  $c \neq 0$ 
  - (i)  $c$  y  $m$  primos entre sí.
  - (ii)  $(a - 1)/p$  es un entero para cada divisor primo,  $p$ , de  $m$ .
  - (iii) Si  $m$  múltiplo de 4 entonces  $(a - 1)$  debe ser divisible por 4.

c).- Representar gráficamente en un plano todos los puntos definidos por las coordenadas  $(\xi_i, \xi_{i+1})$ ,  $i = 0, 1, \dots, 28$ , para los apartados (b) y (c) del ejercicio 1. ¿ Es lo que cabría esperar de un generador aleatorio ?

### 6.3 Aplicar un test de $\chi^2$ a un generador uniforme del siguiente modo:

a) Generar  $N = 10000$  números aleatorios y distribuirlos en un histograma de  $k = 50$  *bines*.

b) Si en cada *bin* tenemos  $N_j$  sucesos construir el estadístico:

$$T = \sum_{j=1}^k \frac{(N_j - \frac{N}{k})^2}{\frac{N}{k}} = \frac{k}{N} \sum_{j=1}^k (N_j - \frac{N}{k})^2$$

- c) La variable aleatoria anterior,  $T$ , sigue una distribución de  $\chi^2$  con  $\nu = 49$  grados de libertad. Para comprobarlo repetir el experimento 100 veces y representar los valores de  $T$  en un histograma superponiendo la distribución de  $\chi^2(T, \nu)$  correspondiente.
- d) Calcular el estadístico  $q$  de Kolmogorov-Smirnov entre el histograma y la función cumulativa de  $\chi^2(T, \nu)$  y mediante las tablas correspondientes estimar la probabilidad de obtener un valor mayor.
- e) ¿ Existen indicios de no uniformidad ?

**6.4.-** Mediante un generador uniforme construir una secuencia de 10000 números enteros en el rango (0, 9). Para ello multiplicar cada número aleatorio,  $\xi$ , por 10 y tomar la parte entera:

$$k = [10 * \xi]$$

Dividir la secuencia en subgrupos de diez números, y de cada subgrupo contar los  $r$  que son impares. Dibujar un histograma con los resultados obtenidos y comparar con una distribución binomial con  $N = 10$  y  $p = 0.5$ , mediante un test de  $\chi^2$ .

**6.5.-** Repetir el ejercicio anterior pero contando los números múltiplos de 3 y comparando con la distribución binomial correspondiente.

# -Métodos estadísticos-

## Problemas Tema 6:

### Métodos Monte Carlo. Generadores.

Fernando Hueso González\*

Valencia, 3 de diciembre de 2011

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

#### 6.1.- La aguja del conde de Buffon

**Enunciado** En 1777, el conde de Buffon encontró una relación entre el número  $\pi$  y la probabilidad de que una aguja de longitud  $l$ , lanzada sobre una superficie plana en la que se han trazado una serie de líneas paralelas separadas una distancia  $d$ , cruzara una de éstas líneas.

- a) Demostrar analíticamente que la probabilidad viene dada por:

$$P = \frac{2l}{\pi d}$$

- b) Programar el método de Buffon para la determinación del número  $\pi$ . ¿Cuántos lanzamientos son necesarios para determinar  $\pi$  con un error de una milésima?

#### Resolución

- a) Sea un plano infinito cubierto enteramente con rectas paralelas (ver figura 1a) equidistantes una distancia  $d$ . Sea una aguja de longitud  $l$ , tal que  $l < d$ . Sea  $y$  la distancia (en valor absoluto) desde el centro de la aguja a la línea más cercana a la misma. Sea  $\theta$  el ángulo (en valor absoluto) que forma la aguja con la horizontal (ver figura 1b). Por la propia definición,  $0 \leq y \leq d/2$ ,  $0 \leq \theta \leq \pi/2$ . La probabilidad diferencial  $dP_y$  de que el centro de la aguja caiga en el intervalo  $(y, y + dy)$  es:

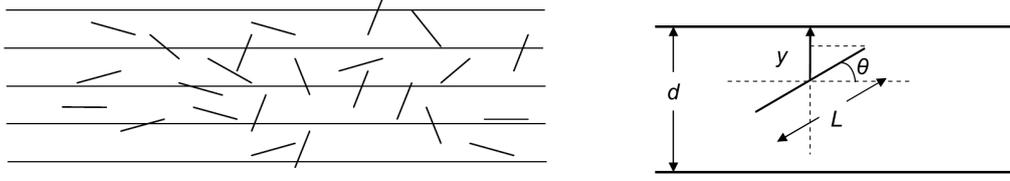
$$dP_y = \frac{dy}{d/2} \quad (1)$$

Tan sólo en el caso de que  $y \leq l/2$  existe la posibilidad de que la aguja corte una línea. Para un valor  $y$  en dicho intervalo, la aguja cortará una línea si cae con un ángulo  $\theta$  mayor que el ángulo límite  $\theta_L$  ( $\theta_L \leq \theta \leq \pi/2$ ,  $0 \leq y \leq l/2$ ), el cual se obtiene a partir del caso en que la punta de la aguja toca la línea. Por trigonometría, tenemos que:

$$\sin(\theta_L) = \frac{y}{l/2} \quad ; \quad 0 \leq \theta_L \leq \pi/2 \quad (2)$$

---

\*ferhue#alumni.uv.es



**Figura 1:** Geometría del problema de la aguja del conde de Buffon

Una condición equivalente (a partir de la ecuación 2) es que  $y \leq \frac{l}{2} \sin \theta$ . La probabilidad de que el ángulo sea mayor que  $\theta_L$ , es decir, de que la aguja corte una línea dado un valor de  $y$  menor que  $l/2$  es:

$$P_\theta = \frac{\pi/2 - \theta_L}{\pi/2} = 1 - \frac{\theta_L}{\pi/2} \quad (3)$$

La probabilidad global de que la aguja corte la línea viene dada por la suma para valores de  $y$  del intervalo multiplicado por  $P_\theta$ :

$$P \equiv P_{\text{corte}} = \int_{y=0}^{y=l/2} dP_y * P_\theta = \frac{2}{d} \int_{y=0}^{y=l/2} dy * \left(1 - \frac{\theta_L}{\pi/2}\right) = \frac{2}{d} \int_{y=0}^{y=l/2} dy * \left(1 - \frac{\arcsin(\frac{y}{l/2})}{\pi/2}\right) \quad (4)$$

Integrando el primer factor, y haciendo un cambio de variable  $\sin(z) = \frac{y}{l/2}$  en el segundo, queda que:

$$P = \frac{l}{d} - \frac{2}{d} \int_{z=0}^{z=\pi/2} \frac{l}{2} \cos(z) dz * \frac{z}{\pi/2} = \frac{l}{d} - \frac{2l}{\pi d} \int_{z=0}^{z=\pi/2} z \cos(z) dz \quad (5)$$

Integrando por partes:

$$\int_{z=0}^{z=\pi/2} z \cos(z) dz = z \sin(z) \Big|_{z=0}^{z=\pi/2} - \int_{z=0}^{z=\pi/2} \sin(z) dz = \frac{\pi}{2} - (-) \cos(z) \Big|_{z=0}^{z=\pi/2} = \frac{\pi}{2} - 1 \quad (6)$$

Sustituyendo en la ecuación 5, obtenemos la fórmula que queríamos demostrar:

$$P = \frac{l}{d} - \frac{2l}{\pi d} \left(\frac{\pi}{2} - 1\right) = \frac{2l}{\pi d} \quad (7)$$

Por tanto, si tiramos  $n$  agujas y obtenemos  $r$  éxitos (formalismo de distribución binomial), entonces podemos estimar  $\hat{P} = r/n$ . La distribución binomial que gobierna este experimento tiene media  $\mu=nP$  y varianza  $\sigma(r)^2=nP(1-P)$ . Despejando  $\pi$  de la ecuación 7:

$$\hat{\pi} = \frac{2l}{d\hat{P}} = \frac{2ln}{dr} \quad (8)$$

Suponiendo que el error de  $l$  y  $d$  es despreciable, entonces el error de  $\hat{\pi}$  según la propagación de errores es:

$$\begin{aligned} \sigma(\hat{\pi}) &= \frac{2l}{d} \sigma(1/\hat{P}) = \frac{2l}{d} \frac{1}{\hat{P}^2} \sigma(\hat{P}) \\ \sigma(\hat{P}) &= \sigma(r/n) = \frac{1}{n} \sigma(r) \simeq \frac{\sqrt{n\hat{P}(1-\hat{P})}}{n} \\ \sigma(\hat{\pi}) &= \frac{2l}{d} \frac{1}{\hat{P}^2} \frac{\sqrt{n\hat{P}(1-\hat{P})}}{n} = \frac{2l}{d} \frac{1}{\hat{P}} \sqrt{\frac{1}{\hat{P}} - 1} * \frac{1}{\sqrt{n}} \propto \frac{1}{\sqrt{n}} \end{aligned} \quad (9)$$

$$\sigma_{\text{teo}} = \frac{2l}{d} \frac{\sqrt{P(1-P)}}{P^2} \frac{1}{\sqrt{n}} = \pi \sqrt{\frac{\pi d}{2l} - 1} * \frac{1}{\sqrt{n}} \quad (10)$$

- b) Mediante la función TRandom1->Rndm() de ROOT genero 2 series de  $n$  datos aleatorios (el valor  $y$  y el ángulo  $\theta$ ). Escojo  $d = 1.5$  y  $l = 1.1$  (en realidad, sólo es importante el cociente  $d/l$ ). Si se cumple la condición  $y \leq \frac{l}{2} \sin(\theta)$ , incremento el número de éxitos  $r$  en una unidad. Realizo el experimento para distintos valores de  $n$ , y se obtienen los siguientes resultados:

$n$	$\hat{\pi}$	$\pi - \hat{\pi}$	$\sigma(\hat{\pi})$
100	2.99320	0.14840	0.30537
1000	3.02405	0.11754	0.09854
10000	3.17460	-0.03301	0.03426
100000	3.14594	-0.00434	0.01064
1000000	3.14397	-0.00238	0.00336
10000000	3.14066	0.00094	0.00106

**Tabla 1:** Estimación del número  $\pi$  mediante el experimento de la aguja del conde de Buffon

El acuerdo entre el error real y el estimado a partir de la propagación de errores es del mismo orden en todos los casos, y se cumple que  $\sigma\sqrt{n} \simeq 3.23 \simeq cte$ . El error se reduce lentamente con el número  $n$  de agujas, lo cual es corriente de los métodos Monte Carlo (dependencia  $1/\sqrt{n}$  de la ecuación 9). Aparte, se comprueba que para que el error sea de una milésima, se necesita un valor de  $n = 10.000.000$ . Este valor se puede comparar con la estimación teórica a partir de la fórmula 10, despejando  $n$  y sustituyendo  $\sigma_{\text{teo}} = 0.001$ :

$$n(\sigma_{\text{teo}}) = \pi^2 \left( \frac{\pi d}{2l} - 1 \right) \frac{1}{\sigma_{\text{teo}}^2} \quad (11)$$

$$n(\sigma_{\text{teo}} = 0.001) \simeq 11.271.039$$

Se comprueba que la estimación teórica es del mismo orden que la experimental.

## 6.2.- Propiedades de los generadores de números aleatorios

### Enunciado

- a) Consideremos algunos ejemplos de generadores congruentes de la forma:

$$X_{n+1} = (a \cdot X_n + c) \bmod m$$

Construir la secuencia hasta completar periodo para los siguientes valores.

- a)  $m = 31$ ,  $c = 0$ ,  $a = 7$  y  $X_0 = 1$ .  
b)  $m = 31$ ,  $c = 0$ ,  $a = 3$  y  $X_0 = 1$ .  
c)  $m = 32$ ,  $c = 1$ ,  $a = 5$  y  $X_0 = 1$ .
- b) Verificar que los casos (b) y (c) del ejercicio anterior satisfacen las condiciones de máximo periodo que vienen dadas por:
- Caso  $c = 0$ 
    - (i)  $m$  es un número primo
    - (ii) Para todos los divisores primos  $p$  de  $(m - 1)$ , se cumple que  $a^{(m-1)/p} \bmod m \neq 1$
  - Caso  $c \neq 0$ 
    - (i)  $c$  y  $m$  primos entre sí
    - (ii)  $(a - 1)/p$  es un entero para cada divisor primo  $p$  de  $m$ .
    - (iii) Si  $m$  es múltiplo de 4, entonces  $(a - 1)$  debe ser divisible por 4.
- c) Representar gráficamente en un plano todos los puntos definidos por las coordenadas  $(\xi_i, \xi_{i+1})$ ,  $i = 0, 1, \dots, 28$ , para los apartados (b) y (c) del ejercicio 1. ¿Es lo que cabría esperar de un generador aleatorio?

### Resolución

- a) Construimos las secuencias hasta completar periodos para el generador congruente citado en el enunciado (ver tabla 2).
- b)
- En los casos (a) y (b),  $c = 0$ . Analizamos cada condición por separado:
    - (i) Tanto (a) como (b) cumplen la primera condición, ya que  $m = 31$  es primo.
    - (ii) Los divisores primos  $p$  de  $(m - 1) = 30$  son 5, 3 y 2.  
En el caso (a),  $7^{30/p} \bmod m = 1, 13$  y 8 respectivamente. Por tanto no cumple la condición de que todos sean  $\neq 1$ .  
En el caso (b)  $3^{30/p} \bmod m = 5, 9$  y 12 respectivamente. Por tanto sí cumple la segunda condición.  
En resumen, el caso (a) no satisface la condición de máximo periodo, pero (b) sí.
  - En el caso (c),  $c \neq 0$ . Analizamos cada condición separadamente:
    - (i) Es trivial comprobar que  $c = 1$  y  $m = 32$  son primos entre sí (no tienen ningún factor primo en común, o equivalentemente, si su máximo común divisor es 1).
    - (ii) El único divisor primo de  $m$  es  $p = 2$ .  $(a - 1)/p = 4/2 = 2$  es entero.
    - (iii)  $m = 32$  es múltiplo de 4 y  $(a - 1) = 4$  es divisible por 4.  
Como cumple las tres condiciones, (c) satisface la condición de máximo periodo.

$n$	$X_n$
0	1
1	7
2	18
3	2
4	14
5	5
6	4
7	28
8	10
9	8
10	25
11	20
12	16
13	19
14	9
15	1

(a)  $m = 31, c = 0, a = 7,$   
 $X_0 = 1$

$n$	$X_n$
0	1
1	3
2	9
3	27
4	19
5	26
6	16
7	17
8	20
9	29
10	25
11	13
12	8
13	24
14	10
15	30
16	28
17	22
18	4
19	12
20	5
21	15
22	14
23	11
24	2
25	6
26	18
27	23
28	7
29	21
30	1

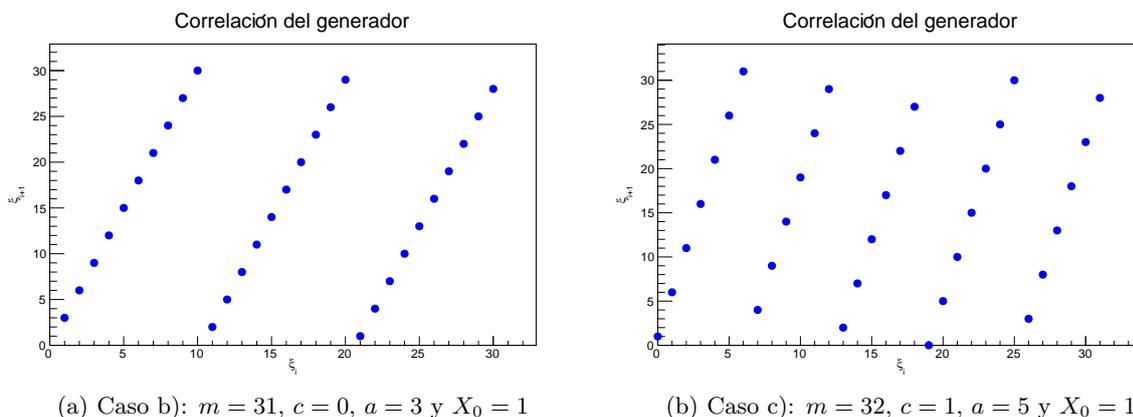
(b)  $m = 31, c = 0, a = 3$  y  
 $X_0 = 1$

$n$	$X_n$
0	1
1	6
2	31
3	28
4	13
5	2
6	11
7	24
8	25
9	30
10	23
11	20
12	5
13	26
14	3
15	16
16	17
17	22
18	15
19	12
20	29
21	18
22	27
23	8
24	9
25	14
26	7
27	4
28	21
29	10
30	19
31	0
32	1

(c)  $m = 32, c = 1, a = 5$  y  
 $X_0 = 1$

**Tabla 2:** Secuencias completas de un generador congruente.

- c) Representamos los puntos definidos por las coordenadas  $(\xi_i, \xi_{i+1})$ , desde  $i = 0$  hasta el último par de la secuencia para los apartados (b) y (c) del ejercicio 1. Se comprueba una correlación entre números pseudoaleatorios consecutivos, ya que los puntos se agrupan según rectas (efecto Marsaglia) y no como una nube aleatoria uniforme (idealmente). Un generador será mejor cuanto mayor granularidad (o número de hiperplanos, si representamos n-tuplas) presente.



**Figura 2:** Correlación del generador pseudoaleatorio congruente

### 6.3.- Test de Kolmogorov-Smirnov a un generador uniforme

**Enunciado** Aplicar un test  $\chi^2$  a un generador uniforme del siguiente modo:

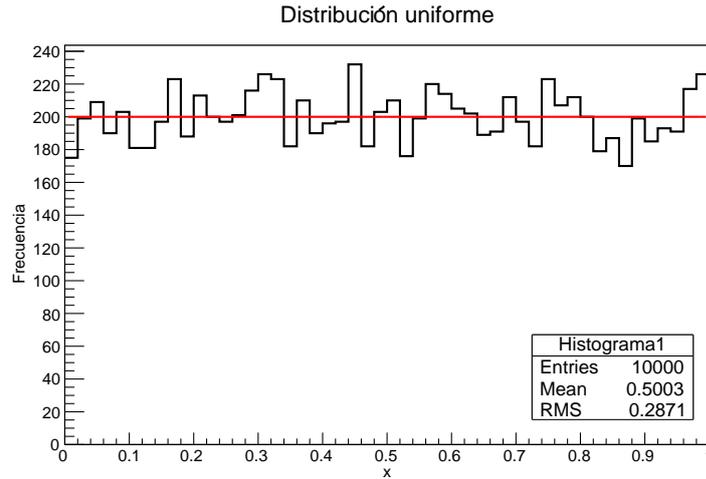
- Generar  $N = 10000$  números aleatorios y distribuirlos en un histograma de  $k = 50$  *bines*.
- Si en cada *bin* tenemos  $N_j$  sucesos, construir el estadístico:

$$T = \sum_{j=1}^k \frac{(N_j - \frac{N}{k})^2}{\frac{N}{k}} = \frac{k}{N} \sum_{j=1}^k \left( N_j - \frac{N}{k} \right)^2$$

- La variable aleatoria anterior,  $T$ , sigue una distribución de  $\chi^2$  con  $\nu = 49$  grados de libertad. Para comprobarlo, repetir el experimento 100 veces y representar los valores de  $T$  en un histograma superponiendo la distribución de  $\chi^2(T, \nu)$  correspondiente.
- Calcular el estadístico  $q$  de Kolmogorov-Smirnov entre el histograma y la función acumulativa de  $\chi^2(T, \nu)$  y mediante las tablas correspondientes estimar la probabilidad de obtener un valor mayor.
- ¿Existen indicios de no uniformidad?

#### Resolución

- Genero  $N = 10000$  números aleatorios con la función TRandom1->Rndm() de ROOT y los represento en un histograma con  $k = 50$  *bines*. Cabe señalar que las fluctuaciones son acordes a lo esperado, como ya se comprobó en ejercicios de temas anteriores.

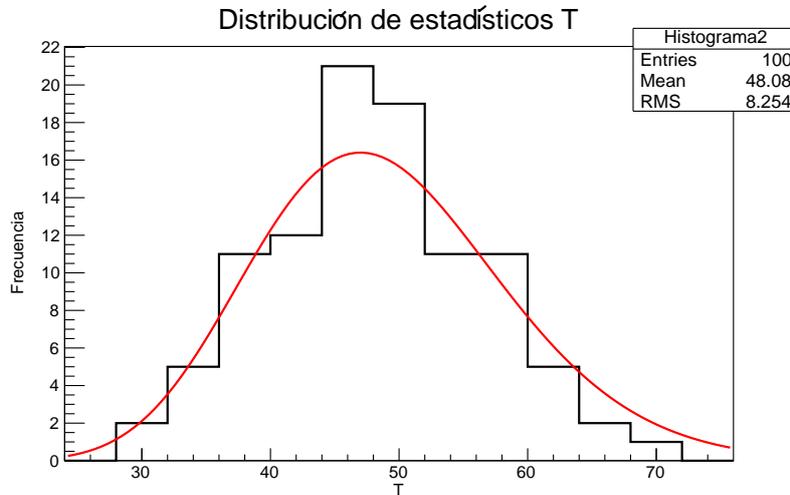


**Figura 3:** Histograma de una distribución uniforme de números pseudoaleatorios. La línea roja es el valor esperado medio en cada bin  $N/k$

b) El estadístico  $T$  para el diagrama anterior es:

$$56.5$$

c) Repito el experimento  $Nt = 100$  veces y represento los valores de  $T$  en un nuevo histograma con  $k_t = 13$  bins entre  $T = 24$  y  $T = 76$  superponiendo la distribución teórica  $\chi^2(T, \nu)$ , con  $\nu = k - 1 = 49$ .



**Figura 4:** Histograma de la distribución de estadísticos  $T$  de  $N_t = 100$  distribuciones uniformes, cada uno con  $N = 10000$  números pseudoaleatorios. La línea roja es la distribución teórica  $\chi^2(T, \nu)$

Se comprueba que la distribución obtenida es compatible con la teórica  $\chi^2(T, \nu)$ , con fluctuaciones del orden de la raíz del número de cuentas en la mayoría de los bins. Cabe señalar que el número de repeticiones del experimento  $N_t$  es pequeño, por lo que he tenido que escoger un número de bins  $k_t$  pequeño para que hubiese suficientes cuentas en cada uno de ellos. En cierta manera, el acuerdo visual con la distribución teórica depende de una adecuada elección de  $k_t$ .

- d) Calculo el estadístico  $q$  de Kolmogorov-Smirnov entre la distribución acumulativa de  $T$  (ordeno los  $N_t$  valores de  $T$  de menor a mayor en un vector  $\vec{T}$ ) y la función acumulativa de  $\chi^2(T, \nu)$ , que viene dado por la expresión siguiente:

$$q = \sqrt{N_t} \max \left| F_N((\vec{T})_i) - F_0((\vec{T})_i) \right| = \sqrt{N_t} \max \left| i/N_t - \chi_{cdf}^2((\vec{T})_i, \nu = 49) \right|$$

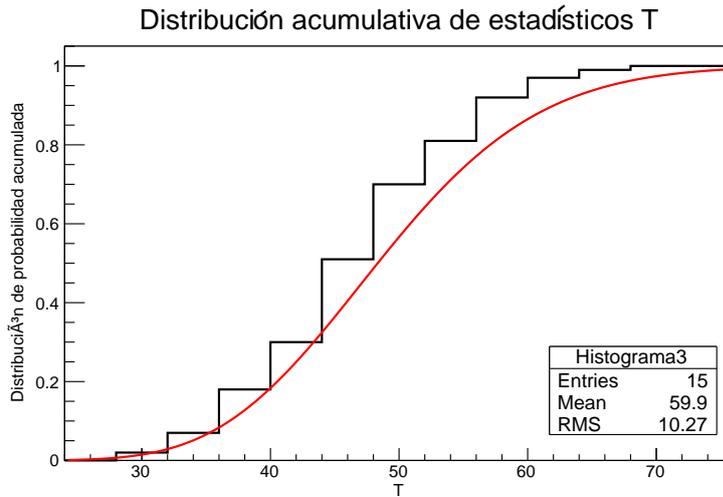
donde  $i$  va de 1 a  $N_t$ . Se obtiene el siguiente valor de  $q$ :

$$q = 0.843973$$

La probabilidad  $f$  de obtener un valor de  $q'$  igual o mayor a  $q$  si repetimos todo el procedimiento, (suponiendo cierto que los datos vienen gobernados por dicha distribución teórica), se calcula mediante la función acumulativa  $F$  de la distribución de Kolmogorov como  $f(q' \geq q) = 1 - F_K(q)$ , expresión que viene implementada en ROOT con la función `TMath::KolmogorovProb(q')`:

$$f(q' \geq q) = -2 \sum_{j=1}^{\infty} (-1)^j e^{-2j^2 q^2} \simeq 47.5\%$$

Cabe indicar que este test depende del número de veces que repetimos el experimento  $N_t$  y del número de bins  $k$  con el que se calcula  $T$ , pero no del número de bins  $kt$  (el histograma es sólo para visualizar la distribución, no para el análisis numérico). Por completitud, representamos el histograma acumulativo con  $kt$  bins:



**Figura 5:** Histograma acumulativo de la distribución de estadísticos  $T$ . La línea roja es la distribución acumulativa teórica  $\chi_{cdf}^2(T, \nu)$

El acuerdo experimento teoría es bueno, pues la forma de ambas distribuciones es similar. En este histograma, parece que haya una desviación sistemática en todos los bins, pero la causa puede estar en el exceso o defecto en uno solo de ellos (uno del principio), que genera un desacuerdo visual sistemático en los restantes (algo acusado en este caso al ser  $Nt$  pequeño).

- e) Por los resultados vistos anteriormente, no se aprecian indicios de no uniformidad. Las fluctuaciones son acordes con lo esperado y el valor  $f$  (nivel de confianza) es del 48%, con lo que podemos afirmar que el generador de números pseudoaleatorios uniformes es fiable y que la distribución de valores de  $T$  viene gobernada por la distribución  $\chi^2$ .

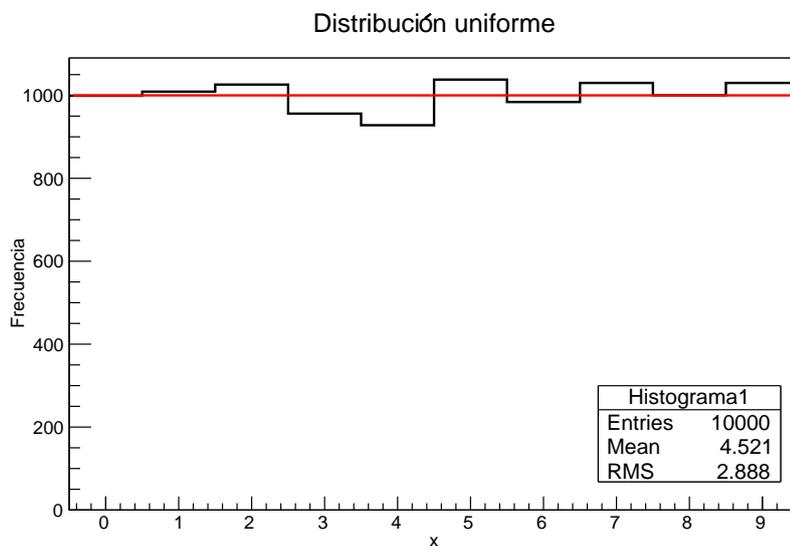
## 6.4.- Test $\chi^2$ de una distribución binomial

**Enunciado** Mediante un generador uniforme construir una secuencia de  $N = 10000$  números enteros en el rango  $(0, 9)$ . Para ello, multiplicar cada número aleatorio,  $\xi$ , por 10 y tomar la parte entera:

$$k = E[10 * \xi]$$

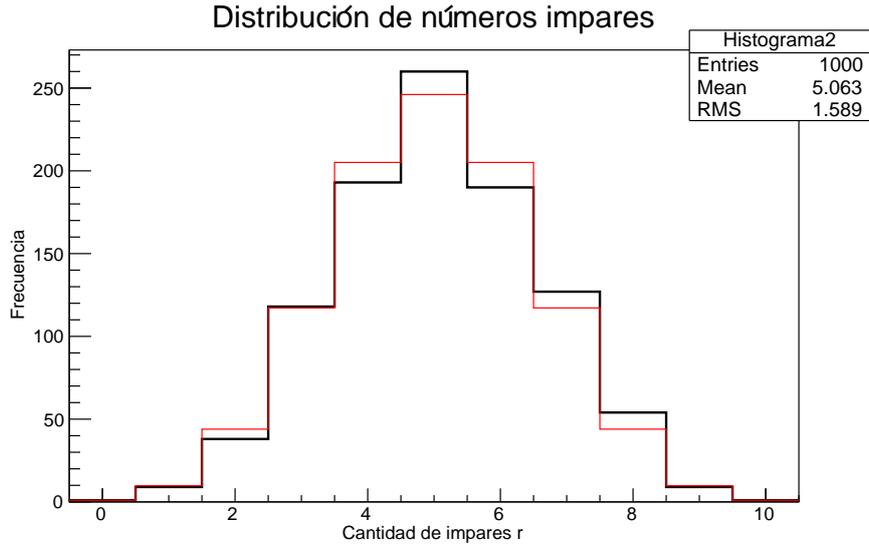
Dividir la secuencia en subgrupos de diez números, y de cada subgrupo contar los  $r$  que son impares. Dibujar un histograma con los resultados obtenidos y comparar con una distribución binomial con  $N = 10$  y  $p = 0.5$  mediante un test de  $\chi^2$ .

**Resolución** Genero los  $N = 10000$  números enteros según la fórmula del enunciado, con el generador de ROOT TRandom->Rndm(), y obtengo el siguiente histograma (con 10 bins, 1 por cada número entero) donde se aprecia la uniformidad:



**Figura 6:** Histograma de la distribución uniforme entre 0 y 9. La línea roja es el valor esperado medio en cada bin  $N/10 = 1000$

Agrupo los valores obtenidos de 10 en 10 y cuento la cantidad  $r$  de números impares hay en cada grupo. Histograma los valores de  $r$  con de nuevo 10 bins, uno por cada entero:



**Figura 7:** Histograma de la distribución de la cantidad de números impares en grupos de 10 números pseudoaleatorios entre 0 y 9. La línea roja es la distribución binomial  $B(r, 0.5, 10)$

Se comprueba visualmente que el ajuste entre la distribución pseudoaleatoria ( $b$ ) y la teórica ( $B$ ) es bueno. Para cuantificar esta observación, calculamos el test de  $\chi^2$  dado por:

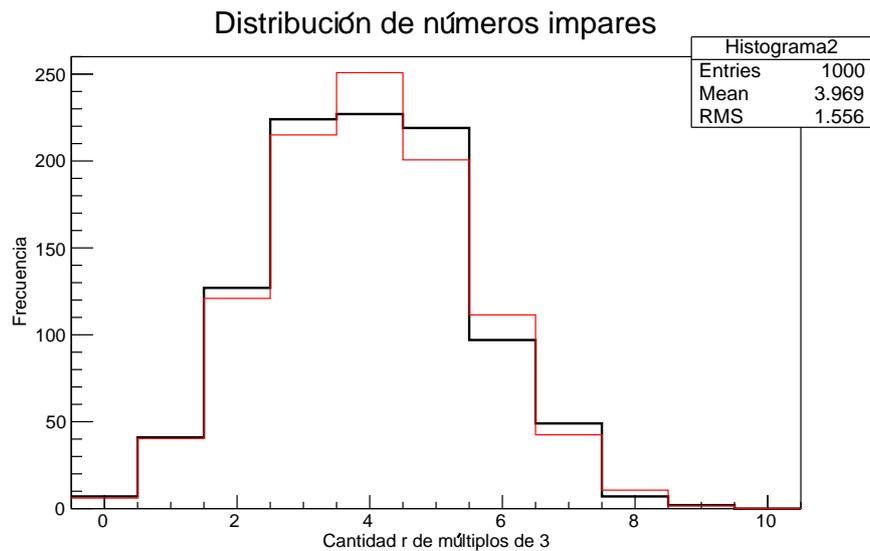
$$\chi^2 = \sum_{i=0}^{i=9} \frac{(b(r) - B(r, 0.5, 10))^2}{B(r, 0.5, 10)} = 6.65905$$

El nivel de confianza es  $CL = P(\chi'^2 \geq \chi^2) \simeq 82.6\%$  según los valores tabulados para  $\nu = 11$  grados de libertad (o bien con la función de ROOT `TMath::Prob( $\chi^2, \nu$ )`), con lo que podemos concluir que el ajuste es bueno y la distribución binomial gobierna estos sucesos.

## 6.5.- Test $\chi^2$ de una distribución binomial (múltiplos de 3)

**Enunciado** Repetir el ejercicio anterior pero contando los números múltiplos de 3 y comparando con la distribución binomial correspondiente.

**Resolución** Empleo la misma secuencia de números pseudoaleatorios que en el ejercicio anterior. En los grupos de 10, cuento la cantidad  $r$  de números múltiplos de 3 (incluido el 0) que en cada grupo. Histograma los valores de  $r$  con de nuevo 10 bins, uno por cada entero. En este caso, la distribución binomial tendrá una probabilidad  $p = 0.4$  (de diez valores entre 0 y 9, cuatro son múltiplos de 3 (el 0,3,6,9)).



**Figura 8:** Histograma de la distribución de la cantidad de números múltiplos de 3 en grupos de 10 números pseudoaleatorios entre 0 y 9. La línea roja es la distribución binomial  $B(r, 0.4, 10)$

Calculamos el test de  $\chi^2$  análogamente al ejercicio anterior:

$$\chi^2 = 9.12121$$

El nivel de confianza es  $CL = P(\chi'^2 \geq \chi^2) \simeq 61.1\%$  según los valores tabulados para  $\nu = 11$  grados de libertad, con lo que podemos concluir que el ajuste es bueno y la distribución binomial gobierna estos sucesos.

## Tema 7.- Muestreo de distribuciones

**7.1.-** Mediante un generador de números aleatorios uniforme obtener dos números  $\xi_1$  y  $\xi_2$  y construir la variable  $z = \xi_1 + \xi_2$ . Construir un histograma con 10000 entradas y comprobar que la variable  $z$  se distribuye según la función triangular.

**7.2.-** Obtener tres números aleatorios de una distribución uniforme y ordenarlos de manera que  $\xi_1 \leq \xi_2 \leq \xi_3$ . Tomar entonces el valor central,  $\xi_2$ , y construir un histograma con 10000 entradas. Comprobar empírica y analíticamente que dicha variable se distribuye según la función densidad de probabilidad  $f(x) = 6x(1-x)$  en el dominio  $[0, 1]$ .

**7.3.-** Obtener cinco números aleatorios de una distribución uniforme y obtener la variable aleatoria  $z = \max \{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$ . Construir un histograma de la variable  $z$  con 10000 entradas. Mediante el método de la transformación inversa construir un generador de números aleatorios distribuidos de acuerdo con la función  $f(x) = x^4$  en el dominio  $[0, 1]$  y construir un histograma con 10000 entradas (no olvidar que la pdf ha de estar normalizada). Comparar las distribuciones de los apartados anteriores y sacar conclusiones.

**7.4** Diseñar un programa que genere sucesos distribuidos según Poisson mediante cualquiera de los métodos estudiados en clase.

- a) Aplicar el programa para generar tres muestras de 100000 sucesos distribuidos según Poisson con  $\mu = 0.1$ ,  $\mu = 1$  y  $\mu = 10$ . Representar gráficamente las distribuciones y obtenidas y la función de Poisson correspondiente.
- b) Para cada una de las muestras anteriores, obtener la distribución exponencial correspondiente a partir de las distancias de separación entre sucesos consecutivos. Ajustar dichas distribuciones y obtener una estimación del parámetro  $\lambda$  en cada caso.

**7.5.-** En un experimento para determinar la estabilidad de una partícula se mide el cociente  $R$  entre dos secciones eficaces. En términos de las cantidades medidas en el experimento  $R$  puede expresarse como:

$$R = \frac{a}{\frac{d}{ke}(b-c) - 2(1 - \frac{kd}{e})a}$$

donde :

$$\begin{aligned} a &= 3.84 \pm 1.33 \\ b &= 74 \pm 4 \\ c &= 9.5 \pm 3 \\ d &= 0.112 \pm 0.009 \\ e &= 0.32 \pm 0.02 \\ k &= 0.89 \end{aligned}$$

- a) Calcular el valor de  $R$  con su error utilizando propagación de errores y suponiendo que las variables son independientes.

- 
- b) Según cálculos teóricos, la partícula objeto de este análisis es inestable si  $R < 0.42$ . ¿Qué podemos concluir del resultado anterior? Según el análisis de errores realizado ¿Qué probabilidad hay de que la partícula sea estable?
- c) La fórmula de propagación de errores solo es válida para errores pequeños, sin embargo dos de las cantidades anteriores presentan errores bastante elevados. Realizar un análisis del error mediante Monte Carlo del siguiente modo: generar valores de  $a$ ,  $b$ ,  $c$ ,  $d$  y  $e$  aleatoriamente suponiendo que son variables gaussianas (utilizar un generador de números gaussiano basado en el método de Box-Mueller) y calcular el valor de  $R$ . Obtener de esta forma 10000 valores de  $R$  y construir el correspondiente histograma.
- d) ¿Qué porcentaje de valores de  $R$  es superior a 0.42? ¿Podemos sacar las mismas conclusiones que con el análisis anterior?
- e) Reducir los errores de  $a$  y  $c$  en un 100 % y volver a ejecutar el Monte Carlo ¿Qué conclusiones podemos extraer?

# -Métodos estadísticos-

## Problemas Tema 7:

### Muestreo de distribuciones

Fernando Hueso González\*

Valencia, 3 de enero de 2012

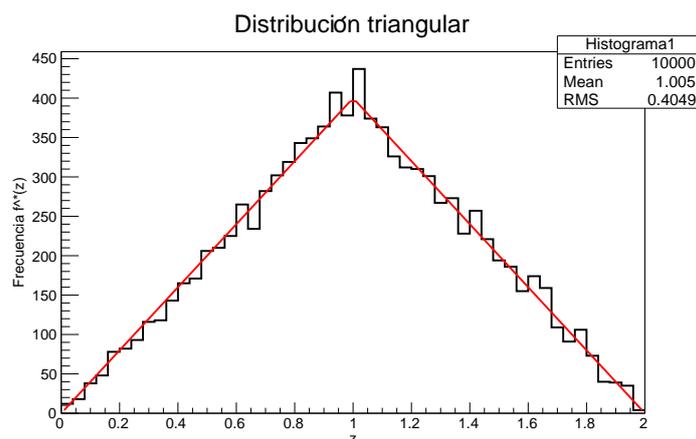
MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

#### 7.1.- Función triangular

**Enunciado** Mediante un generador de números aleatorios uniforme obtener dos números  $\xi_1$  y  $\xi_2$  y construir la variable  $z = \xi_1 + \xi_2$ . Dibujar un histograma con 10000 entradas y comprobar que la variable  $z$  se distribuye según la función triangular.

**Resolución** Generamos  $N = 10000$  parejas  $\xi_1, \xi_2$  de números pseudoaleatorios con la función `TRandom->Rndm()` de ROOT. Obtenemos  $N = 10000$  valores de  $z = \xi_1 + \xi_2$  y los histogramamos entre 0 y 2 con  $n_{bin} = 50$  bins, y lo comparamos con la función triangular teórica, dada por:

$$f^*(z) = N \frac{2}{n_{bin}} * \begin{cases} z & 0 \leq z < 1 \\ 2 - z & 1 \leq z \leq 2 \end{cases} \quad (1)$$



**Figura 1:** Histograma de una distribución triangular a partir de la suma de dos números pseudoaleatorios uniformes entre 0 y 1. La línea roja es la función triangular teórica (ecuación 1)

\*ferhue#alumni.uv.es

Comprobamos que la variable  $z$  se distribuye según la función triangular, con fluctuaciones respecto a la curva teórica acordes con la estadística.

## 7.2.- Valor central

**Enunciado** Obtener tres números aleatorios de una distribución uniforme y ordenarlos de manera que  $\xi_1 \leq \xi_2 \leq \xi_3$ . Tomar entonces el valor central,  $\xi_2$ , y construir un histograma con 10000 entradas. Comprobar empírica y analíticamente que dicha variable se distribuye según la función densidad de probabilidad  $f(x) = 6x(1 - x)$  en el dominio  $[0, 1]$ .

**Resolución** Generamos  $N = 10000$  tríos  $x_1, x_2, x_3$  de números pseudoaleatorios con la función TRandom1  $\rightarrow$  Rndm() de ROOT. Los ordenamos de menor a mayor ( $\xi_1 \leq \xi_2 \leq \xi_3$ ) y histogramamos los  $N = 10000$  valores de  $\xi_2$  con  $n_{bin} = 50$ , para compararlo con la función densidad de probabilidad teórica en el dominio  $[0, 1]$ . A continuación se demuestra analíticamente la fórmula.

Si se cumple  $x_1 \leq x_2 \leq x_3$ , donde cada  $x_i$  viene descrita por la pdf uniforme  $f(x_i) = 1$  y su cdf  $F(x_i) = x$ , la probabilidad es:

$$P(x_1 \leq x_2 \leq x_3) = P(x_1 \geq x_2) * P(x_3 \geq x_2) = F(x_2) * (1 - F(x_2)) = x_2(1 - x_2) \quad (2)$$

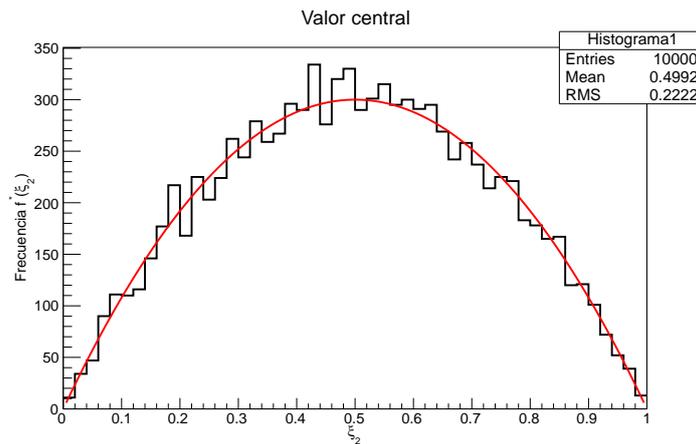
Sin embargo, podría ser  $x_1$  mayor que  $x_2$ , ó  $x_3$ , etc. Las distintas ordenaciones ( $\xi$ ) de tres valores de  $x_i$  son  $3!$ , con lo que para el valor intermedio  $\xi_2$  ya sorteado, queda:

$$f(\xi_2) = 6\xi_2(1 - \xi_2) \quad (3)$$

Es inmediato comprobar que esta pdf está debidamente normalizada a 1. Para comparar con el histograma, debemos multiplicar por una constante:

$$f^*(\xi_2) = \frac{N}{n_{bin}} 6\xi_2(1 - \xi_2) \quad (4)$$

El histograma obtenido es:



**Figura 2:** Histograma de una distribución triangular a partir de la suma de dos números pseudoaleatorios uniformes entre 0 y 1. La línea roja es la función teórica (ecuación 4)

Comprobamos que la variable  $\xi_2$  se distribuye según la función teórica, con fluctuaciones respecto a la curva teórica acordes con la estadística.

### 7.3.- Valor máximo

**Enunciado** Obtener cinco números aleatorios de una distribución uniforme y obtener la variable aleatoria  $z = \max\{\xi_1, \xi_2, \xi_3, \xi_4, \xi_5\}$ . Construir un histograma de la variable  $z$  con 10000 entradas. Mediante el método de la transformación inversa, construir un generador de números aleatorios distribuidos de acuerdo con la función  $f(x) = x^4$  en el dominio  $[0, 1]$  y construir un histograma con 10000 entradas (no olvidar que la pdf ha de estar normalizada). Comparar las distribuciones de los apartados anteriores y sacar conclusiones.

**Resolución** Estudiemos la distribución de probabilidad de obtener un valor mayor que los cuatro restantes (los cinco pseudoaleatorios uniformes entre 0 y 1).  $f$  es la pdf de la distribución uniforme y  $F$  la cdf. Supongamos que  $x_5$  es el mayor de todos:

$$\begin{aligned} P(x_1, x_2, x_3, x_4 \leq x_5) &= P(x_1 \leq x_5) * P(x_2 \leq x_5) * P(x_3 \leq x_5) * P(x_4 \leq x_5) \\ &= F(x_5) * F(x_5) * F(x_5) * F(x_5) = F(x_5)^4 = x_5^4 \end{aligned} \quad (5)$$

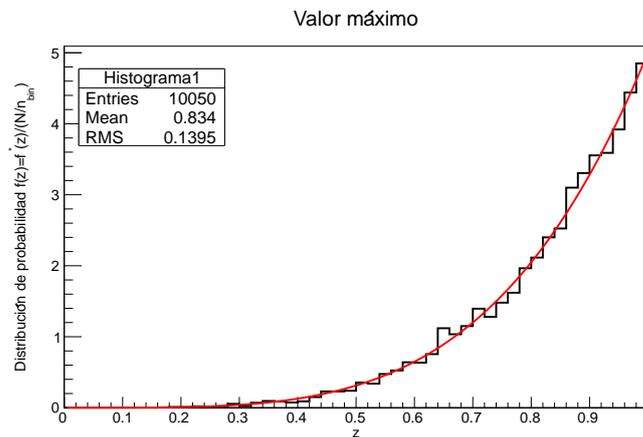
Ahora bien, el mayor de ellos puede ser  $x_5$  o cualquiera de los otros 4. Puesto que el orden relativo entre los menores no es importante, tenemos 5 combinaciones equivalentes, con lo que si  $z = \max\{x_i\}$  es el mayor de todos:

$$f(z) = f(\max\{x_i\}) = 5z^4 \quad (6)$$

que está debidamente normalizado a la unidad. La cdf es  $F(z) = z^5$ .

Generamos  $N = 10000$  quintetos  $x_1, x_2, x_3, x_4, x_5$  de números pseudoaleatorios con la función TRandom1 -> Rndm() de ROOT. Elegimos el valor máximo ( $z = \max\{x_i\}$ ) y histogramamos los  $N = 10000$  valores de  $z$  con  $n_{bin} = 50$ , para compararlo con la función densidad de probabilidad teórica en el dominio  $[0, 1]$ . Normalizamos el histograma a la unidad, dividiendo la frecuencia observada  $f^*$  entre  $N/n_{bin}$ , y obtenemos:

El histograma obtenido es: Comprobamos que la variable  $z$  se distribuye según la función



**Figura 3:** Histograma a partir del valor máximo de 5 números pseudoaleatorios uniformes entre 0 y 1. La línea roja es la función teórica (ecuación 6)

teórica, con fluctuaciones respecto a la curva teórica acordes con la estadística.

## 7.4.- Poisson

**Enunciado** Diseñar un programa que genere sucesos gobernados por la distribución de Poisson mediante cualquiera de los métodos estudiados en clase.

- a) Aplicar el programa para generar tres muestras de 100.000 sucesos distribuidos según Poisson con  $\mu = 0.1$ ,  $\mu = 1$ ,  $\mu = 10$ . Representar gráficamente las distribuciones obtenidas y la función de Poisson correspondiente.
- b) Para cada una de las muestras anteriores, obtener la distribución exponencial  $e^{-\mu s}$  correspondiente a partir de las distancias  $s$  de separación entre sucesos consecutivos. Ajustar dichas distribuciones y obtener una estimación del parámetro  $\mu$  en cada caso.

**Resolución** Diseñar un programa para generar sucesos  $x$  distribuidos según la distribución de Poisson  $P$ :

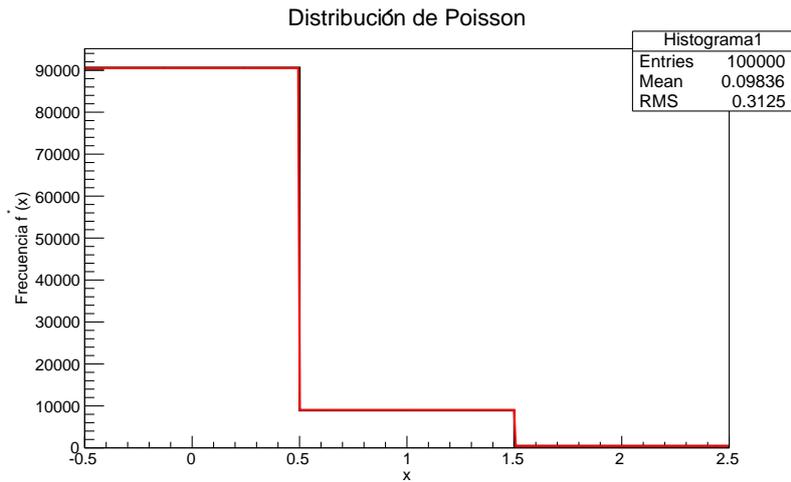
$$P(x, \mu) = \frac{e^{-\mu} \mu^x}{x!}, \quad x \in \mathbb{N}_0 \quad (7)$$

La frecuencia observada  $f^*(x, \mu)$ :

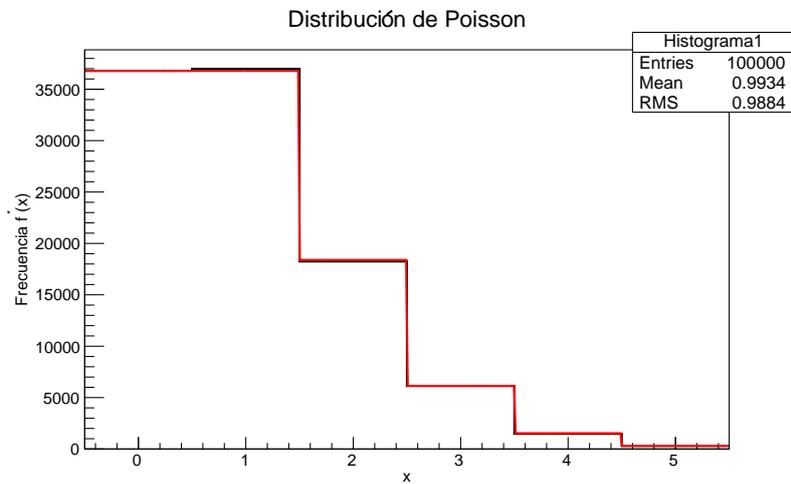
$$f^*(x, \mu) = N * P(x, \mu) \quad (8)$$

Utilizamos el siguiente algoritmo:

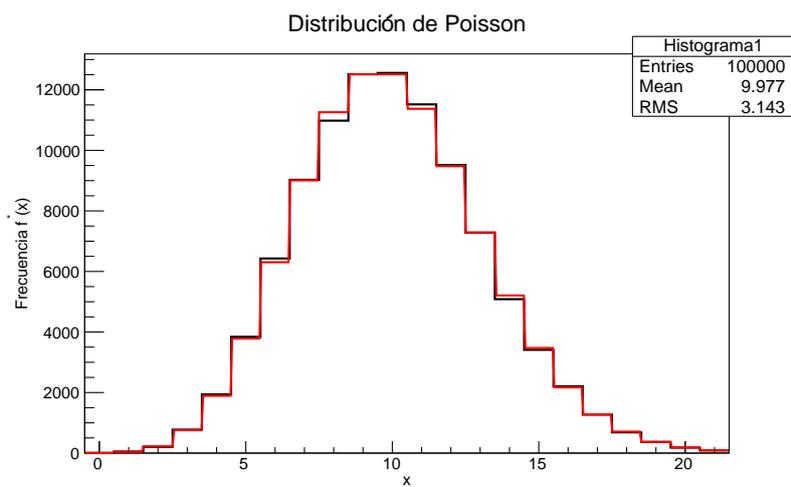
- I. Generamos  $\xi \in [0, 1]$  mediante `TRandom1->Rndm()` de ROOT.
  - II.  $k \leftarrow \xi$
  - III.  $j \leftarrow 0$ 
    1.  $k \leftarrow k - P(j, \mu)$
    2. Si  $k \leq 0$ ,  $x \leftarrow j$ , volver a I.
    3. Si  $k > 0$ :  $j \leftarrow j + 1$ , volver a 1.
- a) Aplico el algoritmo con  $N = 100.000$  cuentas y construyo el histograma, que superpongo con la distribución teórica de Poisson. Para los tres casos de  $\mu$  se obtienen los histogramas de la figura 4. Se comprueba que la variable  $x$  se distribuye según la función teórica, con pequeñas fluctuaciones estadísticas. En el caso de  $\mu$  entero, se comprueba que en los bins  $x = \mu$  y  $x = \mu - 1$ , el número de cuentas es similar, como se puede demostrar analíticamente con facilidad.



(a)  $\mu = 0.1$



(b)  $\mu = 1$



(c)  $\mu = 10$

**Figura 4:** Generador de sucesos pseudoaleatorios gobernados por una distribución de Poisson para diferentes  $\mu$ . La línea roja es la distribución teórica (ecuación 8)

- b) Para cada una de las muestras anteriores, contamos las distancias entre sucesos con  $x \geq 1$ , es decir, cuántos sucesos  $s$  consecutivos dan todos ellos ( $x = 0$ ). Teóricamente se espera que se distribuyan según una ley exponencial:

$$f^*(s) = Ae^{-\mu s} \quad (9)$$

Si realizamos un ajuste de los histogramas de la figura 5, obtenemos valores de  $\mu_{fit}$  compatibles con el  $\mu$  real:

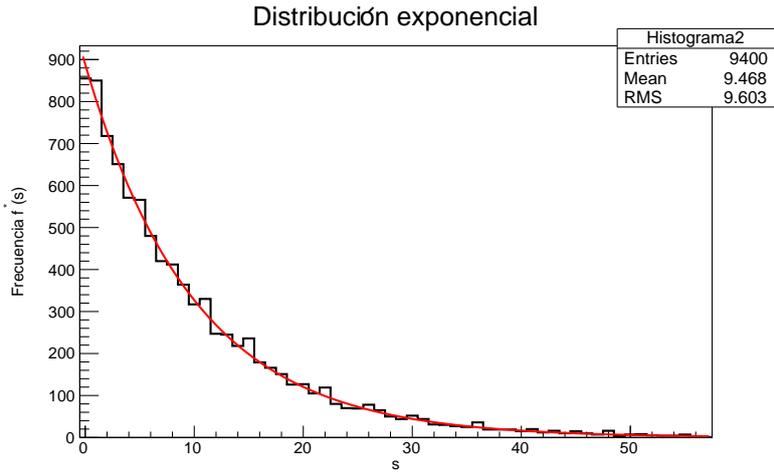
Figura 5a:  
 $A = 888 \pm 13$   
 $\mu_{fit} = 0.0999 \pm 0.0011$

Figura 5b:  
 $A = 39967 \pm 188$   
 $\mu_{fit} = 1.001 \pm 0.004$

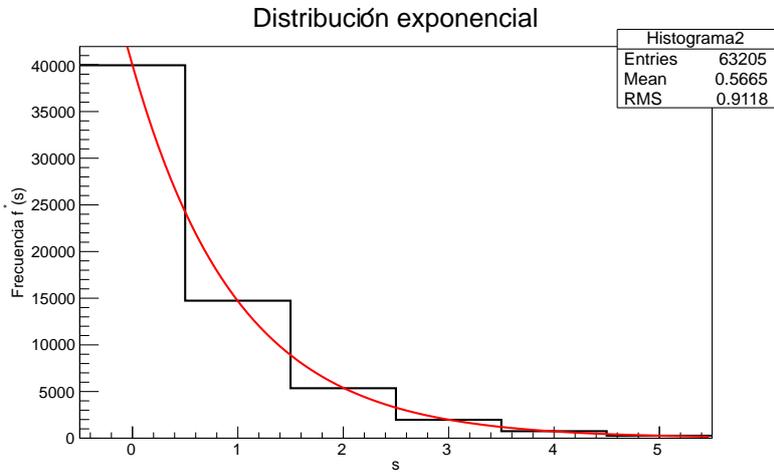
Figura 5c:  
 $A = 99990 \pm 316$   
 $\mu_{fit} = 9.9 \pm 0.4$

Se comprueba que la variable  $s$  se distribuye según la función exponencial, y que los parámetros  $\mu_{fit}$  obtenidos mediante el ajuste de ROOT son plenamente compatibles en todos los casos con los valores originales  $\mu$ .

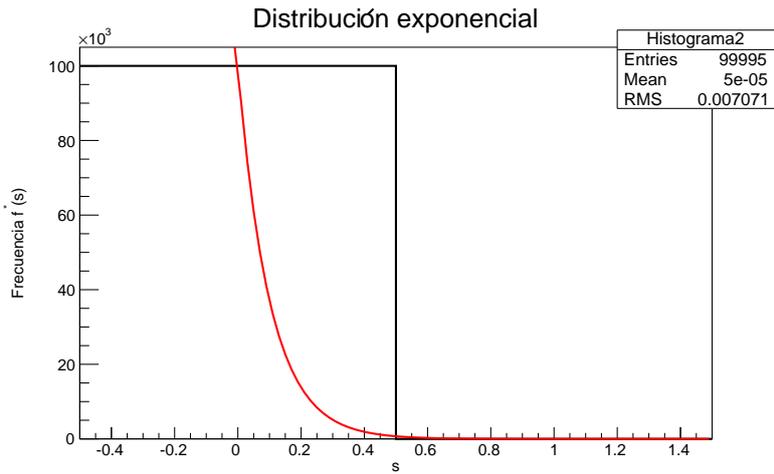
Cabe señalar que en el caso de la figura 5c, al estar la mayoría de cuentas (casi 100.000) en el bin  $s = 0$ , y sólo 5 en el bin  $s = 1$  (apenas había medidas con 0 cuentas), el ajuste es delicado y sólo ha funcionado tras darle unos valores iniciales para el *fit* a los parámetros de ajuste muy cercanos a los valores reales conocidos.



(a)  $\mu = 0.1$



(b)  $\mu = 1$



(c)  $\mu = 10$

**Figura 5:** Generador de sucesos pseudoaleatorios gobernados por una distribución de Poisson para diferentes  $\mu$ . La línea roja es el ajuste a la distribución teórica (ecuación 9) dejando  $A$  y  $\mu$  como parámetros a determinar.

## 7.5.- Análisis de errores

**Enunciado** En un experimento para determinar la estabilidad de una partícula se mide el cociente  $R$  entre dos secciones eficaces. En términos de las cantidades medidas en el experimento,  $R$  puede expresarse como:

$$R = \frac{a}{\frac{d}{ke}(b-c) - 2(1 - \frac{kd}{e})a} \equiv \frac{a}{D} \quad (10)$$

donde:

$$\begin{aligned} a &= 3.84 \pm 1.33 \\ b &= 74 \pm 4 \\ c &= 9.5 \pm 3 \\ d &= 0.112 \pm 0.009 \\ e &= 0.32 \pm 0.02 \\ k &= 0.89 \end{aligned} \quad (11)$$

- Calcular el valor de  $R$  con su error utilizando propagación de errores y suponiendo que las variables son independientes.
- Según cálculos teóricos, la partícula objeto de este análisis es inestable si  $R < 0.42$ . ¿Qué podemos concluir del resultado anterior? Según el análisis de errores realizado, ¿qué probabilidad hay de que la partícula sea estable?
- La fórmula de propagación de errores sólo es válida para errores pequeños, sin embargo, dos de las cantidades anteriores presentan errores bastante elevados. Realizar un análisis del error mediante Monte Carlo del siguiente modo: generar valores de  $a$ ,  $b$ ,  $c$ ,  $d$  y  $e$  aleatoriamente suponiendo que son variables gaussianas (utilizar un generador de números gaussiano basado en el método de Box-Müller) y calcular el valor de  $R$ . Obtener de esta forma 10000 valores de  $R$  y construir el correspondiente histograma.
- ¿Qué porcentaje de valores de  $R$  es superior a 0.42? ¿Podemos sacar las mismas conclusiones que con el análisis anterior?
- Reducir los errores de  $a$  y  $c$  en un 100% y volver a ejecutar el Monte Carlo. ¿Qué conclusiones podemos extraer?

### Resolución

- El valor de  $R$  se calcula según las ecuaciones 10 y 11. El error  $\delta(R)$  mediante propagación cuadrática de errores (variables independientes) es:

$$\begin{aligned} \delta(R)^2 = & \left[ \frac{1}{D^2} (1 * D - (-a) * 2(1 - \frac{kd}{e})) \delta(a) \right]^2 + \left[ \frac{-a}{D^2} \frac{d}{ke} \delta(b) \right]^2 + \left[ \frac{a}{D^2} \frac{d}{ke} \delta(c) \right]^2 \\ & + \left[ \frac{-a}{D^2} \left( \frac{b-c}{ke} + \frac{2ka}{e} \right) \delta(d) \right]^2 + \left[ \frac{-a}{D^2} \left( -\frac{d(b-c)}{ke^2} - \frac{2kda}{e^2} \right) \delta(e) \right]^2 \end{aligned} \quad (12)$$

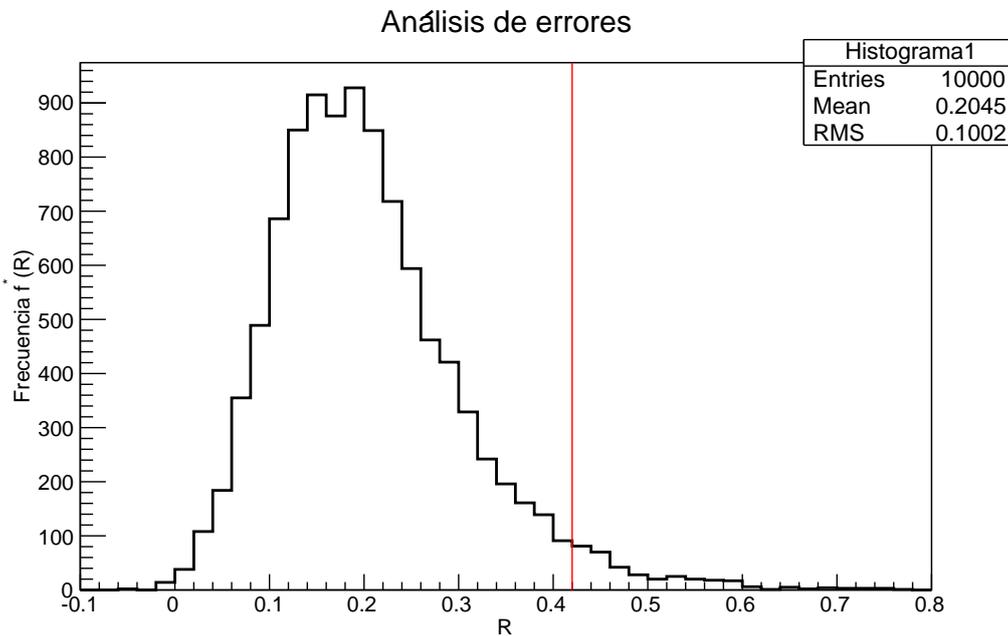
Obtenemos el siguiente resultado:

$$R = 0.19 \pm 0.09$$

- b) Según los cálculos del apartado anterior, la partícula objeto es inestable, pues  $R < 0.42$  y dicho umbral está a 2.5 desviaciones estándar. Cuando damos el valor de  $R \pm \delta(R)$  con su error, suponemos que al medirlo muchas veces obtendremos una distribución gaussiana centrada en  $\mu = R$  y con desviación  $\sigma = \delta(R)$ . La probabilidad de que sea estable es, según los valores tabulados del área de la gaussiana (o mediante la función error complementaria de ROOT TMath::Erfc( $|R - 0.42|/\delta(R)/\sqrt{2}$ )/2 bastante pequeña:

$$P(R > 0.42) \simeq 0.55\%$$

- c) Generamos valores aleatorios de  $a$ ,  $b$ ,  $c$ ,  $d$  y  $e$  aleatoriamente suponiendo que son variables gaussianas (centradas en la media y con desviación estándar dadas por la ecuación 11) mediante el método de Box-Müller y calculamos el valor de  $R$ . Repetimos el proceso 10000 veces y construimos el correspondiente histograma:



**Figura 6:** Histograma de valores de  $R$  según un sorteo aleatorio gaussiano (método de Box-Müller) de los parámetros. La línea roja marca el límite de estabilidad  $R = 0.42$

El valor medio y la desviación estándar muestral son:

$$\begin{aligned}\bar{R} &= 0.2048 \\ \sigma(R) &= 0.10\end{aligned}$$

- d) Si calculamos el porcentaje de valores de  $R$  superior a 0.42 a partir del histograma de la figura 6, se obtiene:

$$P(R > 0.42) \simeq 3.5\%$$

En este caso, estamos a escasas 2 desviaciones estándar de  $R = 0.42$  y la probabilidad de que la partícula sea estable es mayor que en el apartado b), donde quedaba prácticamente descartado (probabilidad menor del 1%). En cambio, aquí la probabilidad es apreciablemente mayor (7 veces más), con lo que no podemos tomar las mismas conclusiones con tanto nivel de confianza o seguridad.

Es decir, un análisis de errores completo, apto para errores grandes en los parámetros a diferencia de la propagación cuadrática de errores (aproximación para errores relativos pequeños), permite conocer realmente el nivel de confianza con que uno determina una magnitud física en función de sus errores. Nótese que la función error para la desviación estándar muestral daría una probabilidad diferente a la calculada mediante el histograma, pues presupone que la distribución de  $R$  es gaussiana.

- e) Las variables  $a$  y  $c$  son las que mayor error relativo tienen. Supongamos que  $a$  y  $c$  tuviesen error despreciable y repitamos todos los apartados anteriores.

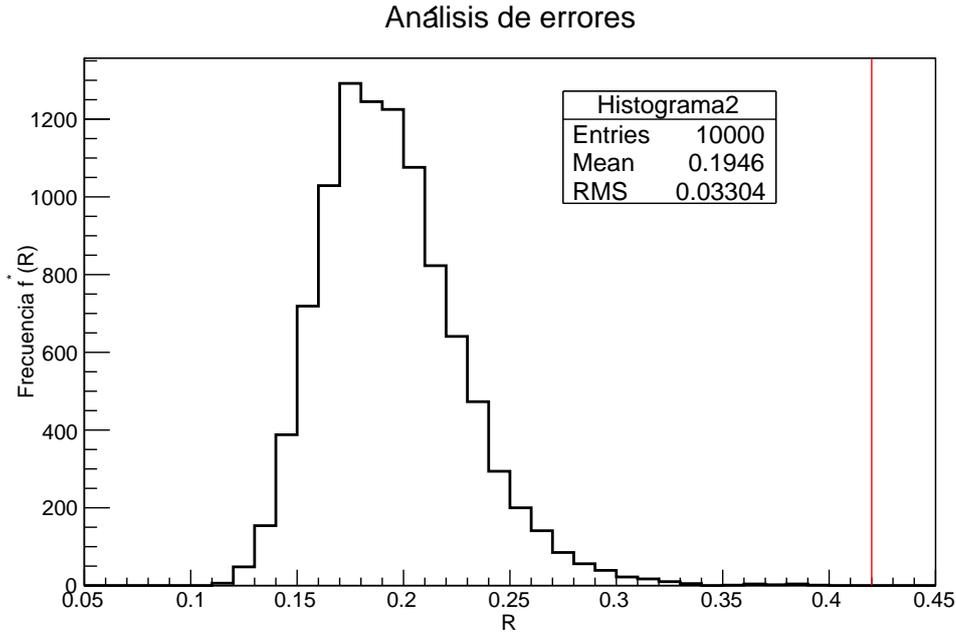
Mediante propagación cuadrática de errores (ecuación 12) se obtiene:

$$R = 0.19 \pm 0.03$$

Por tanto, la partícula objeto es inestable, pues  $R < 0.42$  y dicho umbral está a más de 7 desviaciones estándar. La probabilidad de que sea estable es, según la función error complementaria de ROOT TMath::Erfc( $|R - 0.42|/\delta(R)/\sqrt{2}$ )/2:

$$P(R > 0.42) \simeq 5.7e^{-12}\%$$

Si realizamos la simulación de Monte Carlo, se obtiene el siguiente histograma:



**Figura 7:** Histograma de valores de  $R$  según un sorteo aleatorio gaussiano (método de Box-Müller) de los parámetros, donde se ha considerado  $a$  y  $c$  sin error. La línea roja marca el límite de estabilidad  $R = 0.42$

El valor medio y la desviación estándar muestral son:

$$\begin{aligned} \bar{R} &= 0.1946 \\ \sigma(R) &= 0.03 \end{aligned}$$

Cabe señalar que en este caso el error mediante el método de Monte Carlo y la propagación cuadrática son iguales, dado que hemos eliminado los errores de los parámetros que presentaban un mayor error relativo ( $a$  y  $c$ ).

El porcentaje de valores de  $R$  superior a 0.42 a partir del histograma de la figura 7 es:

$$P(R > 0.42) = 0.0\%$$

En este caso, estamos a más de 7 desviaciones estándar de  $R = 0.42$  y la probabilidad de que la partícula sea estable es nula (al menos para esta simulación y el valor de  $N$  elegido). Como el  $N$  de la simulación de Monte Carlo es suficientemente grande (respecto al número de parámetros con error), podemos concluir que la partícula es inestable con una confianza del 100% (siempre que los errores experimentales se hayan determinado adecuadamente).

## Tema 8.- Integración Montecarlo

8.1.- Realizar la integral

$$\int_0^1 g(x)dx = \int_0^1 e^x dx$$

mediante Monte Carlo crudo para 100, 1000, 10000 y 100000 de puntos. Estimar no solo la integral sino el error de la estimación. ¿Disminuye el error con el número de puntos como es de esperar?

Repetir el cálculo de la integral 100 veces utilizando 1000 puntos diferentes cada vez y representar los resultados en un histograma. ¿La anchura de la distribución es la que esperamos?

Evaluar la integral también por los siguientes métodos:

- a) Montecarlo crudo.
- b) Utilizando el método de muestreo por importancia con la función de muestreo  $f(x) = 2(1+x)/3$ .
- c) Seleccionando valores aleatorios  $\xi$  en el intervalo  $[0, 1]$  y utilizando el estimador:

$$\bar{g} = [g(\xi) + g(1 - \xi)]/2$$

- d) Combinando los dos métodos anteriores, es decir, utilizar la función de importancia  $f(x)$  con el estimador  $\bar{g}$ .
- e) Igual que en el apartado (b) pero utilizando como función de importancia el desarrollo en serie de potencias de  $\bar{g}$  en torno a  $x = 1/2$  hasta el término cuadrático. Normalizar adecuadamente esta función para que pueda ser utilizada como una pdf.
- f) Combinando el método anterior con el método del apartado (c).

En todos los casos realizar el cálculo con  $N = 10000$  y obtener una estimación aproximada del error de la integral. Comparar con el resultado analítico.

8.2.- Sea la integral en seis dimensiones de la función:

$$g(\mathbf{x}, \mathbf{y}) = e^{-\mathbf{x}^2 - \mathbf{y}^2 - (\mathbf{x} - \mathbf{y})^2 / 2}$$

cuyo valor exacto en todo el dominio de integración  $[-\infty, \infty]$  es  $I = 10.9626$

- a) Realizar una estimación de la integral y su error mediante la técnica de Montecarlo crudo, con  $N = 1000000$ .
- b) Utilizando como función de importancia la función  $e^{-\mathbf{x}^2 - \mathbf{y}^2}$ .

Realizar ambas estimaciones en un dominio de integración  $[-R, R]$  con  $R = 5$ .

# -Métodos estadísticos-

## Problemas Tema 8:

### Integración Monte Carlo

Fernando Hueso González\*

Valencia, 15 de febrero de 2012

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

#### 8.1.- Integral de la función exponencial

**Enunciado** Realizar la integral

$$\int_0^1 g(x)dx = \int_0^1 e^x dx$$

mediante Monte Carlo crudo para 100, 1.000, 10.000 y 100.000 puntos. Estimar no sólo la integral sino el error de la estimación. ¿Disminuye el error con el número de puntos como es de esperar?

Repetir el cálculo de la integral 100 veces utilizando 1.000 puntos diferentes cada vez y representar los resultados en un histograma. ¿La anchura de la distribución es la que esperamos?

Evaluar la integral para  $N = 10.000$ , obtener una estimación aproximada del error de la integral y comparar con el resultado analítico para los siguientes métodos:

- Monte Carlo crudo
- Utilizando el método de muestreo por importancia con la función de muestreo  $f(x) = 2(1+x)/3$ .
- Seleccionando valores aleatorios  $\xi$  en el intervalo  $[0, 1]$  y utilizando el estimador:

$$\bar{g} = [g(\xi) + g(1 - \xi)]/2$$

- Combinando los dos métodos anteriores, es decir, utilizar la función de importancia  $f(x)$  con el estimador  $\bar{g}$ .
- Igual que en el apartado (b), pero utilizando como función de importancia el desarrollo en serie de potencias de  $\bar{g}$  en torno a  $x = 1/2$  hasta el término cuadrático. Normalizar adecuadamente esta función para que pueda ser utilizada como una pdf.
- Combinando el método anterior con el método del apartado (c).

---

\*ferhue#alumni.uv.es

**Resolución** Utilizo las fórmulas de Monte Carlo crudo para estimar la integral  $I = \int_0^1 g(x)dx = \int_0^1 e^x dx$  generando  $N$  números aleatorios  $x_i = \xi$  uniformes entre 0 y 1:

$$I \simeq \theta = \frac{1}{N} \sum_{i=1}^N g(x_i) \quad (1)$$

$$\sigma(\theta) \simeq \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^N g^2(x_i) - \left(\frac{1}{N} \sum_{i=1}^N g(x_i)\right)^2} \quad (2)$$

a comparar con los resultados analíticos:

$$I = e - 1 \quad (3)$$

$$\sigma(I) = \frac{1}{\sqrt{N}} ((e^2 - 1)/2 - (e - 1)^2) = \frac{1}{\sqrt{N}} (2e - 3/2 - e^2/2) \quad (4)$$

y obtengo los siguientes resultados:

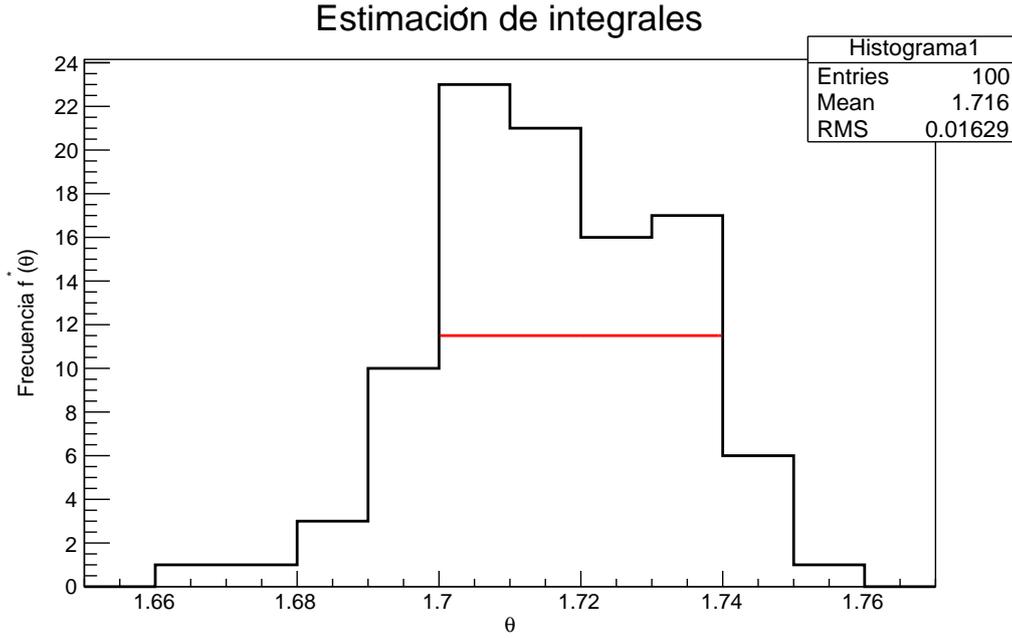
$N$	$\theta$	$\sigma(\theta)$	$\sigma(\theta)\sqrt{N}$	$(I - \theta)/\sigma(\theta)$
100	1.6839	0.0473	0.4733	0.73
1000	1.6723	0.0153	0.4830	3.01
10000	1.7157	0.0049	0.4876	0.53
100000	1.7166	0.0016	0.4913	1.09

**Tabla 1:** Estimación de la integral  $I = 1.718281$ ,  $\sigma(I)\sqrt{N} = 0.491971$

Se comprueba que según aumenta  $N$  nos vamos acercando más al valor analítico de la integral. El error estimado es del orden adecuado, pues es de una a tres desviaciones estándar respecto al valor real (ver columna 5 de la tabla 1).

Asimismo, el error disminuye con el número de puntos como la inversa de la raíz de  $N$ . Esto lo hemos comprobado a partir de la cuarta columna de la tabla anterior, donde el producto de ambas cantidades permanece constante. Además, este valor se va acercando al valor analítico que figura en el título de la tabla.

A continuación, calculamos el valor de la integral para  $N = 1000$  números aleatorios y repetimos el proceso  $n = 100$  veces. Representamos las estimaciones de la integral en el siguiente histograma:



**Figura 1:** Histograma de  $n = 100$  valores de  $\theta$  (estimaciones de la integral) calculados cada uno con  $N = 1000$  números aleatorios uniformes con  $n_{bin} = 12$ . La línea roja marca el *FWHM*.

A primera vista, podemos observar que la anchura de la distribución es del orden de  $FWHM \simeq 0.04$ , que está relacionado con la desviación estándar por la fórmula  $FWHM = 2\sqrt{2 \ln 2} \sigma$ , con lo que  $\sigma_{hist} \approx 0.017$ . Este valor se corresponde con el esperado analíticamente (ver ecuación 4):  $\sigma(I, N) \simeq 0.016$ .

A continuación, evaluamos la integral para  $N = 10.000$ , obtenemos una estimación aproximada de su error de la integral, y comparamos con el resultado analítico por distintos métodos:

- a) Monte Carlo crudo (ver página anterior).
- b) Mediante el método de muestreo por importancia con la función de muestreo  $f(x) = 2(1+x)/3$ . Nótese que esta función ya está normalizada. Construimos  $\bar{g}(x) = g(x)/f(x)$ . El valor esperado de dicha función será la integral, donde la *pdf* que gobierna la misma es la función  $f(x)$ .

$$\theta = E[\bar{g}(x)] = \frac{1}{N} \sum_{i=1}^N \bar{g}(x_i) \quad (5)$$

donde  $x_i \neq \xi_i$  no es uniforme (variable  $\xi$ ) sino que se corresponde con:

$$\int_0^x f(x') dx' = F(x) = (2x + x^2)/3 = \xi \implies x = -1 + \sqrt{1 + 3\xi} \quad (6)$$

La fórmula para estimar el error es similar a la de la ecuación 2:

$$\sigma(\theta) = \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^N g^2(x_i) - \theta^2} \quad (7)$$

El resultado obtenido es compatible con el analítico, con un error muy similar al del apartado a) (ecuación 3):

$$\theta = 1.7165 \pm 0.0016 \leftrightarrow I = 1.7183$$

Nótese que en estos y sucesivos apartados, a diferencia de la ecuación 4 en el caso a), no obtenemos una expresión analítica  $\sigma(I)$  para la varianza de la distribución al ser una pdf más elaborada y no simplemente la distribución uniforme.

- c) El siguiente método consiste en seleccionar valores aleatorios  $\xi$  en el intervalo  $[0, 1]$  y utilizar el estimador:

$$\bar{g} = [g(x(\xi)) + g(x(1 - \xi))]/2$$

Nótese que en este caso,  $x = \xi$ , pero en general  $x$  tendrá una dependencia distinta en  $\xi$ . Por ello, se especifica que lo que hay que cambiar es la  $\xi \rightarrow 1 - \xi$  y no la  $x \rightarrow 1 - x$  (aunque en este caso coincida). Las fórmulas empleadas son análogas a 1, 2 con tan sólo cambiar la expresión de  $g$  por  $\bar{g}$ . El resultado obtenido es compatible con el analítico (ecuación 3) y presenta menor error que en los casos a) y b):

$$\theta = 1.7180 \pm 0.0006 \leftrightarrow I = 1.7183$$

- d) Para combinar los dos métodos anteriores, debemos utilizar la función de importancia  $f(x)$ , que divide a la función  $g$ , y el nuevo estimador será  $\bar{g} = (g/f(x(\xi)) + g/f(x(1 - \xi)))/2$  (el promedio se hace sobre el cociente, no sobre la  $g$  original). Se obtienen los siguientes resultados:

$$\theta = 1.7185 \pm 0.0002 \leftrightarrow I = 1.7183$$

Se observa que el error es menor que en cada uno de los casos anteriores, al haber utilizado una combinación de ambos métodos (a costa de aumentar muy poco el volumen de cálculo).

- e) De manera análoga al apartado b), definimos una función de importancia a partir del desarrollo en serie de potencias de la exponencial hasta el término cuadrático en torno a  $x = 1/2$  (recuérdese que la función debe normalizarse):

$$f(x) = ke^x|_{2^\circ \text{ orden}} \approx ke^{0.5} (1 + (x - 0.5) + (x - 0.5)^2/2) \quad (8)$$

$$\int_0^1 f(x) dx = ke^{0.5} \left[ x + \frac{(x - 0.5)^2}{2!} + \frac{(x - 0.5)^3}{3!} \right]_0^1 = ke^{0.5} * 25/24 = 1 \Rightarrow k = \frac{24}{25} e^{-0.5} \quad (9)$$

$$f(x) = \frac{12}{25} \left( \frac{5}{4} + x + x^2 \right) \quad (10)$$

Debemos sortear los números aleatorios de acuerdo con  $f(x)$ , que se utilizarán para calcular el valor esperado de  $\bar{g} = g/f$ . Para ello, debemos utilizar el método de composición de variables (reescribimos  $f$  haciendo explícito los factores de peso  $\beta$  y funciones normalizadas  $h(x)$  en cada uno de los tres términos):

$$f(x) = \frac{3}{5} + \frac{6}{25} 2x + \frac{4}{25} 3x^2 = \sum_{i=1}^3 \beta_i h_i(x) \quad (11)$$

Generaré dos números aleatorios  $\xi_1, \xi_2$ .

$$\begin{cases} \text{Si } \xi_1 \in [0, 3/5] & \Rightarrow x = \xi_2 \\ \text{Si } \xi_1 \in (3/5, 21/25] & \Rightarrow x = \xi_2^{1/2} \\ \text{Si } \xi_1 \in (21/25, 1] & \Rightarrow x = \xi_2^{1/3} \end{cases} \quad (12)$$

Es decir, se aplica la transformada inversa en una de las tres funciones, que se selecciona con otro número aleatorio de acuerdo a su factor de peso relativo.

El resultado obtenido es compatible con el analítico, con menor error que en los casos anteriores, al haber hecho un desarrollo hasta segundo orden (a diferencia de la función de muestra del caso b), que es lineal):

$$\theta = 1.71836 \pm 0.00013 \leftrightarrow I = 1.71828$$

f) Si combinamos el método anterior con el del apartado c), análogamente a lo realizado en el apartado d) con la función de muestra de b), se obtienen los siguientes resultados:

$$\theta = 1.71835 \pm 0.00005 \leftrightarrow I = 1.71828$$

De nuevo el error es notablemente menor que en los casos anteriores, al haber combinado ambos métodos.

Por último, presentamos los resultados de todos los apartados en una misma tabla:

Apartado	$\theta$	$\sigma(\theta)$	$(I - \theta)/\sigma(\theta)$
a)	1.7166	0.0016	1.09
b)	1.7165	0.0016	1.08
c)	1.7180	0.0006	0.43
d)	1.7185	0.0002	-0.83
e)	1.71836	0.00013	-0.61
f)	1.71835	0.00005	-1.20

**Tabla 2:** Estimación de la integral  $I = 1.71828$  mediante diferentes métodos.

## 8.2.- Integral multidimensional

**Enunciado** Sea la integral en seis dimensiones  $I = \int_{V_6} g(\vec{x}, \vec{y}) d^3 \vec{x} d^3 \vec{y}$  de la función:

$$g(\vec{x}, \vec{y}) = e^{-\vec{x}^2 - \vec{y}^2 - (\vec{x} - \vec{y})^2 / 2}$$

cuyo valor exacto en todo el dominio de integración  $[-\infty, +\infty]$  es  $I = 10.9626$

a) Realizar una estimación de la integral y su error mediante la técnica de Monte Carlo crudo, con  $N = 1.000.000$

b) Utilizando como función de importancia la función  $e^{-\vec{x}^2 - \vec{y}^2}$

Realizar ambas estimaciones en un dominio de integración  $[-R, R]$  con  $R = 5$ .

## Resolución

- a) La integración mediante Monte Carlo crudo es similar a la de una dimensión, pero hay que transformar el volumen de integración a un hipercubo unidad de dimensión  $d = 6$  mediante el cambio de variable  $x_j(\xi_j) = -R + 2R\xi_j$ ,  $y_j(\xi_j) = -R + 2R\xi_j$  (con  $j = 1, \dots, 3$ ), con lo que habrá un factor adicional de volumen  $V_d = (2R)^6$  (jacobiano de la transformación):

$$\theta = V_d \frac{1}{N} \sum_{i=1}^N g(x_1^i(\xi_1^i), \dots) \quad ; \quad \sigma(\theta) \simeq \frac{1}{\sqrt{N}} \sqrt{\frac{V_d^2}{N} \sum_{i=1}^N g^2(x_1^i(\xi_1^i), \dots) - \theta^2} \quad (13)$$

El resultado obtenido es compatible con el analítico:

$$\theta = 10.3 \pm 1.0 \leftrightarrow I = 10.9626$$

- b) Si utilizamos como función de importancia  $f(x) = e^{-x^2} e^{-y^2} / \pi^3$  (nótese que hemos introducido un factor de normalización  $1/\sqrt{\pi}$  por cada dimensión), debemos generar números aleatorios según una distribución normal (método de Box-Müller o la función implementada en ROOT) con la regla  $x_j, y_j = G(0, 1)/\sqrt{2}$  para  $x$  e  $y$  por separado, pero también para cada componente del vector por separado (justificado al ser la covarianza nula), donde  $G$  es el generador de números gaussianos de media 0 y sigma 1. Esto se puede ver fácilmente a partir de la generación de números aleatorios de una distribución multinormal mediante la descomposición de Cholesky de la matriz de covarianza (diagonal en este caso).

Posteriormente, se calcula el valor esperado de  $g(x)$  dividido entre la función de importancia normalizada (en este caso no hay factores de volumen  $V_d$  en la expresión de  $\theta$  ni podemos delimitar el intervalo, ya que no hemos realizado ningún cambio de variable). Las fórmulas serían:

$$\theta = \frac{1}{N} \sum_{i=1}^N g(x_1^i, \dots) / f(x_1^i, \dots) \quad (14)$$

$$\sigma(\theta) \simeq \frac{1}{\sqrt{N}} \sqrt{\frac{1}{N} \sum_{i=1}^N (g/f)^2(x_1^i, \dots) - \theta^2} \quad (15)$$

$$x_j, y_j = G(0, 1)/\sqrt{2} = G(0, 1/\sqrt{2}) \quad , \quad j = 1, 2, 3 \quad (16)$$

El resultado obtenido es compatible con el analítico y hemos reducido el error en dos órdenes de magnitud en comparación con el método a), a costa de introducir una nueva función que cuesta algo más de muestrear (gaussiana frente a distribución uniforme):

$$\theta = 10.958 \pm 0.008 \leftrightarrow I = 10.963$$

Por último, presentamos los resultados de ambos apartados en una tabla:

Apartado	$\theta$	$\sigma(\theta)$	$(I - \theta)/\sigma(\theta)$
a)	10.3	1.0	0.64
b)	10.958	0.008	0.61

**Tabla 3:** Estimación de la integral  $I = 10.9626$  mediante ambos métodos.

## Problemas Métodos Estadísticos y Monte Carlo en Física (Curso 2011-2012)

Para la evaluación de la asignatura tenéis que entregar dos problemas resueltos, para lo cual os doy las dos opciones siguientes:

- Elegir dos de entre todos los problemas propuestos a continuación.
- Elegir uno de los problemas propuestos y un segundo problema libre que tenga que ver con vuestra investigación y en el que se utilicen de alguna manera las técnicas Monte Carlo que hemos visto en el curso.

### Problema 1

#### Factor geométrico de una Cámara de Niebla de Difusión

El objetivo del problema es calcular el factor geométrico de una DCC,  $\epsilon_G$ , (*Difusion Cloud Chamber*) para las partículas  $\alpha$  producidas por las desintegraciones del radón en el aire. Es decir, queremos saber cuantas de las desintegraciones  $\alpha$  observadas se han originado en el volumen efectivo del detector. Para ello suponemos que el volumen efectivo de la cámara consiste en una placa infinita de espesor igual a 0.5 cm. La longitud de las trazas  $\alpha$  es de 5 cm y éstas solo pueden detectarse procedentes desde una de las caras del detector pues en la otra se ha colocado una plancha metálica que impide el paso de las partículas  $\alpha$ .

Para calcular el factor geométrico se generan trazas uniformemente tanto dentro como fuera del volumen efectivo de la cámara y de las que atraviesen el detector, contamos tanto las que se han originado dentro como las que se han originado fuera.

- (a) Calcular dicho factor geométrico,  $\epsilon_G$ , definido como el cociente entre el número de trazas que han atravesado el detector y se han originado en el interior del mismo, dividido por el número total de trazas que han atravesado la cámara independientemente de donde se han originado.
- (b) Repetir el cálculo del factor geométrico pero solo para las trazas cuya longitud en la cámara,  $d$ , es mayor que un determinado valor. Calcular los valores de  $\epsilon_G$  para  $d = 1, 2, 3, 4$  y  $5$  y representarlo gráficamente.

Nota.- Podéis observar las trazas producidas por las partículas  $\alpha$  en la DCC del IFIC situada en el vestíbulo de la nave experimental. Las partículas  $\alpha$  son muy ionizantes y depositan toda su energía al interaccionar con la materia. En la DCC son trazas que se caracterizan por tener un grosor de varios milímetros de diámetro y una longitud de unos pocos cm.

### Problema 2

#### Distribución de Breit-Wigner y Resonancias

La distribución de Breit-Wigner que describe la forma de una resonancia, viene dada por la p.d.f:

$$f(x) = B(\Gamma, x_0) = \frac{1}{\pi} \frac{\Gamma}{(\Gamma^2 + (x - x_0)^2)}$$

donde  $\Gamma > 0$  es la anchura de la resonancia y  $x_0 > 0$  se interpreta como la masa de la misma.

- Aplicando el método de la transformación inversa diseñar un algoritmo para generar una distribución de Breit-Wigner cualquiera.
- Se puede demostrar que si  $z_1$  y  $z_2$  son dos variables aleatorias normales de  $N(0, 1)$ , la variable aleatoria  $y = z_1/z_2$  se distribuye según una Breit-Wigner  $B(1, 0)$ . Además, si  $y$  es una variable  $B(1, 0)$  entonces  $z = \Gamma y + x_0$  es una variable  $B(\Gamma, x_0)$ . Construir un algoritmo que genere una distribución de Breit-Wigner cualquiera de acuerdo con este algoritmo.
- Reproducir mediante un programa Monte Carlo un espectro con una resonancia de masa  $M_0 = 150$  MeV y anchura  $\Gamma = 12$  MeV en la que tenemos un fondo que se comporta según  $b(E) = 1/(a + bE)^3$  con  $a = 0.5$  y  $b = 0.005$  hasta un valor de la energía de 300 MeV. Para generar sucesos de fondo, normalizar su pdf entre 0 y 300 MeVs. Suponer que aproximadamente uno de cada tres sucesos es de fondo. Utilizar cualquiera de los algoritmos anteriores.
- Ajustar el espectro obtenido en el ejercicio anterior y comparar los parámetros ajustados con los parámetros utilizados en la generación. Tener en cuenta que para el ajuste hay que incluir dos parámetros de normalización uno para cada contribución y que hay que dar un valor inicial de los parámetros para comenzar el ajuste que sea similar pero no idéntico a los utilizados en la simulación.

### Problema 3

#### Desintegración nuclear a diferentes canales

El núcleo atómico  $^{226}\text{Ac}$ , que tiene un periodo de semidesintegración de  $T_{1/2} = 29$  h, se desintegra mediante emisión  $\beta$  un 83 % de las veces y mediante captura electrónica un 17 %.

- Obtener las vidas medias parciales de cada canal de desintegración  $\tau_A$  y  $\tau_B$ .
- Simular la desintegración del núcleo generando dos variables exponenciales aleatorias  $t_A$  y  $t_B$  de acuerdo con las distribuciones de cada modo de desintegración:

$$\frac{1}{\tau_A} e^{-t/\tau_A} \quad \frac{1}{\tau_B} e^{-t/\tau_B}$$

y obtener el espectro de la variable  $t = \min \{t_A, t_B\}$

- Mediante un ajuste, comprobar que dicho espectro corresponde a una distribución exponencial de la forma:

$$\frac{1}{\tau} e^{-t/\tau}$$

donde  $\tau$  es la vida media del  $^{226}\text{Ac}$ .

- Contar las veces que se produce en el Montecarlo cada modo de desintegración, calcular los porcentajes correspondientes y relacionarlos con los *branching ratios* de cada canal.
- Demostrar analíticamente que la variable aleatoria definida como  $\xi = \min \{\xi_1, \xi_2\}$ , donde  $\xi_1$  y  $\xi_2$  son dos variables exponenciales distribuidas según  $\lambda_1 e^{-\lambda_1 x}$  y  $\lambda_2 e^{-\lambda_2 x}$  respectivamente, se distribuye también exponencialmente como  $\lambda e^{-\lambda x}$ . Encontrar una expresión de  $\lambda$  en términos de  $\lambda_1$  y  $\lambda_2$
- Comentar todos los resultados obtenidos.

### Problema 4

#### Dispersión de electrones no relativistas en la materia

Queremos simular el paso de un haz incidente de electrones a través de un trozo de material homogéneo y calcular los porcentajes de trasmisión, backscattering y absorción para diferentes

grosos, energías y elementos. El modelo que supondremos es el denominado *continuous slowing down approximation*. Supongamos que en un momento dado, el electrón tiene una energía  $E_n$  y una dirección  $\mathbf{v}_n$  definida por los ángulos  $(\theta_n, \phi_n)$ . Cuando sufre una colisión elástica, el electrón es dispersado un ángulo  $\theta$  y  $\phi$  de su dirección original cambiando a una nueva dirección  $\mathbf{v}_{n+1}$  definida por los ángulos  $(\theta_{n+1}, \phi_{n+1})$ . El electrón continúa dentro del material ionizando átomos y perdiendo su energía de manera continua y constante hasta que sufre una nueva colisión elástica en la que cambia de nuevo su dirección. El proceso acaba cuando se cumple alguna de estas condiciones: el electrón logra atravesar todo el material (transmisión), el electrón es rebotado y sale por donde ha entrado (*backscattering*) o bien, el electrón pierde toda su energía y se absorbe (absorción).

- **Pérdida de energía**

La pérdida de energía por unidad de longitud depende de la energía cinética del electrón,  $E$ , y del número atómico del material,  $Z$ , y viene dada por la expresión de Bethe:

$$\frac{dE}{ds} = -\frac{2\pi e^4}{E} NZ \ln\left(\frac{2E(\text{eV})}{11.5Z}\right) \quad (1)$$

donde  $e$  es la carga del electrón y  $N$  es la densidad del blanco en átomos/cm<sup>3</sup>. Sin embargo, para facilitar los cálculos conviene expresar la pérdida de energía en unidades de KeV/ $\mu$ m con lo que la expresión anterior se convierte en:

$$\frac{dE}{ds} = -7.84\left(\frac{\rho Z}{AE}\right) \ln\left(\frac{173.9E}{Z}\right)\left(\frac{\text{KeV}}{\mu\text{m}}\right) \quad (2)$$

donde  $\rho$  es la densidad del material en g/cm<sup>3</sup>,  $A$  es la masa atómica en g/mol, y  $E$  es la energía cinética en KeV.

- **Sección eficaz diferencial**

Para la sección eficaz diferencial supondremos un potencial de Coulomb apantallado:

$$\frac{d\sigma}{d\Omega} = \frac{Z(Z+1)e^4}{p^2v^2} \frac{1}{(1-\cos\theta+2\beta)^2} \quad (3)$$

donde  $\beta = 0.25(1.12\lambda_0 h/(2\pi p))^2$ ,  $\lambda_0 = Z^{1/3}/(0.885a_0)$ ,  $p = mv$  y  $a_0$  es el radio de Bohr. El ángulo  $\theta$  representa el ángulo de dispersión o *scattering*. En términos de la energía cinética  $E$  y en unidades de KeV se obtiene  $\beta = 5.448 \times 10^{-3} Z^{2/3}/E$ . Integrando para todo el ángulo sólido se obtiene la sección eficaz total:

$$\sigma_T = \int_0^\pi \int_0^{2\pi} d\Omega \left(\frac{d\sigma}{d\Omega}\right) = \frac{\pi}{\beta(1+\beta)} \frac{Z(Z+1)e^4}{p^2v^2} \quad (4)$$

y por tanto el recorrido libre medio, que viene dado por:

$$\lambda = \frac{A}{\rho N_0 \sigma_T} = \frac{1.02\beta(1+\beta)AT^2}{Z(Z+1)\rho} \quad (5)$$

donde  $N_0$  es el número de Avogadro.

Deducir las expresiones 2, 4 y 5. A continuación construir un programa que simule el paso del haz incidente electrón a electrón del siguiente modo:

1. Dada una colisión, el ángulo  $\theta$  después de la colisión se puede obtener a partir de la expresión 3 utilizando el método de la transformación inversa.

2. Para obtener la dirección saliente tras el choque  $\mathbf{v}_{n+1}$  necesitamos un vector unitario arbitrario  $\mathbf{u}$  perpendicular a  $\mathbf{v}_n$ . Para obtenerlo, generamos un vector unitario arbitrario cualquiera  $\mathbf{q}$  isotropicamente, y calculamos el vector  $\mathbf{u} = \mathbf{v}_n \times \mathbf{q}$ , que por construcción será unitario y perpendicular a la dirección inicial  $\mathbf{v}_n$ . Finalmente, la nueva dirección vendrá dada por  $\mathbf{v}_{n+1} = \cos \theta \mathbf{v}_n + \sin \theta \mathbf{u}$ .
3. Generar la distancia  $s_n$  hasta la próxima colisión a partir del valor de  $\lambda$  de la expresión 5 para la energía del electrón en ese momento.
4. Calcular la energía con que llegará hasta la próxima colisión según la fórmula de Bethe:

$$E_{n+1} = E_n - \left| \frac{dE}{ds} \right| s_n$$

5. Comprobar si el electrón ha atravesado el material,  $z > D$ , siendo  $D$  el grosor del material, ha rebotado,  $z < 0$ , o bien ha sido absorbido completamente en el interior del material, ( $E_{n+1} < 0.5$  KeV).
6. Si ninguna de estas condiciones se da, volver al primer paso. En caso contrario aumentar el contador correspondiente y comenzar con un nuevo electrón desde el principio.

Una vez realizado el programa realizar las siguientes estimaciones:

- (a) Calcular el porcentaje de transmisión  $T$ , *backscattering*  $R$  y absorción  $A$  para un haz de electrones de 20 KeV que incide perpendicularmente sobre una lámina de aluminio de  $0.5 \mu\text{m}$ .
- (b) Repetir el apartado anterior para una lámina de oro de  $0.1 \mu\text{m}$ .
- (c) Obtener el espectro de energía de los electrones transmitidos en ambos casos.
- (d) Repetir los apartados (a) y (b) para diferentes grosores equiespaciados de Al y Au y representar los valores de  $R$  y  $T$  en función del grosor. Comentar los resultados.

Ayuda.- El ejercicio propuesto está basado en el siguiente artículo:

W. Williamson and G. C. Duncan, *Monte Carlo simulation of nonrelativistic electron scattering*, Am. J. Phys. **54**(3), 1986

## Problema 5

### Integración Monte Carlo

Realizar la integral

$$I = \int_0^1 g(x) dx = \int_0^1 x^3 dx$$

mediante Monte Carlo crudo para 100, 1000, 10000 y 100000 de puntos. Estimar no solo la integral sino el error de la estimación. ¿Disminuye el error con el número de puntos como es de esperar?

Repetir el cálculo de la integral 100 veces utilizando 1000 puntos diferentes cada vez y representar los resultados en un histograma. ¿La anchura de la distribución es la que esperamos?

Evaluar la integral también por los siguientes métodos:

- a) Mediante la técnica de rechazo (*hit or miss*).
- b) Utilizando el método de muestreo por importancia con la función de muestreo  $f(x) = x^2$ .
- c) Utilizando el método de muestreo estratificado dividiendo la región de integración en dos intervalos  $[0, 0.5]$  y  $[0.5, 1]$ , y utilizando  $f \cdot 10000$  puntos en  $[0, 0.5]$  y  $(1 - f) \cdot 10000$  en  $[0.5, 1]$ , donde  $f = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$  y  $0.9$ . Representar el error de  $I$  en función de  $f$ .

- d) Como en el apartado b) pero con  $f = 0.5$  y variando el tamaño de los intervalos  $[0, c]$  y  $[c, 1]$ , para los valores de  $c$  siguientes:  $c = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8$  y  $0.9$ . Representar el error de  $I$  en función de  $c$ .
- d) Seleccionando valores aleatorios  $\xi$  en el intervalo  $[0, 1]$  y utilizando el estimador:

$$\bar{g} = [g(\xi) + g(1 - \xi)]/2$$

En todos los casos donde no se especifique realizar el cálculo con  $N = 10000$  y obtener una estimación aproximada del error de la integral. Comparar con el resultado analítico.

# -Métodos estadísticos- Problemas Obligatorios

Fernando Hueso González\*

*Valencia, 5 de marzo de 2012*

MÁSTER DE FÍSICA AVANZADA 2011-2012 - *Itinerario de Física Nuclear y de Partículas*

## Problema 1: Factor geométrico de una cámara de niebla de difusión

### Enunciado

El objetivo del problema es calcular el factor geométrico de una DCC,  $\epsilon_G$ , (Diffusion Cloud Chamber) para las partículas  $\alpha$  producidas por las desintegraciones del radón en el aire. Es decir, queremos saber cuántas de las desintegraciones  $\alpha$  observadas se han originado en el volumen efectivo del detector. Para ello, suponemos que el volumen efectivo de la cámara consiste en una placa infinita de espesor igual a 0.5 cm. La longitud de las trazas  $\alpha$  es de 5 cm y éstas sólo pueden detectarse procedentes desde una de las caras del detector, pues en la otra se ha colocado una plancha metálica que impide el paso de las partículas  $\alpha$ .

Para calcular el factor geométrico se generan trazas uniformemente tanto dentro como fuera del volumen efectivo de la cámara y de las que atraviesan el detector, contamos tanto las que se han originado dentro como las que se han originado fuera.

- (a) Calcular dicho factor geométrico  $\epsilon_G$ , definido como el cociente entre el número de trazas que han atravesado el detector y se han originado dentro del mismo, dividido por el número total de trazas que han atravesado la cámara independientemente de dónde se han originado.
- (b) Repetir el cálculo del factor geométrico, pero sólo para las trazas cuya longitud en la cámara  $d$ , es mayor que un determinado valor. Calcular los valores de  $d = 1, 2, 3, 4$  y  $5$  y representarlos gráficamente.

Nota.- Podéis observar las trazas producidas por las partículas  $\alpha$  en la DCC del IFIC situada en el vestíbulo de la nave experimental. Las partículas  $\alpha$  son muy ionizantes y depositan toda su energía al interactuar con la materia. En la DCC son trazas que se caracterizan por tener un grosor de varios milímetros de diámetro y una longitud de unos pocos cm.

### Resolución

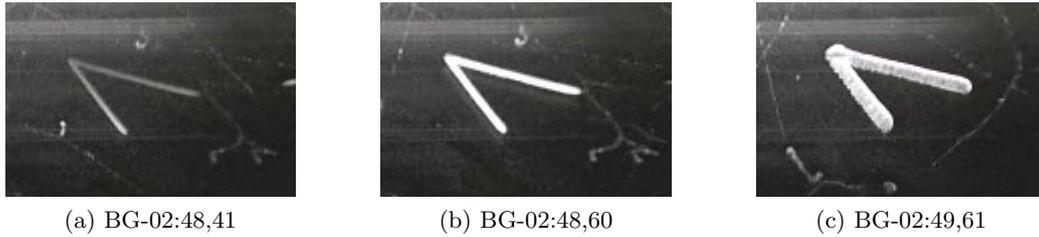
- (a) En la cámara de difusión, podemos observar partículas  $\alpha$  provenientes principalmente de los rayos cósmicos y de la serie del  $^{238}\text{U}$ , a causa de la presencia del gas radón (parte de dicha cadena) en la atmósfera. Para distinguir uno u otro origen, debemos tener en cuenta la energía cinética de las partículas  $\alpha$ , mucho mayor en la radiación cósmica que en la desintegración

---

\*ferhue#alumni.uv.es

de la serie mencionada. Por tanto, sólo se tiene en cuenta las partículas  $\alpha$  con una traza menor a 7 cm, que es la distancia máxima que recorre en el medio activo la partícula  $\alpha$  más energética de las desintegraciones en cadena del  $^{222}\text{Rn}$  (ver [2]). Aparte, no sólo la distancia sino también la topología del suceso nos puede ayudar a distinguir su origen: si se observa una traza con forma de V, será con casi total certeza debida a la cadena y no a los rayos cósmicos.

Cada punto o diferencial de volumen en la atmósfera emite con cierta probabilidad una partícula  $\alpha$  debida a la presencia del gas radón de forma isotrópica. Dichas partículas recorren una distancia  $L$  (cálculo detallado en apartado 2.4.2 de [2]), que es la media arimética para los tres hijos de la cadena.



**Figura 1:** Partículas  $\alpha$  en forma de V sin retroceso del núcleo intermedio (parte de la cadena del radón). Extraído de [2].

Sea la concentración de radón  $C_{rad}$  (desintegraciones por unidad de tiempo y volumen). El número de desintegraciones  $N_0$  que se **originan** en la cadena del radón ( $N_{iso}$  isótopos) en el interior del medio activo  $V_0$  en un tiempo  $T$  viene dado por:

$$N_0 = N_{iso}C_{rad}TV_0 \quad (1)$$

Nótese que multiplicamos por los distintos hijos de la cadena del radón  $N_{iso}$  que producen partículas  $\alpha$ .

El número de trazas  $N$  que seleccionamos sobre el total de las que se originan dentro del medio activo  $V_0$  se define por:

$$N = N_{iso}C_{rad}TV \quad (2)$$

En este apartado  $N = N_0$ ,  $V = V_0$ , pero no así en el apartado (b). Por otro lado, el número de trazas  $N'$  que **alcanzan** el medio activo independientemente de dónde se hayan originado lo definimos como:

$$N' = N_{iso}C_{rad}TV' = N_{iso}C_{rad}T(V + \hat{V}) \quad (3)$$

donde  $\hat{V} > 0$  ya que algunas trazas pueden proceder del aire situado por encima del medio activo  $V_0$  (de dimensiones  $a_0$ ,  $b_0$  y  $c_0$ ), es decir, que su trayectoria alcanza el mismo antes de depositar toda su energía (por debajo, suponemos que la placa metálica es gruesa y que el aire bajo ella emite partículas  $\alpha$  que se frenan antes de llegar al medio activo).

El factor geométrico  $\epsilon_G$ , según el enunciado es:

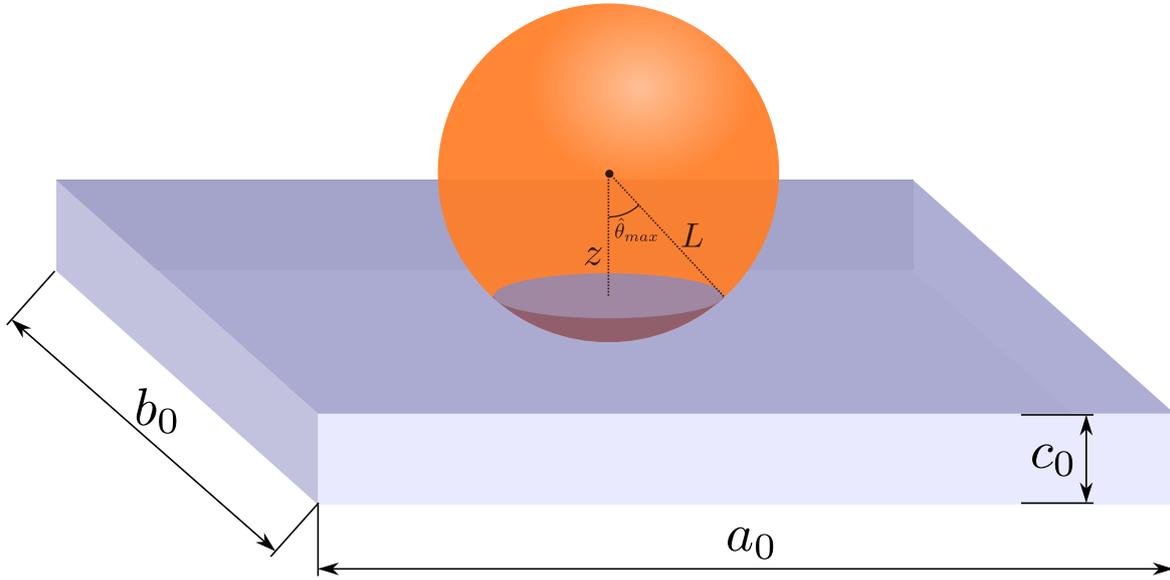
$$\epsilon_G \equiv \frac{N}{N'} = \frac{V}{V'} = \frac{V}{V + \hat{V}} = \frac{1}{1 + \hat{V}/V} \quad (4)$$

Lo deduciremos analíticamente a continuación suponiendo que el medio activo de anchura infinita ( $a_0, b_0 \gg c_0$ , plano infinito de espesor  $c_0$ , sólo hay dependencia en la dirección vertical, que denominamos  $z$ ), y que también verificaremos con una simulación con el método de Monte Carlo.

Sin embargo, el volumen  $\hat{V}$  no es directamente el volumen  $\bar{V} = a_0 b_0 L$  que **contiene** a los puntos externos a  $V_0$  que originan trazas que alcanzan el medio activo. Esto se debe a que este volumen también genera trazas que no lo alcanzan, a diferencia de  $V_0$ , donde todas las generadas están dentro somos capaces de identificarlas.

Por tanto, hay que ponderar este volumen  $\bar{V}$  por un factor  $\hat{\chi}(z)$  que tenga en cuenta qué porcentaje de desintegraciones **originadas** en un punto contenido en  $\bar{V}$  **alcanzan** el medio activo. Como el porcentaje es menor a la unidad, es fácil ver que  $\hat{V} < \bar{V}$ . Más aún, como por encima del medio activo, se pierden al menos la mitad de las desintegraciones (un punto justo por encima del medio activo sólo emite hacia el medio activo la mitad de veces; la otra mitad, en promedio, tiene dirección hacia arriba y no llega al medio), una cota superior sencilla a  $\hat{V}$  es:  $\hat{V} < \bar{V}/2$ .

El factor  $\hat{\chi}(z)$  se obtiene calculando<sup>1</sup> qué porcentaje de la radiación de todo el ángulo sólido entra en el detector. Para ello utilizamos el esquema de la figura 2.



**Figura 2:** Esquema del cálculo del factor de corrección geométrico

Como se observa, para un punto de coordenada  $z$  sobre el medio activo, la proporción detectada corresponde al área  $S_{int}$  de la esfera de radio  $L$  que intersecta con el medio activo respecto al área total  $S$ . Una analogía visual sería la situación de una pelota flotando en el agua que podemos hundir más o menos (coordenada  $z$ ) respecto al nivel del agua. Lo que tratamos de calcular es qué porcentaje ( $S_{int}/S$ ) de la superficie de la pelota está mojada.

El área  $S_{int}$  se calcula mediante la integración en coordenadas esféricas locales, centradas en el punto de coordenada  $z$  y con el eje hacia abajo:

$$z = L \cos(\hat{\theta}_{max}) \quad (5)$$

$$S = 4\pi L^2 \quad (6)$$

$$S_{int} = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\hat{\theta}_{max}} L^2 \sin(\theta) d\theta d\phi = 2\pi L^2 [1 - \cos(\hat{\theta}_{max})] \quad (7)$$

<sup>1</sup>Este cálculo es análogo al problema de la aguja del conde de Buffon, pero extendido a 3 dimensiones.

$$\hat{\chi} = S_{int}/S = \frac{1 - \cos(\hat{\theta}_{max})}{2} \quad (8)$$

Se comprueba que para los valores extremos, el peso se corresponde con la visión intuitiva del problema.  $\hat{\chi}(\hat{\theta}_{max} = 0) = 0$ : emisión se produce a una distancia  $L$  del medio activo y por tanto no lo alcanza y la contribución al volumen efectivo es nula.  $\hat{\chi}(\hat{\theta}_{max} = \pi/2) = 1/2$ : emisión justo en el borde del medio activo, sólo contribuye la mitad del ángulo sólido.

Para calcular el volumen efectivo de exceso  $\hat{V}$ , supondremos que:

$$\hat{V} = \hat{a}\hat{b}\hat{c} = a_0b_0\hat{c} \quad (9)$$

donde  $\hat{c}$  es la altura del volumen efectivo que contribuye a la radiación. El valor de  $\hat{c}$  se obtiene mediante integración de la ecuación 8 para valores de  $z$  desde 0 hasta  $L$ , es decir, pesando cada diferencial de  $z$  por la proporción que contribuye a la radiación:

$$\hat{c} = \int_{z=0}^L \hat{\chi}(\hat{\theta}_{max}) dz \quad (10)$$

Utilizando la relación 5 para sustituir  $\hat{\theta}_{max}$ :

$$\hat{c} = \int_{z=0}^L \frac{1 - z/L}{2} dz = \frac{z - \frac{z^2}{2L}}{2} \Big|_0^L = \frac{L}{2} - \frac{L}{4} = \frac{L}{4} \quad (11)$$

Y vemos que cumple la cota superior intuitiva a primera vista:

$$\hat{V} = ab\hat{c} = a_0b_0L/4 = \bar{V}/4 < \bar{V}/2 \quad (12)$$

Por consiguiente, el factor de corrección geométrico  $\epsilon_G$  coincide con la obtenida en [1] y es:

$$\epsilon_G = \frac{1}{1 + \frac{\hat{c}}{c}} = \frac{1}{1 + \frac{L/4}{c_0}} < 1 \quad (13)$$

De acuerdo con [2], el valor de  $L$  promedio para las distintas partículas  $\alpha$  de la cadena del radón a  $T=0^\circ\text{C}$  (medio activo y sus proximidades están a esa temperatura) es de  $5, 0 \pm 1, 2$  cm. El factor geométrico que se obtiene, conocido (ver [1]) el valor de  $c_0 = 0, 50 \pm 0, 25$  cm, es:

$$\epsilon_G = 0.29 \pm 0.11 = (29 \pm 11)\%$$

Finalmente, si se miden  $N'$  partículas  $\alpha$  (debidas exclusivamente a la desintegración del radón) en un cierto intervalo de tiempo  $T_0$ , despejamos de 3 y la concentración  $C_{rad}$  de radón por unidad de volumen y tiempo es:

$$C_{rad} = \frac{N'}{T_0V'} = \frac{N'}{N_{iso}T_0} \frac{\epsilon_G}{V} \quad \left[ \frac{\text{Bq}}{\text{m}^3} \right] \quad (14)$$

También se puede añadir a esta expresión otros factores (que en general son cercanos a la unidad): un factor de eficiencia de detección (incluyendo tiempo muerto) y de observación  $\epsilon_{obs}$  (mayor o menor que 1 según si sobreestimas o infraestimas el número de trazas al decidir si provienen o no de rayos cósmicos<sup>2</sup>). Los resultados experimentales en la práctica que realicé el año pasado se pueden consultar en [2].

<sup>2</sup>En [1],  $C_{rad} \propto \epsilon_{obs}$ , por tanto si sobreestimas  $N_0$  debes asociar un  $\epsilon_{obs} < 1$ , mientras que si lo subestimas,  $\epsilon_{obs} > 1$ . En mi opinión (quizá sea una errata en el artículo), sería más coherente con el lenguaje si  $\epsilon_{obs}$  estuviese dividiendo en lugar de multiplicando  $N_0$  de forma que si subestimas el conteo, la eficiencia sea menor que la unidad.

Por otro lado, comprobamos dicho factor con una simulación de Monte Carlo (anchura infinita). Para ello, a tenor de la ecuación 12, definimos  $\kappa = \hat{V}/\bar{V} = \frac{ab\hat{c}}{a_0b_0L} = \hat{c}/L$  y tenemos en cuenta que:

$$\epsilon_G = \frac{V}{V + \hat{V}} = \frac{1}{1 + \kappa\bar{V}/V} \quad (15)$$

En nuestro caso  $\bar{V} = a_0b_0L$  (ya definido anteriormente).  $\kappa = \hat{c}/L = 1/4$  (según el cálculo analítico) es equivalente al cociente del número de desintegraciones  $N_{det}$  originadas por encima del medio activo (en  $\bar{V}$ ) y que alcanzan el medio activo entre el número de desintegraciones  $\bar{N}$  generadas por encima del medio activo (en  $\bar{V}$ ) independientemente de si lo alcanzan o no. Matemáticamente, a partir de 4:

$$\epsilon_G \equiv \frac{N}{N'} = \frac{N}{N + N_{det}} = \frac{1}{1 + N_{det}/N} \quad (16)$$

Si genero  $\bar{N}$  desintegraciones en el volumen  $\bar{V}$ , por una simple regla de tres, en  $V_0 = V$  habrá  $N_0 = N = \bar{N}V/\bar{V}$ , con lo que la ecuación anterior queda:

$$\epsilon_G = \frac{1}{1 + \frac{N_{det}}{\bar{N}V/\bar{V}}} = \frac{1}{1 + \frac{N_{det}}{\bar{N}}\bar{V}/V} \quad (17)$$

Comparando con 15, hemos estimado  $\kappa = N_{det}/\bar{N}$  a partir de número de trazas que se pueden simular por Monte Carlo. El algoritmo de Monte Carlo, para anchura infinita, se basará en:

1.  $\bar{N} = 100.000$ ,  $i = 1$ ,  $N_{det} = 0$
2. Genero un número aleatorio uniforme  $\xi_z \in [0, 1]$  y calculo una posición aleatoria en el interior de  $\bar{V}$ :  $z = L\xi_z$ .
3. Calculo el ángulo polar  $\theta \in [0, \pi]$  mediante otro número aleatorio  $\xi_\theta \in [0, 1]$  y el método de la transformada inversa:  $\cos \theta = 1 - 2\xi_\theta$ .  $\phi$  es uniforme entre 0 y  $2\pi$  (no hace falta calcularlo al haber simetría azimutal)
4. Si  $L \cos(\theta) > z \Rightarrow N_{det}++$
5.  $i++$ . Si  $i \leq \bar{N}$ , volver al paso 2. Si no, voy al paso siguiente.
6. Calcular  $\kappa_{MC} = N_{det}/\bar{N} \pm \sqrt{N_{det}/\bar{N}}$  y comparar con el valor analítico  $\kappa_{teo} = 1/4$  (extraído de ecuación 13 y 15).

Otra opción sería:

1.  $N_{tot} = 100.000$ ,  $i = 1$ ,  $N' = 0$ ,  $N = 0$
2. Genero un número aleatorio uniforme  $\xi_z \in [0, 1]$  y calculo una posición aleatoria en el interior de  $(V_0 + \bar{V})$ :  $z = (L + c_0)\xi_z - c_0$ .
3. Si  $z < 0$ ,  $N'++$ ,  $N++$ , y saltar 6. Si no, ir al paso siguiente.
4. Calculo el ángulo polar  $\theta \in [0, \pi]$  mediante otro número aleatorio  $\xi_\theta \in [0, 1]$  y el método de la transformada inversa:  $\cos \theta = 1 - 2\xi_\theta$
5. Si  $L \cos(\theta) > z \Rightarrow N'++$
6.  $i++$ . Si  $i \leq N_{tot}$ , volver al paso 2. Si no, voy al paso siguiente.
7. Calcular  $\epsilon_{G,MC} = N'/N \pm \epsilon_{G,MC} \sqrt{1/N' + 1/N}$  y comparar con el valor analítico (ecuación 13).

En esta segunda variante se calcula directamente  $N'$  y  $N$ , con lo que puede resultar más intuitiva e igualmente válida.

El resultado obtenido mediante programación en ROOT para la primera variante de esta simulación de Monte Carlo es:

$$\kappa_{MC} = 0.2494 \pm 0.0010 \leftrightarrow \kappa_{teo} = 0.25$$

Comprobamos por tanto que ambos resultados son compatibles (a mayor  $\bar{N}$ , mejor estimación tendríamos).

Para la segunda variante:

$$\epsilon_{G,MC} = (28.6 \pm 0.3)\% \leftrightarrow \epsilon_{G,teo} = 28.6\%$$

Comprobamos por tanto que ambos resultados son compatibles (a mayor  $N_{tot}$ ), mejor estimación tendríamos. No obstante, en este cálculo sólo se ha tenido en cuenta el error estadístico y no el error de las magnitudes  $L$  y  $c_0$ , que para este caso introduce un error adicional de un 11% según propagación cuadrática de errores (caso analítico).

- (b) A la hora de contar trazas de partículas  $\alpha$ , puede haber errores a la hora de identificar trazas muy cortas (incluso pueden aparecer como puntos si cruzan el plano perpendicularmente), efecto que se incluye en un factor de eficiencia de observación  $\epsilon_{obs}$ . Para evitar estas trazas cortas que inducen a confusión, podemos recalculer el factor geométrico especificando que sólo tendremos en cuenta aquellas trazas cuya longitud (en el medio activo) sea mayor que una cota  $d$ .

Recapitulamos de manera compacta las fórmulas generales utilizadas hasta ahora, formalismo que aplicaremos para este caso:

$$\begin{aligned} \epsilon_G &= \frac{1}{1 + \hat{V}/V} = \frac{1}{1 + \hat{c}/c} \\ c &= \int dz \chi(z) \quad \text{con } z \in V_0 \quad ; \quad \hat{c} = \int dz \hat{\chi}(z) \quad \text{con } z \in \bar{V} \\ \chi(z) &= \frac{S_{int}}{S} = \frac{\Omega_{int}}{\Omega} = \frac{2\pi \int_{\theta_{min}}^{\theta_{max}} \sin \theta d\theta}{4\pi} = \frac{1}{2}(\cos \theta_{min} - \cos \theta_{max}) \quad ; \quad \text{idem con } \hat{\chi} \end{aligned} \quad (18)$$

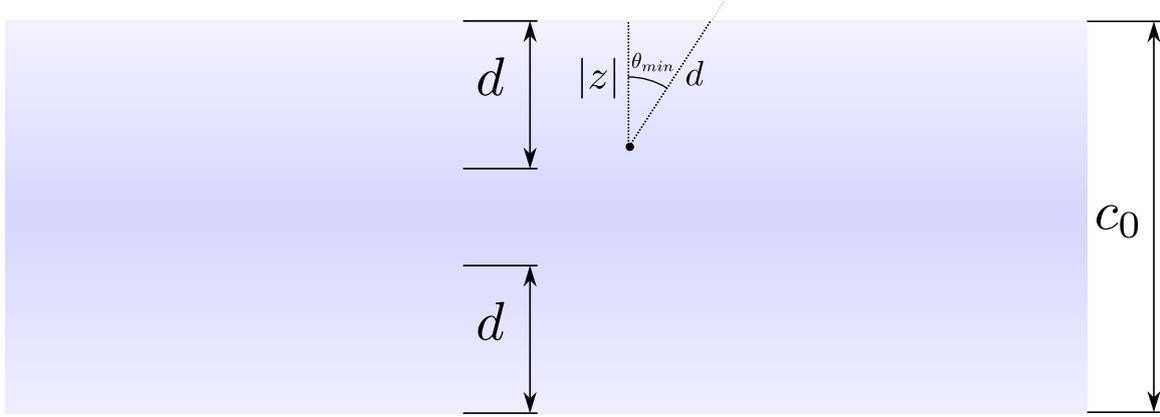
Al haber simetría azimutal, integramos la parte en  $\phi$  directamente.  $\chi$  se define como el ángulo sólido que contribuye a trazas que son **seleccionadas** en nuestro contaje dividido entre  $4\pi$ . Nótese que no hay que tener en cuenta aquellas que lo alcanzan pero no cumplen las condiciones exigidas. En el apartado anterior no exigíamos ninguna condición adicional.

Aparte,  $V_0 \equiv a_0 b_0 c_0$ ,  $V \equiv abc = a_0 b_0 c \neq V_0$  en general. En el apartado (a) sí coinciden porque  $\chi(z) = 1 \rightarrow c = c_0$  (al no haber exigido la condición extra de tener traza mayor que  $d$ ).

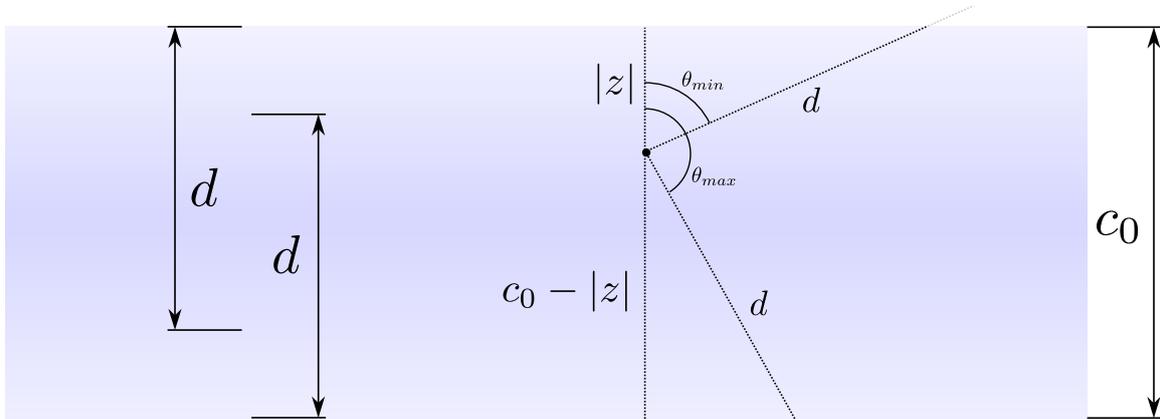
$\theta_{max,min}(z)$  son los ángulos extremos para los que se cumple la condición de selección de traza, y dependerán en general de  $z$ . La integral en  $z \in V_0$  se realiza entre  $z = -c_0$  y  $z = 0$ , mientras que en  $\bar{V}$  se realiza entre  $z = 0$  y  $z = z_{max}$ , donde  $z_{max}$  es aquel valor que cumple  $\theta_{min}(z_{max}) = \theta_{max}(z_{max})$ . El origen de  $z$  está justo en la frontera entre  $V_0$  y  $\bar{V}$ . Aparte,  $\hat{V} = \hat{a}\hat{b}\hat{c} = a_0 b_0 \hat{c} \neq \bar{V}$ , con  $\bar{V} \equiv a_0 b_0 z_{max}$ .

Es inmediato comprobar que se recupera el resultado del apartado (a) particularizando  $\theta_{min} = \hat{\theta}_{min} = 0$ ,  $\theta_{max} = \pi$ ,  $\cos \hat{\theta}_{max} = z/L \Rightarrow \chi = 1 \Rightarrow c = c_0$ ,  $\hat{\chi}(z) = (1 - z/L)/2$ ,  $z_{max} = L \Rightarrow \hat{c} = L/4$ . Pasamos ahora a analizar este apartado, distinguiendo el caso en que  $d < c_0$  y  $d > c_0$  (y  $d < L$  siempre).

**Caso  $d < c_0$ .** Sabemos que  $L > d$ , pero no importa si  $L < c_0$  ó  $L > c_0$ . El esquema es el siguiente:



(a)  $c_0 \geq 2d$



(b)  $2d \geq c_0 > d$

**Figura 3:** Esquema del cálculo del factor  $\chi$  (caso  $c_0 > d$ )

En el interior de  $V_0$ , de acuerdo con las convenciones que utilizo, el origen de  $z$  (muda) está en la cara superior del medio activo y apunta hacia arriba, y defino el ángulo polar respecto al punto de emisión y la dirección  $z$  positiva. Se observa que en la capa  $z \in [-d, 0]$ :

$$\begin{aligned} \cos \theta_{min} &= \frac{|z|}{d} = -\frac{z}{d} \\ \theta_{max} &= \pi \\ \chi(z) &= (1 - z/d)/2 \end{aligned} \quad (19)$$

Por otro lado, por simetría, la capa  $z \in [-c_0, -c_0 + d]$  contribuirá idénticamente a la anterior. Por último, en la capa  $z \in [-c_0 + d, -d]$ ,  $\theta_{min} = 0$ ,  $\theta_{max} = \pi \Rightarrow \chi(z) = 1$ . Juntando las contribuciones:

$$\begin{aligned} c &= \int_{z=-c_0}^{z=0} \chi(z) dz = \int_{z=-c_0+d}^{-d} dz + 2 \int_{z=-d}^0 (1 - z/d) dz / 2 \\ &= c_0 - 2d + d + d/2 = c_0 - d/2 \end{aligned} \quad (20)$$

La deducción a partir de la figura 3a es en principio sólo válida cuando  $c_0 \geq 2d$ . Ahora voy a comprobar que el valor de  $\hat{c}$  no cambia cuando  $c_0 \leq 2d$  utilizando la figura 3b.

Se observa que en la capa  $z \in [-c_0 + d, 0]$ :

$$\begin{aligned}\cos \theta_{min} &= \frac{|z|}{d} = -\frac{z}{d} \\ \theta_{max} &= \pi \\ \chi(z) &= (1 - z/d)/2\end{aligned}\tag{21}$$

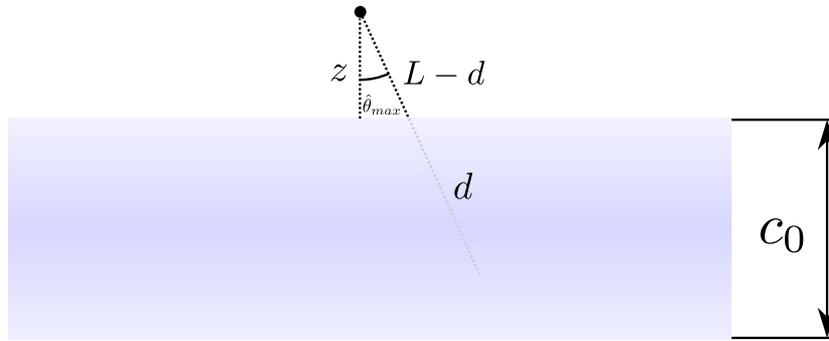
Por otro lado, por simetría, la capa  $z \in [-c_0, -d]$  contribuirá idénticamente a la anterior. Por último, en la capa  $z \in [-d, -c_0 + d]$ :

$$\begin{aligned}\cos \theta_{min} &= \frac{|z|}{d} = -\frac{z}{d} \\ \cos(\pi - \theta_{max}) &= -\cos \theta_{max} = \frac{c_0 - |z|}{d} = \frac{c_0 + z}{d} \\ \chi(z) &= \frac{c_0}{2d}\end{aligned}\tag{22}$$

Juntando las contribuciones, comprobamos que el resultado es idéntico:

$$\begin{aligned}c &= \int_{z=-c_0}^{z=0} \chi(z) dz = \int_{z=-d}^{-c_0+d} \frac{c_0}{2d} dz + 2 \int_{z=-c_0+d}^0 (1 - z/d) dz / 2 \\ &= \frac{c_0}{2d} (2d - c_0) + c_0 - d + \frac{c_0^2}{2d} + \frac{d^2}{2d} - \frac{c_0 d}{d} = c_0 - d/2\end{aligned}\tag{23}$$

Por otro lado, de acuerdo con la figura 4, en el volumen por encima del medio activo<sup>3</sup> (aquí el dibujo ya es independiente de si  $c_0 > 2d$  o no):



**Figura 4:** Esquema del cálculo del factor  $\hat{\chi}$  (caso  $c_0 > d$ )

Se observa que, para  $z > 0$ :

$$\begin{aligned}\hat{\theta}_{min} &= 0 \\ \cos \hat{\theta}_{max} &= \frac{z}{L-d} \\ z_{max} &= L-d \\ \hat{\chi}(z) &= \left(1 - \frac{z}{L-d}\right) / 2\end{aligned}\tag{24}$$

<sup>3</sup>Según mi convención, los ángulos polares los defino hacia arriba o hacia abajo según si estoy en el medio activo o por encima de él respectivamente.

$$\begin{aligned}\hat{c} &= \int_{z=0}^{z=L-d} \hat{\chi}(z) dz = \frac{1}{2} \int_0^{L-d} \left(1 - \frac{z}{L-d}\right) dz = \\ &= \frac{1}{2} \left(L-d - \frac{L-d}{2}\right) = \frac{L-d}{4}\end{aligned}\quad (25)$$

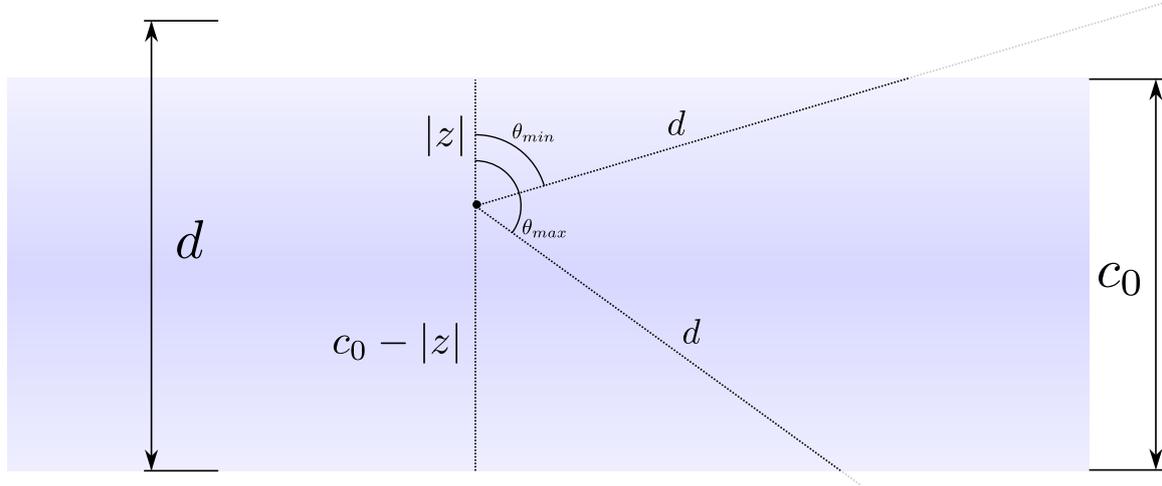
Y el factor geométrico queda:

$$\epsilon_G = \frac{1}{1 + \frac{L-d}{4(c_0 - d/2)}} \quad (26)$$

En el caso  $d = 0$  recuperamos exactamente el resultado del apartado (a).

**Caso  $d > c_0$ .** En este caso, la condición  $L > d$  implica  $L > d > c_0 \Rightarrow L > c_0$  siempre.

Por otro lado, de acuerdo con la figura 5, en el volumen del medio activo:



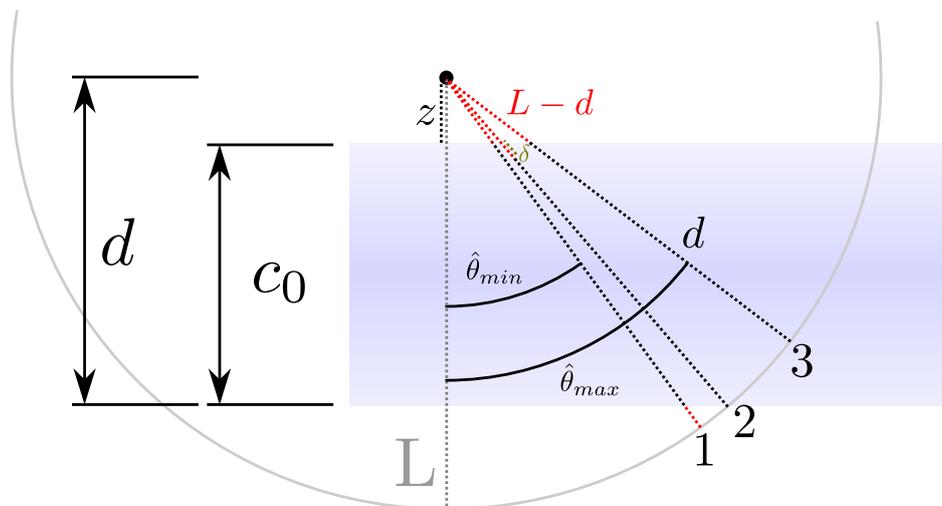
**Figura 5:** Esquema del cálculo del factor  $\chi$  (caso  $c_0 > d$ )

Se observa que, para  $z < 0$ :

$$\begin{aligned}\cos \theta_{min} &= -\frac{z}{d} \\ \cos(\pi - \theta_{max}) &= -\cos \theta_{max} = \frac{c_0 + z}{d} \\ \chi(z) &= \frac{c_0}{2d}\end{aligned}\quad (27)$$

$$c = \int_{z=-c_0}^{z=0} \chi(z) dz = \int_{-c_0}^0 \frac{c_0}{2d} dz = \frac{c_0^2}{2d} \quad (28)$$

Para  $z > 0$ , tenemos un esquema más complicado (figura 6, código de colores<sup>4</sup>).



**Figura 6:** Esquema del cálculo del factor  $\chi$  (caso  $c_0 > d$ )

A una distancia  $z$  arbitraria, la traza de mayor longitud corresponde a la de la línea 2, por trigonometría. Las líneas 1 y 3 correspondrán a los ángulos límite, donde la traza es justo  $d$ . Por tanto, añadiendo las debidas líneas auxiliares a la figura 6, podemos deducir las siguientes relaciones:

$$\begin{aligned}
 \cos \hat{\theta}_{min} &= \frac{c_0}{d} \\
 \cos \hat{\theta}_{max} &= \frac{z}{L-d} \\
 \hat{\chi}(z) &= \left( \frac{c_0}{d} - \frac{z}{L-d} \right) / 2 \\
 z_{max} &= (L-d) \frac{c_0}{d}
 \end{aligned} \tag{29}$$

Otra manera de deducir  $z_{max}$  es exigir que  $\delta \geq 0$  (figura 6) para que el problema tenga solución, pues la línea 2 corresponde a la traza de mayor longitud ( $d + \delta$ ). Como  $\cos \hat{\theta}_2 = \frac{c_0}{d+\delta} = \frac{z+c_0}{L} \Rightarrow \delta = \frac{c_0 L}{c_0+z} - d \geq 0 \Rightarrow z \leq \frac{c_0}{d}(L-d) \equiv z_{max}$ .

Y el factor geométrico queda:

$$\begin{aligned}
 \hat{c} &= \int_{z=0}^{z=c_0(L-d)/d} \hat{\chi}(z) dz = \frac{1}{2} \int_0^{c_0(L-d)/d} \left( \frac{c_0}{d} - \frac{z}{L-d} \right) dz = \\
 &= \frac{1}{2} \left( \frac{c_0^2}{d^2}(L-d) - \frac{c_0^2}{2d^2}(L-d) \right) = \frac{c_0^2}{4d^2}(L-d) \\
 \epsilon_G &= \frac{1}{1 + \frac{L-d}{2d}} = \frac{2d}{d+L}
 \end{aligned} \tag{30}$$

<sup>4</sup>Nótese el código de colores en las trazas dibujadas: el gris corresponde a una esfera de radio  $L$ , el negro a la distancia  $d$  y el rojo a la distancia  $L-d$ , aun cuando esté partido en 2 contribuciones (línea 1). El verde lima corresponde a  $\delta$ .

Es inmediato comprobar que en  $d = c$ , ambos casos calculados en el apartado (b) coinciden. En nuestro montaje experimental, se da que  $d > c$ , con lo que habrá que emplear la segunda variante a la hora de plantear algoritmo de Monte Carlo.

Comparamos con la simulación de Monte Carlo, mediante el siguiente algoritmo (análogo a la segunda variante del apartado anterior):

1.  $N_{tot} = 100.000$ ,  $i = 1$ ,  $N' = 0$ ,  $N = 0$
2. Genero un número aleatorio uniforme  $\xi_z \in [0, 1]$  y calculo una posición aleatoria en el interior de  $(V_0 + \bar{V})$ :  $z = (\frac{c_0}{d}(L - d) + c_0)\xi_z - c_0$ .
3. Calculo el ángulo polar  $\theta \in [0, \pi]$  mediante otro número aleatorio  $\xi_\theta \in [0, 1]$  y el método de la transformada inversa:  $\cos \theta = 1 - 2\xi_\theta$ .
4. Si  $z < 0$ , ir al paso siguiente. Si no, saltar a paso 6.
5. Si  $\cos \theta \in [-(z + c_0)/d, -z/d] \Rightarrow N++, N'++$ . Saltar al paso 7 en cualquier caso.
6. Si  $\cos \theta \in [z/(L - d), c_0/d] \Rightarrow N'++$ . Ir al paso siguiente.
7.  $i++$ . Si  $i \leq N_{tot}$ , volver al paso 2. Si no, ir al paso siguiente.
8. Calcular  $\epsilon_{G,MC} = N'/N \pm \epsilon_{G,MC} \sqrt{1/N' + 1/N}$  y comparar con el valor analítico (ecuación 13).

*Nota: este algoritmo es el más eficiente para este problema, pero ha requerido de una base analítica elaborada previamente. Se podría hacer un algoritmo más rudimentario, en el que simplemente generásemos desintegraciones en  $V$  y en un volumen menos restrictivo, por ejemplo  $\bar{V} = a_0 b_0 (L - d)$ , calculásemos en todos los casos el vector o segmento asociado a la traza, sus intersecciones con el medio activo (si las hay) y la longitud entre los puntos de intersección. Si esta longitud es mayor que  $d$ , se tiene en cuenta el suceso. Este algoritmo no requiere haber hecho la deducción analítica anterior, y saldrán resultados equivalentes, pero es menos eficiente y preciso (para un mismo valor de  $N_{tot}$ ) ya que muchos de estos sucesos no son útiles al no haber restringido suficientemente  $\bar{V}$ , y además la selección o no del suceso es más costosa.*

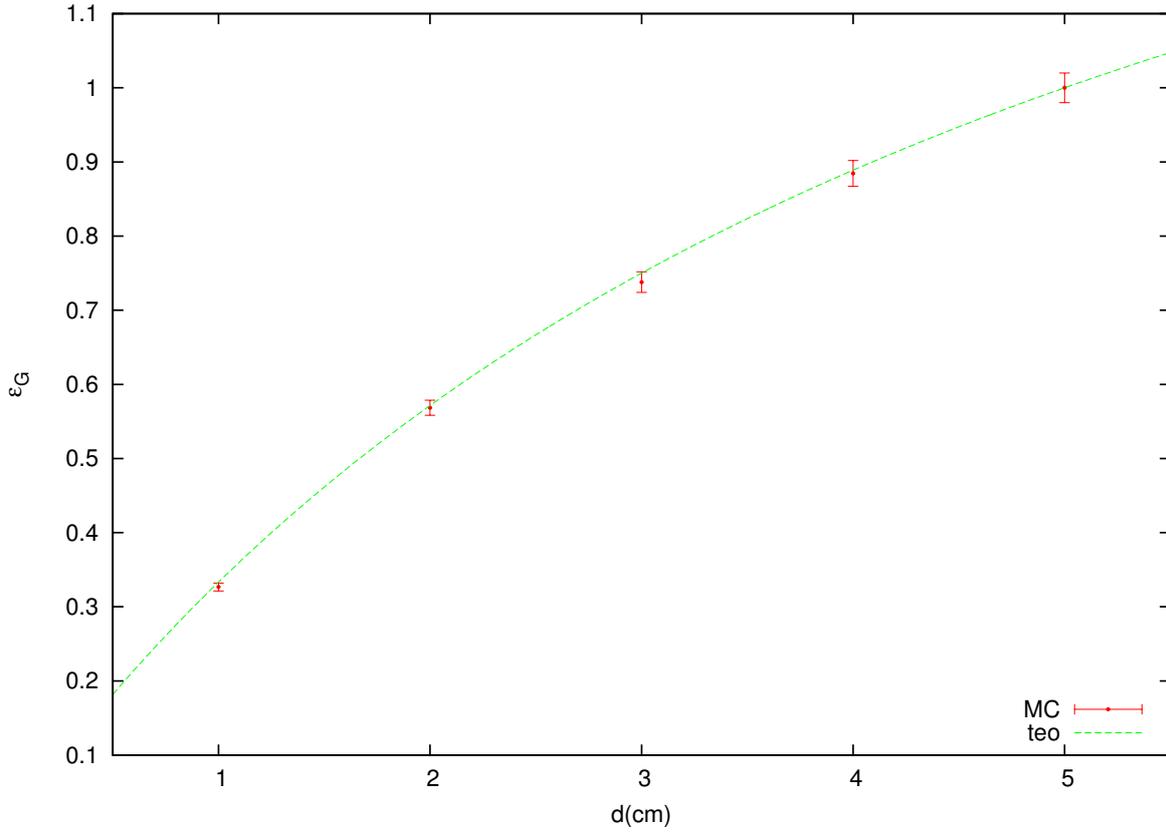
Obtenemos los siguientes resultados:

$d$	$\epsilon_{G,MC}$	$\epsilon_{G,teo}$
1	$0.327 \pm 0.005$	0.333
2	$0.569 \pm 0.010$	0.571
3	$0.738 \pm 0.014$	0.750
4	$0.885 \pm 0.017$	0.889
5	$1.00 \pm 0.02$	1.00

**Tabla 1:** Comparación del factor geométrico  $\epsilon_G$  calculado analíticamente (*teo*) y mediante la simulación (*MC*). No se tienen en cuenta los errores de  $d$ ,  $L$  y  $c_0$ .

Se comprueba en todos los casos que los valores simulados son compatibles con los analíticos (dentro del margen de incertidumbre).

Finalmente, representamos la tabla anterior en una gráfica:



**Figura 7:** Comparación del factor geométrico  $\epsilon_G$  calculado analíticamente (*teo*) y mediante la simulación (*MC*). No se tienen en cuenta los errores de  $d$ ,  $L$  y  $c_0$ .

Se comprueba que el factor geométrico tiende a la unidad. Esto se debe a que cuanto más estrictos somos a la hora de seleccionar las trazas, menos proceden del volumen exterior a  $V$  y  $N'$  tiende a  $N$ . Esto no es especialmente deseable ya que en tal caso estaríamos despreciando un gran número de trazas y perdemos por tanto estadística.

Por último, cabe comentar que los efectos de borde se podrían tener en cuenta mediante una simulación de Monte Carlo más elaborada, pero no analíticamente. En cualquier caso, estos efectos son despreciables pues  $c_0 \ll a_0, b_0$  en nuestro montaje experimental, y el error del valor de  $c_0$  introduce mucha más incertidumbre que la anchura finita de la placa, por lo que no vale la pena complicar el desarrollo.

## Problema 4: Dispersión de electrones no relativistas en la materia

### Enunciado

Queremos simular el paso de un haz incidente de electrones a través de un trozo de material homogéneo y calcular los porcentajes de transmisión, backscattering y absorción para diferentes grosores, energías y elementos. El modelo que supondremos es el denominado *continuous slowing down approximation*.

Supongamos que en un momento dado, el electrón tiene una energía  $E_n$  y una dirección  $\mathbf{v}_n$  definida por los ángulos  $(\theta_n, \phi_n)$ . Cuando sufre una colisión elástica, el electrón es dispersado un ángulo  $\theta$  y  $\phi$  de su dirección original cambiando a una nueva dirección  $\mathbf{v}_{n+1}$  definida por los ángulos  $(\theta_{n+1}, \phi_{n+1})$ .

El electrón continúa dentro del material ionizando átomos y perdiendo su energía de manera continua y constante hasta que sufre una nueva colisión elástica en la que cambia de nuevo su dirección. El proceso acaba cuando se cumpla alguna de estas condiciones: el electrón logra atravesar todo el material (transmisión), el electrón es rebotado y sale por donde ha entrado (backscattering), o bien el electrón pierde toda su energía y se absorbe (absorción).

- **Pérdida de energía**

La pérdida de energía por unidad de longitud depende de la energía cinética del electrón,  $E$ , y del número atómico del material,  $Z$ , y viene dada por la expresión de Bethe (sistema cgs-gaussiano):

$$\frac{dE}{ds} = -\frac{2\pi e^4}{E} N Z \ln \left( \frac{2E(\text{eV})}{11.5Z} \right) \quad (31)$$

donde  $e$  es la carga del electrón y  $N$  es la densidad del blanco en átomos/cm<sup>3</sup>. Sin embargo, para facilitar los cálculos conviene expresar la pérdida de energía en unidades de keV/ $\mu\text{m}$  con lo que la expresión anterior se convierte en:

$$\frac{dE}{ds} = -7.84 \frac{\rho Z}{AE} \ln \left( \frac{173.9E}{Z} \right) \quad \left[ \frac{\text{keV}}{\mu\text{m}} \right] \quad (32)$$

donde  $\rho$  es la densidad del material en g/cm<sup>3</sup>,  $A$  es la masa atómica en g/mol, y  $E$  es la energía cinética en keV.

- **Sección eficaz diferencial**

Para la sección eficaz diferencial supondremos un potencial de Coulomb apantallado:

$$\frac{d\sigma}{d\Omega} = \frac{Z(Z+1)e^4}{p^2v^2} \frac{1}{(1 - \cos\theta + 2\beta)^2} \quad (33)$$

donde  $\beta = 0.25(1.12\lambda_0 h/(2\pi p))^2$ ,  $\lambda_0 = Z^{1/3}/(0.885a_0)$ ,  $p = mv$  y  $a_0$  es el radio de Bohr. El ángulo  $\theta$  representa el ángulo de dispersión o *scattering*. En términos de la energía cinética  $E$  y en unidades de keV, se obtiene  $\beta = 5.448 \times 10^{-3} Z^{2/3}/E$ . Integrando para todo el ángulo sólido se obtiene la sección eficaz total:

$$\sigma_T = \int_0^\pi \int_0^{2\pi} d\Omega \left( \frac{d\sigma}{d\Omega} \right) = \frac{\pi}{\beta(1+\beta)} \frac{Z(Z+1)e^4}{p^2v^2} \quad (34)$$

y por tanto el recorrido libre medio, que viene dado por:

$$\lambda = \frac{A}{\rho N_0 \sigma_T} = \frac{1.02\beta(1+\beta)AT^2}{Z(Z+1)\rho} \quad (35)$$

donde  $N_0$  es el número de Avogadro.

Deducir las expresiones 32, 34 y 35 a partir de 31 y 33. A continuación, construir un programa que simule el paso del haz incidente electrón a electrón del siguiente modo:

1. Dada una colisión, el ángulo  $\theta$  después de la colisión se puede obtener a partir de la expresión 33 utilizando el método de la transformación inversa.
2. Para obtener una dirección saliente tras el choque  $\mathbf{v}_{n+1}$  necesitamos un vector unitario arbitrario  $\mathbf{u}$  perpendicular a  $\mathbf{v}_n$ . Para obtenerlo, generamos un vector unitario arbitrario cualquiera  $\mathbf{q}$  isotrópicamente, y calculamos el vector  $\mathbf{u} = \mathbf{v}_n \times \mathbf{q} / |\mathbf{v}_n \times \mathbf{q}|$  que por construcción será unitario y perpendicular a la dirección inicial  $\mathbf{v}_n$ . Finalmente, la nueva dirección vendrá dada por  $\mathbf{v}_{n+1} = \cos \theta \mathbf{v}_n + \sin \theta \mathbf{u}$ .
3. Generar la distancia  $s_n$  hasta la próxima colisión a partir del valor de  $\lambda$  de la expresión 35 para la energía del electrón en ese momento.
4. Calcular la energía con que llegará hasta la próxima colisión según la fórmula de Bethe:

$$E_{n+1} = E_n - \left| \frac{dE}{ds} \right| s_n \quad (36)$$

5. Comprobar si el electrón ha atravesado el material,  $z > D$ , siendo  $D$  el grosor del material, ha rebotado,  $z < 0$ , o bien ha sido absorbido completamente en el interior del material ( $E_{n+1} < 0.5 \text{keV}$ ).
6. Si ninguna de estas condiciones se da, volver al primer paso. En caso contrario, aumentar el contador correspondiente y comenzar con un nuevo electrón desde el principio.

Una vez realizado el programa, realizar las siguientes estimaciones:

- (a) Calcular el porcentaje de transmisión  $T$ , *backscattering*  $R$  y absorción  $A$  para un haz de electrones de 20 keV que incide perpendicularmente sobre una lámina de aluminio de 0.5  $\mu\text{m}$ .
- (b) Repetir el apartado anterior para una lámina de oro de 0.1  $\mu\text{m}$ .
- (c) Obtener el espectro de energía de los electrones transmitidos en ambos casos.
- (d) Repetir los apartados (a) y (b) para diferentes grosores equiespaciados de Al y Au y representar los valores de  $R$  y  $T$  en función del grosor. Comentar los resultados.

Ayuda.- El ejercicio propuesto está basado en el siguiente artículo:

W. Williamson and G. C. Duncan, *Monte Carlo simulation of nonrelativistic electron scattering*, Am. J. Phys. **54**(3), 1986

## Resolución

**Deducción de la expresión 32** Partimos de la fórmula de Bethe (ecuación 31), donde en el argumento del logaritmo la energía está expresada en eV. Como el logaritmo es adimensional, esto es equivalente a decir que hay un factor  $E(\text{eV}) \equiv E/1\text{eV}$ . Si queremos que aparezca  $E(\text{keV}) \equiv E/1\text{keV}$ , comprobamos que están relacionados por un factor 1000:

$$E(\text{eV}) \equiv \frac{E}{1\text{eV}} = \frac{E}{1\text{eV}} \frac{1\text{keV}}{1\text{keV}} = \frac{E}{1\text{keV}} \frac{1\text{keV}}{1\text{eV}} = \frac{E}{1\text{keV}} 1000 = 10^3 \frac{E}{1\text{keV}} \equiv 10^3 E(\text{keV}) \quad (37)$$

Sustituyendo el resultado anterior en el logaritmo de la expresión 31 y calculando el factor numérico global, su argumento nos queda (con un decimal significativo):

$$173.913 * E(\text{keV})/Z \approx 173.9 * E(\text{keV})/Z \quad (38)$$

Por otro lado, el factor que precede al logaritmo está expresado en unidades del sistema cgs-gaussiano (carga  $e$  expresada en  $esu$ ,  $E$  en  $\text{g}\cdot\text{cm}^2\text{s}^{-2}$  y  $N$  en  $\text{cm}^{-3}$ ). Sabiendo que la carga del electrón  $e = 4.803 \times 10^{-10} esu$  y que  $1esu = 1\text{g}^{1/2}\text{cm}^{3/2}\text{s}^{-1}$ , queda que, en el sistema cgs:

$$\frac{dE}{ds}(\text{cgs}) = -3.344 \times 10^{-37} \frac{NZ}{E}(\text{cgs}) \ln \left( \frac{173.9E(\text{keV})}{Z} \right) \quad (39)$$

Cambiamos las energías que todavía están expresadas en el sistema cgs a unidades de keV:

$$\begin{aligned} E(\text{cgs}) &\equiv \frac{E}{\text{gcm}^2\text{s}^{-2}} = 10^{3+4} \frac{E}{\text{kgm}^2\text{s}^{-2}} = 10^7 \frac{E}{\text{J}} = 10^7 \frac{E}{\text{keV} \frac{\text{J}}{\text{keV}}} \\ &= 10^7 \frac{E}{\text{keV} \frac{\text{J}}{1.602 \times 10^{-16} \text{J}}} = 1.602 \times 10^{-9} \frac{E}{\text{keV}} \equiv 1.602 \times 10^{-9} E(\text{keV}) \end{aligned} \quad (40)$$

Y la distancia  $s$  de cm a  $\mu\text{m}$ :

$$s(\text{cgs}) \equiv s/1\text{cm} = s/10^4 \mu\text{m} \equiv 10^{-4} s(\mu\text{m}) \quad (41)$$

Si sustituimos las relaciones 40 y 41 en 39, queda que:

$$\begin{aligned} \frac{dE}{ds}(\text{keV}/\mu\text{m}) &= -\frac{10^{-4}}{(1.602 \times 10^{-9})^2} * 3.344 \times 10^{-37} \frac{NZ}{E(\text{keV})} \ln \left( \frac{173.9E(\text{keV})}{Z} \right) \\ &= -1.303 \times 10^{-23} \frac{NZ}{E(\text{keV})} \ln \left( \frac{173.9E(\text{keV})}{Z} \right) \end{aligned} \quad (42)$$

Por último, sabemos que el número de átomos por  $\text{cm}^3$   $N$  se puede reescribir en función de la densidad  $\rho$  ( $\text{g}/\text{cm}^3$ ), el peso atómico molar  $A$  ( $\text{g}/\text{mol}$ ) y la constante de Avogadro  $N_0$  (átomos/mol) como:

$$N(\text{cm}^{-3}) = \frac{N_{\text{atomos}}}{\text{Vol}(\text{cm}^{-3})} = \frac{N_{\text{atomos}}}{m(\text{g})/\rho(\text{gcm}^{-3})} = \rho \left( \frac{\text{g}}{\text{cm}^3} \right) \frac{N_{\text{atomos}}}{1\text{mol}} \frac{1\text{mol}}{m(\text{g})} = \rho \left( \frac{\text{g}}{\text{cm}^3} \right) \frac{N_0}{A(\text{g/mol})} \quad (43)$$

Sustituyendo 43 y  $N_0 = 6.022 \times 10^{23}$  en 42, queda que:

$$\frac{dE}{ds}(\text{keV}/\mu\text{m}) \approx -7.84 \frac{\rho(\text{gcm}^{-3})Z}{A(\text{g/mol})E\text{keV}} \ln \left( \frac{173.9E(\text{keV})}{Z} \right) \quad (44)$$

con lo que hemos demostrado detalladamente la deducción de la expresión 32 a partir de la 31. Nótese que  $E$  es la energía cinética (equivale a  $T$ ).

**Deducción de la expresión 34** Se obtiene integrando la ecuación 33:

$$\begin{aligned} \sigma_T &= \int_0^\pi \int_0^{2\pi} d\Omega \frac{Z(Z+1)e^4}{p^2v^2} \frac{1}{(1-\cos\theta+2\beta)^2} = \frac{Z(Z+1)e^4}{p^2v^2} \int_0^\pi \frac{2\pi \sin\theta d\theta}{(1-\cos\theta+2\beta)^2} \\ &= \frac{2\pi Z(Z+1)e^4}{p^2v^2} \left[ -\frac{1}{1-\cos\theta+2\beta} \right]_0^\pi = \frac{2\pi Z(Z+1)e^4}{p^2v^2} \left( \frac{1}{2\beta} - \frac{1}{2(1+\beta)} \right) \\ &= \frac{\pi Z(Z+1)e^4}{p^2v^2\beta(1+\beta)} \quad \text{q.e.d} \end{aligned} \quad (45)$$

donde hemos integrado directamente  $\int_0^{2\pi} d\phi = 2\pi$  al haber simetría azimutal.

**Deducción de la expresión 35** El recorrido libre medio se puede expresar como el valor esperado de la distancia absoluta recorrida  $x$  (por definición, será positiva) por cada electrón según una determinada distribución (pdf)  $f(s)$  que dependerá del proceso físico:

$$\lambda = E[x] = \int_0^{\infty} s f(s) ds \quad (46)$$

Es muy usual suponer que la atenuación de los electrones es de tipo exponencial (ley de Beer-Lambert), con lo que la pdf debidamente normalizada sería:

$$f(s) = \mu e^{-\mu s} \quad (47)$$

donde  $\mu$  es el coeficiente lineal de atenuación que se puede expresar como  $\mu = N\sigma_T = \sigma_T \rho N_0/A$ . Integrando por partes, queda que:

$$\lambda = \mu \int_0^{\infty} s e^{-\mu s} = \mu \left[ -\frac{1}{\mu} s e^{-\mu s} \Big|_0^{\infty} + \frac{1}{\mu} \int_0^{\infty} e^{-\mu s} ds \right] = \frac{1}{\mu} \quad (48)$$

En este caso particular,  $1/\mu$  coincide con  $\lambda$ , con lo que podíamos haber acertado la deducción, pero podría ser diferente si modificamos la pdf añadiendo otros efectos, para lo cual sí es útil la fórmula general 46.

Al sustituir  $N$ ,  $\sigma_T$  y tener en cuenta que la energía cinética  $T = p^2/2m = mv^2/2 \rightarrow p^2 v^2 = 4T^2$  queda que:

$$\lambda = \frac{4T^2}{\pi N_0 e^4} \frac{A\beta(1+\beta)}{\rho Z(Z+1)} \quad (49)$$

expresado en el sistema cgs-gaussiano. Realizando los cambios 41 en  $\lambda$ , 40 en  $T$  y sustituyendo el valor de  $e$  en  $esu$  y la constante de Avogadro  $N_0$ , queda que:

$$\lambda(\mu m) 10^{-4} = \frac{4 * (T(keV) 1.602 \times 10^{-9})^2}{\pi * 6.022 \times 10^{23} (4.803 \times 10^{-10})^4} \frac{A(\text{g/mol})\beta(1+\beta)}{\rho(\text{gcm}^{-3})Z(Z+1)} \quad (50)$$

$$\lambda(\mu m) \approx \frac{1.02\beta(1+\beta)A(\text{g/mol})(T(\text{keV}))^2}{Z(Z+1)\rho(\text{gcm}^{-3})} \quad \text{q.e.d} \quad (51)$$

Otra vía (quizá más elegante) para transformar estas ecuaciones en el sistema es mediante la constante de estructura fina  $\alpha = e^2/4\pi \approx 1/137$  y  $\hbar c = 197,326 \text{ fm MeV}$  (energías en distancias).

Por último, podemos simplificar la expresión de  $\beta$  sustituyendo  $p^2 = 2mT$  y los valores de las constantes en el sistema cgs ( $a_0 = 0.5291 \times 10^{-8} \text{ cm}$ ,  $h = 6.626 \times 10^{-27} \text{ erg} \cdot \text{s}$ ,  $m = 9.109 \times 10^{-28} \text{ g}$ ):

$$\begin{aligned} \beta &= 0.25 \left( \frac{1.12 Z^{1/3} h}{2 * 0.885 a_0 \pi p} \right)^2 = \frac{0.25}{2m} \left( \frac{1.12 h}{2 * 0.885 a_0 \pi} \right)^2 \frac{Z^{2/3}}{T} \\ &= 8.728 \times 10^{-10} \frac{Z^{2/3}}{T(\text{cgs})} = 0.00545 \frac{Z^{2/3}}{T(\text{keV})} \end{aligned} \quad (52)$$

**Consideraciones geométricas** El ángulo de dispersión del electrón  $\theta$  está definido como el ángulo que forman los vectores velocidad directores (unitarios) antes y después de la colisión (aparte, habrá un ángulo azimutal arbitrario  $\phi$  uniforme entre 0 y  $2\pi$  asociado a la rotación de  $\mathbf{v}_{n+1}$  en el cono definido por  $\mathbf{v}_n$  el ángulo  $\theta$ , que define el eje  $z'$  en la figura 8a):

$$\mathbf{v}_n \mathbf{v}_{n+1} = \cos \theta \quad (53)$$

Nótese que  $\theta$  es un ángulo relativo entre dos vectores  $\mathbf{v}_n$ ,  $\mathbf{v}_{n+1}$  y es distinto de los ángulos de coordenadas esféricas  $\theta_n$ ,  $\phi_n$  respecto a los ejes en el sistema de referencia del laboratorio (trasladado al punto de colisión), tal y como se observa en la figura 8.

Por tanto, tenemos que en el SL (ejes fijos independientemente de la colisión) y en el sistema propio SP (eje  $z'$  definido por  $\mathbf{v}_n$ , se reorienta en cada colisión):

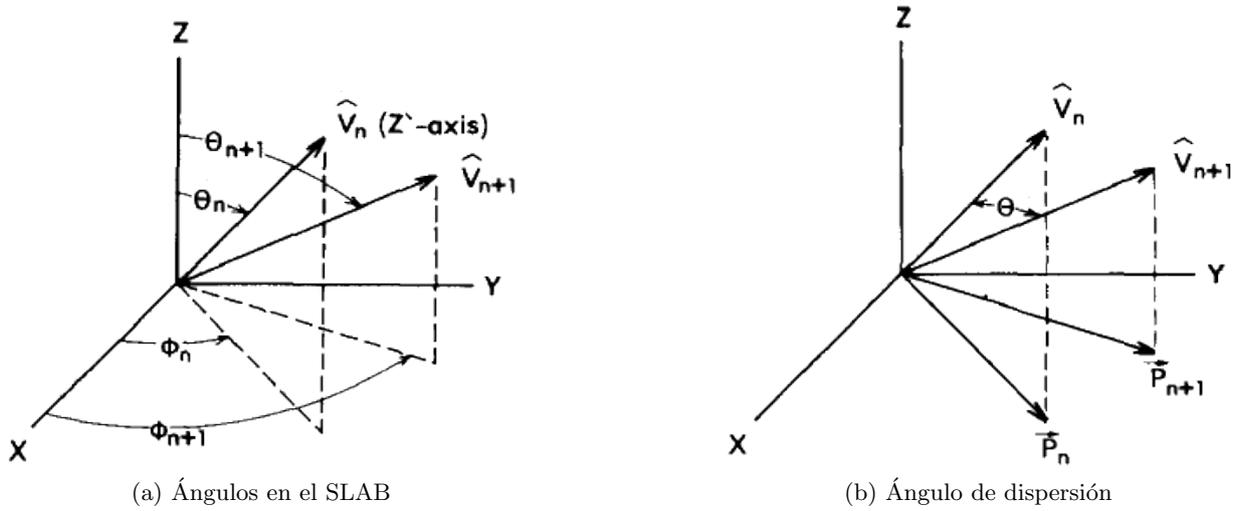
$$\begin{aligned}\mathbf{v}_n &= (\cos \phi_n \sin \theta_n, \sin \phi_n \sin \theta_n, \cos \theta_n)_{SL} = (0, 0, 1)_{SP} \\ \mathbf{v}_{n+1} &= (\cos \phi_{n+1} \sin \theta_{n+1}, \sin \phi_{n+1} \sin \theta_{n+1}, \cos \theta_{n+1})_{SL} \\ &= (\cos \phi \sin \theta, \sin \phi \sin \theta, \cos \theta)_{SP}\end{aligned}\tag{54}$$

La definición del ángulo  $\phi$  es arbitraria. Elegimos que el eje  $y'$ , cuya única restricción es que sea perpendicular a  $\mathbf{v}_n$  (eje  $z'$ ), sea perpendicular también al eje  $z$ , y el eje  $x'$  vendrá fijado por los otros dos. Los vectores directores del sistema propio serán<sup>5</sup>:

$$\begin{aligned}\mathbf{k}' = \mathbf{v}_n &= (\cos \phi_n \sin \theta_n, \sin \phi_n \sin \theta_n, \cos \theta_n)_{SL} \\ \mathbf{j}' = \mathbf{k}' \times \mathbf{k} / |\mathbf{k}' \times \mathbf{k}| &= (\sin \phi_n \sin \theta_n, -\cos \phi_n \sin \theta_n, 0)_{SL} / \sin \theta_n \\ &= (\sin \phi_n, -\cos \phi_n, 0)_{SL} \\ \mathbf{i}' = \mathbf{j}' \times \mathbf{k}' &= (-\cos \phi_n \cos \theta_n, -\sin \phi_n \cos \theta_n, \sin \theta_n)_{SL}\end{aligned}\tag{55}$$

---

<sup>5</sup>En [3], hay una errata en la ecuación 7b, debería figurar  $\sin \phi_n$  en lugar de  $\sin \theta_n$ .



**Figura 8:** Diferencia entre el ángulo relativo de dispersión entre dos vectores velocidad directores (antes y después de la colisión  $n$ ) y los ángulos en coordenadas esféricas que describen la orientación del vector en el Sistema Laboratorio (SLAB), cuyos ejes están fijos y su origen se traslada al punto de la colisión  $n$ -ésima (consultar más detalles en [3]).

El objetivo es obtener los ángulos  $(\theta_{n+1}, \phi_{n+1})$  a partir de los ángulos  $(\theta_n, \phi_n), (\theta, \phi)$ . Para ello es útil realizar los siguientes productos escalares (sigo la notación de [3]), y se obtiene desarrollando la expresión de los vectores tanto en el SL como en SP<sup>6</sup>:

$$\begin{aligned}
 \mathbf{k}\mathbf{v}_{\mathbf{n}+1} &= \cos\theta_{n+1} = \cos\phi \sin\theta(\mathbf{i}' \cdot \mathbf{k}) + \sin\phi \sin\theta(\mathbf{j}' \cdot \mathbf{k}) + \cos\theta(\mathbf{k}' \cdot \mathbf{k}) \\
 &= \cos\phi \sin\theta \sin\theta_n + \sin\phi \sin\theta * 0 + \cos\theta \cos\theta_n \\
 &= \cos\theta_n \cos\theta + \sin\theta_n \sin\theta \cos\phi
 \end{aligned} \tag{56}$$

$$\begin{aligned}
 \mathbf{v}_{\mathbf{n}+1}\mathbf{v}_{\mathbf{n}} &= \cos\theta = \cos\phi_{n+1} \sin\theta_{n+1} \cos\phi_n \sin\theta_n + \sin\phi_{n+1} \sin\theta_{n+1} \sin\phi_n \sin\theta_n + \cos\theta_{n+1} \cos\theta_n \\
 &= \sin\theta_n \sin\theta_{n+1} (\cos\phi_{n+1} \cos\phi_n + \sin\phi_{n+1} \sin\phi_n) + \cos\theta_{n+1} \cos\theta_n \\
 &= \cos\theta_n \cos\theta_{n+1} + \sin\theta_n \sin\theta_{n+1} \cos(\phi_{n+1} - \phi_n)
 \end{aligned} \tag{57}$$

Si queremos que aparezca la tangente en lugar de el coseno, hay que utilizar las proyecciones  $\mathbf{P}$  de la figura 8. En el artículo original se puede consultar la derivación completa. En resumen:

$$\begin{aligned}
 \cos\theta_{n+1} &= \cos\theta_n \cos\theta + \sin\theta_n \sin\theta \cos\phi \\
 \cos(\phi_{n+1} - \phi_n) &= \frac{\cos\theta - \cos\theta_{n+1} \cos\theta_n}{\sin\theta_n \sin\theta_{n+1}} \\
 \sin(\phi_{n+1} - \phi_n) &= \frac{\sin\theta \sin\phi}{\sin\theta_{n+1}} \\
 \tan(\phi_{n+1} - \phi_n) &= \frac{\sin\theta_n \sin\theta \sin\phi}{\cos\theta - \cos\theta_{n+1} \cos\theta_n}
 \end{aligned} \tag{58}$$

En el enunciado se sugiere un camino alternativo, igualmente válido, en el que los ejes del sistema propio se definen a partir de un vector unitario  $\mathbf{q}$  con orientación aleatoria isotrópica. Se

<sup>6</sup>En [3], hay otra errata en la ecuación 11b, debería figurar  $\cos(\phi_{n+1} - \phi_n)$  y  $\sin\theta_{n+1}$  por comparación con 8c.

genera un vector unitario perpendicular a  $\mathbf{v}_n$  mediante  $\mathbf{u} = \mathbf{v}_n \times \mathbf{q} / |\mathbf{v}_n \times \mathbf{q}|$ . Finalmente, la nueva dirección viene dada por  $\mathbf{v}_{n+1} = \cos \theta \mathbf{v}_n + \sin \theta \mathbf{u}$ .

Yo optaré por el método descrito en el artículo.

### Construcción del programa

1. En cada colisión genero aleatoriamente el ángulo de dispersión  $\theta$  a partir de la expresión 33 mediante el método de la transformación inversa  $F(\theta) = \xi = \int_0^\theta f(\theta') d\theta'$ . Para ello, hay que normalizar adecuadamente dicha expresión dividiendo entre la sección eficaz total (34):

$$f(\theta) = \frac{1}{\sigma_T} \frac{d\sigma}{d\theta} = \frac{1}{\sigma_T} \int_0^{2\pi} \sin \theta d\phi \frac{d\sigma}{d\Omega} = \frac{1}{\sigma_T} 2\pi \frac{\sin \theta d\sigma}{d\Omega} = \frac{2\beta(1+\beta) \sin \theta}{(1 - \cos \theta + 2\beta)^2} \quad (59)$$

$$F(\theta) = \int_0^\theta f(\theta') d\theta' = \int_0^\theta \frac{2\beta(1+\beta) \sin \theta'}{(1 - \cos \theta' + 2\beta)^2} d\theta' = \left[ -\frac{2\beta(1+\beta)}{1 - \cos \theta' + 2\beta} \right]_0^\theta$$

$$= \left( 1 + \beta - \frac{2\beta(1+\beta)}{1 - \cos \theta + 2\beta} \right) = \xi_\theta \quad (60)$$

$$\cos \theta = 1 + 2\beta - \frac{2\beta(1+\beta)}{1 - \xi_\theta + \beta} = 1 - \frac{2\beta\xi_\theta}{1 + \beta - \xi_\theta} \quad (61)$$

Y el ángulo azimutal es uniforme:

$$\phi = 2\pi\xi_\phi \quad (62)$$

2. La dirección saliente tras el choque  $\mathbf{v}_{n+1}$  vendrá dada por la ecuación 58.
3. Genero la distancia  $s_n$  hasta la próxima colisión a partir de la ecuación 47 y 51 para la energía del electrón en el punto de la colisión anterior:

$$F(s) = \int_0^s \frac{1}{\lambda} e^{-s'/\lambda} = 1 - e^{-s/\lambda} = \xi_s \quad (63)$$

$$s = -\lambda \ln(1 - \xi_s) \quad (64)$$

4. La energía que tendrá el electrón en la próxima colisión viene dada por la ecuación 36 (aproximación lineal,  $dE/ds$  constante en el intervalo de integración, integras directamente  $ds$ ), que a su vez depende de 44 y 64.
5. Posteriormente se comprueba si el electrón ha atravesado el material,  $z > D$ , siendo  $D$  el grosor del material, ha rebotado,  $z < 0$ , o bien ha sido absorbido completamente en el interior del material ( $E_{n+1} < 0.5\text{keV}$ ).
6. Si ninguna de estas condiciones se da, se vuelve al primer paso. En caso contrario, se aumenta el contador correspondiente y se comienza con un nuevo electrón desde el principio.

Podemos hacer un formulario compacto. Se lanzan electrones desde  $\mathbf{R}_0 = 0$  ( $T_0 = 20\text{keV}$ ) con incidencia normal. Se genera la posición de la primera colisión ( $n = 1$ ):  $\mathbf{R}_1 = (0, 0, s_0)$ ,  $\theta_1 = 0$ ,

$\phi_1 = 0$  (en realidad no está definido<sup>7</sup>), punto al que llega con  $T_1 = T_0 - s_0 7.84 \frac{\rho Z}{AT_0} \ln\left(\frac{173.9T_0}{Z}\right)$ . A partir de aquí, para  $n \geq 1$ , partiendo de una posición  $\mathbf{R}_n$ , dirección  $\theta_n$ ,  $\phi_n$  y energía cinética  $T_n$  del electrón tras la colisión  $n$ -ésima en un material dado ( $A$ ,  $Z$ ,  $\rho$ ):

- (i) Calculo  $\beta = 0.00545 \frac{Z^{2/3}}{T_n}$  (adimensional).
- (ii) Genero 3 números aleatorios uniformes entre 0 y 1:  $\xi_\theta, \xi_\phi, \xi_s$  (función TRandom en ROOT).
- (iii) Calculo  $\theta = \arccos[1 - 2\beta\xi_\theta/(1 + \beta - \xi_\theta)]$ ;  $\phi = 2\pi\xi_\phi$ .
- (iv) Calculo  $\theta_{n+1} = \arccos[\cos\theta_n \cos\theta + \sin\theta_n \sin\theta \cos\phi]$ .
- (v) Calculo  $\phi_{n+1} = \phi_n + \arccos\left[\frac{\cos\theta - \cos\theta_{n+1} \cos\theta_n}{\sin\theta_n \sin\theta_{n+1}}\right]$  Y<sup>8</sup> alternativamente:  

$$\phi_{n+1} = \phi_n + \arcsin\left[\frac{\sin\theta \sin\phi}{\sin\theta_{n+1}}\right]$$

$$\phi_{n+1} = \phi_n + \arctan\left[\frac{\sin\theta_n \sin\theta \sin\phi}{\cos\theta - \cos\theta_{n+1} \cos\theta_n}\right]$$
- (vi) Calculo  $s_n = -\frac{1.02\beta(1+\beta)AT_n^2}{Z(Z+1)\rho} \ln(1 - \xi_s)$ ;  
 $\mathbf{R}_{n+1} = \mathbf{R}_n + s_n(\cos\phi_{n+1} \sin\theta_{n+1}, \sin\phi_{n+1} \sin\theta_{n+1}, \cos\theta_{n+1})_{SL}$  (ver figura 10).
- (vii) Calculo  $T_{n+1} = T_n - s_n 7.84 \frac{\rho Z}{AT_n} \ln\left(\frac{173.9T_n}{Z}\right)$  (aproximación lineal)
- (viii) Testear:
  - A) Si se ha transmitido el electrón, aumentar el contador  $N_t$ , almacenar  $T_n$  y volver al paso (i), es decir, lanzar el siguiente electrón.
  - B) Si se ha retrodispersado (backscatter), aumentar el contador  $N_r$ .
  - C) Si su energía es menor que 0.5keV, aumentar el contador  $N_a$  (absorbido).  
Si ninguna se cumple, todavía necesita más iteraciones/colisiones para que suceda alguna de las anteriores: volver al paso (i). Si no almacenar en un vector distinto cada colisión, debes renombrar  $\theta_{n+1} \rightarrow \theta_n, \phi_{n+1} \rightarrow \phi_n, \mathbf{R}_{n+1} \rightarrow \mathbf{R}_n, T_{n+1} \rightarrow T_n$  antes de volver al paso (i).
- (ix) Una vez hayas simulado todos los  $N_{sim}$  electrones deseados, calcular  $R = (N_r \pm \sqrt{N_r})/N_{sim}$ ;  $T = (N_t \pm \sqrt{N_t})/N_{sim}$ ;  $A = (N_a \pm \sqrt{N_a})/N_{sim}$  y el espectro energético de los electrones transmitidos. Se puede repetir todo el proceso para otros materiales o grosores. También se puede calcular el ángulo promedio de dispersión  $\bar{\theta} \pm \sigma(\bar{\theta})$ , el número promedio de colisiones  $\bar{n}_{col} = (N_{tot,col} \pm \sqrt{N_{tot,col}})/N_{sim}$ , etc.

$\rho$  (g/cm<sup>3</sup>),  $Z$ ,  $A$  (g/mol) son constantes durante toda la simulación (dependen del material). Las unidades para  $T$  y  $s$  son keV y  $\mu m$  respectivamente.

El tratamiento de la función arcocoseno, acotada entre 0 y  $\pi$  es delicado. En el caso de  $\theta_n$  no es problemático porque el ángulo polar está también acotado entre 0 y  $\pi$ . El problema surge en la resta de ángulos azimutales en la ecuación 58, pues cada uno de ellos va de 0 a  $2\pi$ . Por eso, en general,  $\chi \equiv (\phi_{n+1} - \phi_n) \in [-2\pi, 2\pi]$ , y conocemos  $C \equiv \cos\chi \in [-1, 1]$ , que nos da analíticamente 4 soluciones distintas (ver figura 9), mientras que el ordenador sólo daría una:  $\alpha \equiv \arccos(C) \in [0, \pi]$ .

<sup>7</sup>Si la incidencia  $\theta_n$  es normal, hay que ir con cuidado a la hora de programar, porque en la ecuación 58 (o 18d de [3]) habría una indeterminación. En ese caso habría que usar la ecuación 11c) de [3]. En el caso de la incidencia inicial,  $\phi_1$  no está definido, y según 11c) tendríamos  $\theta_2 = \theta$  y  $\phi_2 = \phi + \phi_1$ , con lo que estaría indeterminado. Por ello, en este caso particular inicial, calculamos  $\phi_2$  aleatoriamente entre 0 y  $2\pi$  (distribución uniforme). Se puede ver como una ruptura aleatoria de simetría.

<sup>8</sup>Ver aclaración en página siguiente



## Resultados

Coloco el sistema de referencia SLAB al comienzo de cada lámina (que supongo de anchura infinita y espesor  $d$ ), con el eje  $z$  perpendicular al plano definido por la lámina, y los electrones incidiendo normalmente (aunque es inmediato cambiar este valor para hacerlos incidir con un ángulo arbitrario. Si la coordenada  $z$  del electrón es mayor que  $d$ , se habrá transmitido, si es menor que 0, se habrá retrodispersado. Si está entre 0 y  $d$  y su  $T$  es menor que 0.5keV, se habrá absorbido. Si está entre 0 y  $d$  y su  $T$  es mayor que 0.5keV, se habrá absorbido.

El número de electrones que genero para cada caso es  $N_{sim} = 50.000$ .

Obtenemos la siguiente tabla de valores de  $T$ ,  $R$  y  $A$  para diferentes grosores  $d$  de la lámina, tanto para aluminio como oro:

$\frac{A}{Z}X$	$d(\mu\text{m})$	$T$	$R$	$A$	$\bar{n}_{col}$	$\bar{\theta}(^\circ)$
${}_{13}^{27}\text{Al}$	0.1	0.994 ± 0.004	0.0060 ± 0.0003	0 ± 0	3.112 ± 0.008	6.76 ± 0.02
	0.2	0.987 ± 0.004	0.0129 ± 0.0005	0 ± 0	6.611 ± 0.011	6.800 ± 0.015
	0.3	0.976 ± 0.004	0.0238 ± 0.0007	0.00004 ± 0.00003	10.584 ± 0.015	6.858 ± 0.012
	0.4	0.965 ± 0.004	0.0344 ± 0.0008	0.00032 ± 0.00008	15.238 ± 0.017	6.943 ± 0.010
	0.5	0.949 ± 0.004	0.0498 ± 0.0010	0.00164 ± 0.00018	20.65 ± 0.02	7.085 ± 0.009
	0.6	0.927 ± 0.004	0.0678 ± 0.0012	0.0055 ± 0.0003	27.33 ± 0.02	7.352 ± 0.008
	0.7	0.907 ± 0.004	0.0816 ± 0.0013	0.0117 ± 0.0005	34.47 ± 0.03	7.596 ± 0.008
	0.8	0.879 ± 0.004	0.0972 ± 0.0014	0.0238 ± 0.0007	43.25 ± 0.03	7.937 ± 0.007
	0.9	0.847 ± 0.004	0.1098 ± 0.0015	0.0431 ± 0.0009	53.19 ± 0.03	8.326 ± 0.007
	1	0.814 ± 0.004	0.1199 ± 0.0015	0.0661 ± 0.0012	63.75 ± 0.04	8.657 ± 0.006
${}_{79}^{197}\text{Au}$	0.1	0.606 ± 0.003	0.386 ± 0.003	0.0075 ± 0.0004	77.64 ± 0.04	13.540 ± 0.008
	0.2	0.295 ± 0.002	0.500 ± 0.003	0.205 ± 0.002	271.19 ± 0.07	20.081 ± 0.006
	0.3	0.0945 ± 0.0014	0.500 ± 0.003	0.405 ± 0.003	407.45 ± 0.09	22.258 ± 0.005
	0.4	0.0148 ± 0.0005	0.505 ± 0.003	0.480 ± 0.003	449.20 ± 0.09	22.915 ± 0.005
	0.5	0.00120 ± 0.00015	0.506 ± 0.003	0.493 ± 0.003	454.86 ± 0.10	23.017 ± 0.005
	0.6	0.00002 ± 0.00002	0.501 ± 0.003	0.499 ± 0.003	458.73 ± 0.10	23.055 ± 0.005
	0.7	0 ± 0	0.502 ± 0.003	0.498 ± 0.003	458.25 ± 0.10	23.050 ± 0.005
	0.8	0 ± 0	0.501 ± 0.003	0.499 ± 0.003	458.50 ± 0.10	23.049 ± 0.005
	0.9	0 ± 0	0.499 ± 0.003	0.501 ± 0.003	459.90 ± 0.10	23.063 ± 0.005
	1	0 ± 0	0.505 ± 0.003	0.495 ± 0.003	456.59 ± 0.10	23.024 ± 0.005

**Tabla 2:** Valores  $T$ ,  $R$  y  $A$  para diferentes grosores  $d$  de la lámina de aluminio ( $\rho_{Al} = 2.70$ ) y oro ( $\rho_{Au} = 19.3$ ). Por completitud, se muestra también el número promedio de colisiones antes  $\bar{n}_{col}$  hasta ser transmitido, retrodispersado o absorbido y el ángulo promedio de dispersión  $\bar{\theta}(^\circ)$  en cada colisión.

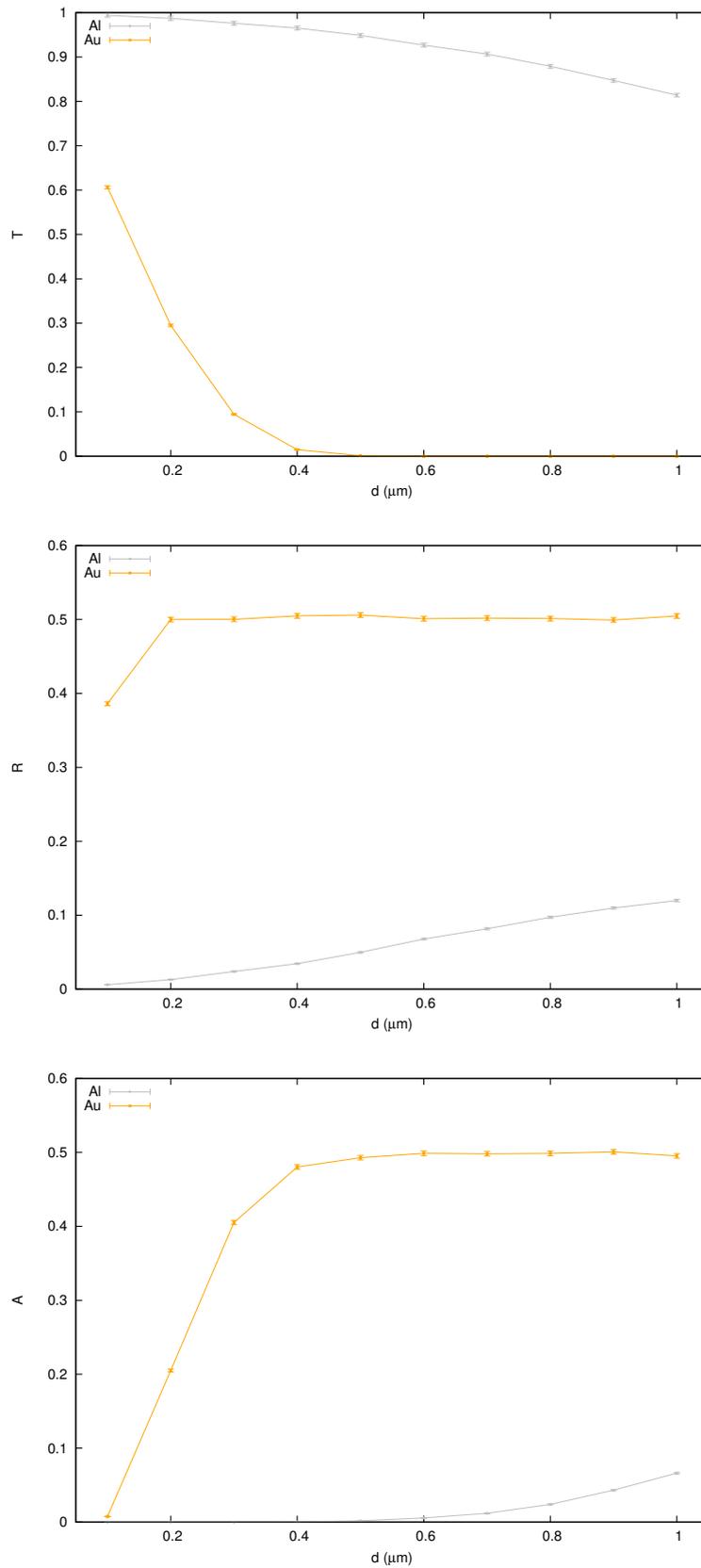
Estos resultados son compatibles con los presentados en el artículo [3], cuyas barras de error son mayores al haber generado menor número de electrones. El tiempo de cálculo global ha sido de 2 horas<sup>9</sup> en un PC Pentium 4.

Representamos los valores de  $R$ ,  $T$  y  $A$  en función del grosor en la figura 11.

Se observa, como es de esperar, un mayor coeficiente de transmisión en el aluminio que en el oro, al ser el primero más ligero (menor densidad) y tener menor valor de  $Z$  y  $A$ . En consecuencia, el backscatter (y absorción) es mayor en el oro que en el aluminio.

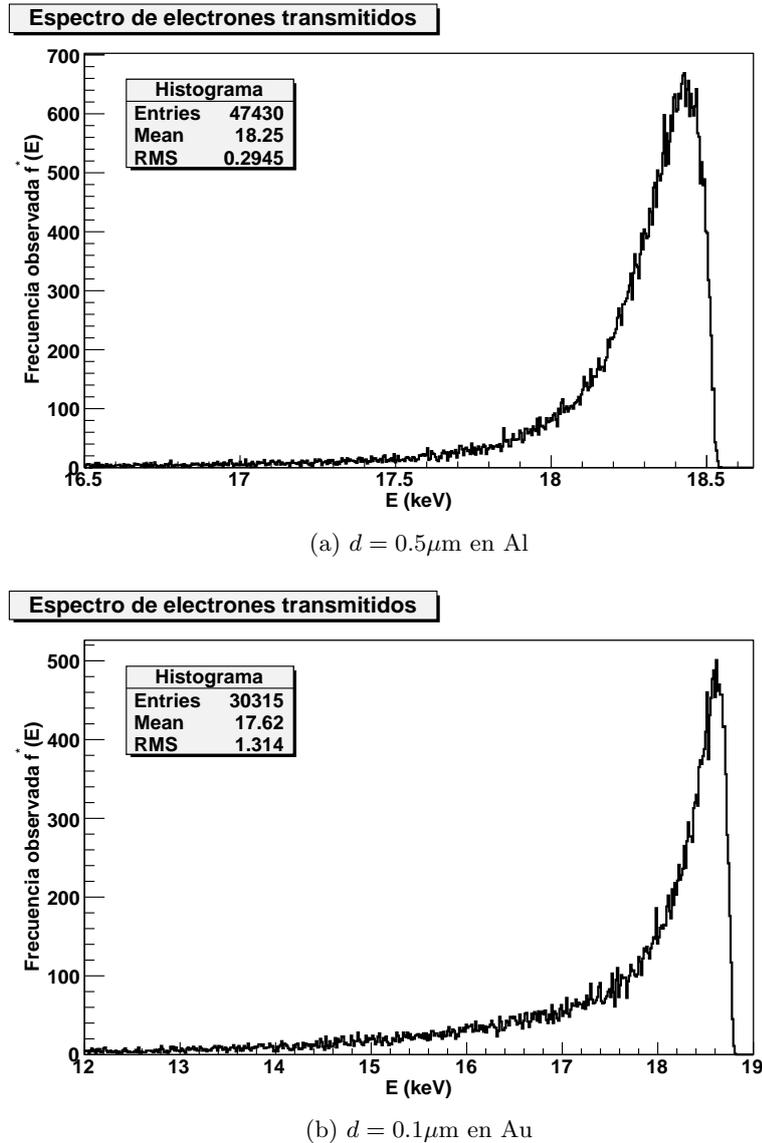
Además, a mayor grosor, menor transmisión y mayor retrodispersión (y absorción), aunque a suficiente grosor,  $R$  y  $A$  se saturan y acaban por valer ambos 1/2 (con  $T = 0$ ). Por esta razón, para blindajes frente a radiación se utilizan elementos pesados como plomo.

<sup>9</sup>La eficiencia se puede mejorar considerablemente si en lugar de calcular ángulos mediante funciones arco- trabajamos directamente con los cosenos=  $C$ , senos= $\sqrt{1 - C^2}$ , etc. modificando adecuadamente alguna expresión del algoritmo, aunque como se trata de un ejercicio académico no lo he estudiado con mayor profundidad. Otra opción más eficiente también es la variante que comenta en el enunciado de generar una dirección arbitraria extra  $\mathbf{q}$ .



**Figura 11:** Valores de transmisión  $T$ , retrodispersión  $R$  y absorción  $A$  en función del grosor y el material. La suma de coeficientes es la unidad.

Para los casos particulares de  $d = 0.5\mu\text{m}$  en aluminio y  $d = 0.1\mu\text{m}$  en oro, representamos el espectro de energía de los electrones transmitidos<sup>10</sup>. Vemos una acumulación a energías altas

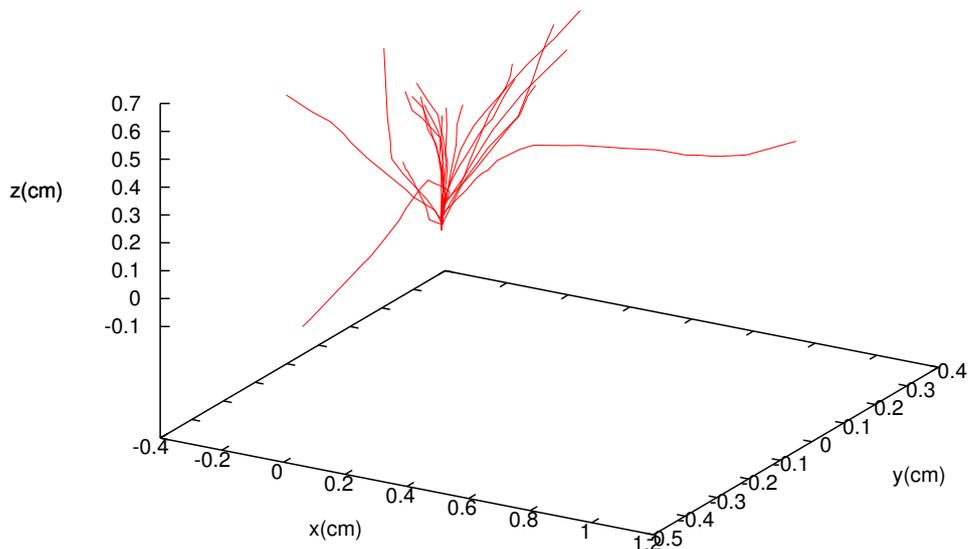


**Figura 12:** Espectro energético de los electrones transmitidos

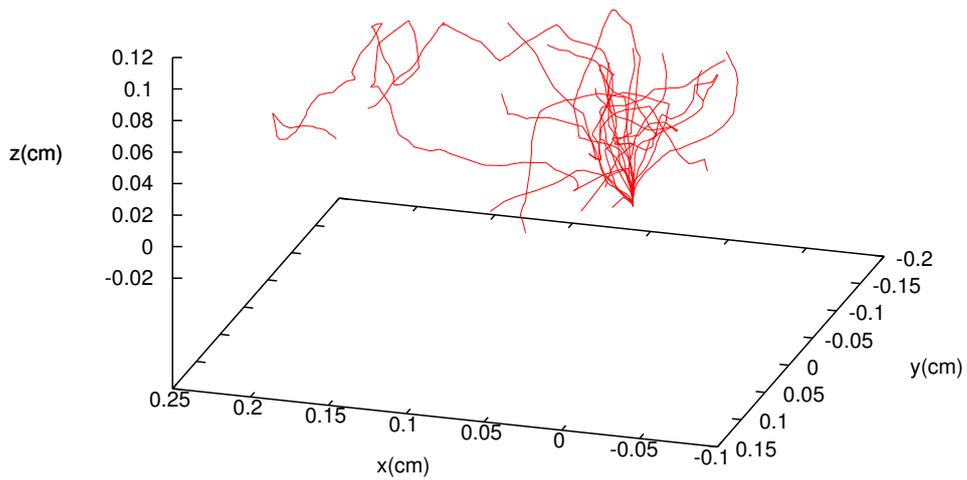
(cercanas a la inicial), es decir, la mayoría de los que se transmiten han sufrido pocas colisiones (habrán cruzado casi en línea recta). En el caso del aluminio (con mayor grosor), se observa que el pico de los electrones está ligeramente a menor energía que en el caso del oro, aunque el valor medio de energías es más bajo en el caso del oro al tener mayor densidad, número atómico y másico, con lo que los electrones transmitidos han perdido en promedio más energía (salían con 20 keV). En consecuencia, el espectro en oro es bastante más ancho, (se vería bien si los pintásemos a escala).

Por completitud, dibujamos la traza de los 20 primeros  $e^-$  de los 2 casos particulares escogidos:

<sup>10</sup>Hay que señalar que hay un pequeño sesgo en la energía de los electrones, ya que tras la última colisión, se supone que siguen perdiendo energía durante el trayecto hasta la hipotética futura colisión, ya fuera del material, con lo que si la última colisión dentro del material ha sido en el borde, el error será mayor. Se podría programar fácilmente que sólo perdiese energía hasta el borde del material, pero al ser una gráfica ilustrativa, por simplicidad lo omitimos en el código.



(a)  $d = 0.5\mu\text{m}$  en Al



(b)  $d = 0.1\mu\text{m}$  en Au

**Figura 13:** Trazas de los 20 primeros electrones (se emiten en el origen del sistema de coordenadas e inciden perpendicularmente sobre una lámina plana, de anchura infinita y espesor  $d$  (entre  $z = 0$  y  $z = d$ ). En el oro, hay más colisiones y las trayectorias son más retorcidas

## Bibliografía

- [1] *Measuring radon concentration in air using a diffusion cloud chamber*, R. Cases, E. Ros, and J. Zúñiga, American Journal of Physics – September 2011 – Volume 79, Issue 9, pp. 903
- [2] *Visualizaciónn de trazas y medida de la concentración de radón atmosférico con la cámara de niebla de difusión*, E. Alcusa, F. Hueso, Práctica de Laboratorio de Física Nuclear y de Partículas (4º de Grado)
- [3] W. Williamson and G. C. Duncan, *Monte Carlo simulation of nonrelativistic electron scattering*, Am. J. Phys. **54**(3), 1986