Lección 1: Introducción a la Teoría de Probabilidades

Definición de variables estocásticas

Son variables que toman un determinado valor con una cierta probabilidad. Entonces, una variable estocástica X viene definida por el conjunto de posibles valores que puede tener (rango, conjunto de estados, espacio muestral, o espacio fásico) y una distribución de probabilidad sobre este conjunto.

- El conjunto de posibles valores puede ser *discreto:* cara o cruz en una moneda; número de electrones en la banda de conducción de un semiconductor; o *continuo* en un intervalo dado: componentes de la velocidad de una partícula Browniana; energía de la partícula, etc. La variable estocástica puede tener un carácter multidimensional en cuyo caso la notaramos en forma vectorial $\boldsymbol{X}$: por ejemplo las tres componentes de la velocidad de una partícula Browniana. Por simplicidad trabajaremos con estados discretos o con rango continuo en dimensión 1.

- La distribución de probabilidad cumple para variables continuas:
  - $P(x) \geq 0$,
  - $\displaystyle\int P(x)dx = 1$ (Normalización). El producto $P(x)dx$ es la probabilidad de que la variable estocástica X tenga un valor entre x y $x + dx$.

- Para variables discretas:
  - $X = \{x_1, \ldots, x_N\}$; $p_n \equiv P(x_n) \geq 0$ con
  - $\displaystyle\sum_n p_n = 1$ (normalización)

- Puede haber situaciones en las que en un intervalo continuo se tengan funciones delta,

$$P(x) = \sum_n p_n \delta(x - x_n) + \tilde{P}(x). \tag{1}$$

  La condición de normalización queda en este caso como

$$\sum_n p_n + \int \tilde{P}(x)dx = 1. \tag{2}$$

  Esto se puede interpretar como que hay un conjunto discreto de estados x_n con probabilidad p_n embebida en un rango contínuo.

- Una forma alternativa de describir una distribución de probabilidad es mediante la función $\mathcal{P}(x)$ definida como la probabilidad total de que X tenga un valor $\leq x$:

$$\mathcal{P}(x) = \int_{-\infty}^{x} P(y)dy \tag{3}$$

que los matemáticos llaman *función distribución de probabilidad*, mientras que utilizan el término densidad de probabilidad para $P(x)$. Los físicos llaman a $\mathcal{P}(x)$ *función de distribución de probabilidad acumulada.* Nosotros como físicos vamos a usar la densidad de probabilidad $P(x)$ y nos referiremos a ella como "distribución de probabilidad" o simplemente "la probabilidad".

Promedios

- El conjunto de posibles valores y su probabilidad definen una variable estocástica. Además podemos definir el *promedio* o *valor esperado* de una función $f(X)$ como

$$\langle f(X) \rangle = \int f(x)P(x)\, dx \tag{4}$$

- Definimos el momento de orden m de la variable estocástica X como

$$\mu_m \equiv \langle X^m \rangle = \int x^m P(x)\, dx \tag{5}$$

Se tiene entonces que μ_1 es el promedio o *media* de X y que

$$\sigma^2 = \langle (X - \langle X \rangle)^2 \rangle = \mu_2 - \mu_1^2 \tag{6}$$

es la *variancia* o *dispersión,* que el cuadrado de la desviación estándar σ

- Definimos la *función característica* de X cuyo rango es I como

$$G(k) = \langle e^{ikX} \rangle = \int_I e^{ikx} P(x)\, dx \tag{7}$$

y es tal que existe para todo número real k y verifica $G(0) = 1$ y $|G(k)| \leq 1$. Es tal que es la función generadora de los momentos de $P(X)$,

$$G(k) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \mu_m, \tag{8}$$

de tal forma que

$$\mu_n = \frac{1}{i^n} \left. \frac{d^n G(k)}{dk^n} \right|_{k=0}. \tag{9}$$

También sirve para generar los *cumulantes* κ_m de la forma

$$\log G(k) = \sum_{m=1}^{\infty} \frac{(ik)^m}{m!} \kappa_m. \tag{10}$$

Se tiene que $G(k)$ es la tranformada de Fourier de $\bar{P}(x)$,

$$\bar{P}(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} G(k) e^{-ikx}\, dk \tag{11}$$

con

$$\bar{P}(x) = \begin{cases} P(x) & x \in I \\ 0 & x \notin I \end{cases} \tag{12}$$

Esta distinción entre P y $\bar{P}$ es importante para el caso de variables estocásticas discretas. Supongamos que X toma sólo valores enteros $n = \ldots, -2, -1, 0, 1, 2, \ldots$ con probabilidades p_n. Para construir la función característica escribimos primero $\bar{P}(x)$ sobre todos los números reales:

$$\bar{P}(x) = \sum_n p_n \delta(x - n), \tag{13}$$

de donde se tiene

$$G(k) = \sum_n p_n e^{ikn}, \tag{14}$$

con k un número real. Además uno puede obtener p_n haciendo la TF sobre un periodo de la forma

$$p_n = \frac{1}{2\pi} \int_0^{2\pi} G(k) e^{-ikn}\, dk \tag{15}$$

- Una distribución cuyo rango son los puntos na con $a > 0$ y n entero se llama una *distribución reticular.* Para estas distribuciones $G(k)$ es periódica con periodo $2\pi/a$ de forma que $G(k)$ no toma sólo el valor 1 en $k = 0$. para cualquier otro tipo de distribución se tiene

$$|G(k)| < 1 \quad \text{for } k \neq 0. \tag{16}$$

- En muchos casos se define $s = ik$ y se usa $G(s) = \langle e^{sX} \rangle$. Cuando X toma sólo valores positivos tiene algunas ventajas el utilizar $\langle e^{-sX} \rangle$. Cuando X toma sólo valores enteros es más util utilizar $F(z) = z^X$ que está unívocamente definida para todo z en el círculo unidad $|z| = 1$.

Funciones de distribución de varias variables

Sea una variable estocástica con r componentes $\boldsymbol{X} = \{X_1, \ldots, X_r\}$ y con densidad de probabilidad $P_r(x_1, \ldots, x_r)$ o *distribución de probabilidad conjunta* de las r variables $X_1, \ldots, X_r$.

- Dado un subconjunto $s < r$ de variables $X_1, \ldots, X_s$ se define la *probabilidad marginal* del subconjunto como

$$P_s(x_1, \ldots, x_s) = \int P_r(x_1, \ldots, x_s, x_{s+1}, \ldots, x_r)\ dx_{s+1} \ldots dx_r. \quad (17)$$

  y es tal que el producto

$$P_s(x_1, \ldots, x_s)\ dx_1 \ldots dx_s \quad (18)$$

  es la probabilidad de que X_1 esté entre x_1 y $x_1 + dx_1$,... y X_s esté entre x_s y $x_s + dx_s$.

- Podemos también suponer que $X_{s+1}, \ldots X_r$ toman valores fijos $x_{s+1} \ldots x_r$. Entonces podemos preguntarnos por la distribución de probabilidad del resto de las componentes. Esta probabilidad se llama *probabilidad condicional* de $X_1, \ldots X_s$, supuesto que $X_{s+1}, \ldots X_r$ valen $x_{s+1} \ldots x_r$ y se escribe:

$$P_{s|r-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_r) \quad (19)$$

- De las definiciones anteriores se tiene

$$P_r(x_1, \ldots, x_r) = P_{s|r-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_r) P_{r-s}(x_{s+1}, \ldots, x_r) \quad (20)$$

  de donde se sigue la fórmula de Bayes para la probabilidad condicional:

$$P_{s|r-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_r) = \frac{P_r(x_1, \ldots, x_r)}{P_{r-s}(x_{s+1}, \ldots, x_r)} \quad (21)$$

- Supongamos que las r componentes estocásticas se pueden dividir en dos subconjuntos $(X_1, \ldots, X_s)$ y $(X_{s+1}, \ldots, X_r)$ tales que P_r factoriza:

$$P_r(s_1, \ldots, x_r) = P_s(x_1, \ldots, x_s) P_{r-s}(x_{s+1}, \ldots, x_r) \quad (22)$$

  entonces de dice que los dos conjuntos son estadísticamente independientes y se tiene:

  - P_s es la probabilidad marginal de $(X_1, \ldots, X_s)$
  - $P_{s|r-s}(x_1, \ldots, x_s | x_{s+1}, \ldots, x_r) = P_s(x_1, \ldots, x_s)$

- Los momentos de una distribución de varias variables son

$$\langle X_1^{m_1} \dots X_r^{m_r} \rangle = \int x_1^{m_1} \dots x_r^{m_r} \, dx_1 \dots dx_r. \tag{23}$$

- La función característica es:

$$G(k_1, \dots, k_r) = \langle e^{i(k_1 X_1 + \dots + k_r X_r)} \rangle \tag{24}$$

que, en términos de los momentos, se puede poner

$$G(k_1, \dots, k_r) = \sum_0^{\infty} \frac{(ik_1)^{m_1} \dots (ik_r)^{m_r}}{m_1! \dots m_r!} \langle X_1^{m_1} \dots X_r^{m_r} \rangle \tag{25}$$

y los cumulantes (denotados por $\langle\langle \cdot \rangle\rangle$) vienen definidos por

$$\log G(k_1, \dots, k_r) = \sum_0^{\infty}{}' \frac{(ik_1)^{m_1} \dots (ik_r)^{m_r}}{m_1! \dots m_r!} \langle\langle X_1^{m_1} \dots X_r^{m_r} \rangle\rangle \tag{26}$$

donde la prima indica la ausencia del término con todos los $m's$ iguales a zero.

- Se define la matriz covariante como

$$\langle\langle X_i X_j \rangle\rangle = \langle (X_i - \langle X_i \rangle)(X_j - \langle X_j \rangle) \rangle = \langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \tag{27}$$

cuyos elementos diagonal son las varianzas, y sus elementos fuera de la diagonal se llaman *covarianzas*. Cuando dichos elementos están normalizados se les llaman coeficientes de correlación:

$$\rho_{ij} = \frac{\langle\langle X_i X_j \rangle\rangle}{\sqrt{\langle\langle X_i^2 \rangle\rangle \langle\langle X_j^2 \rangle\rangle}} \tag{28}$$

- Para $r = 2$ la independencia estadística de la variables estocásticas X_1 y X_2 se expresa por alguno de los siguientes criterios:
  - Todos los momentos factorizan: $\langle X_1^{m_1} X_2^{m_2} \rangle = \langle X_1^{m_1} \rangle \langle X_2^{m_2} \rangle$
  - La función característica factoriza: $G(k_1, k_2) = G_1(k_1) G_2(k_2)$.
  - Los cumulantes $\langle\langle X_1^{m_1} X_2^{m_2} \rangle\rangle$ son nulos cuando ambos m_1 y m_2 son distintos de zero. Diremos que las variables X_1 y X_2 están descorrelacionadas cuando su covariancia es cero.

Suma de variables estocásticas

Sean dos variables estocásticas X_1 y X_2 con probabilidad conjunta $P_X(x_1, x_2)$. La probabilidad de que la variable $Y = X_1 + X_2$ tenga un valor entre y y $y + \Delta y$ es:

$$P_Y(y)\Delta y = \iint\limits_{y \,<\, x_1 + x_2 \,<\, y + \Delta y} P_X(x_1, x_2)\; dx_1\; dx_2 \tag{29}$$

que es equivalente a

$$\begin{aligned} P_Y(y) &= \iint \delta(x_1 + x_2 - y) P_X(x_1, x_2)\; dx_1\; dx_2 \\ &= \int P_X(x_1, y - x_1)\; dx_1 \end{aligned} \tag{30}$$

- Si X_1 y X_2 son independientes entonces

  $$P_Y(y) = \int P_{X_1}(x_1) P_{X_2}(y - x_1)\; dx_1, \tag{31}$$

  esto es la convolución de las dos probabilidades individuales. Además se cumple que

  - $\langle Y \rangle = \langle X_1 \rangle + \langle X_2 \rangle$,
  - $\langle\langle Y^2 \rangle\rangle = \langle\langle X_y^2 \rangle\rangle + \langle\langle X_2^2 \rangle\rangle$, si X_1 y X_2 están descorrelacionadas
  - La función característica de Y es

    $$G_Y(k) = G_{X_1 X_2}(k, k) \tag{32}$$

    Si X_1 y X_2 son independientes

    $$G_Y(k) = G_{X_1}(k) G_{X_2}(k) \tag{33}$$

- Como ejemplo particular vamos a ver el caso del caminante aleatorio con tiempo discreto. Simularía por ejemplo el caminar de un borracho que se mueve a lo largo de una línea de forma que en cada paso temporal se mueva hacia la derecha o hacia la izquierda con igual probabilidad. Su posible posición son $-\infty < n < \infty$. Cual es la probabilidad de que esté en la posición n después de r pasos, esto es $p_n(r)$, supuesto que en $t = 0$ estaba en $n = 0$.

- En cada paso temporal tenemos una variable estocástica que puede tomar los valores $+1$ o -1 con probabilidad $1/2$. La posición después de r pasos es

$$Y = X_1 + \ldots + X_r \tag{34}$$

Según los resultados anteriores se tiene $\langle Y \rangle = 0$ y

$$\langle Y^2 \rangle = r \langle X^2 \rangle = r \tag{35}$$

comportamiento típico de los procesos de difusión. Por otra parte el desplazamiento por unidad de tiempo, esto es la velocidad, tiene una varianza

$$\left\langle \left(\frac{Y}{r} \right)^2 \right\rangle = \frac{1}{r} \text{as } r \to \infty. \tag{36}$$

- La función característica es

$$G_Y(k, r) = [G_X(k)]^r = [\frac{1}{2} e^{ik} + \frac{1}{2} e^{-ik}]^r. \tag{37}$$

Desarrollando el binomio y teniendo en cuenta la definición de función característica se tiene

$$p_n(r) = \frac{1}{2^r} \begin{pmatrix} r \\ (r-n)/2 \end{pmatrix} \tag{38}$$

que es la distribución de probabilidad binomial.

Ejercicio: Demostrar que para $r \to \infty$, introduciendo la densidad

$$P(x, r) \Delta x = \sum_{x < n r^{-1/2} < x + \Delta x} P_n(r) \tag{39}$$

se tiene

$$\lim_{r \to \infty} P(x, r) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \tag{40}$$

Ayuda: Utilizar la fórmula de Stirling.

Transformación de variables

En esta sección vamos a determinar la distribución de probabilidad de una variable estocástica que es función de otras variables estocásticas cuya distribución es conocida

- Sea $Y = f(X)$. Por ejemplo cuando dibujamos en escala logaritmica ($Y = \log X$) y la transformación de frecuencias a longitud de ondas (Y=1/X). En general el rango de Y difiere del rango de X. La probabilidad que Y tenga un valor entre y y $y + \Delta y$ es

$$P_Y(y)\Delta y = \sum_{y<f(x)<y+\Delta y} P_X(x)\; dx \tag{41}$$

o lo que es lo mismo:

$$P_Y(y) = \int \delta[f(x) - y] P_X(x)\; dx \tag{42}$$

de donde se tiene para la función característica:

$$G_Y(k) = \langle e^{ikf(X)} \rangle. \tag{43}$$

- Estas expresiones son válidas para X teniendo r componentes e Y s componentes, donde s puede ser o no igual a r. Consideremos como ejemplo $r = 3$ y $s = 1$ para la distribución de Maxwell de las velocidades $\boldsymbol{X}= (v_x, v_y, v_z)$ y queremos encontrar la distribución de la energía $E = \frac{1}{2}m(v_x^2 + v_y^2 + v_z^2)$,

$$\begin{aligned} P(E) &= \int \delta(\tfrac{1}{2}mv^2 - E) \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT}\; dv_x dv_y dv_z \\ &= 2\pi^{-1/2}(kT)^{-3/2}E^{1/2}e^{-E/kT} \end{aligned} \tag{44}$$

- Si $r = s$ uno podría hayar la inversa de la función f y obtener $X = g(Y)$. En este caso se tiene

$$P_Y(y) = P_X(x)|J| \tag{45}$$

con $J = \det(dx/dy)$ el Jacobiano de la transformación. En este caso se tiene

$$P_Y(y)d^r y = P_X(x)d^r x \tag{46}$$

La distribución de probabilidad Gaussiana

Ya la hemos visto anteriormente como límite de la distribución de un caminante aleatorio. Su forma general es:

$$P(x) = Ce^{-\frac{1}{2}Ax^2 - Bx} \quad (-\infty < x < \infty) \tag{47}$$

con $A > 0$, que determina la anchura, B que indica la posición del pico y C que es la normalización de la distribución:

$$C = \left(\frac{A}{2\pi}\right)^{1/2} e^{-B^2/2A} \tag{48}$$

- En términos del promedio $\mu_1 = -B/A$ y de la variancia $\sigma^2 = 1/A$ la distribución queda

$$P(x) = (2\pi\sigma^2)^{-1/2} \exp\left[-\frac{(x-\mu_1)^2}{2\sigma^2}\right] \tag{49}$$

- Su función característica es:

$$G(k) = e^{i\mu_1 k - \frac{1}{2}\sigma^2 k^2} \tag{50}$$

- Los cumulantes son

$$\kappa_1 = \mu_1, \quad \kappa_2 = \sigma^2, \quad \kappa_3 = \kappa_4 = \ldots = 0 \tag{51}$$

- Para r variables Gaussianas mutuamente independientes $X_1, \ldots, X_r$ su suma Y también es Gaussiana. y se tiene

$$\begin{aligned} \langle Y \rangle &= \sum_i \langle X_i \rangle \\ \sigma_Y^2 &= \sum_i \sigma_{X_i}^2 \end{aligned} \tag{52}$$

- La distribución Gaussiana r dimensional se escribe de forma general como

$$P(x) = C \exp\left[-\frac{1}{2}\sum_{i,j=1}^{r} A_{ij} x_i x_j - \sum_{i=1}^{r} B_i x_i\right] \tag{53}$$

  con $\boldsymbol{A}$ simétrica y definida positiva. En notación vectorial se escribe

$$P(\boldsymbol{x}) = C \exp[-\tfrac{1}{2}\boldsymbol{x} \cdot \boldsymbol{A} \cdot \boldsymbol{x} - \boldsymbol{B} \cdot \boldsymbol{x}] \tag{54}$$

  C se encuentra mediante una transformación de variables en la que $\boldsymbol{A}$ es diagonal

$$C = \left(\frac{\det \boldsymbol{A}}{(2\pi)^r}\right)^{1/2} \exp[-\tfrac{1}{2}\boldsymbol{B} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{B}] \tag{55}$$

- La función característica es

$$G(\boldsymbol{k}) = \exp[-\tfrac{1}{2}\boldsymbol{k} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{k} - i\boldsymbol{k} \cdot \boldsymbol{A}^{-1} \cdot \boldsymbol{B}] \tag{56}$$

de donde se tiene, expandiendo en potencias de $\boldsymbol{k}$ que

$$\begin{aligned} \langle X_i \rangle &= -\textstyle\sum_j (\boldsymbol{A}^{-1})_{ij} B_j, \\ \langle\langle X_i X_j \rangle\rangle &= (\boldsymbol{A}^{-1})_{ij} \end{aligned} \tag{57}$$

luego la matriz covariante es justo la inversa de la matriz $\boldsymbol{A}$. Así, la distribución conjunta gaussiana de r variables viene completamente determinada por los promedios de las componentes y por la matriz covariante. Supuesto que conocemos la distribución gaussiana conjunta, entonces si las componentes están descorrelacionadas son estadísticamente independientes.

Ejercicio: Usando la función característica demostrar que para $\boldsymbol{B} = 0$ los momentos de la distribución Gaussiana de r componentes se pone

$$\langle X_i X_j X_k \ldots \rangle = \sum \langle X_p X_q \rangle \langle X_u X_v \rangle \ldots \tag{58}$$

Teorema del límite central

Sean $X_1, \ldots, X_r$ un conjunto de r variables estocásticas independientes, cada una con la misma densidad de probabilidad Gaussiana $P_X(x)$ con promedio cero y varianza σ^2. Su suma Y tiene densidad de probabilidad

$$P_Y(y) = [2\pi r\sigma^2]^{-1/2} \exp\left[-\frac{y^2}{2r\sigma^2}\right]. \tag{59}$$

entonces $\langle Y^2 \rangle = r\sigma^2$.

Por otra parte se tiene

$$\left\langle \left(\frac{Y}{r}\right)^2 \right\rangle = \frac{\sigma^2}{r}. \tag{60}$$

Podemos entonces definir la suma escalada de la siguiente forma:

$$Z = \frac{X_1 + \ldots + X_r}{\sqrt{r}} \tag{61}$$

que tiene varianza σ^2 y distribución de probabilidad

$$P_Z(z) = [2\pi\sigma^2]^{-1/2} \exp\left[-\frac{z^2}{2\sigma^2}\right]. \tag{62}$$

El teorema del límite central establece que para $P_X(x)$ general (no necesariamente Gaussiana) con promedio nulo y varianza finita σ^2, el anterior resultado es cierto en el límite de $r \to \infty$. La demostración es sencilla. Construyamos la función característica para cada una de las variables X_i. Usando la definición de $G_X(k)$ y dado que $\langle X_i \rangle = 0$ se tiene

$$G_X(k) = 1 - \frac{1}{2}\sigma^2 k^2 + \ldots . \tag{63}$$

Teniendo en cuenta que Z es una suma escalada de las variables estocásticas X_i se tiene

$$G_Z(k) = \left[G_X\left(\frac{k}{\sqrt{r}}\right) \right]^2 = \left[1 - \frac{\sigma^2 k^2}{2r} \right]^r \to e^{-\frac{1}{2}\sigma^2 k^2} \tag{64}$$

que el la función característica de una distribución de probabilidad Gaussiana con media nula (TF inversa)

Ley de los grandes números

Mucho del contenido de la ley de los grandes números está contenida en el teorema del límite central, sin embargo es más general que dicho teorema.

- La Ley de los grandes números se aplica a un conjunto de r experimentos independientes: Si un evento A tiene una probabilidad p de ocurrir, entonces la frecuencia de ese evento tiende a p cuando $r \to \infty$. Para demostrar esta ley primero derivaremos la desigualdad de *Tchebycheff.*

- La desigualdad de *Tchebycheff* establece la relación entre la varianza y la probabilidad de que una variable estocástica X difiera de su promedio $\langle X \rangle$ en una cantidad ϵ. Por definición se tiene

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \langle X \rangle)^2 P_X(x)\, dx \tag{65}$$

de donde se tiene

$$\sigma^2 \geq \int_{-\infty}^{\langle X \rangle - \epsilon} (x - \langle X \rangle)^2 P_X(x)\, dx + \int_{\langle X \rangle + \epsilon}^{\infty} (x - \langle X \rangle)^2 P_X(x)\, dx \tag{66}$$

al quitar la parte de la integral en la que $|x - \langle X \rangle| \leq \epsilon$. Por tanto las integrales son tales que $|x - \langle X \rangle| \geq \epsilon$, y podemos sustituir en los integrandos $(x - \langle X \rangle)^2$ por ϵ^2 sin que la desigualdad cambie:

$$\sigma^2 \geq \epsilon^2 \left[\int_{-\infty}^{\langle X \rangle - \epsilon} P_X(x)\, dx + \int_{\langle X \rangle + \epsilon}^{\infty} P_X(x)\, dx \right] = \epsilon^2 P(|X - \langle X \rangle| \geq \epsilon) \tag{67}$$

donde $P(|X-\langle X\rangle| \geq \epsilon)$ es la probabilidad de que la variable estocástica X difiera de su promedio en más de $\pm\epsilon$. Así

$$P(|X - \langle X\rangle| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad \text{Desigualdad de } \textit{Tchebycheff.} \tag{68}$$

- Si consideramos ahora r experimentos independientes para la variable estocástica X podemos definir el valor medio de los resultados

$$Y = \frac{x_1 + \ldots + x_r}{r}. \tag{69}$$

Entonces la ley de los grandes números establece que

$$\lim_{r\to\infty} P(|Y - \langle X\rangle| \geq \epsilon) = 0. \tag{70}$$

Para demostrarlo, usamos el hecho de que por ser los eventos independientes se tiene

$$\begin{aligned} \langle Y\rangle &= \langle X\rangle \\ \sigma_Y^2 &= \frac{\sigma^2}{r} \end{aligned} \tag{71}$$

y usando la desigualdad de *Tchebycheff* se tiene:

$$P(|Y - \langle X\rangle| \geq \epsilon) \leq \frac{\sigma_Y^2}{\epsilon^2} = \frac{\sigma^2}{r\epsilon^2} \tag{72}$$

de donde asumiendo que σ^2 es finita se tiene

$$\lim_{r\to\infty} P(|Y - \langle X\rangle| \geq \epsilon) \to 0 \tag{73}$$