

Teorema del límite central

Desigualdades de Chebyshev y Markov
y ley de los grandes números

Fabián Abarca Calderón

IE0405 – Modelos Probabilísticos de Señales y Sistemas

12

Tema III



UNIVERSIDAD DE
COSTA RICA

EIE

Escuela de
Ingeniería Eléctrica

El teorema del límite central establece que la función de distribución probabilística de la **suma** o de las **medias muestrales** de un número grande de variables aleatorias aproxima a una **distribución gaussiana** (normal).

El teorema es clave en la teoría de probabilidad porque implica que los métodos probabilísticos y estadísticos que funcionan para distribuciones **normales** pueden ser aplicables a muchos problemas que involucran **otros tipos de distribuciones**¹.

También ayuda a explicar por qué la distribución normal es tan...normal.

¹*Stark, H. (2011). Probability, Statistics, and Random Processes for Engineers. Pearson.*

Teorema del límite central para sumas

Definición de la convergencia en la distribución

Considérese X_1, \dots, X_N variables aleatorias *independientes e idénticamente distribuidas* **IID**, con media común $\mu_{X_i} = \mu$ y desviación estándar $\sigma_{X_i} = \sigma$. Sea además

$$S_N = X_1 + \dots + X_N \quad (1)$$

la **suma** de ellas. Según el teorema del límite central, $S_N \sim \mathcal{N}(N\mu, N\sigma^2)$ conforme $N \rightarrow \infty$. Alternativamente, si Z es una distribución *normalizada* de S_N ,

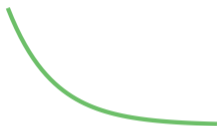
$$Z = \frac{S_N - \mu_{S_N}}{\sigma_{S_N}} = \frac{S_N - N\mu}{\sigma\sqrt{N}} \quad (2)$$

Z tiene una distribución $f_Z(z)$ que aproxima a $\mathcal{N}(0, 1)$ conforme $N \rightarrow \infty$.

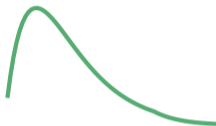
$\mathcal{N}(0, 1)$ es una función de distribución gaussiana de media 0 y varianza 1, donde \mathcal{N} alude a “normal”.

Visualización del teorema del límite central para sumas

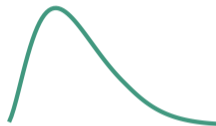
Caso 1: convolución de la función de densidad $f_{S_N}(x_1, \dots, x_N) = f_{X_1}(x_1) * \dots * f_{X_N}(x_N)$



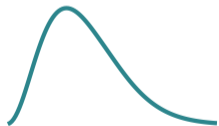
(a) X_i



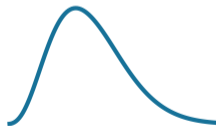
(b) $\sum_{i=1}^2 X_i$



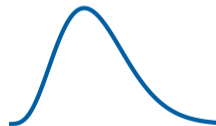
(c) $\sum_{i=1}^3 X_i$



(d) $\sum_{i=1}^4 X_i$



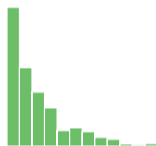
(e) $\sum_{i=1}^5 X_i$



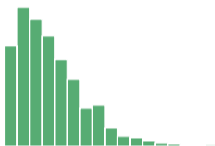
(f) $\sum_{i=1}^6 X_i$

Visualización del teorema del límite central para sumas

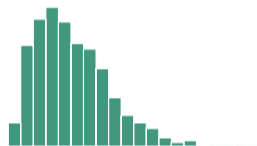
Caso 2: simulación de la suma de datos aleatorios generados de N v.a $X_i \sim \text{exponencial}(1)$



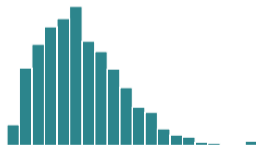
(a) X_i



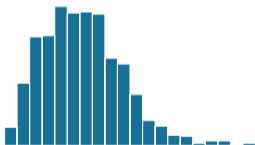
(b) $\sum_{i=1}^2 X_i$



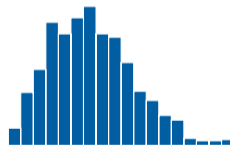
(c) $\sum_{i=1}^3 X_i$



(d) $\sum_{i=1}^4 X_i$



(e) $\sum_{i=1}^5 X_i$



(f) $\sum_{i=1}^6 X_i$

Los resistores tienen una resistencia *nominal* y un porcentaje de *tolerancia*. Por ejemplo, un resistor de 330 ohm con una tolerancia del 5 % se espera que tenga una resistencia entre 313,5 y 346,5 ohm.

Considérense cinco resistores de 330 ohm, escogidos aleatoriamente de una población con 5 % de tolerancia, y modélese la resistencia de cada una como una distribución uniforme en $[313,5 \quad 346,5]$. Si son conectados en serie, ¿cuál es la distribución de la resistencia R del sistema, dada por $R = X_1 + \dots + X_5$? Los X_i son los valores de resistencia **IID**.

Una variable aleatoria uniformemente distribuida en $[a \ b]$ tiene media $\mu = (a + b)/2$ y desviación estándar $\sigma = (b - a)/\sqrt{12}$. Para **cada resistencia**:

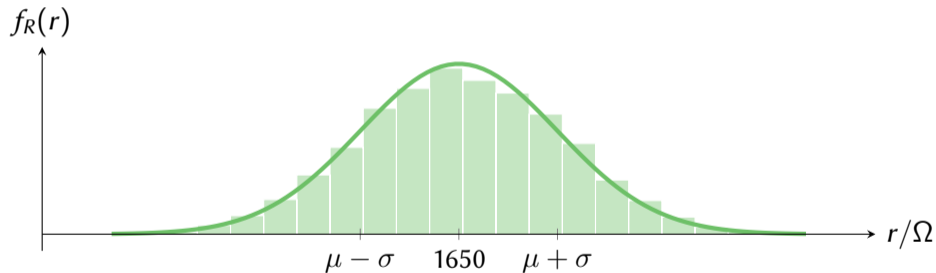
- la media es $E[X_i] = (313,5 + 346,5)/2 = 330 \ \Omega$, es decir, la resistencia nominal
- la desviación estándar es $\sigma = (346,5 - 313,5)/\sqrt{12} = 9,529 \ \Omega$

La resistencia **del sistema en serie** tiene una media y desviación estándar de

$$E[R] = N\mu = 5 \cdot 330 = 1650 \ \Omega$$

$$SD[R] = \sqrt{N}\sigma = \sqrt{5} \cdot 9,529 = 21,3 \ \Omega$$

¿Cómo es la distribución de probabilidad de $R = X_1 + \dots + X_5$? ¿Es R también distribuido uniformemente? Una simulación de 10 000 instancias distintas de R muestra:



que es una muy buena aproximación de $\mathcal{N}(\mu, \sigma) = \mathcal{N}(1650, 21,3)$.

Hay 40 formularios por revisar. Por los años de experiencia, la persona que los revisa sabe que el tiempo requerido para revisar cada uno es una variable aleatoria con un valor esperado de 6 minutos y una desviación estándar de 6 minutos. Si los tiempos de revisión son independientes y la persona inicia a las 7:50 a.m. revisando de forma continua, ¿cuál es la probabilidad de que termine antes de las 12:00 m.d.?

- Recordar que $Z = \frac{S_N - N\mu}{\sigma\sqrt{N}}$
- También que $f_Z(z) \rightarrow \mathcal{N}(0, 1)$
- ¿Cuánto es N ?
- ¿Cuánto es μ ?
- ¿Cuánto es σ ?
- ¿Cuánto es el tiempo disponible?

Ejemplo de la revisión de formularios antes del mediodía II

Para este problema, se debe aplicar el teorema del límite central para sumas, puesto que se trabajará con la suma del tiempo de revisión de todos los formularios. De la información del enunciado:

- 40 formularios: $N = 40$
- Tiempo promedio de revisión por formulario: $\mu = 6$
- Desviación estándar: $\sigma = 6$

La persona comienza a las 7:50 a.m. y para terminar antes de las 12:00 m.d. deberá hacerlo en menos de 250 minutos. El objetivo es, por tanto, encontrar $P(S_N \leq 250)$, o de forma equivalente, $P(Z \leq Z_0)$, donde

$$Z_0 = \frac{S_0 - N\mu}{\sigma\sqrt{N}} = \frac{250 - 40 \cdot 6}{6\sqrt{40}} = 0,263$$

Es decir,

$$P\left(\frac{S_N - 40 \cdot 6}{6\sqrt{40}} \leq \frac{250 - 40 \cdot 6}{6\sqrt{40}}\right)$$

$$P(Z \leq 0,263)$$

Utilizando la tabla para probabilidades acumulativas para valores positivos de Z , el valor más cercano a 0,263 está dado para 0,26.

$$P(Z \leq 0,263) \approx P(Z \leq 0,26) = 0,6026 \Rightarrow 60,26 \%$$

La persona que revisa tiene un 60,26 % de probabilidad de completar la revisión de formularios antes del mediodía.

Teorema del límite central para medias de muestras

El teorema del límite central para medias de muestras

Definición de la convergencia en la media

Si $\{X_i\}_{i=1}^N$ es una **muestra** de N elementos de una **población**, su **media muestral** es $\overline{X}_N = S_N/N = \mu_N$, donde $S_N = X_1 + X_2 + \dots + X_N$. Se puede considerar (2) para hacer

$$\frac{S_N - N\mu}{\sigma\sqrt{N}} \cdot \frac{1/N}{1/N} = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}$$

Definición de convergencia en la media

Sean X_1, \dots, X_N una muestra de variables aleatorias **IID** de una **población** con media común μ y desviación estándar σ y con una **media de la muestra** \overline{X}_N . Entonces

$$Z = \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}} \quad (3)$$

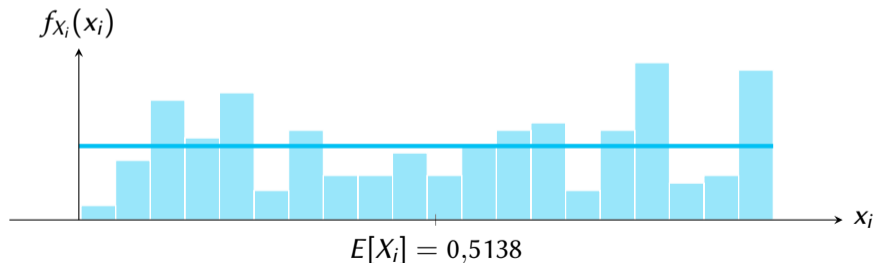
aproxima $\mathcal{N}(0, 1)$ conforme $N \rightarrow \infty$. Equivalentemente, $\overline{X}_N \sim \mathcal{N}(\mu, \sigma^2/N)$.

Visualización del teorema del límite central para medias de muestras I

Aunque una **población** tenga una distribución con media $\mu = E[X]$, una **realización** o **muestra** de esta distribución tendrá casi siempre un valor ligeramente distinto.

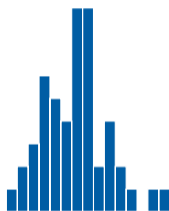
Ejemplo de una distribución uniforme

Sea $X \sim \text{unif}(0, 1)$ con $\mu = E[X] = 0,5$. Sea además X_i una **muestra** de esta distribución con 500 elementos y con una media estadística de $\mu_{X_i} = 0,5138 \neq 0,5$.

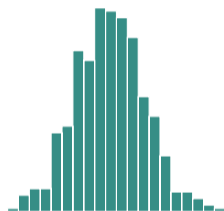


Visualización del teorema del límite central para medias de muestras II

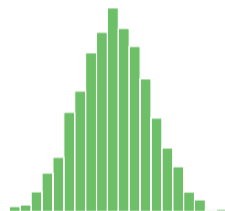
Al hacer una simulación de N muestras se obtienen N medias distintas μ_{x_i} . ¿Cómo se distribuyen estos valores alrededor de μ y cómo cambia la distribución según N ?



(a) $N = 50$



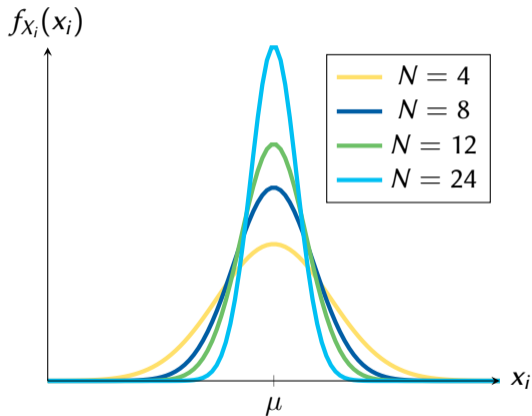
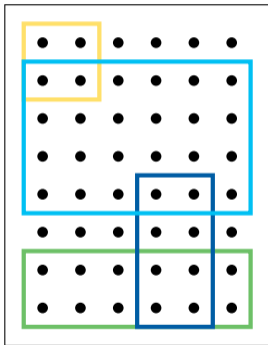
(b) $N = 500$



(c) $N = 5000$

Entre más grande es N más “agrupados” están los valores de la media de la muestra $\overline{X}_N = \mu_N$ alrededor de la “media verdadera” de la población, μ .

Visualización del teorema del límite central para medias de muestras III



Ejemplo del número de visitas mensuales al cajero automático I

Suponga que el número de veces que un cliente utiliza el cajero automático de un banco en un mes es una variable aleatoria con un valor medio de 3,2 y una desviación estándar de 2,4. El banco conoce estos datos con exactitud pues puede monitorear cada visita de la **población** de sus miles de clientes.

Si se selecciona aleatoriamente una muestra de 100 clientes, ¿qué tan probable es que el promedio de veces que el cajero es utilizado en la **muestra** exceda 3,5?

Ejemplo del número de visitas mensuales al cajero automático II

La probabilidad solicitada es $P(\overline{X}_N > 3,5)$, donde \overline{X}_N es el valor medio de la muestra. La muestra es grande ($N = 100$ clientes) y por tanto la distribución de \overline{X}_N se puede aproximar a una distribución normal.

$$P(\overline{X}_N > 3,5) \approx P\left(Z > \frac{3,5 - \mu}{\sigma/\sqrt{N}}\right) = P\left(Z > \frac{3,5 - 3,2}{0,24}\right) = 1 - F_Z(1,25) = 0,1056$$

La probabilidad es pequeña porque la muestra es grande y la desviación estándar de la muestra es muy pequeña, de solo 0,24, de forma tal que la media de una muestra de 100 personas se acerca “bastante” a la media de la población de quizá miles de personas.

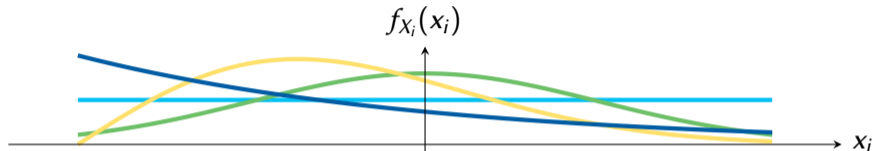
Desigualdades o límites superiores de probabilidad

Sin conocer exactamente la distribución de un fenómeno aleatorio, es posible establecer estimaciones de la probabilidad en un rango conociendo solo su varianza o media.

Desigualdad de Chebyshev

Premisas para la desigualdad de Chebyshev

Sea W una variable aleatoria con media 0. Esta es **cualquier** VA.

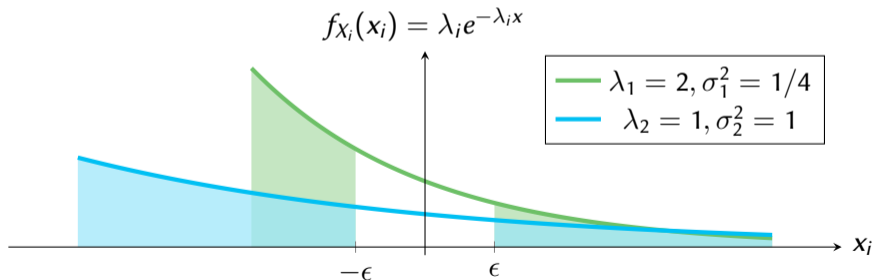


- La media de W es 0, pero cualquier realización simple de W puede estar bastante alejada de 0.
- La varianza es una medida de la dispersión de los valores de W alrededor de 0.
- Entre mayor el valor de la varianza de W , más probable es que el valor de W puede estar lejos de 0.

Dada la varianza σ^2 , ¿qué tan cercanos a $\mu = 0$ los valores de W podrían estar?

Fíjese un número $\epsilon > 0$ y búsqese la probabilidad de que W está más alejada que ϵ de su media $\mu = 0$. Supóngase que W tiene una función de densidad $f_W(w)$, entonces:

$$P(|W| \geq \epsilon) = \int_{|w| \geq \epsilon} f_W(w) dw = \int_{w^2 \geq \epsilon^2} \frac{\epsilon^2}{\epsilon^2} f_W(w) dw$$



Es esperable que la probabilidad $P(|W| \geq \epsilon)$ debería hacerse más grande conforme σ^2 se hace más grande, puesto que los valores de W están más dispersos.

$$\begin{aligned} \int_{w^2 \geq \epsilon^2} \frac{\epsilon^2}{\epsilon^2} f_W(w) \, dw &\leq \int_{w^2 \geq \epsilon^2} \frac{w^2}{\epsilon^2} f_W(w) \, dw \\ &\leq \int_{-\infty}^{\infty} \frac{w^2}{\epsilon^2} f_W(w) \, dw \\ &= \frac{1}{\epsilon^2} \int_{-\infty}^{\infty} w^2 f_W(w) \, dw \\ &= \frac{E[W^2]}{\epsilon^2} \\ P(|W| \geq \epsilon) &\leq \frac{\sigma^2}{\epsilon^2} \end{aligned}$$

- La primera desigualdad es porque el intervalo de integración contiene los puntos w donde $w^2 \geq \epsilon^2$ y, por lo tanto, el integrando será mayor si ϵ^2 se reemplaza por w^2 .
- La segunda desigualdad viene de aumentar el intervalo de integración de los puntos w donde $w^2 \geq \epsilon^2$, a la recta numérica de $-\infty$ a $+\infty$.
- $E[W^2]$ (el segundo momento ordinario) es igual en este caso a la varianza σ^2 (el segundo momento central) porque la media es cero y $\sigma^2 = E[W^2] - E^2[W]$.

Desigualdad de Chebyshev

Si $E[W] = 0$ y dado cualquier número positivo ϵ , el evento que W difiera en por lo menos ϵ de cero, está acotado por la probabilidad:

$$P(|W| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad (4)$$

Si $\mu = E[X] \neq 0$ pero $W = X - \mu$, entonces $E[W] = 0$, y el desarrollo anterior aplica a X :

Desigualdad de Chebyshev generalizada

Sea X una VA con media finita μ y varianza finita σ^2 . Entonces para $\epsilon > 0$ un número fijo, la probabilidad que X difiera en a lo menos ϵ de su media, está acotada:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2} \quad (5)$$

O en términos del evento complementario: $P(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$.

Comentario: Este es un límite “laxo” en el sentido de que **no** es muy restrictivo y por tanto no muy preciso o informativo.

Si X tiene tres posibles valores: $\{-1, 0, 1\}$, con probabilidades $\{\frac{1}{18}, \frac{8}{9}, \frac{1}{18}\}$, respectivamente. ¿Cuál es la probabilidad $P(|X - \mu| \geq 3\sigma)$ y cómo se compara con el límite de Chebyshev?

- Recordar que

$$E[X] = \sum_{i=1}^N x_i P(x_i)$$

- También que

$$\sigma_X^2 = E[(X - \bar{X})^2] = E[X^2] - E^2[X]$$

- Siendo que

$$P(|W| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

- Pero una forma equivalente es

$$P(|X - \mu_X| \geq k\sigma_X) \leq \frac{1}{k^2}$$

La media y la varianza de la VA discreta se obtienen de la siguiente forma:

$$E[X] = \sum_{i=1}^3 x_i P(x_i) = (-1)\frac{1}{18} + (0)\frac{8}{9} + (1)\frac{1}{18} = 0$$

$$\text{Var}[X] = \sum_{i=1}^3 (x_i - E[X])^2 P(x_i) = (-1)^2 \frac{1}{18} + (0)^2 \frac{8}{9} + (1)^2 \frac{1}{18} = \frac{1}{9}$$

Utilizando la definición provista de la desigualdad de Chebyshev, se obtiene

$$P(|X - \mu_X| \geq k\sigma_X) = P(|X - 0| \geq 3\frac{1}{3}) \leq \frac{1}{3^2} = \frac{1}{9}$$

Mientras tanto, utilizando la PDF propiamente, se puede encontrar la probabilidad $P(|X| \geq 1)$ solicitada. Considerando que solo hay tres valores posibles de X , $\{-1, 0, 1\}$, los elementos de interés son $\{-1, 1\}$ cuyas probabilidades son $1/18 + 1/18 = 1/9$, igual que con Chebyshev.

En general, la desigualdad de Chebyshev será mucho menos restrictiva que el análisis de la PDF, pero en este caso de ejemplo resultaron iguales.

Desigualdad de Markov

Desigualdad de Markov

Si X es una VA con $f_X(x) = 0$ para $x < 0$, entonces X es llamada una VA no-negativa, para la cual aplica la desigualdad de Markov:

$$P(X \geq \epsilon) \leq \frac{E[X]}{\epsilon} \quad (6)$$

Comentario: En contraste con el límite de Chebyshev, que involucra tanto la media como la varianza, este límite requiere únicamente de la media de X .

La consideración es ahora en relación con la definición del valor esperado (el *momento ordinario de primer orden*):

$$\begin{aligned} E[X] &= \int_0^{\infty} x f_X(x) dx \geq \int_{\epsilon}^{\infty} x f_X(x) dx \\ &\geq \int_{\epsilon}^{\infty} \epsilon f_X(x) dx = \epsilon \int_{\epsilon}^{\infty} f_X(x) dx \\ &= \epsilon P(X \geq \epsilon) \end{aligned}$$

Es posible asumir que en la manufactura de resistores eléctricos de baja calidad de 1000Ω , la resistencia promedio es en efecto de 1000Ω , según se determina por un análisis estadístico de mediciones, pero hay una gran variación alrededor de este valor. Si todos los resistores por encima de 1500Ω deben descartarse, ¿cuál es la fracción máxima de resistores que terminarían por fuera?

- Recordar que

$$P(X \geq \epsilon) \leq \frac{E[X]}{\epsilon}$$

En este problema tenemos una distribución no-negativa (no hay valores negativos de resistencia eléctrica), y disponemos únicamente de la media y no de la varianza de la distribución, entonces aplica la desigualdad de Markov.

$$P(X \geq \epsilon) = P(X \geq 1500) \leq \frac{E[X]}{\epsilon} = \frac{1000}{1500} = \frac{2}{3}$$

Es posible, aun así, que la proporción de resistencias descartadas sea menor. Bastaría con conocer la distribución para tener más certeza.

Ley de los grandes números

Sea $\{X_i\}_{i=1}^N$ una muestra aleatoria. Los X_i son variables aleatorias idénticamente distribuidas e independientes **IID** con media común μ y varianza σ^2 . Se espera intuitivamente que la media de la muestra, \overline{X}_N , debería ser cercana a la media de la población μ para N grande. Esto es

$$\lim_{N \rightarrow \infty} \overline{X}_N = \lim_{N \rightarrow \infty} \frac{S_N}{N} = \mu$$

donde $S_N = X_1 + \dots + X_N$ es la suma.

Sea $\epsilon > 0$ un número fijo. Por la independencia de la secuencia X_1, \dots, X_N en la muestra aleatoria, se cumple que:

$$E[S_N] = N\mu \quad \sigma_{S_N}^2 = N\sigma^2$$

La desigualdad de Chebyshev aplicada a S_N establece que:

$$P(|S_N - N\mu| \geq N\epsilon) \leq \frac{\sigma_{S_N}^2}{(N\epsilon)^2} = \frac{N\sigma^2}{N^2\epsilon^2} = \frac{\sigma^2}{N\epsilon^2}$$

En términos de la media $\overline{X_N}$ de la muestra,

$$P(|\overline{X_N} - \mu| \geq \epsilon) = P\left(\left|\frac{S_N}{N} - \mu\right| \geq \epsilon\right) = P(|S_N - N\mu| \geq N\epsilon) \leq \frac{\sigma^2}{N\epsilon^2}$$

Para $N \rightarrow \infty$ la cota de la derecha en la última ecuación tiende a 0 y entonces:

Ley de los grandes números

Sea $\{X_i\}_{i=1}^N$ una muestra aleatoria con media común μ y varianza σ^2 . Sea

$$S_N = X_1 + \cdots + X_N$$

Entonces

$$P\left(\left|\frac{S_N}{N} - \mu\right| \geq \epsilon\right) \rightarrow 0 \quad (7)$$

conforme $N \rightarrow \infty$ para cualquier $\epsilon > 0$.

En aquel Imperio, el arte de la cartografía logró tal perfección que el mapa de una sola provincia ocupaba toda una ciudad, y el mapa del Imperio, toda una provincia. Con el tiempo, estos mapas desmesurados no satisficieron y los colegios de cartógrafos levantaron un mapa del Imperio, que tenía el tamaño del Imperio y coincidía puntualmente con él.

Menos adictas al estudio de la cartografía, las generaciones siguientes entendieron que ese dilatado mapa era inútil y no sin impiedad lo entregaron a las inclemencias del sol y los inviernos. En los desiertos del oeste perduran despedazadas ruinas del mapa, habitadas por animales y por mendigos; en todo el país no hay otra reliquia de las disciplinas geográficas.