1 Desigualdades y teoremas límites

Las desigualdades y el comportamiento asintótico de secuencias de variables aleatorias es un parte importante de la teoria de las probabilidades. El principal contexto involucra una secuencia $X_1, X_2, \ldots$ de variables aleatorias independientes idénticamente distribuidas con media μ y varianza σ^2.

Existen importantes desigualdades que se pueden aplicarse en este escenario: la desigualdad de Markov, utiliza sólo la esperanza de una variable aleatoria. La desigualdad de Chebychev, utiliza tanto la esperanza como la varianza de una variable aleatoria y normalmente genera cotas más estrictas que el resultado de Markov.

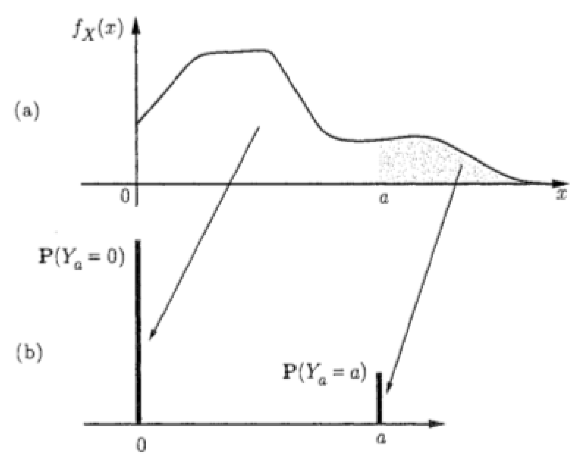
1.1 Desigualdad de Markov

Teorema 1.1 Sea X es una variable no negativa y supongamos que $\mathbb{E}(X)$ existe. Entonces se cumple:

$$\mathbb{P}(X > a) \leq \frac{\mathbb{E}(X)}{a}.$$

Para algún $a > 0$.

Este resultado es ilustrado mediante el siguiente gráfico:



Donde la parte a) muestra el `función densidad de probabilidad` de una variable aleatoria no negativa X. La parte b) muestra la `función de masa de probabilidad` de una variable Y_a, que se construye de la siguiente manera: Toda la masa de probabilidad en el PDF de X que se encuentra entre 0 y a se le asigna 0 y toda la masa que se encuentra por encima de a se le asigna a a, puesto que la masa se desplaza a la izquierda, la esperanza sólo puede disminuir.

Vayamos con la prueba.

Fijemos un número positivo a y consideremos una variable aleatoria Y_a definida como:

$$Y_a = \begin{cases} 0, & \text{si } X < a \\ a, & \text{si } X \geq a \end{cases}$$

Luego:

$$Y_a \leq X$$

Lo que implica, si tomamos esperanzas en ambos lados:

$$\mathbb{E}(Y_a) \leq \mathbb{E}(X)$$

$$\mathbb{E}(X) \geq \mathbb{E}(Y_a) = a\mathbb{P}(Y_a = a) = a\mathbb{P}(X \geq a)$$

Por tanto :

$$a\mathbb{P}(X \leq a) \leq \mathbb{E}(X).$$

Ejemplo 1.1 Sea X la variable aleatoria que denota la edad (en años) de un niño de un determinado colegio. Si la edad promedio de edad de ese colegio es 12.5 años, entonces usando la desigualdad de Markov, la probabilidad de que un niño es tiene por lo menos 20 años es:

$$\mathbb{P}(X \geq 20) \leq 12.5/20 = 0.6250.$$

En general si X es una variable aleatoria y h una función no decreciente y no negativa. La esperanza $h(X)$ asumiendo que existe, es dada por:

$$\mathbb{E}(h(X)) = \int_{-\infty}^{\infty} h(u) f_X(u) du,$$

y escribimos:

$$\int_{-\infty}^{\infty} h(u) f_X(u) du \geq \int_{a}^{\infty} h(u) f_X(u) du \geq h(a) \int_{t}^{\infty} f_X(u) du = h(a)\mathbb{P}(X \geq a).$$

Esto conduce a la desigualdad de Markov: $\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(h(X))}{h(a)}$ para $a > 0$.

1.2 Desigualdad de Chebyshev

La desigualdad de Chebyshev, dice que si una variable aleatoria tiene una pequeña varianza, entonces la probabilidad que toma un valor grande lejos de la media es también pequeño.

Teorema 1.2 Sea $\mu = \mathbb{E}(X)$ y $\sigma^2 = \text{Var}(X)$. Entonces para un $t > 0$

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2} \qquad y \qquad \mathbb{P}(|Z| \geq k) \leq \frac{1}{k^2}.$$

donde $Z = (X - \mu)/\sigma$. En particular $\mathbb{P}(|Z| > 2) \leq 1/4$ y $\mathbb{P}(|Z| > 3) \leq 1/9$.

Aplicando la desigualdad de Markov, podemos concluir que:

$$\mathbb{P}((X - \mu)^2 \leq t^2) \leq \frac{\mathbb{E}(X - \mu)^2}{t^2} = \frac{\sigma^2}{t^2}.$$

Para completar el evento $(X - \mu)^2 \geq t^2$ es idéntico al evento $|X - \mu| \geq t$, así:

$$\mathbb{P}(|X - \mu| \leq t) = \mathbb{P}((X - \mu)^2 \leq t^2) \leq \frac{\sigma^2}{t^2}.$$

Para la segunda parte se coloca $t = k\sigma$.

Ejemplo 1.2 Volvamos al ejemplo (1.1) de los niño de la escuela y recalculamos el límite usando la desigualdad de Chebychev. En este caso también necesitamos la varianza de las edades de los niños, que tomamos como 3. Buscamos la probabilidad de que un niño en la escuela pueda ser mayor que 20. Primero debemos poner esto en la forma requerida por La ecuación de Chebychev:

$$\mathbb{P}(X \geq 20) = \mathbb{P}[(X - \mathbb{E}(X)) \geq (20 - \mathbb{E}(X))] = \mathbb{P}[(X - \mathbb{E}(X)) \geq 7.5)].$$

Sin embargo:

$$\mathbb{P}[(X - 12.5) \geq 7.5] \neq \mathbb{P}[|X - 12.5| \geq 7.5)] = \mathbb{P}(5 \leq X \geq 20).$$

Así no podemos aplicar la desigualdad de Chebyshev para calcular directamente $\mathbb{P}(X \geq 20)$. En efecto, calculamos, una cota para $\mathbb{P}(5 \leq X \geq 20)$ y obtenemos:

$$\mathbb{P}[|X - \mathbb{E}(X)| \geq 7.5)] \leq \frac{3}{(7.5)^2} = 0.0533,$$

que sigue siendo un límite mucho más estricto que el obtenido anteriormente.

1.3 Desigualdad de Hoeffding

Empecemos con un resultado importante:

Proposición 1.1 Supongamos que $\mathbb{E}(X) = 0$ y que $a \leq x \leq b$. Entonces:

$$\mathbb{E}(e^{tX}) \leq e^{t^2(b-a)^2/8}.$$

Para desarrollar la desigualdad de Hoeffding, necesitamos el `método de Chernoff.`

Proposición 1.2 (Método de Chernoff) Sea X una variable aleatoria. Entonces:

$$\mathbb{P}(X > \epsilon) \leq \inf_{t \geq 0} e^{-t\epsilon} \mathbb{E}(e^{tX}).$$

Este método puede ser derivado de la desigualdad de Markov con la función $h(x) = e^{tx}$ que es la función generadora de momentos $\mathbb{E}(e^{tX})$.

Para algún $t > 0$

$$\mathbb{P}(X > \epsilon) = \mathbb{P}(e^X > e^\epsilon) = \mathbb{P}(e^{tX} > e^{t\epsilon}) \leq e^{-t\epsilon}\mathbb{E}(e^{tX}).$$

Desde que esto es cierto para $t \geq 0$ el resultado sigue.

Teorema 1.3 (Desigualdad de Hoeffding) Sean $Y_1, Y_2, \ldots, Y_n$ observaciones indepedientes, tal que $\mathbb{E}(Y_i) = 0$ y $a_i \leq Y_i \leq b_i$. Sea $\epsilon > 0$. Entonces, para algún $t > 0$

$$\mathbb{P}(|\overline{Y}_n - \mu| \geq \epsilon) \leq 2e^{-2n\epsilon^2/(b-a)^2}.$$

Sin pérdida de generalidad, supongamos $\mu = 0$. Entonces se tiene:

$$\begin{aligned}\mathbb{P}(|\overline{Y}_n| \geq \epsilon) &= \mathbb{P}(\overline{Y}_n \geq \epsilon) + \mathbb{P}(\overline{Y}_n| \leq -\epsilon) \\ &= \mathbb{P}(\overline{Y}_n \geq \epsilon) + \mathbb{P}(-\overline{Y}_n \geq \epsilon)\end{aligned}$$

Usamos el método de chernoff. Para algún $t > 0$, desde la desigualdad de Markov, que:

$$\begin{aligned}\mathbb{P}(\overline{Y}_n \geq \epsilon) &= \mathbb{P}\Big(\sum_{i=1}^{n} Y_i \geq n\epsilon\Big) = \mathbb{P}(e^{\sum_{i=1}^{n} y_i} \geq e^{n\epsilon}) \\ &= \mathbb{P}\Big(e^{t\sum_{i=1}^{n} y_i} \geq e^{tn\epsilon}\Big) \leq e^{-tn\epsilon}\mathbb{E}\Big(e^{t\sum_{i=1}^{n} y_i}\Big) \\ &= e^{-tn\epsilon}\prod_i (e^{tY_i}) = e^{-tn\epsilon}(\mathbb{E}(e^{tY_i}))^n\end{aligned}$$

Por la proposición (1.1): $\mathbb{E}(e^{tY_i}) \leq e^{t^2(b-a)^2/8}$. Así:

$$\mathbb{P}(\overline{Y}_n \geq \epsilon) \leq e^{-tn\epsilon}e^{t^2 n(b-a)^2/8}$$

Esto se reduce al mínimo cuando $t = 4\epsilon/(b-a)^2$ dando el resultado:

$$\mathbb{P}(\overline{Y}_n \geq \epsilon) \leq e^{-2n\epsilon^2/(b-a)^2}$$

Aplicando el mismo argumento a $\mathbb{P}(-\overline{Y}_n \geq \epsilon)$ se produce el resultado.

Ejemplo 1.3 Sea $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$. Desde la desigualdad Hoeffding se tiene,

$$\mathbb{P}(|\overline{X}_n - p| > \epsilon) \leq 2\epsilon^{-2n\epsilon^2}.$$

donde $\overline{X}_n = n^{-1}\sum_{i=1}^{n} X_i$

Sea $X_1, X_2, \ldots, X_n \sim \text{Bernoulli}(p)$. Sea $n = 100$ y $\epsilon = .2$. Por la desigualdad de Chebyshev, tenemos

$$\mathbb{P}(|\overline{X}_n - p| > \epsilon) \leq .0625$$

De acuerdo a la desigualdad de Hoeffding:

$$\mathbb{P}(|\overline{X}_n - p| > .2) \leq 2e^{-2(100)(.2)^2} = .00067$$

que es mucho más pequeño que el resultado obtenido por la desigualdad de Chebyshev .0625.

El siguiente resultado extiende la desigualdad de Hoeffding a funciones más generales $g(x_1, \ldots, x_n)$. Consideramos la desigualdad de McDiarmind.

1.4 Desigualdad de McDiarmind

Teorema 1.4 Teorema de McDiarmid Sea $X_1, X_2, \ldots X_n$ variables aleatorias independientes. Suponganse que

$$\sup_{x_1,\ldots,x_n,x_i'} \Big| g(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - g(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n) \Big| \leq c_i$$

para $i = 1, \ldots, n$. Entonces

$$\mathbb{P}\Big(g(X_1, \ldots, X_n) - \mathbb{E}(g(X_1, \ldots, X_n)) \geq \epsilon\Big) \leq \exp\Big\{-\frac{2\epsilon^2}{\sum_{i=1}^{n} c_i^2}\Big\}.$$

Si tomamos $g(x_1,\ldots,x_n) = n^{-1}\sum_{i=1}^{n} x_i$, es la desigualdad de Hoeffding.

Ejemplo 1.4 Supongamos que lanzamos m pelotas en n recipientes. ¿ Qué fracción de los recipientes están vacios?.

Sea Z el número de recipientes vacios y sea $F = Z/n$ la fracción de recipientes vacios. Podemos escribir $Z = \sum_{i=1}^{n} Z_i$, donde $Z_i = 1$ de un recipiente i está vacío y $Z_i = 0$ en otro caso. Entonces:

$$\mu = \mathbb{E}(Z) = \sum_{i=1}^{n} \mathbb{E}(Z_i) = n(1-1/n)^m = ne^{m\log(1-1/n)} \approx ne^{-m/n}.$$

y $\theta = \mathbb{E}(F) = \mu/n \approx e^{-m/n}$. ¿ Cuán cerca está Z de μ?. Se debe notar que las variables Z_i no son independientes, por lo que no se puede aplicar la desigualdad de Hoeffding. En este caso procedemos de otra forma:

Definamos las variables $X_1,\ldots,X_m$ donde $X_s = i$ si la pelota s cae en el recipiente i. Entonces $Z = g(X_1,\ldots,X_m)$. Si movemos una pelota a un recipiente diferente, entonces Z puede cambiar a lo más 1. Así del Teorema de McDiarmind con $c_i = 1$ tenemos:

$$\mathbb{P}(|Z-\mu| > t) \leq 2e^{-2t^2/m}.$$

Como las fracción de recipientes vacios es $F = Z/m$ con media $\theta = \mu/n$. Tenemos:

$$\mathbb{P}(|F-\theta| > t) = \mathbb{P}(|Z-\mu| > nt) \leq 2e^{-2n^2t^2/m}.$$

1.5 Cotas para los valores esperados

Teorema 1.5 Desigualdad de Cauchy -Schwartz Si X y Y tienen varianza finita, entonces:

$$\mathbb{E}|XY| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y^2)}.$$

La desigualdad de Cauchy-Schwarz puede ser escrita como:

$$\mathrm{Cov}^2(X,Y) \leq \delta_X^2\delta_Y^2.$$

Teorema 1.6 Desigualdad de Jensen Si g es convexa, entonces:

$$\mathbb{E}g(X) \geq g(\mathbb{E}X)$$

Si g es cóncava

$$\mathbb{E}g(X) \leq g(\mathbb{E}X).$$

En efecto, sea $L(x) = a + bx$ una línea tangente a $g(x)$ en el punto $\mathbb{E}(X)$. Desde que g es convexa, está por encima de la línea $L(X)$. Así:

$$\mathbb{E}g(X) \geq \mathbb{E}L(X) = \mathbb{E}(a+bX) = a + b\mathbb{E}(X) = L(\mathbb{E}(X)) = g(\mathbb{E}X).$$

De la desigualdad de Jensen, se cumple $\mathbb{E}(X^2) \geq (\mathbb{E}X)^2$ y si X es positiva, entonces $\mathbb{E}(1/X) \geq 1/\mathbb{E}(X)$. Desde que log es cóncava, $\mathbb{E}(\log X) \leq \log\mathbb{E}(X)$.

2 La ley de promedios

Si lanzamos un dado 5 millones de veces y se mantiene un registro de los datos. El promedio de los números los cuales fueron lanzados es 3.500867. Desde que la media de cada lanzamiento es $\frac{1}{6}(1+2+\cdots+6) = 3\frac{1}{2}$, este resultado muestra que para un x_i el i ésimo lanzamiento, el promedio definido como:

$$a_n = \frac{1}{n}(x_i + x_2 + \cdots + x_n)$$

se acerca a la media $3\frac{1}{2}$, cuando $n \to \infty$.

En general si tenemos una secuencia $X_1, X_2, \ldots$ de variables aleatorias identicamente distribuidas e independientes cada una teniendo una media μ, deberiamos probar que el promedio:

$$\overline{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$$

converge cuando $n \to \infty$ a la media μ.

Definición 2.1 Decimos que la secuencia $X_1, X_2, \ldots$ de variables aleatorias `converge en media cuadrada a la variable X` si se cumple:

$$\mathbb{E}([X_n - X]^2) \to 0 \text{ cuando } n \to \infty$$

y se escribe $X_n \to x$ `en media cuadrada cuando` $n \to \infty$.

Ejemplo 2.1 Sea X_n una secuencia de variables aleatorias discretas con una función de masa de probabilidad:

$$\mathbb{P}(X_n = 1) = \frac{1}{n}, \quad \mathbb{P}(X_n = 2) = 1 - \frac{1}{n}.$$

Entonces X_n converge a la variable aleatoria 2 en media cuadrada cuando $n \to \infty$, desde que

$$\begin{aligned} \mathbb{E}([X_n - 2]^2) &= (1-2)\frac{1}{n} + (2-2)^2\Big(1 - \frac{1}{n}\Big) \\ &= \frac{1}{n} \to 0 \quad \text{cuando} \;\; n \to \infty. \end{aligned}$$

2.1 Primera versión de la Ley de los Grandes Números

Teorema 2.1 Sea una secuencia $X_1, X_2, \ldots$ una secuencia de variables independientes cada una con media μ y varianza σ^2. El promedio de los primeras n variables aleatorias X_i satisfacen cuando $n \to \infty$ que:

$$\frac{1}{n}(X_1 + X_2 + \cdots + X_n) \to \mu \quad \text{en media cuadrada.}$$

Sea $S_n = X_1 + X_2 + \cdots X_n$, la n-ésima suma parcial de los X_i.

Entonces:

$$\mathbb{E}\Big(\frac{1}{n}S_n\Big) = \frac{1}{n}\mathbb{E}(X_1 + X_2 + \cdots + X_n) = \frac{1}{n}n\mu = \mu.$$

y así:

$$\begin{aligned}\mathbb{E}\Big(\Big[\frac{1}{n}S_n-\mu\Big]^2\Big) &= \mathrm{Var}\Big(\frac{1}{n}S_n\Big)\\ &= \frac{1}{n^2}\mathrm{Var}(X_1+X_2+\cdots+X_n)\\ &= \frac{1}{n^2}(\mathrm{Var}X_1+\cdots+\mathrm{Var}X_n)\\ &= \frac{1}{n}n^2\sigma^2=\frac{1}{n}\sigma^2\to 0 \quad \text{cuando} \quad n\to\infty\end{aligned}$$

y así, $n^{-1}S_n\to\mu$ en media cuadrática cuando $n\to\infty$.

Definición 2.2 Decimos que la secuencia de variables aleatoria $X_1, X_2, \ldots$ de variables aleatorias `converge en probabilidad` a X si,

$$\texttt{Para todo } \epsilon>0 \quad \mathbb{P}(|X_n-X|>\epsilon)\to \quad \text{cuando} \quad n\to\infty$$

Si esto se cumple, escribimos $X_n\to X$ `en probabilidad cuando` $n\to\infty$

Teorema 2.2 Si $X_1, X_2, \ldots$ es una secuencia de variables aleatorias y $X_n\to X$ en media cuadrada cuando $n\to\infty$ entonces $X_n\to X$ en probabilidad.

Para hacer la prueba de este teorema, escribamos la `la desigualdad de Chebyshev` de la siguiente manera

Si Y es una variable aleatoria y $\mathbb{E}(Y^2)<\infty$, entonces:

$$\mathbb{P}(|Y|\geq t)\leq\frac{1}{t^2}\mathbb{E}(Y^2) \quad \text{para} \quad t>0$$

Luego aplicando la desigualdad de Chebyshev a la variable aleatoria $Y=X_n-X$ para encontrar que:

$$\mathbb{P}(|X_n-X|>\epsilon)\leq\frac{1}{\epsilon^2}\mathbb{E}([X_n-X]^2) \quad \epsilon>0.$$

Si $X_n\to X$ en media cuadrada cuando $n\to\infty$, el lado derecho de la desigualdad tiende a 0 cuando $n\to\infty$ y así tiende a 0 para todo $\epsilon>0$, como es requerido. El caso contrario no se cumple.

2.2 Ley Débil de los Grandes Números

Teorema 2.3 Sea $X_1, X_2, \ldots$ una secuencia de variables aleatorias, cada una con media μ y varianza σ^2. El promedio de los primeros n de los X_i satisface, cuando $n\to\infty$,

$$\frac{1}{n}(X_1+X_2+\cdots+X_n)\to\mu \quad \text{en probabilidad.}$$

Ejemplo 2.2 Sea X_n una secuencia de variables aleatorias que convergen en probabilidad, mostremos que no converge en media cuadrada. Si las variables aleatorias tienen una función de masa de probabilidad:

$$\mathbb{P}(X_n=0)=1-\frac{1}{n}, \quad \mathbb{P}(X_n=n)=\frac{1}{n}.$$

Entonces, para $\epsilon > 0$ y para un n dado, tenemos:

$$\mathbb{P}(|X_n| > \epsilon) = \mathbb{P}(X_n = n) = \frac{1}{n} \to 0 \quad \text{cuando} \quad n \to \infty$$

dando que $X_n \to 0$ en probabilidad. Por otro lado,

$$\begin{aligned}\mathbb{E}([X_n - 0]^2) = \mathbb{E}(X_n^2) &= 0\Big(1 - \frac{1}{n}\Big) + n^2\frac{1}{n} \\ &= n \to \infty \ \text{ cuando } \ n \to \infty.\end{aligned}$$

así X_n no converge a 0 en media cuadrada.

Ejemplo 2.3 Considere el lanzamiento de una moneda donde la probabilidad de obtener cara es p. Sea X_i el resultado de un lanzamiento $(0-1)$. Por lo tanto, $p = \mathbb{P}(X_i = 1) = \mathbb{E}(X_i)$. La fracción de caras después de n lanzamientos es $\overline{X}_n$. De acuerdo con la ley de los grandes números, $\overline{X}_n$ converge en probabilidad a p. Esto no quiere decir que $\overline{X}_n$ sea numéricamente igual p. Esto significa que, cuando n es grande, la distribución de $\overline{X}_n$ está `concentrado` alrededor p.

Supongamos que $p = 1/2$, encontremos el valor de n para que $\mathbb{P}(.4 \leq \overline{X}_n \leq .6) \geq .7$.

Para ello tenemos que $\mathbb{P}(\overline{X}_n) = p = 1/2$ y $\text{Var}(\overline{X}_n) = \sigma^2/n = p(1-p)/n = 1/n$. Por la desigualdad de Chebyshev, se tiene:

$$\begin{aligned}\mathbb{P}(.4 \leq \overline{X}_n \leq .6) &= \mathbb{P}(|\overline{X}_n - \mu) \leq .1) \\ &= 1 - \mathbb{P}(|\overline{X}_n - \mu| > .1) \\ &\geq 1 - \frac{1}{4n(.1)^2} = 1 - \frac{25}{n}\end{aligned}$$

Esta expresión es mayor que .7 si $n = 84$.

2.3 Teorema del Límite Central

Teorema 2.4 Sea $X_1, X_2, \ldots$ variables aleatorias identicamente distribuidas e independientes cada una con media μ y varianza distinta de cero σ^2. Por la ley de los grandes números, la suma $S_n = X_1 + X_2 + \cdots + X_n$ es casi tan grande que $n\mu$ para valores de n muy grandes. Un siguiente paso es determinar el orden de la diferencia $S_n - n\mu$, que resulta ser de orden $\sqrt{n}$.

Se define la versión estándar de S_n

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{Var}(S_n)}}$$

Esta es una función lineal $Z_n = a_n S_n + b_n$ de S_n, donde a_n y b_n tienen que ser elegida de manera que $\mathbb{E}(Z_n) = 0$ y $\text{Var}(Z_n) = 1$. Además:

$$\begin{aligned}\mathbb{E}(S_n) &= \mathbb{E}(X_1) + \mathbb{E}(X_2) + \cdots + \mathbb{E}(X_n) \\ &= n\mu\end{aligned}$$

También,

$$\begin{aligned}\operatorname{Var}(S_n) &= \operatorname{Var}(X_1) + \operatorname{Var}(X_2) + \cdots + \operatorname{Var}(X_n) \\ &= n\sigma^2\end{aligned}$$

y así,

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

Muchas de las propiedades de Z_n se pueden sintetizar en el teorema de límite central:

Sean $X_1, X_2, \ldots$ variables aleatorias identicamente distribuidas e independiente, cada una con media μ y varianza distinta de cero σ^2. La versión estándar:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

de la suma $S_n = X_1 + X_2 + \cdots + X_n$, satisface cuando $n \to \infty$:

$$\mathbb{P}(Z_n \leq x) \to \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \Phi(x) \quad \text{para} \;\; x \in \mathbb{R}.$$

El lado derecho de la última de la ecuación es sólo la función de distribución de la distribución normal con media 0 y varianza 1, así que esta expresión se puede escribir como:

$$\mathbb{P}(Z_n \leq x) \to \mathbb{P}(Y \leq x) \quad \text{para} \;\; x \in \mathbb{R}.$$

donde Y es la variable aleatoria con esta distribución normal estándar. También podemos escribir el teorema de la siguiente forma:

$$\lim_{n\to\infty} \mathbb{P}(Z_n \leq x) = \Phi(x) \;\; \text{para cada } x.$$

El teorema del límite central, nos permite calcular probabilidades relacionadas a Z_n como si Z_n fuese normal. Desde que la normalidad es preservada bajo transformaciones lineales, esto es equivalente a tratar S_n como una distribución normal con media $n\mu$ y varianza $n\sigma^2$.

Sea $S_n = X_1 + X_2 + \cdots + X_n$ donde las X_i son variables aleatorias idénticamente distribuidas e independientes cada una con media μ y varianza distinta de cero σ^2. Si n es muy grande, la probabilidad $\mathbb{P}(S_n \leq c)$ puede ser aproximada considerando S_n como si fuese normal de acuerdo al siguiente procedimiento:

- Calculamos la media $n\mu$ y la varianza $n\sigma^2$ de S_n.
- Calculamos el valor normalizado $z = (c - n\mu)/\sigma\sqrt{n}$.
- Usamos la aproximación,

  $$\mathbb{P}(S_n \leq c) \approx \Phi(z),$$

  donde $\Phi(z)$ es disponible desde la tabla de función densidad acumulativa de la distribución estándar.

Ejemplo 2.4 Cargamos en un avión 100 paquetes cuyos pesos son variables aleatorias independientes que se distribuyen uniformemente entre 5 y 50 libras. ¿Cuál es la probabilidad de que el peso total es superior a 3000 libras?.

No es fácil calcular la función densidad acumulativa del peso total y la probabilidad deseada, pero una respuesta aproximada se puede obtener usando el teorema del límite central:

Calculemos $\mathbb{P}(S_{100} > 3000)$, donde S_{100} es la suma de los pesos de los 100 paquetes. En este caso la media y la varianza de los pesos de un único paquete es:

$$\mu = \frac{5+50}{2} = 27.5 \quad \sigma^2 = \frac{(50-5)^2}{12} = 168.75.$$

basados en la fórmulas para la media y la varianza de la función densidad de una distribución uniforme, calculemos el valor normalizado :

$$z = \frac{3000 - 100 \cdot 27.5}{\sqrt{168.75 - 100}} = \frac{250}{129.9} = 1.92$$

y usando las tablas normal estándar para obtener la aproximación tenemos:

$$\mathbb{P}(S_{100} \leq 3000) \approx \Phi(1.92) = 0.9726.$$

y así la probabilidad pedida es:

$$\mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \leq 3000) \approx 1 - 0.97626 = 0.0274.$$

2.4 Distribuciones que no cumplen el teorema del límite central

No todas las distribuciones obedecen al teorema del límite central. Un ejemplo es la variable aleatoria de Cauchy, que tiene como función densidad de probabilidad:

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

La esperanza y todos los momentos superiores de una variable aleatoria de Cauchy no están definidos. En particular, no tiene una varianza finita y por lo tanto no satisface las condiciones del teorema del límite central.