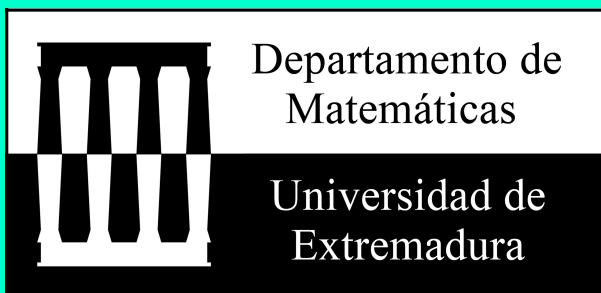




PROBABILIDAD II

Jesús Montanero Fernández



ÍNDICE GENERAL

0. Introducción	1
0.1. Un poco de Teoría de la Medida	1
0.1.1. El área como ejemplo de medida	1
0.1.2. Integral	2
0.1.3. La medida cardinal	3
0.1.4. Los resultados básicos	3
0.1.5. Casos continuo y discreto	5
0.1.6. Distribución de probabilidad	6
0.2. Fenómeno aleatorio	6
1. Caracterización de distribuciones	11
1.1. Función de densidad o probabilidad	11
1.2. Momentos de una distribución	11
1.3. Función de distribución	12
1.4. Función generatriz de momentos	14
1.5. Función característica	14
1.5.1. El teorema de inversión	15
1.6. Caracterización multidimensional	17
2. Principales modelos de distribución	21
2.1. Modelos de distribución discretos	21
2.1.1. Degenerada en un punto	21
2.1.2. Distribución binomial	21
2.1.3. Distribución multinomial	22
2.1.4. Distribución de Poisson	23
2.1.5. Distribución binomial negativa	25
2.1.6. Distribución hipergeométrica	25
2.2. Modelos de distribución continuos	26
2.2.1. Distribución normal	26
2.2.2. Distribución uniforme	29
2.2.3. Distribución Gamma	29
2.2.4. Distribución exponencial	30
2.2.5. Distribución χ^2	31
2.2.6. Distribución F -Snedecor	33

2.2.7. Distribución Beta	34
2.2.8. Distribución t de Student	35
2.2.9. Distribución log-normal	36
2.2.10. Distribución de Cauchy	36
3. Convergencias	39
3.1. Resultados básicos	39
3.1.1. Límites superior e inferior	39
3.1.2. Teoremas de convergencia	40
3.2. Tipos de convergencias	42
3.2.1. Convergencia casi seguro	43
3.2.2. Convergencia en L^2	43
3.2.3. Convergencia en probabilidad	43
3.2.4. Convergencia en distribución	45
3.3. Convergencias y operaciones	46
4. Principales teoremas límite	51
4.1. Leyes de los Grandes Números	51
4.2. Teorema Central del Límite	54
4.2.1. Aplicaciones del TCL	56
4.2.2. Extensiones del TCL	60
5. Problemas propuestos	63

0. INTRODUCCIÓN

Prólogo

Los objetivos de este manual son estudiar las principales distribuciones de probabilidad y dar sentido a la axiomática de Kolmogorov mediante la demostración de la Ley de los Grandes Números (LGN) y el Teorema Central del Límite (TCL). No nos referimos a sendos teoremas sino más bien a colecciones de resultados que se engloban bajo esas denominaciones. Dado que no se asume el suficiente dominio de la Teoría de la Medida por parte del lector, el estudio de estos resultados no se abordará con completo rigor, sino que se realizará de una forma selectiva y en ocasiones heurística, obviando las demostraciones más técnicas. Se propone al lector Gut(2013) como referencia avanzada y, como intermedia, Rincon(2007), disponibles en el aula virtual. Antes de afrontar estos problemas debemos contemplar aspectos anteriores en la teoría de la Probabilidad, empezando por un repaso de conceptos básicos sean o no familiares.

0.1. Un poco de Teoría de la Medida

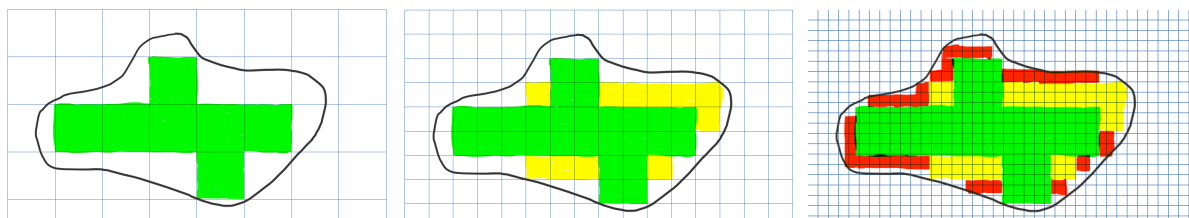
Como dijimos antes, LGN y TCL dan sentido al concepto axiomático de Probabilidad como medida P de extensión 1, pero no debemos malinterpretar esta afirmación. Veamos los dos ejemplos más importantes de medida, el de área y el cardinal, y recordemos las definiciones y resultados más básicos de Teoría de la Medida. Algunos de ellos se probarán en el capítulo 3.

0.1.1. El área como ejemplo de medida

Efectivamente, el concepto de área de una superficie puede ser el más adecuado para entender la definición general de medida μ . ¿Cómo se define el área de cierta superficie en \mathbb{R}^2 ? Si se trata de un rectángulo de lados a y b se define como el producto ab . El área de la unión de varios rectángulos se define como la suma de sus áreas. Ahora bien, ¿cómo definir el área de un subconjunto cualquiera de \mathbb{R}^2 ? La figura (1) debería servir para ilustrar, no sólo el concepto, sino las claves de la Teoría de la Medida.

Ciertamente, el área de la figura puede entenderse como el límite, a medida que la cuadrícula se hace más fina, de las áreas de los cuadrados contenidos en la misma. Pero esta definición no es viable para todos los subconjuntos de \mathbb{R}^2 , sino sólo para una familia $\mathcal{B} \subset \mathcal{P}(\mathbb{R}^2)$ considerados medibles. Además y para que la teoría tenga sentido, \mathcal{B} debe estar dotada de la propiedad de σ -álgebra, es decir, $\mathbb{R}^2 \in \mathcal{B}$; si $B \in \mathcal{B}$ entonces $\bar{B} \in \mathcal{B}$; si $(B_n)_{n \in \mathbb{N}} \subset \mathcal{B}$ entonces $\bigcup_{n=1}^{\infty} B_n \in \mathcal{B}$.

Figura 1: Cálculo de áreas



En nuestro caso, \mathcal{B} se define como la menor σ -álgebra que contiene a los rectángulos. Esta definición extendida a \mathbb{R}^k conduce al concepto de σ -álgebra de Borel. En general, solemos utilizar la letra Ω para designar al conjunto considerado, \mathbb{R}^k o el que sea, y la letra \mathcal{A} para una σ -álgebra cualquiera del mismo. El espacio de medida está constituido entonces por el espacio medible (Ω, \mathcal{A}) junto con la medida μ definida sobre el mismo. En general, una medida es una aplicación positiva sobre \mathcal{A} que verifique estas dos propiedades, típicas del área: $\mu(\emptyset) = 0$ y $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$, si $(A_n)_{n \in \mathbb{N}}$ son elementos disjuntos de \mathcal{A} . Si Ω puede expresarse como unión numerable de conjuntos medibles de medida finita, se dice que μ es σ -finita. En el caso del ejemplo, el espacio medible es $(\mathbb{R}^2, \mathcal{B})$ y la medida μ es el área, que es σ -finita. Puede extenderse a una dimensión menor o mayor, como el cálculo de longitudes y volúmenes, respectivamente. En general, se denomina de Lebesgue a este tipo de medidas en \mathbb{R}^k .

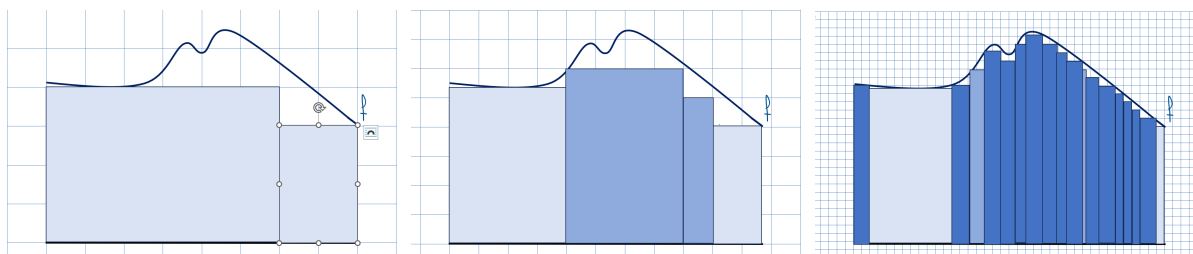
0.1.2. Integral

Consideremos una función f definida sobre $(\Omega, \mathcal{A}, \mu)$ con valores en $(\mathbb{R}^k, \mathcal{B})$ para algún k o en un conjunto a lo sumo numerable (que puede identificarse igualmente con un subconjunto de \mathbb{R}). La función f se dice medible si conserva la medibilidad, es decir, si $f^{-1}(B) \in \mathcal{A}$ para todo $B \in \mathcal{B}$. Es un concepto que extiende el de σ -álgebra en el sentido de que cualquier elemento $A \in \mathcal{A}$ se identifica con la función medible I_A , que vale 1 en A y 0 en su complementario, denominada indicador. También son medibles las funciones simples, definidas como combinaciones lineales de indicadores, así como las funciones positivas que se obtienen como límite de una sucesión creciente de funciones simples. Recíprocamente, puede probarse que cualquier función medible positiva es límite de una sucesión creciente $(X_n)_{n \in \mathbb{N}}$ de funciones simples del tipo $X_n = \sum_{i=1}^{k_n} \xi_{ni} I_{A_{ni}}$, con $A_{ni} \in \mathcal{A}$.

Basándonos en ello, se define entonces su integral como $\sup_n \sum_i \xi_{ni} \mu(A_{ni})$ (figura 2). Se trata pues de un concepto que extiende al de medida y que, de nuevo, consiste en el límite de sumas de elementos medibles y que son subyacentes a f . Se denota por $\int f(x) d\mu$. Si nos restringimos a un subconjunto medible $B \in \mathcal{B}$ estaremos hablando de $\int_B f d\mu$. Cuando la función f está definida en $(\mathbb{R}^k, \mathcal{B})$ con la medida de Lebesgue se denota $\int_B f(x) dx$.

Cualquier función real medible f descompone en la resta $f = f^+ - f^-$, donde $f^+ = \max\{f, 0\}$ y $f^- = \max\{-f, 0\}$, ambas no negativas. Se dice que f no es integrable si $\int f^+ d\mu = \int f^- d\mu = +\infty$. En caso contrario, su integral existe y es concretamente $\int f d\mu = \int f^+ d\mu - \int f^- d\mu \in \overline{\mathbb{R}}$. Si las integrales de f^+ y f^- son finitas también lo es la de f , en cuyo caso se dice integrable.

Figura 2: Integral en (a, b) de f



0.1.3. La medida cardinal

Como caso particular, si contamos con un conjunto finito o numerable $X = (x_n)_{n \in \mathbb{N}}$ y no establecemos en principio diferencias entre sus elementos, la medida de un subconjunto cualquiera $A \in \mathcal{P}(X)$ es su cardinal $\#A$ o cantidad de elementos que lo componen. Esa misma idea podría aplicarse al cálculo de áreas en la figura 1 si los conjuntos a medir pudieran identificarse perfectamente mediante una cuadrícula, es decir, si la imagen se viera pixelada, como en la figura 3.

La integral respecto a la medida cardinal de una función f definida sobre los elementos de X se reduce a $\sum_{n=1}^{\infty} f(x_n)$.

0.1.4. Los resultados básicos

Enunciamos aquí de manera heurística los resultados fundamentales de Teoría de la Medida.

Teorema de Radon-Nykodim: Dadas dos medidas ν_1, ν_2 definidas sobre un mismo espacio medible, (Ω, \mathcal{A}) , se dice que ν_2 domina a ν_1 si se verifica $[\nu_2(A) = 0 \Rightarrow \nu_1(A) = 0]$. En tal caso, si ν_2 es σ -finita, existe una densidad $f = d\nu_1/d\nu_2$, lo cual significa que $\nu_1(A) = \int_A f d\nu_2$ para todo $A \in \mathcal{A}$. Esta propiedad puede extenderse a cualquier función integrable g , de manera que

$$\int g d\nu_1 = \int g \cdot f d\nu_2 \quad (1)$$

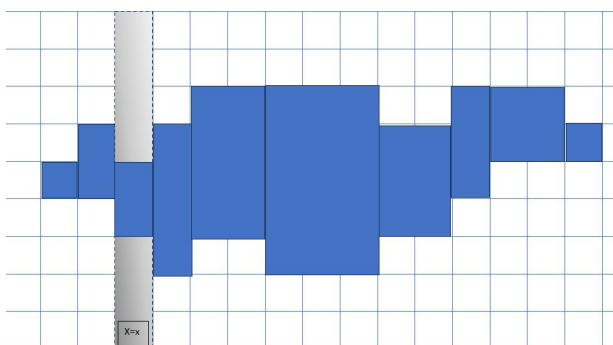
Teorema de la Medida Imagen: Una función medible $X : (\Omega, \mathcal{A}, \mu) \rightarrow (\mathbb{R}^k, \mathcal{B})$ induce sobre el espacio imagen la medida definida para cada $B \in \mathcal{B}$ mediante $\mu^X(B) = \mu(X^{-1}(B))$. Esta propiedad se extiende de los elementos de \mathcal{B} a cualquier función medible g , en la σ -álgebra \mathcal{B} , de manera que

$$\int g d\mu^X = \int g \circ X d\mu \quad (2)$$

Teorema de Dynkin: Se trata de un resultado muy utilizado pues permite extender propiedades que se verifican en cierta familia a la menor σ -álgebra que las contiene. De esta forma, podemos generalizar propiedades que se cumplen en intervalos de la forma $(-\infty, a]$ a \mathcal{B} , por ejemplo.

Teorema de Fubini: En un espacio medible producto $(\Omega_X \otimes \Omega_Y, \mathcal{A}_X \otimes \mathcal{A}_Y)$ se define la medida producto $\mu_x \otimes \mu_y$, como la que asigna a cada subconjunto medible tipo $B_1 \times B_2$ la medida $\mu_X(B_1)\mu_Y(B_2)$. Es precisamente el teorema de Dynkin el que garantiza que la medida producto queda así perfectamente determinada. Imaginemos un conjunto medible A en un espacio producto. En lo que sigue supongamos, por simplificar, que se trata de un conjunto de \mathbb{R}^2 como el de la figura 3. Si suponemos que cada cuadrado mide 1 de lado, el cálculo de su medida (área) total es

Figura 3: Esquema del teorema de Fubini



$$\mu(A) = (1 \times 1) + (2 \times 1) + (2 \times 1) + (4 \times 1) + (4 \times 2) + (5 \times 3) + (3 \times 2) + (3 \times 1) + (2 \times 2) + (1 \times 1)$$

Es decir, $\mu(A) = \sum_x \mu_Y(A_x) \mu_X(x)$, donde

$$A_x = \{y \in \mathbb{R} : (x, y) \in A\} \quad (3)$$

En general, la medida de A y, por extensión, la integral de una función medible definida sobre el mismo, se calcula de la forma

$$\int_A f(x, y) d\mu_x \otimes \mu_y = \int_{\mathbb{R}} \left(\int_{A_x} f(x, y) d\mu_y \right) d\mu_x \quad (4)$$

En particular, si $f(x, y) = f_1(x)f_2(y)$, entonces

$$\int f(x, y) d\mu_x \otimes \mu_y = \int f_1(x) d\mu_x \cdot \int f_2(y) d\mu_y \quad (5)$$

Densidades marginal y condicional: En el espacio producto (seguimos suponiendo que es \mathbb{R}^2 por simplificar), μ_X y μ_Y se denominan medidas marginales y pueden obtenerse como imagen de la medida producto a través de las respectivas proyecciones mediante

$\mu_X(B) = \mu_X \otimes \mu_Y(B \times \mathbb{R})$. Lo mismo ocurre con μ_Y . En el caso continuo y extendiendo lo anterior, la densidad f_X de la marginal de X se relaciona con la conjunta f mediante

$$f_X(x) = \int_{\mathbb{R}} f(x, y) d\mu_Y(y) \quad (6)$$

Y lo mismo ocurre con la de μ_Y . Por otra parte, la figura 3 invita también a considerar, dado x fijo (franja gris), la medida condicional $\mu_{Y|X=x}$, cuya densidades en los casos continuo y discreto son, respectivamente,

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}, \quad \mu(Y = y|X = x) = \frac{\mu(X = x, Y = y)}{\mu(X = x)} \quad (7)$$

De esta igualdad se deduce la Regla de Bayes si desarrollamos la expresión anterior invirtiendo los papeles de X e Y :

$$f_{Y|X=x}(y) = \frac{f_{X|Y=y}(x)f_Y(y)}{\int_{\mathbb{R}} f_{X|Y=z}(x)f_Y(z) d\mu_Y(z)}, \quad \mu(Y = y|X = x) = \frac{\mu(X = x|Y = y)\mu(Y = y)}{\sum_z \mu(X = x|Y = z)\mu(Y = z)} \quad (8)$$

Teoremas de la convergencia monótona (TCM) y dominada (TCD): Son fundamentales para conocer la integral de un límite de funciones medibles. Primeramente, si $(f_n)_n$ son no negativas y tales que $f_n \nearrow f$, entonces $\int f_n d\mu \nearrow \int f d\mu$. Por otra parte, si $(g_n)_n$ está acotada por una función integrable y $g_n \rightarrow g$, entonces $\int g_n d\mu \rightarrow \int g d\mu$. Estos resultados se probarán en el capítulo 3.

Teorema del Cambio de Variables: Dada una función integrable $f : (g(a), g(b)) \rightarrow \mathbb{R}$ con $g : (a, b) \rightarrow (g(a), g(b))$ biyectiva, de manera que tanto g como g^{-1} son derivables, se verifica que

$$\int_{g(a)}^{g(b)} f(x) dx = \int_a^b f \circ g(y) g'(y) dy \quad (9)$$

La versión multidimensional afirma lo siguiente: si A y B son abiertos con medida positiva en \mathbb{R}^k , $g : A \rightarrow B$ es un difeomorfismo de clase 1 y f es integrable en B , entonces

$$\int_B f dx_1 \dots x_n = \int_A f \circ g \cdot |J_g| dy_1 \dots y_n, \quad \text{donde } J_g = \det \left(\frac{\partial(g_1 \dots g_n)}{\partial(y_1 \dots y_n)} \right) \quad (10)$$

0.1.5. Casos continuo y discreto

Estamos muy interesados en dos tipos de medidas: aquéllas que están dominadas por la medida de Lebesgue en \mathbb{R}^k , para algún $k \geq 1$, y las que están dominadas por la medida cardinal en algún conjunto numerable. Las primeras se denominarán (absolutamente) continuas y tendrán una función de densidad f respecto a la medida de Lebesgue.

Por ejemplo, si una medida μ sobre $(\mathbb{R}^k, \mathcal{B})$ es continua debe asignar un valor nulo a cada punto, pues la longitud (así como el área y el volumen) del mismo lo es. En consecuencia, cualquier subconjunto numerable de puntos tendrá medida nula. Además, si h es una función integrable, se verifica que $\int h(x) d\mu = \int h(x)f(x) dx$.

Las medidas dominadas por la medida cardinal se denominan discretas. En ese caso, existe un subconjunto finito o numerable X cuyo complementario tiene medida nula. Sólo

los puntos $(x_n)_{n \in \mathbb{N}}$ que componen X pueden tener una medida positiva $\nu(x_n)$. La densidad respecto a la medida cardinal $\#$ es la llamada función de probabilidad, que asigna a cada uno de esos puntos su medida ν . Se verifica entonces que $\int h(x) d\nu = \sum_n h(x_n)\nu(x_n)$.

0.1.6. Distribución de probabilidad

La axiomática de Kolmogorov consiste considerar la probabilidad como una medida P sobre (Ω, \mathcal{A}) tal $P(\Omega) = 1$. Los elementos de \mathcal{A} se denominan sucesos. Dada una función medible $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}^k, \mathcal{B})$ (en este nuevo contexto se denomina variable aleatoria), se denomina distribución de P respecto de X a la medida imagen P^X . Las distribuciones, ya sean unidimensionales o multidimensionales, serán caracterizadas en el siguiente capítulo. En el capítulo posterior se enumerarán los modelos más importantes distinguiendo entre continuas y discretas.

Independencia: Dos sucesos A, B se dicen independientes cuando se verifica que $P(A \cap B) = P(A)P(B)$. Esta definición puede extenderse a un vector Y compuesto por n variables aleatorias X_1, \dots, X_n sobre (Ω, \mathcal{A}, P) , que se dicen independientes cuando $P^Y = \prod_{i=1}^n P^{X_i}$, es decir, $P(\cap_i \{X_i \in B_i\}) = \prod_i P(\{X_i \in B_i\})$. Por otra parte, se dicen que son idénticamente distribuidas cuando existe una distribución Q tal que $P^{X_i} = Q$ para todo i . Se denomina muestra aleatoria simple de tamaño n de una distribución Q a un conjunto de n variables reales independientes e idénticamente distribuidas (iid) según Q definidas en (Ω, \mathcal{A}, P) .

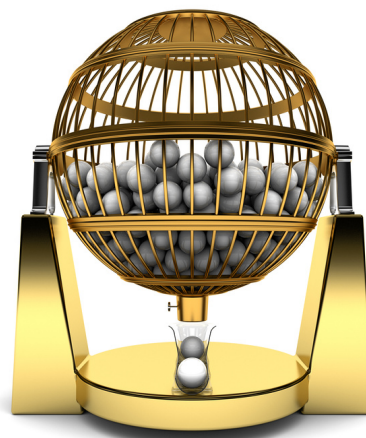
Espacio n -muestral: Para comprender bien esta teoría conviene partir de un espacio de probabilidad original $(\Omega_0, \mathcal{A}_0, P_0)$, entendiéndose que P_0 rige de alguna forma un fenómeno aleatorio puro. Dada una variable aleatoria $X : (\Omega_0, \mathcal{A}_0, P_0) \rightarrow \mathbb{R}$, la forma natural de obtener una muestra aleatoria de tamaño n de P_0^X es considerar en el espacio $(\Omega_0^n, \mathcal{A}_0^n, P_0^n)$ las variables $X_i = X \circ \pi_i$, $1 \leq i \leq n$, donde π_i denota la proyección i -ésima. El espacio producto anterior se denomina n -muestral. La axiomática de Kolmogorov va pues más allá, pues parte realmente de una probabilidad $P = P_0^n$ que rige n réplicas de un fenómeno aleatorio. Dado que en nuestra teoría consideraremos incluso infinitas réplicas del fenómeno, nuestro espacio en los capítulos 3 y 4 será $(\Omega_0^{\mathbb{N}}, \mathcal{A}_0^{\mathbb{N}}, P_0^{\mathbb{N}})$.

0.2. Fenómeno aleatorio

Considerar la probabilidad como una medida P_0 replicable hasta el infinito en $P_0^{\mathbb{N}}$ permite que el Cálculo de Probabilidades se desarrolle en el marco de la Teoría de la Medida, alcanzando así el prestigio científico que las Matemáticas en general otorgan. Pero, si no sabemos qué mide exactamente dicha probabilidad, el Cálculo de Probabilidades no debería traspasar el ámbito de lo abstracto, cosa que sí hace con una creciente influencia en el devenir diario. Nosotros sabemos de qué hablamos en los ejemplos anteriores de medidas, es decir, cuando mencionamos el área, el volumen o la cantidad de elementos. ¿Sucede lo mismo con la probabilidad? No lo parece pues, de hecho, podemos encontrarnos con al menos dos posturas enfrentadas.

El planteamiento bayesiano: Supone que cualquier fenómeno imaginable puede entenderse como aleatorio, siendo \mathcal{A}_0 el espacio de los sucesos susceptibles de acaecer tras la ejecución del mismo y P_0 la función que asigna a cada suceso la medida de su grado de posibilidad entre 0 (imposible) y 1 (seguro). Las probabilidades se obtendrán en última instancia como resultado de un modelo. Con frecuencia, dicho resultado es el estado de parada de un algoritmo iterativo en el que se mezclan la información disponible y otra que se supone *a priori*. Un ejemplo de esto podemos verlo en una retransmisión deportiva cuando un futbolista se hace con el balón en determinada posición cercana a la portería contraria. En pantalla pueden mostrarnos el resultado de un algoritmo bayesiano que indica la probabilidad de que marque gol. ¿Cómo interpretamos esa probabilidad?

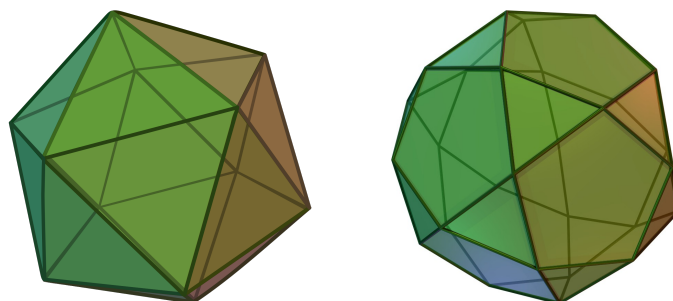
Figura 4: A la izquierda, Robben ante Casillas en la final del Mundial 2010; a la derecha, el espacio clásico $(\Omega_0, \mathcal{A}_0, P_0)$ por antonomasia.



El planteamiento clásico: Entiende el azar como simetría pura, lo cual conduce al concepto de equiprobabilidad entre los sucesos elementales, que son los elementos del espacio original Ω_0 , que debe ser finito. Así pues, la probabilidad de que ocurra cualquier suceso o subconjunto $A \subset \Omega_0$ es el cociente $\#A/\#\Omega_0$. Si el fenómeno se replica n veces en idénticas condiciones, la simetría se extiende al espacio Ω_0^n y la probabilidad correspondiente equivale al producto P_0^n . Un ejemplo de esto puede ser una secuencia de $n = 5$ lanzamientos de un icosaedro, que es un sólido platónico de 20 caras, si suponemos que no se deforma tras cada lanzamiento. Así pues, se contemplan 20^5 resultados posibles y equiprobables tras la serie de cinco lanzamientos.

Ley del azar o de los Grandes Números: Es un hecho empírico que, en condiciones de supuesta simetría, como en el caso de un dado no deformable y, en principio, equilibrado, si se lanza un gran número de veces, n , la proporción de resultados favorables a cualquier suceso A , $p_n(A)$, denominada frecuencia relativa, suele aproximarse a la probabilidad del mismo calculada tal como indicamos antes. Eso da sentido a la idea de probabilidad como grado de posibilidad, al menos en este tipo de fenómenos. Este hecho empírico tiene un reflejo matemático en la denominada Ley de los Grandes Números que se formula en $(\Omega_0^{\mathbb{N}}, \mathcal{A}_0^{\mathbb{N}}, P_0^{\mathbb{N}})$ mediante

Figura 5: Icosaedro (izquierda) e icosidodecaedro (derecha)



$$\text{LGN:} \quad \forall A \in \mathcal{A}, \quad P_0^{\mathbb{N}} \left[\lim_{n \rightarrow \infty} p_n(A) = P_0(A) \right] = 1 \quad (11)$$

Es decir, si existe una medida P_0 que rige un fenómeno aleatorio, debe ser coherente con las frecuencias relativas observadas tras un gran número de replicas en idénticas condiciones.

Punto de vista frecuentista: En consonancia con la LGN, este punto de vista considera como fenómenos aleatorios aquéllos que pueden replicarse indefinidamente, aparentemente en idénticas condiciones, y en los que se aprecian estabilizaciones en las frecuencias relativas tras muchas ejecuciones. Engloba el punto de vista clásico pues en los fenómenos caracterizados por su simetría se describe esta situación, aunque abarca otros fenómenos en los que no se aprecia una simetría evidente. Son pues planteamientos compatibles. En ambos casos se exige que el fenómeno sea replicable tantas veces como se quiera y en idénticas condiciones. Que eso implique una estabilización de frecuencias relativas no se da por descontado si no partimos de una clara simetría. Se exige cierto grado de evidencia empírica.

Podemos preguntarnos si el lanzamiento de un icosidodecaedro, compuesto por 20 caras triangulares y 12 pentagonales (parte derecha de la figura 5) es un fenómeno aleatorio. No está claro desde el punto de vista clásico porque no hay una simetría evidente. Desde el punto de vista frecuentista es plausible, aunque antes deberíamos lanzarlo bastantes veces y analizar cómo evolucionan las frecuencias relativas de los triángulos y pentágonos. Desde el punto de vista Bayesiano siempre es un fenómeno aleatorio.

Por contra, el resultado de un disparo en un partido de fútbol no es un fenómeno aleatorio desde el punto de vista clásico porque no se dan las condiciones de simetría; tampoco lo es desde el punto de vista frecuentista porque el fenómeno no es siquiera replicable; por supuesto que lo es desde el punto de vista Bayesiano.

Crítica a los tres planteamientos: El planteamiento clásico es el más satisfactorio desde el punto de vista formal pues se basa en última instancia en la idea geométrica de simetría, pero es demasiado restrictivo. No obstante, hay que tener en cuenta el muestreo estadístico es un fenómeno aleatorio tipo lotería, por lo que alcanza para desarrollar los métodos estadísticos clásicos de Inferencia estadística, no así los Bayesianos.

El planteamiento frecuentista es algo más laxo pues considera aleatorios otros tipos de fenómenos físicos así como errores de medición, pero sigue siendo mucho más restrictivo que el planteamiento Bayesiano y, además, es ambiguo respecto al número n de repeticiones necesarias y al grado de estabilidad que se precisa en la frecuencia relativa. En la práctica y dado que la simetría pura es un concepto ideal (no podemos construir un dado ni una ruleta completamente simétricos) no suele establecerse una diferencia clara entre el planteamiento clásico y el frecuentista, constituyendo ambos un marco estadístico común.

El planteamiento bayesiano otorga un gran margen de maniobra y puede resultar muy elegante desde el punto de vista matemático, pero deja sin responder la pregunta inicial: ¿Qué mide la probabilidad? Si la respuesta es el grado de posibilidad de un suceso, habría que preguntar: ¿Qué es eso? ¿Cómo puede llegar a aproximarse si ni siquiera puede repetirse el fenómeno en las mismas condiciones? De nuevo, si la respuesta es que es simplemente la salida de un modelo, habría que preguntar: ¿Cómo conocemos los parámetros de entrada que exige el modelo y la fiabilidad del mismo si seguimos sin saber exactamente qué estamos midiendo? Si la respuesta es que no importa mucho porque el algoritmo es iterativo y tiende a mejorar unas probabilidades de entrada arbitrarias, puede que estemos ante un planteamiento inquietante que no todos los estadísticos están dispuestos a asumir sin más, aunque en muchas ocasiones, no todas, resulte útil.

La solución en falso: Es bastante común pensar que la antigua confrontación entre frecuentistas y bayesianos quedó zanjada al asumir la definición axiomática de Kolmogorov, es decir, la probabilidad como medida de extensión 1. Como dijimos al principio, eso es una mala interpretación de la LGN. Este resultado no permite formalizar mediante una probabilidad cualquier fenómeno imaginable, sino que describe qué debe observarse necesariamente en promedio y a la larga en un fenómeno formalizable mediante una ley de probabilidad $P_0^{\mathbb{N}}$, es decir, aleatorio. En otras palabras, permite caracterizar el fenómeno aleatorio como determinista en promedio.

Quizás la raíz del malentendido es definir como aleatorio cualquier fenómeno no determinista. Deberíamos hablar mejor de fenómenos inciertos. Puede que cualquier fenómeno incierto descomponga en una componente determinista y otra aleatoria pura o ruido. Es en cierta manera el uso que se da al concepto de probabilidad a la hora de medir magnitudes físicas. En todo caso la controversia sigue ahí. Si la ignoramos, estaremos dando por bueno cualquier modelo cuyo resultado se exprese en términos probabilísticos y acataremos cualquier predicción sobre lo que ocurrirá en 2050, por ejemplo, si viene de un ordenador y está expresada con decimales.

NOTAS DE CLASE

ESPACIO MEDIBLE

Consideremos un conjunto cualquiera Ω y una familia de subconjuntos de Ω , \mathcal{A} , con las siguientes propiedades:

1. $\Omega \in \mathcal{A}$
2. Si $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
3. Si $B_n \in \mathcal{A}, n \geq 1 \Rightarrow \bigcup_{n=1}^{\infty} B_n \in \mathcal{A}$

Mostraremos a (Ω, \mathcal{A}) "espacio medible"

y a la familia \mathcal{A} "σ-álgebra"

Pero no solo de "probabilidades" vive el estadístico.

Hay otras medidas que no dan a Ω probabilidad 1 (ni siquiera finitas)

Centrémos en el caso real para ilustrar dos medidas que utilizaremos insistentemente: La medida de Lebesgue en \mathbb{R}^n y la medida cardinal para Ω 's finitos o numerables. La medida de Lebesgue en $\Omega = \mathbb{R}$ es la medida que asigna a los intervalos su longitud, la denotaremos " λ ".

En $\Omega = \mathbb{R}^2$, es la medida que asigna a los rectángulos su área...

La teoría se completa con la σ-álgebra de Borel. En \mathbb{R} es la menor que contiene a los intervalos, \mathcal{B} , en \mathbb{R}^2 , la menor que contiene a los rectángulos, \mathcal{B}^2 ... El teorema de Dynkin nos permitirá generalizar resultados que podemos probar para los rectángulos (o intervalos) a toda la σ-álgebra.

Si $\Omega = \mathbb{R}^2$, hablaremos de área... Insistimos un poco más en esta situación para ilustrar las claves de la teoría de la medida.

¿Cómo medimos un conjunto cualquiera en \mathcal{B}^2 ?

El área del conjunto A de las figuras adjuntas, podrá entenderse como el límite, a medida que la cuadrícula se hace más fina, de las áreas de los cuadrados contenidos en A .

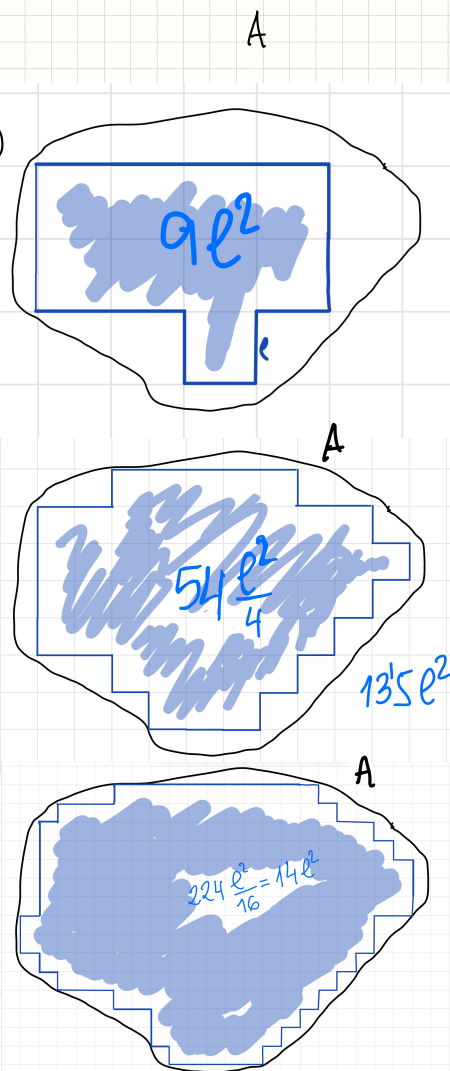
LA MEDIDA

Queremos medir los conjuntos de \mathcal{A} . La medida que utilizaremos es "la probabilidad", P .

P es una función real de conjuntos:

$$P: \mathcal{A} \longrightarrow \mathbb{R}$$
$$A \longmapsto P(A) \geq 0 \quad \text{tal que}$$

1. $P(\Omega) = 1$
2. Si $A_n \in \mathcal{A}, n \geq 1$, disjuntos dos a dos, entonces $\lambda\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \lambda(A_n)$



Otra medida que no es una probabilidad la utilizaremos, a veces, cuando Ω es finito o numerable.

En este caso utilizaremos la σ -álgebra $\mathcal{P}(\Omega)$, partes de Ω , es decir, la familia de todos los subconjuntos de Ω .

En este espacio medible $(\Omega, \mathcal{P}(\Omega))$, consideraremos la medida cardinal que asigna a cada conjunto $A \in \mathcal{P}(\Omega)$, el no. de elementos que tiene. La denotaremos como $\#$.

Llamaremos a $(\Omega, \mathcal{P}(\Omega), \#)$ "espacio discreto".

ESPACIO DISCRETO
(medida Cardinal)

VARIABLES ALEATORIAS

Una función $X: (\Omega, \mathcal{A}) \longrightarrow (\mathbb{R}, \mathcal{B})$ es una variable aleatoria (real, en este caso) si

$$\forall B \in \mathcal{B}, X^{-1}(B) \in \mathcal{A}.$$

$$X^{-1}(B) = \{ \omega \in \Omega / X(\omega) \in B \}$$

EJEMPLOS SENCILLOS

① Sea $A \in \mathcal{A}$. $I_A(\omega) = \begin{cases} 0 & \text{si } \omega \notin A \\ 1 & \text{si } \omega \in A \end{cases}$
 \uparrow
Indicador de A .

② Sean A_1, \dots, A_n una partición de Ω , entonces $A_1 \cup \dots \cup A_n = \Omega$ y $A_i \cap A_j = \emptyset$ si $i \neq j$.
 $S(\omega) = \sum_{i=1}^n x_i I_{A_i}(\omega)$, $x_i \in \mathbb{R}$, $i=1, \dots, n$ es una variable aleatoria que llamaremos SIMPLE.

DISTRIBUCIÓN DE UNA VARIABLE ALEATORIA

Dada una v.a. $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$, la distribución de probabilidad de X es una probabilidad inducida por X en $(\mathbb{R}, \mathcal{B})$. La denotaremos P^X y se define así: $P(B) = P(X^{-1}(B))$ para cada $B \in \mathcal{B}$.

CONEXIÓN ENTRE LAS VARIABLES ALEATORIAS Y LAS DOS MEDIDAS MENCIONADAS

Tipos de variables aleatorias con las que trabajaremos.

1. ABSOLUTAMENTE CONTINUAS

$$X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}, P^X)$$

Así, en $(\mathbb{R}, \mathcal{B})$ tenemos dos medidas P^X y λ (la de Lebesgue). Existe una relación entre ellas, P^X está dominada por λ , $P^X \ll \lambda$,

esto es, si $B \in \mathcal{B}$ tal que $\lambda(B) = 0 \Rightarrow P^X(B) = 0$.

Se puede probar (teorema de Radon-Nikodym) que existe f (función de densidad de X) tal que

$$P^X(B) = \int_B f(x) dx \quad (\text{integral de Lebesgue})$$

INTEGRAL

Sea $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$ v.a.

Definición (Integral)

$$* \int_{\Omega} X dP = P(A) \text{ si } X = I_A, A \in \mathcal{A}.$$

$$* \int_{\Omega} X dP = \sum_{i=1}^n x_i P(A_i) \text{ si } X = \sum_{i=1}^n x_i I_{A_i}, \quad (\text{simple})$$

* Si $X \geq 0 \Rightarrow \exists X_n, n \geq 1$, simples, ≥ 0 , no decreciente, tal que

$$X(\omega) = \lim_{n \rightarrow \infty} X_n(\omega), \quad \omega \in \Omega. \text{ Así, } \int_{\Omega} X dP = \lim_{n \rightarrow \infty} \int_{\Omega} X_n dP$$

$$* \text{ Si } X \in \mathbb{R} \Rightarrow X = X^+ - X^-, \quad X^+, X^- \geq 0$$

$$\text{Así, } \int_{\Omega} X dP = \int_{\Omega} X^+ dP - \int_{\Omega} X^- dP, \text{ supuesto que alguna de las dos es finita}$$

Sea $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B}, P^X)$

Por definición: $P^X(B) = P(X^{-1}(B))$

Si $g: (\mathbb{R}, \mathcal{B}, P^X) \longrightarrow (\mathbb{R}, \mathcal{B})$, entonces

$$\int g dP^X = \int g \circ X dP \quad (\text{Th. medida imagen})$$

* En caso de X sea absolutamente continua con función de densidad f , tendremos que

$$\int g dP^X = \int g(x) f(x) dx$$

* Y si X es discreta con función de probabilidad f , tendremos

$$\int g dP^X = \sum g(x) f(x)$$

En este caso, $f(x)$ acabará siendo $P(X=x)$:

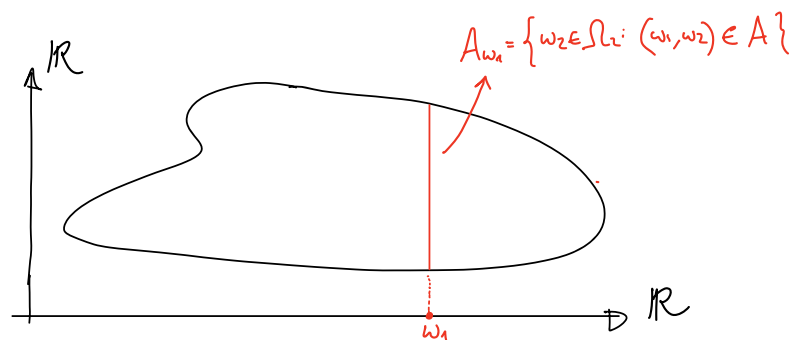
$$\int g dP^X = \sum g(x) \cdot P(X=x)$$

Introducción

En general, dados $(\Omega_1, \mathcal{A}_1, P_1)$ y $(\Omega_2, \mathcal{A}_2, P_2)$ esp. medidos, llamamos ESPACIO PRODUCTO a $(\Omega_1 \otimes \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, P_1 \otimes P_2)$, donde $\Omega_1 \otimes \Omega_2$ es el producto cartesiano, $\mathcal{A}_1 \otimes \mathcal{A}_2$ es la más pequeña σ -álgebra que contiene a los conjuntos de la forma $B_1 \times B_2$, $B_1 \in \mathcal{A}_1$, $B_2 \in \mathcal{A}_2$ y $P_1 \otimes P_2$ la probabilidad que, aplicada a $B_1 \times B_2$, nos da $P_1(B_1) \cdot P_2(B_2)$

Otro modo de observar la medida de un conjunto A en \mathbb{R}^2 es la siguiente:

$$P_1 \otimes P_2(A) = \int_A d(P_1 \otimes P_2) = \int_{\Omega_1 \otimes \Omega_2} I_A d(P_1 \otimes P_2) = \int_{\Omega_1} \left(\int_{A_{w_1}} dP_1 \right) dP_2$$



Es más, generalizando, si $f: (\Omega_1 \otimes \Omega_2, \mathcal{B} \otimes \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$ es medible, entonces

$$\int f(w_1, w_2) d(P_1 \otimes P_2) = \int_{\Omega_1} \left(\int_{A_{w_1}} f(w_1, w_2) dP_1 \right) dP_2 \quad (\text{teorema de Fubini})$$

Por otra parte, $P_1(B_1) = P_1 \otimes P_2(B_1 \times \Omega_2)$. Igual ocurre con P_2 .

A P_1 y P_2 las llamaremos probabilidades marginales

VECTORES ALEATORIOS

Supongamos que P_1 y P_2 son las distribuciones de probabilidad de determinadas variables aleatorias X e Y reales definidas en el espacio de medida (Ω, \mathcal{A}, P) :

$$(X, Y): (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^2, \mathcal{B}^2, P^X \otimes P^Y)$$

llamemos f a la función de densidad de $P^X \otimes P^Y$ respecto a la medida de Lebesgue (densidad conjunta). Igualmente, sean f_X y f_Y sean las densidades de X e Y , respectivamente. Anteriormente escribimos: Dado $A \in \mathcal{B}^2$,

$$\begin{aligned} P^X \otimes P^Y(A) &= \int_A d(P^X \otimes P^Y) = \left(\int_{\mathbb{R}^2} \mathbb{I}_A(x, y) d(P^X \otimes P^Y) \right) = \\ &= \int_{\mathbb{R}} \left(\int_{A_x} dP^Y \right) dP^X \quad (\text{Fubini}) \end{aligned}$$

También podemos considerar las densidades de $P^X(f_X)$ y $P^Y(f_Y)$, (las marginales):

$$f_X(x) = \int_{\mathbb{R}} f(x, y) dy$$

$$f_Y(y) = \int_{\mathbb{R}} f(x, y) dx$$

También podemos considerar,

fijado un valor x ,

la probabilidad $P^Y|X=x$

(probabilidad de Y condicionado a que $X=x$)

Si densidad es:

$$f_{Y|X=x}(y) = \frac{f(x, y)}{f_X(x)}$$

(X, Y) será un vector aleatorio 2-dimensional

(De manera similar introducimos el concepto de vector aleatorio n -dimensional).

INDEPENDENCIA

* Dado (Ω, \mathcal{A}, P) y $A_1, A_2 \in \mathcal{A}$, diremos que A_1 y A_2 son independientes si

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2)$$

* Si $X, Y: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{B})$

Diremos que X e Y son independientes si, dados $B_1, B_2 \in \mathcal{B}$

$$\begin{aligned} P(\{\omega: X(\omega) \in B_1\} \cap \{\omega: Y(\omega) \in B_2\}) &= \\ &= P(X(\omega) \in B_1) \cdot P(Y(\omega) \in B_2) \end{aligned}$$

$$(P(X \in B_1 \cap Y \in B_2) = P(X \in B_1) \cdot P(Y \in B_2))$$

FUNCIÓN DE DISTRIBUCIÓN

Si $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \beta)$ v.a.
Llamamos función de distribución de X a una
función $F: \mathbb{R} \longrightarrow [0,1]$, tal que
$$F(x) = P(X \leq x) = P(X \in (-\infty, x]) =$$
$$= P(X^{-1}((-\infty, x]))$$

1. Es no decreciente: si $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$

2. Es continua por la derecha, i.e.,

$$\lim_{x \rightarrow a, x > a} F(x) = F(a)$$

$$\left(\bigcap_{n \in \mathbb{N}} \{a < X \leq a + 1/n\} \right) = \emptyset.$$

Puntos de discontinuidad: Aquellos con probabilidad
positiva, en una cantidad numerable.

Si X es absolutamente continua $\Rightarrow F$ continua

3. $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$

4. X absolutamente continua $\Rightarrow F' = f$.

5. $F_1 = F_2 \Leftrightarrow P^{X_1} = P^{X_2}$

1. CARACTERIZACIÓN DE DISTRIBUCIONES

En este capítulo estudiaremos fundamentalmente cinco formas diferentes de caracterizar o intentar caracterizar una distribución unidimensional o multidimensional, ya sea discreta o continua, incluyendo la ya conocida función de densidad o probabilidad. Haremos especial hincapié en la denominada función característica, por ser más novedosa y compleja y, sobre todo, por ser esencial en la demostración del TCL. Empezaremos caracterizando las distribuciones unidimensionales para dedicar después la última sección a las multidimensionales.

1.1. Función de densidad o probabilidad

Ya vimos en el capítulo anterior que una distribución absolutamente continua P^X , es decir, dominada por la medida de Lebesgue, se caracterizan por su función de densidad f , de manera que $P(X \in A) = \int_A f(x) dx$ para todo $A \in \mathcal{B}$. Por su parte, una distribución discreta, es decir, dominada por la medida cardinal $\#$ definida sobre un subconjunto finito o numerable $(x_n)_n$, queda asimismo caracterizada por su densidad respecto a $\#$, que es la función que asigna a cada posible valor x_n su probabilidad $P(X = x_n)$, por lo que se denomina comúnmente función de probabilidad. En todo caso, la función de densidad o probabilidad, según corresponda, es la herramienta más inmediata para caracterizar una distribución de probabilidad.

En la figura (1.1) podemos apreciar las funciones de densidad de las distribuciones $N(0, 1)$ y Fejér-de la Vallée-Poussin (FJV), así como la función de probabilidad de $B(5, 0.5)$.

Para calcular la función de densidad de la transformada de una distribución tendremos en cuenta el Teorema del Cambio de Variables.

1.2. Momentos de una distribución

No se trata de funciones sino de números o parámetros que pretenden resumirla y que, en ciertas condiciones, la caracterizan por completo. El primero y más importante es la media o esperanza. Dado $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B})$, se define $E[X] = \int X dP$, que equivale, por el Teorema de la Medida Imagen (2), a $\int x dP^X$. Si P^X tiene densidad f respecto a

MOMENTOS

μ "Medio o Esperanza": $E(X) = \int_{\mathcal{L}} X dP = \int_{\mathbb{R}} x P^X$

$$(\Omega, \mathcal{A}, P) \xrightarrow{X} (\mathbb{R}, \mathcal{B}, P^X) \xrightarrow{g(x)=x} (\mathbb{R}, \mathcal{B})$$

MOMENTO DE ORDEN 1 (μ_1)

$$\int g dP^X = \int g \circ X dP$$

Si f es la función de densidad: $E(X) = \int_{\mathbb{R}} x f(x) dx$

Si X es discreta $\Rightarrow E(X) = \sum_n x_n P(X=x_n)$

No tiene por qué ser finito ni siquiera existir. y X = nº de lanzamientos hasta obtener la primera cara
(p , prob de éxito, la hacemos tender a 0).
($E(X) = 1/p$ $p \rightarrow 0$ se va a ∞)

DISTRIBUCIÓN CONDICIONAL $P^{Y|X=x}$

Si existen, son funciones de x ($P^{Y|X=x}$)

* $E(Y|X)$ (momento de orden 1) (Esperanza condicional)

μ_K

Idem ...

Momento de orden K .

σ^2

VARIANZA

$$\text{Var}(X) = E((X - E(X))^2) = \mu_2 - \mu_1^2$$

PROPIEDADES

$$E(a+bX) = a + bE(X)$$

$$\text{Var}(a+bX) = b^2 \text{Var}(X)$$

Una transformación habitual: $X \rightarrow \frac{X - E(X)}{\sqrt{\text{Var}(X)}}$

$$\gamma_1 \text{ ASIMETRÍA } \gamma_1 = \frac{E((X-\mu)^3)}{\sigma^3}$$

$$\gamma_2 \text{ CURTOSIS } \gamma_2 = \frac{E((X-\mu)^4)}{\sigma^4}$$

FUNCIÓN GENERATRIZ DE MOMENTOS

Dada la v.a. $X: (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$,

llamamos función generatriz de P^X

a la función $g(t) = E(e^{tX})$.

En el caso absolutamente continuo, si f es la densidad de P^X

$$g(t) = \int e^{tx} f(x) dx$$

En el caso discreto, $g(t) = \sum_{n \in \mathbb{Z}} e^{tx_n} P(X=x_n)$.

La función generatriz de momentos g_X , es una función real de variable real. $g_X: \mathbb{R} \rightarrow \mathbb{R}$
A veces puede que no esté bien definida, pues para ciertos valores de t , la esperanza $E(e^{tX})$ no de 0 sea infinita.

Como vemos, cuando está definida en un entorno de 0, la g_X caracteriza la distribución:

TEOREMA Supongamos que $g(t) = E(e^{tX})$ está definida en $|t| < \delta$, $\delta > 0$. Entonces:

- $E(|X|^r) < \infty$, $\forall r > 0$
- $E(X^n) = g^{(n)}(0)$, $n=0,1,\dots$ (derivada n-ésima en 0)
- $g(t) = \sum_{n=0}^{\infty} \frac{t^n}{n!} E(X^n)$, $|t| < \delta$

Sea $r > 0$, $|t| < \delta$. Tenemos que $\frac{|x|^r}{e^{tx}}$ $\downarrow 0$ si $x \uparrow \infty$.

Sea $a > 0$ si $|x| > ar \Rightarrow |x|^r \leq e^{tx}$. Entonces

$$E(|X|^r) = E(|X|^r \cdot I_{|X| \leq ar}) + E(|X|^r \cdot I_{|X| > ar}) \leq ar + E(e^{tx} I_{|X| > ar}) \leq ar + E(e^{tx}) < \infty$$

pues $E(e^{tx}) < \infty$ para los t 's considerados

TEOREMA Sean X, Y v.v.aa en (Ω, \mathcal{A}, P) a valores en $(\mathbb{R}, \mathcal{B})$. Supongamos que existen $g_X(t)$ y $g_Y(t)$ cuando $|t| < \delta$, para algún $\delta > 0$.
 $g_X(t) = g_Y(t)$, para $|t| < \delta \Leftrightarrow F_X = F_Y$ ($X \stackrel{d}{=} Y$)

PROPOSICIÓN Sea X una v.a. con función generatriz g_X y sea $Y = a + bX$ otra v.a. entonces $g_Y = e^{at} g_X(bt)$

$$E(e^{tY}) = E(e^{(a+bX)t}) = E(e^{at} \cdot e^{btX}) = e^{at} E(e^{btX}) = e^{at} g_X(bt)$$

EJEMPLOS

- ① Sea X una v.a. $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \beta)$ tal que $P^X \sim B(n, p)$. Sabemos que X puede tomar los valores $0, 1, \dots, n$ y $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$.
Calculemos la función generatriz de momentos de X .

$$\begin{aligned} \text{Dado } t \in \mathbb{R}, \quad g_X(t) &= E(e^{tX}) = \sum_{k=0}^n e^{tk} P(X=k) = \\ &= \sum_{k=0}^n e^{tk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^n \binom{n}{k} (p \cdot e^t)^k (1-p)^{n-k} = \\ &= (pe^t + (1-p))^n \end{aligned}$$

- ③ Considere una v.a. X con distribución de Poisson de parámetro $\lambda > 0$. Puede tomar los valores $k=0, 1, \dots$

$$\text{y} \quad P(X=k) = \frac{e^{-\lambda} \lambda^k}{k!}$$

$$\begin{aligned} \text{Si } t \in \mathbb{R}, \quad g_X(t) &= E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \frac{e^{-\lambda} \lambda^k}{k!} = \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} e^{\lambda e^t} = e^{\lambda(e^t - 1)} \end{aligned}$$

($e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$)

$$\text{Así, } E(X) = g'_X(0) = \lambda \quad (g'_X(t) = \lambda e^t e^{\lambda(e^t - 1)})$$

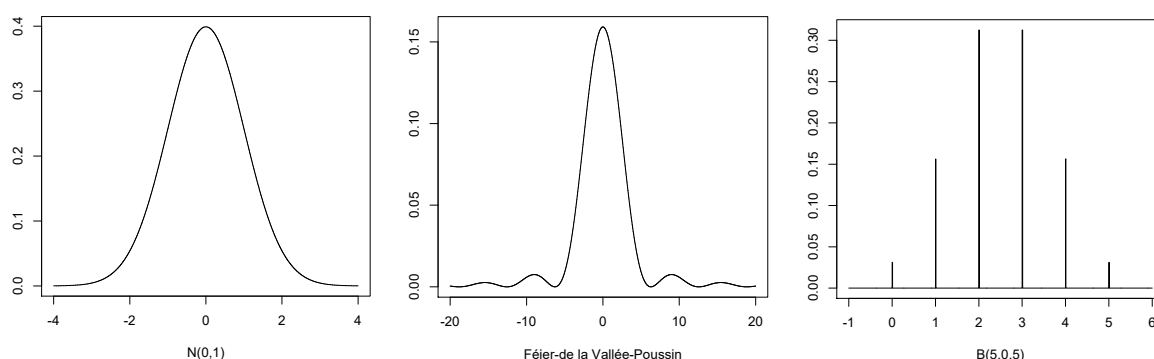
$$E(X^2) = g''_X(0) = \lambda + \lambda^2 \quad (g''_X(t) = \lambda e^t e^{\lambda(e^t - 1)} + \lambda e^t \lambda e^t e^{\lambda(e^t - 1)})$$

$$\begin{aligned} \text{Así, } \text{var}(X) &= E(X^2) - (E(X))^2 = \\ &= \lambda \end{aligned}$$

- ② Considere una variable aleatoria degenerada, esto es, para un cierto $a \in \mathbb{R}$, $P(X=a) = 1$, $P(X \neq a) = 0$.
Dado $t \in \mathbb{R}$, $g_X(t) = E(e^{tX}) = e^{ta} \cdot P(X=a) = e^{ta}$

$$\begin{aligned} \text{Si } \lim_{n \rightarrow \infty} a_n = l^{\infty} &\Rightarrow \\ \lim_{n \rightarrow \infty} a_n^{l_n} &= e^{\lim_{n \rightarrow \infty} (a_n - 1) l_n} \end{aligned}$$

Figura 1.1: Ejemplos de función de densidad y distribución



la medida de Lebesgue se calcula según (1) mediante $\int x f(x) dx$. Si está dominada por la medida cardinal se calcula mediante $\sum_n x_n P(X = x_n)$. Nótese que la media no tiene por qué ser un número finito y puede incluso no estar bien definida. Suele denotarse por la letra μ . En todo caso, pretende ser el valor central de la distribución.

En general, la media se considera el momento de orden 1 o μ_1 de la distribución (aunque la denotaremos por μ , a secas) porque cabe definir el momento de cualquier orden $k \in \mathbb{N}$ como $\mu_k = E[X^k]$. Se calcula mediante $\int x^k f(x) dx$ o $\sum_n x_n^k P(X = x_n)$, según el caso. Con mayor razón, no tiene por qué existir μ_k pero, si existe y es finito, lo son también los momentos de orden inferior. Veremos después que dos distribuciones cuyos momentos son finitos para todos los órdenes y coinciden son iguales.

La varianza de una distribución, que suele denotarse por la letra σ^2 , se define, si el momento de orden 2 es finito, mediante $\text{var}[X] = E[(x - \mu)^2] = \mu_2 - \mu_1^2$. El comportamiento de la media y varianza ante transformaciones afines es el siguiente:

$$E[a + bX] = a + bE[X], \quad \text{var}[a + bX] = b^2 \text{var}[X] \quad (1.1)$$

Por esa razón, la transformación $(X - \mu)/\sigma$, que se denomina tipificación de X , tiene media 0 y varianza 1. Otros parámetros conocidos basados en momentos de orden 3 y 4 son el de asimetría y el de curtosis, que figuran a izquierda y derecha, respectivamente, de (1.2).

$$\gamma_1 = \frac{E[(X - \mu)^3]}{\sigma^3}, \quad \gamma_2 = \frac{E[(X - \mu)^4]}{\sigma^4} \quad (1.2)$$

Podemos definir los momentos asociados a una distribución condicional $P^{Y|X}$ que será, si existen, funciones de x . En particular, la media o esperanza condicional se denota por $E(Y|X)$ y tiene propiedades análogas a las de la esperanza.

1.3. Función de distribución

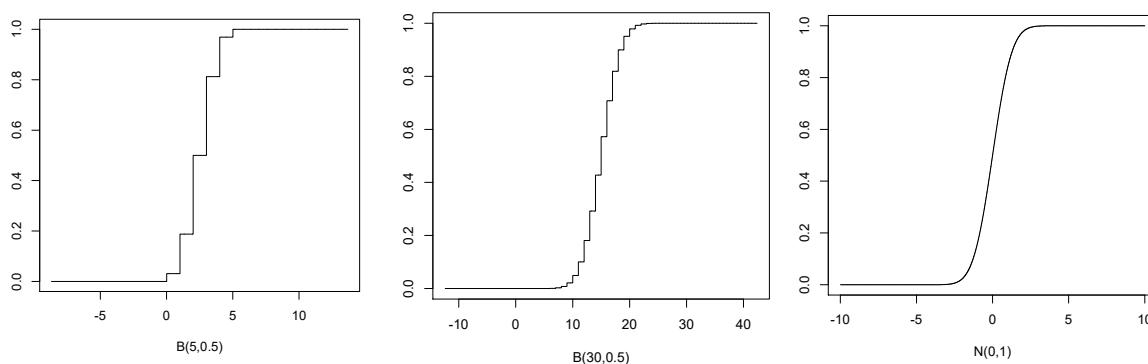
En el caso de una distribución unidimensional P^X (en \mathbb{R}), es la función F de \mathbb{R} en $[0, 1]$ que asigna a cada x el valor $P(X \leq x)$. Sus propiedades elementales son las siguientes:

- (i) Es no decreciente, es decir, si $x_1 < x_2$ $F(x_1) \leq F(x_2)$. Se debe a que $\{X \leq x_1\} \subset \{X \leq x_2\}$.

- (ii) En general, es continua por la derecha, es decir, $\lim_{x \rightarrow a, x > a} F(x) = F(a)$. Se debe a que $\cap_{n \in \mathbb{N}} \{a < X \leq a + 1/n\} = \emptyset$. Las posibles discontinuidades son los puntos de probabilidad positiva, de los cuales hay una cantidad a lo sumo numerable, por lo que los puntos de continuidad son densos en \mathbb{R} . Si la distribución es absolutamente continua, F es continua en todo \mathbb{R} pues $P(X = a) = 0$
- (iii) $\lim_{x \rightarrow -\infty} F(x) = 0$ y $\lim_{x \rightarrow \infty} F(x) = 1$
- (iv) Si P^X es absolutamente continua con densidad f continua, entonces $F' = f$.
- (v) Dos funciones de distribución F_1 y F_2 son idénticas sii los son las distribuciones asociadas. Se debe a que coinciden sobre cualquier intervalo de la forma $(-\infty, x]$. Por el teorema de Dynkin, coinciden entonces en \mathcal{B} .

La ventaja que presenta la función de distribución respecto a la de densidad es que permite resumir las distribuciones continuas mediante una tabla numérica, a partir de la cual se puede calcular, si se desea, la probabilidad de cualquier intervalo. Sin embargo, puede resultar algo anodina desde el punto de vista gráfico. En la figura 1.2 se comparan las funciones de distribución de $B(10, 0.5)$, $B(30, 0.5)$ y $N(0, 1)$. Las funciones de distribu-

Figura 1.2: Ejemplos de funciones de distribución



ción correspondientes a distribuciones continuas, como $N(0, 1)$, presentan un aspecto muy diferente de las discretas, como $B(10, 0.5)$. Los saltos en el caso de $B(10, 0.5)$ se corresponden con los enteros, cuyas probabilidades son positivas. Sin embargo, encontramos mayor similitud entre $N(0, 1)$ y $B(30, 0.5)$, a pesar de las discontinuidades de esta última, lo cual va en la dirección de la tesis del TCL.

Además, el teorema de Glivenko-Cantelli es una extensión de LGN que se enunciará en el capítulo 4 y habla de convergencia uniforme de las funciones de distribución muestrales a la probabilística que da pie a pruebas de bondad de ajuste como la de Kolmogorov-Smirnov-Lilliefors.

Las funciones generatriz y característica, que estudiaremos a continuación, son formulaciones en un marco probabilístico de sendas transformaciones de interés general en Matemáticas y con importantes aplicaciones en Física: las de Laplace y Fourier. Adelantamos que, aunque la primera conserva su interés en ciertas situaciones, queda claramente eclipsada en la teoría asintótica por la segunda debido a una patología que veremos enseguida.

1.4. Función generatriz de momentos

Dada una variable aleatoria $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ se define la función generatriz de P^X como aquella que asigna a cada $t \in \mathbb{R}$ el número $g(t) = E[e^{tX}]$ que, dependiendo de si P^X es continua con densidad h o discreta, se calcula respectivamente mediante

$$g(t) = \int e^{tx} f(x) dx, \quad g(t) = \sum_n e^{tx_n} P(X = x_n) \quad (1.3)$$

El gran problema de $g(t)$ es que no tiene por qué estar bien definida en todo \mathbb{R} , es decir, puede haber valores de t para los cuales la integral o la serie de (1.3) no sea finita. Eso explica que en la Teoría Asintótica quede relegada por la función característica. Sin embargo, cuando está bien definida en un intervalo abierto que contenga a 0, resulta de enorme utilidad y caracteriza perfectamente la distribución.

La transformada del Laplace es de gran interés en el Cálculo Diferencial y, precisamente, en las condiciones indicadas permite obtener los diferentes momentos de la distribución derivando sucesivamente la función g en 0. Efectivamente, dado que

$$e^y = \sum_{k=0}^{\infty} \frac{y^k}{k!}, \quad \forall y \in \mathbb{R} \quad (1.4)$$

se sigue del Teorema de la Convergencia Dominada que, si g está definida si $|t| < \delta$ para algún $\delta > 0$

$$g(t) = \int \sum_{k=0}^{\infty} \frac{(tx)^k}{k!} dP^X = \sum_{k=0}^{\infty} \frac{t^k}{k!} \int x^k dP^X = \sum_{k=0}^{\infty} \mu_k \frac{t^k}{k!} \quad (1.5)$$

Dado que se trata del desarrollo de una serie de potencias en torno a 0, se verifica que

$$\mu_k = g^{(k)}(0) \quad (1.6)$$

En definitiva, en distribuciones de este tipo, podemos afirmar que existen los momentos de todos los órdenes y que pueden obtenerse derivando sucesivamente la función generatriz de momentos, que queda caracterizada por los mismos. De hecho, podemos ir más lejos (razonando con la función característica que veremos a continuación) y afirmar que dos distribuciones de ese tipo cuyos momentos coinciden son idénticas. Es decir, en esas condiciones, tanto los momentos como la función generatriz de momentos caracterizan por completo la distribución. Veremos además que, si g_X está definida en $|t| < \delta$, permite obtener de manera automática la función característica, que definiremos a continuación, en un entorno de 0. Por último, la función generatriz de una transformación afín de X , $a + bX$, se obtiene a partir de g_X mediante

$$g_{a+bX}(t) = e^{at} g_X(bt) \quad (1.7)$$

1.5. Función característica

Dada una variable aleatoria $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$, se define la función característica de P^X como aquella que asigna a cada $t \in \mathbb{R}$ el número $\varphi_X(t) = E[e^{itX}] \in \mathbb{C}$. Puede

descomponerse en parte real e imaginaria mediante

$$\varphi_X(t) = \int e^{itx} dP^X = \int e^{itx} f(x) dx \quad (1.8)$$

$$= \int \cos tx f(x) dx + i \int \sin tx f(x) dx \quad (1.9)$$

$$(1.10)$$

La igualdad (1.10) se verifica sólo cuando P^X es continua con densidad f . Dado que las funciones seno y coseno son acotadas, ambas integrales son finitas y la función está bien definida para todo $t \in \mathbb{R}$. Las propiedades elementales son las siguientes. Su demostración queda propuesta al lector con breves indicaciones:

Proposición 1. (a) $|\varphi_X(t)| \leq 1$ para todo t y vale 1 en $t = 0$.

Indicación: Téngase en cuenta que $\cos^2 y + \sin^2 y = 1$.

(b) $\varphi_X(-t) = \overline{\varphi_X(t)}$.

Indicación: Téngase en cuenta que la función $\cos y$ es simétrica mientras que $\sin y$ es antisimétrica.

(c) Si $a, b \in \mathbb{R}$ con $b \neq 0$ entonces $\varphi_{a+bX}(t) = e^{iat} \varphi_X(bt)$.

(d) φ_X es continua en t .

Indicación: Basta aplicar TCD tanto en la integral del coseno como en la del seno.

Para hacernos una idea de estas propiedades, mostramos en la figura 1.3 las funciones características de las distribuciones $N(0, 1)$, F-V-P y $B(5, 0.5)$. Advertimos que las dos primeras no poseen parte imaginaria, pero la tercera sí. La parte imaginaria aparece cuando la distribución no es simétrica respecto a 0 y el grado de periodicidad está asociado a al de discretización.

1.5.1. El teorema de inversión

La función característica puede utilizarse como en (1.6) para el cálculo de momentos, con la dificultad que añade el hecho de trabajar con números complejos, pero con la ventaja de que no debe exigirse que los momentos de todos los órdenes sean finitos. Concretamente, si μ_n es finito para algún n , se verifica para todo $k \leq n$ que

$$\mu_k = \frac{\varphi_X^{(k)}(0)}{i^k} \quad (1.11)$$

Entonces, considerando un desarrollo de Taylor de φ_X en 0 y si $o(1)$ denota una función de t que converge a 0 cuando $t \rightarrow 0$, se verifica:

Teorema 1. Si μ_n es finito,

$$\varphi_X(t) = \sum_{k=0}^n \varphi_X^{(k)}(0) \frac{t^k}{k!} + t^n o(1) \quad (1.12)$$

TEOREMA Sea $X: (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}, \mathcal{R})$ variable aleatoria
 con función de densidad $f(x)$ y función característica φ_X .
 Si $E(|X|^n) < \infty$ para algún $n \geq 1$, entonces la derivada
 k -ésima ($k \leq n$) de φ_X , $\varphi_X^{(k)}$, existe, es uniformemente
 continua y $\varphi_X^{(k)}(t) = \int_{-\infty}^{\infty} (ik)^k e^{itx} f(x) dx \quad k=1, \dots, n$

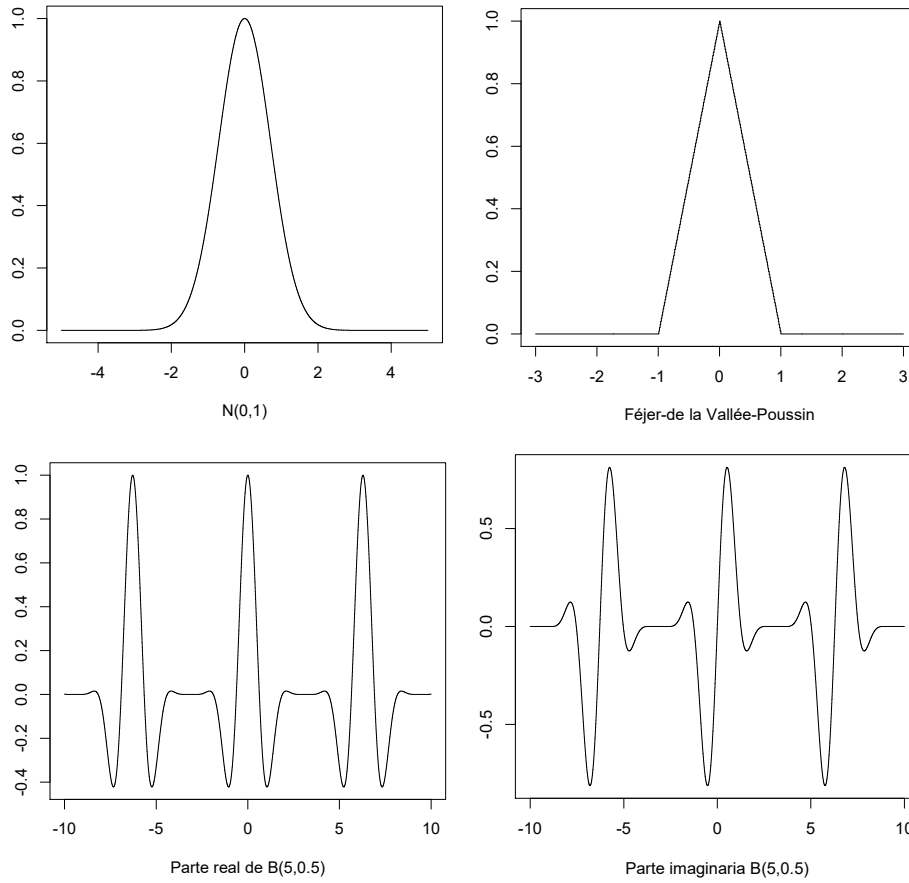
Así, $\varphi_X^{(k)}(0) = i^k \cdot E(X^k)$, $k=1, \dots, n$

$$\text{y } \varphi(t) = 1 + \sum_{k=1}^n \frac{(it)^k}{k!} E(X^k) + o(|t|^n)$$

$$\left(\lim_{t \rightarrow 0} \frac{o(|t|^n)}{|t|^n} = 0 \right)$$

Alas GA
 (Pag 176)

Figura 1.3: Ejemplos de funciones características



En particular, el desarrollo de orden 2 de φ_X se utilizará para probar el TCL. Podemos intuir pues que φ_X caracteriza la distribución P^X . El resultado conocido como Teorema de Inversión de Lévy lo deja claro:

Teorema 2. Sea $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ con función característica F y sean $a + h$ y $a - h$ ($h > 0$) dos puntos tal que $P(X = a + h) = P(X = a - h) = 0$. Entonces,

$$F(a + h) - F(a - h) = \lim_{T \rightarrow \infty} \frac{1}{\pi} \int_{-T}^T \frac{\sin(ht)}{t} e^{ita} \varphi_X(t) dt$$

Como consecuencia de este resultado se tiene el siguiente corolario:

Teorema 3. Se verifica que:

- (a) $[\varphi_X \equiv \varphi_Y] \Leftrightarrow [P^X = P^Y]$
- (b) Si φ_X es integrable, P^X es absolutamente continua con función de densidad

$$f(x) = \frac{1}{2\pi} \int e^{-itx} \varphi_X(t) dt$$

Por lo tanto, no sólo podemos afirmar que la función característica caracteriza la distribución, sino que podemos calcular la función de densidad de la misma (en el caso

continuo) a través de ella como transformada inversa de Fourier (ver (1.10)). Respecto a su cálculo, se sigue de (1.12) que, si $g_X(t)$ está bien definida para $|t| < \delta$, entonces puede extenderse a \tilde{g} definida en $\{z \in \mathbb{C} : |z| < \delta\}$, de manera que

$$\varphi(t) = \sum_{k=0}^{\infty} g_X^{(k)}(0) \frac{(it)^k}{k!} = \tilde{g}_X(it) \quad (1.13)$$

Eso quiere decir que, si somos capaces de calcular la función generatriz en un entorno de 0, obtenemos automáticamente la característica, como se aprecia en la tabla 2.1.

1.6. Caracterización multidimensional

Las definiciones anteriores se extienden de manera natural al caso de un vector aleatorio $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^k$, aunque la notación y los cálculos pueden complicarse. No obstante, surgen nuevas propiedades en relación con el estudio conjunto de las componentes X_1, \dots, X_k de X .

Función de densidad: podemos hablar también de una función de probabilidad en el caso discreto y de una función de densidad en el continuo, que se integrará según indica el Teorema de Fubini:

$$P(X \in A, Y \in B) = \int_{A \times B} f(x, y) \, dxy = \int_A \left(\int_B f(x, y) \, dy \right) dx \quad (1.14)$$

Por lo tanto, X e Y son independientes si la función de densidad conjunta se puede expresar de la forma $f(x, y) = f_1(x)f_2(y)$. Lo mismo puede decirse en el caso discreto de la función de probabilidad. En la figura 2.5 podemos apreciar la función de densidad de una distribución normal bivalente.

Aplicando (4), (6), (9) y (10) podemos calcular la densidades de ciertas transformaciones.

Densidades de transformaciones elementales: Dadas $X, Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ independientes cuyas distribuciones admiten las densidades f_X y f_Y , respectivamente, se verifica:

- La densidad de $X + Y$ es

$$f_{X+Y}(u) = \int f_X(u-v)f_Y(v) \, dv \quad (1.15)$$

Demostración. Dado $B \in \mathcal{B}$ y teniendo en cuenta (10) el cambio de variables $(u, v) = (x + y, y)$ o, equivalentemente, $g(u, v) = (u - v, v)$, con $J_g = 1$, se verifica

$$\begin{aligned} P(X + Y \in B) &= P(X + Y \in B, Y \in \mathbb{R}) \\ &= \int_{g(B \times \mathbb{R})} f_X(u-v)f_Y(v) \, duv \\ &= \int_B \left(\int_{\mathbb{R}} f_X(u-v)f_Y(v) \, dv \right) du \end{aligned}$$

□

La función (1.15) se denomina convolución. Las siguientes propiedades se prueban de manera similar.

- Si $Y > 0$, la densidad de XY es

$$f_{XY}(u) = \int_{\mathbb{R}^+} f_X(u/v) f_Y(v) v^{-1} dx \quad (1.16)$$

Indicación: Razona como antes considerando el cambio de variable $g(u, v) = (u/v, v)$ con $J_g = v^{-1}$.

- Si $Y > 0$, la densidad de X/Y es

$$f_{X/Y}(u) = \int_{\mathbb{R}^+} f_X(uv) f_Y(v) v dx \quad (1.17)$$

Indicación: Considera para el cambio de variable $g(u, v) = (uv, v)$ con $J_g = v$.

- La densidad de $\lambda \cdot X$ es

$$f_{\lambda X}(u) = |\lambda|^{-1} f_X(u/\lambda) \quad (1.18)$$

- Si $X > 0$, la densidad de \sqrt{X} es

$$f_{\sqrt{X}}(u) = 2u f_X(u^2) \quad (1.19)$$

- La densidad de e^X es

$$f_{e^X}(u) = \frac{1}{u} f_X(\log u) I_{(0, \infty)} \quad (1.20)$$

Parámetros multidimensionales: Además de los parámetros unidimensionales de cada una de sus componentes, son de gran interés los relativos a productos de componentes, en especial el de orden 2, es decir, correspondiente al producto entre dos componentes X_i y X_j . Se define pues la covarianza del par (X_i, X_j) mediante

$$\sigma_{ij} = E[X_i X_j]$$

que, por la desigualdad de Cauchy-Swartz, es finita si lo son ambos momentos de orden 2 y con valor absoluto acotado por $\sigma_i \sigma_j$, de manera que dicha cota se alcanza sii existen $\beta_0, \beta_1 \in \mathbb{R}$ tales que $X_j = \beta_0 + \beta_1 X_i$. Esto invita a definir el coeficiente de correlación lineal

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}$$

con valores en $[-1, 1]$ En particular, $\sigma_i^2 = \sigma_{ii}$. En tal caso, las covarianzas se organizan en una matriz simétrica y semidefinida positiva que se denota por Σ . Igualmente, puede definirse una matriz de correlaciones cuya diagonal vale 1.

Función de distribución: Podemos definirla sin problema y con propiedades similares aunque es incómoda de manejar.

Función generatriz: Se define, en principio, como la función que asigna a cada vector $t \in \mathbb{R}^k$ el número

$$g(t) = E[e^{t'X}] = \int \exp \left\{ \sum_{i=1}^k t_i x_i \right\} dP^X$$

De nuevo, el problema de esta función estriba en que no tiene por qué estar bien definida en todo \mathbb{R}^k y la ventaja, en que, si lo está en un entorno de 0, caracteriza la distribución P^X y permite obtener los momentos de cualquier orden de manera análoga a (1.6) mediante el cálculo de derivadas parciales. Así, por ejemplo, la covarianza entre X_i y X_j se obtiene mediante

$$\sigma_{ij} = \frac{1}{2} \frac{\partial^2 g(t_1, \dots, t_n)}{\partial x_i \partial x_j} (0) \quad (1.21)$$

Función característica: Se define como la función que asigna a cada vector $t \in \mathbb{R}^k$ el número complejo $E[e^{it'X}]$. Está bien definida en todo \mathbb{R}^k y las propiedades estudiadas en la sección anterior se extienden al caso multidimensional, incluida una versión en \mathbb{R}^k del teorema de inversión, por lo que caracteriza la distribución también en \mathbb{R}^k . Como consecuencia de dicho teorema junto con el de Fubini, podemos afirmar que las k componentes de X son independientes si $\varphi_X(t) = \prod_{i=1}^k \varphi_{X_i}(t_i)$. Además, si $A \in \mathcal{M}_{p \times k}$ y $b \in \mathbb{R}^p$,

$$\varphi_{AX+b}(t) = e^{ib't} \varphi_X(A't) \quad (1.22)$$

Trabajando con la función generatriz, si está bien definida, se obtienen resultados completamente análogos.

2. PRINCIPALES MODELOS DE DISTRIBUCIÓN

En este capítulo estudiaremos los modelos de distribución fundamentales, tanto discretos como continuos, incluyendo problemas multidimensionales. Los cálculos de las funciones generatriz y característica, así como el de su media y varianza, se dejan como ejercicio para el lector en la mayoría de los casos. En la tabla [2.1](#) mostramos un resumen de los mismos.

2.1. Modelos de distribución discretos

Caracterizaremos inicialmente cada una de las siguiente distribuciones mediante sus respectivas funciones de probabilidad y enunciaremos a continuación sus propiedades fundamentales.

2.1.1. Degenerada en un punto

También conocida como δ de Dirac en a (se denota δ_a), se corresponde con una fenómeno determinista, pues existe un valor $a \in \mathbb{R}$ tal que $P(X = a) = 1$. En ese caso, la media es a y la varianza 0. La función generatriz es $g_X(t) = e^{at}$ y la característica es $\varphi_X(t) = \cos at + i \sin at$. Si $a = 0$, $\varphi_X \equiv 1$.

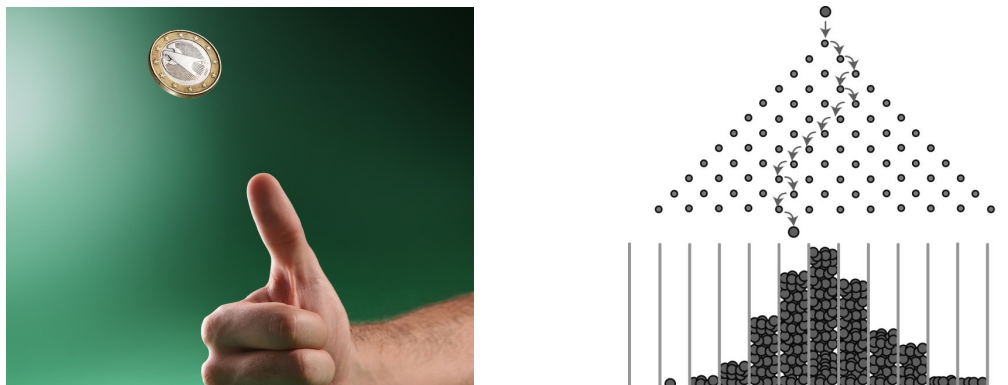
2.1.2. Distribución binomial

Empezaremos definiendo la distribución de Bernoulli de parámetro p pues corresponde al fenómeno aleatorio más simple, aquél con respuesta binaria (0 – 1) y con probabilidad p de obtener 1. Por ejemplo, el lanzamiento de una moneda simétrica debería seguir un modelo $B(0.5)$. Su función de densidad (probabilidad) es $P(X = k) = p^k(1 - p)^{1-k}$ para $k = 0, 1$.

La distribución binomial $B(n, p)$ corresponde a la suma de n réplicas independientes de un fenómeno aleatorio con distribución de Bernoulli p . Es lo que ocurre, por ejemplo, si consideramos la posición en la que cae cada bola en una máquina de Galton (figura [2.1](#)). Su función de probabilidad es

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n \quad (2.1)$$

Figura 2.1: Moneda y máquina de Galton



Las funciones característica y generatriz son, respectivamente,

$$\varphi_X(t) = (1 - p + pe^{it})^n, \quad g_X(t) = (1 - p + pe^t)^n \quad (2.2)$$

Indicación: Calcula φ_X para la distribución de Bernoulli y aplica (1.22) considerando la transformación lineal suma.

Se verifica también que la media y varianza de la distribución $B(n, p)$ son

$$\mu = np, \quad \sigma^2 = np(1 - p) \quad (2.3)$$

Indicación: Deriva dos veces en 0 $g_X(t)$ y aplica (1.6).

2.1.3. Distribución multinomial

Se trata de una extensión multidimensional de la distribución binomial asociada a n réplicas de un fenómeno aleatorio no binario, como puede ser el lanzamiento de un dado, puede que truco. La clave de la distribución se encuentra en la fórmula (2.4), que generaliza el conocido binomio de Newton:

$$\left(\sum_{j=1}^k a_j \right)^n = \sum_{i_1 + \dots + i_k = n} \frac{n!}{i_1! \dots i_k!} a_1^{i_1} \dots a_k^{i_k} \quad (2.4)$$

Dados $p_1, \dots, p_k \geq 0$ tales que $\sum_{j=1}^k p_j = 1$ y $n \in \mathbb{N}$, se dice que $(X_1, \dots, X_k)' : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^k$ sigue un modelo de distribución $M(n; p_1, \dots, p_k)$ cuando admite la siguiente función de probabilidad:

$$P(X_1 = i_1, \dots, X_k = i_k) = \frac{n!}{i_1! \dots i_k!} p_1^{i_1} \dots p_k^{i_k}, \quad \text{donde } i_1, \dots, i_k \geq 0, \sum_{j=1}^k i_j = n$$

Dado que $p_k = 1 - \sum_{j=1}^{k-1} p_j$ y $X_k = n - \sum_{j=1}^{k-1} X_j$, la última componente X_n es redundante, por lo que estamos realmente ante una distribución con $k - 1$ grados de libertad, i.e., que depende de $k - 1$ parámetros. En el caso $k = 2$ coincide con la binomial $B(n, p_1)$. Teniendo

en cuenta (2.4), su función generatriz es la siguiente:

$$\begin{aligned}
 g_{(X_1, \dots, X_{k-1})}(t_1, \dots, t_{k-1}) &= \sum_{i_1, \dots, i_{k-1} \leq n} e^{\sum_j t_j i_j} \frac{n!}{i_1! \dots i_{k-1}! (n - \sum_j i_j)!} p_1^{i_1} \dots p_{k-1}^{i_{k-1}} p_k^{n - \sum_j i_j} \\
 &= \sum_{i_1 + \dots + i_k = n} \frac{n!}{i_1! \dots i_k!} \prod_{j=1}^{k-1} (p_j e^{t_j})^{i_j} \cdot \left(1 - \sum_{j=1}^{k-1} p_j\right)^{i_k} \\
 &= \left(1 - \sum_{j=1}^{k-1} p_j + \sum_{j=1}^{k-1} p_j e^{t_j}\right)^n
 \end{aligned}$$

La función característica se deduce trivialmente de la generatriz. Además, derivando esta última obtenemos fácilmente los diferentes momentos. Concretamente, $E[X_j] = np_j$, $\text{var}[X_j] = np_j(1 - p_j)$ y $\text{cov}[X_j, X_{j'}] = -np_j p_{j'}$.

2.1.4. Distribución de Poisson

Esta distribución discreta tiene dos aplicaciones diferentes: una relacionada con la distribución exponencial, que veremos más adelante, y otra con la binomial, que explicaremos ahora. En todo caso, se dice que $X \sim \mathcal{P}(\lambda)$ con $\lambda > 0$ cuando su función de probabilidad es

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 1, 2, \dots \quad (2.5)$$

Antes de profundizar en esta distribución convendría tener en presentes dos conocidos resultados sobre el número e esenciales a lo largo de nuestra teoría:

Proposición 2. *El número e , definido como $\lim_{n \rightarrow \infty} (1 + 1/n)^n$, verifica:*

$$(a) \quad e^z = \sum_{k=0}^{\infty} \frac{z^k}{k!}, \quad \forall z \in \mathbb{C}$$

(b) *En general, si $\lim_{n \rightarrow \infty} a_n = \infty$, entonces*

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{a_n}\right)^{a_n} = e, \quad \lim_{n \rightarrow \infty} \left(1 - \frac{1}{a_n}\right)^{a_n} = e^{-1} \quad (2.6)$$

De la proposición 2 se deduce que $\sum_{k=0}^{\infty} P(X = k) = 1$, por lo que se trata realmente de una función de probabilidad. Las funciones característica y generatriz son, respectivamente

$$\varphi_X(t) = \exp(\lambda(e^{it} - 1)), \quad g(t) = \exp(\lambda(e^t - 1)) \quad (2.7)$$

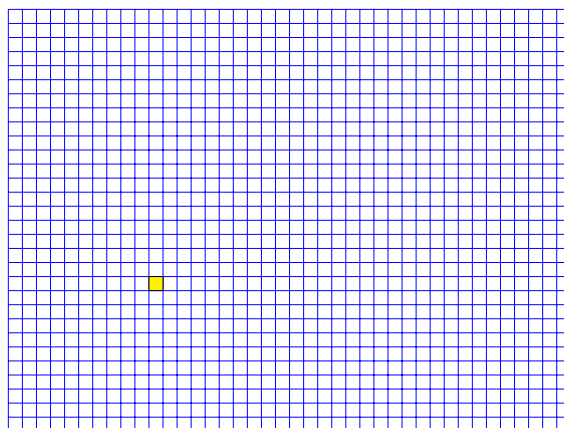
Indicación: Ten en cuenta que

$$\sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!}$$

y aplica la proposición anterior. Tanto la media como la varianza son λ , lo cual puede razonarse a través de la función de probabilidad o bien derivando g_X en 0. Decíamos antes que la distribución de Poisson guarda relación con la binomial. Veamos en qué sentido. En la figura 2.2 observamos una cuadrícula con 192 celdas. Imaginemos que bombardeamos

la cuadrícula de manera que todas las celdas tengan la misma probabilidad de impacto, que será por lo tanto $p = 1/192$. Si se efectúan en total n impactos independientes y nos centramos en la casilla amarilla, el número de impactos sobre la misma seguirá una distribución $B(n, p)$. Al ser p tan pequeño, lo más probable es que reciba 0 impactos, a menos que n sea suficientemente grande, dado que el número esperado de impactos es np . En tal caso, el número de impactos sigue aproximadamente una distribución $\mathcal{P}(np)$.

Figura 2.2: Ejemplo de distribución de Poisson



Teorema 4. Si $p_n \rightarrow 0$ y $np_n \rightarrow \lambda$ cuando $n \rightarrow \infty$, la función de probabilidad (y, por lo tanto, la de distribución) de $B(n, p_n)$ converge a la de $\mathcal{P}(\lambda)$.

Demostración.

$$\begin{aligned} P(X_n = k) &= \binom{n}{k} p_n^k (1 - p_n)^{n-k} \\ &= \frac{(np_n)^k}{k!} \cdot \frac{n \cdot (n-1) \dots (n-k+1)}{n^k} \cdot (1 - p_n)^{n-k} \end{aligned}$$

Denotense respectivamente por x_n , y_n y z_n los tres factores anteriores. En tal caso $x_n \rightarrow \lambda^k/k!$ por hipótesis y $y_n \rightarrow 1$. Respecto a z_n , se verifica

$$\begin{aligned} (1 - p_n)^{n-k} &= \left(1 - \frac{1}{p_n^{-1}}\right)^{p_n^{-1} \cdot np_n} \cdot (1 - p_n)^{-k} \\ &= \left[\left(1 - \frac{1}{p_n^{-1}}\right)^{p_n^{-1}}\right]^{np_n} \cdot (1 - p_n)^{-k} \\ &= \xi_n^{np_n} \cdot (1 - p_n)^{-k} \end{aligned}$$

donde $\lim_{n \rightarrow \infty} \xi_n = e^{-1}$ en virtud de la proposición (2.6). Dado que el segundo factor converge a 1, se tiene que $P(X_n = k)$ converge a $e^{-\lambda} \lambda^k/k!$ \square

Hemos prestado especial atención a este resultado pues es un anticipo del TCL que veremos más adelante.

2.1.5. Distribución binomial negativa

La distribución binomial negativa con $r \in \mathbb{N}$ y $p \in (0, 1)$ se denota por $BN(r, p)$. Generaliza la distribución geométrica, que es la que asigna a cada k la probabilidad de que fenómeno aleatorio binario distribuido según $B(p)$ alcance su primer “éxito” (1) tras k réplicas independientes. La binomial negativa indica la probabilidad de obtener el “éxito” número r tras k réplicas del experimento. Por lo tanto, su función de probabilidad es

$$P(X = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \quad (2.8)$$

Sus propiedades se derivan la fórmula binomial negativa de Newton, que afirma

$$(1+x)^{-n} = \sum_{k=0}^{\infty} \binom{k+n-1}{n-1} x^k \quad (2.9)$$

de la cual se deducen las propiedades fundamentales de la distribución, resumidas en la tabla [2.1](#)

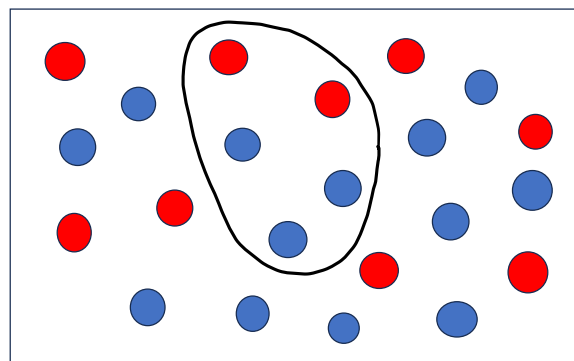
2.1.6. Distribución hipergeométrica

Se trata de un modelo de distribución de gran interés en el muestreo en poblaciones finitas y relacionado, como veremos, con la distribución binomial. Dado $N \in \mathbb{N}$, $n = 1, \dots, n$ y $M \in 1, \dots, N$, se dice que una variable X con valores en $\{0, 1, \dots, M\}$ sigue un modelo $HG(N, M, n)$ cuando

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad \text{donde } \max\{0, n+M-N\} \leq k \leq \min\{n, M\} \quad (2.10)$$

Modeliza un proceso de extracción aleatoria sin reemplazo de n objetos de un total de N posibles donde distinguimos dos categorías: la categoría A con M objetos y la B con $N - M$. X denota entonces el número de objetos de A obtenidos. En la figura [2.3](#) se ilustra el proceso asociado a $HG(22, 9, 5)$ con $k = 2$.

Figura 2.3: Ilustración $HG(22, 9, 5)$ con $k = 2$



DISTRIBUCIÓN BINOMIAL NEGATIVA

Consideremos un experimento con dos posibles resultados, 0 (fracaso) y 1 (éxito). Repetimos el experimento hasta que ocurre el primer éxito.

Consideremos la variable aleatoria X que cuenta el n.º de fracasos antes de obtener 1 éxito. $X \in \{0, 1, 2, 3, \dots\}$

$$\text{Se tiene que } P(X=k) = (1-p)^k p^1$$

Diremos que X sigue una distribución de probabilidad BINOMIAL NEGATIVA de parámetros 1 y p ($X \sim \text{BN}(1, p)$)

$$\text{Comprobemos que } \sum_{k=0}^{\infty} P(X=k) = 1 :$$

$$\sum_{k=0}^{\infty} P(X=k) = \sum_{k=0}^{\infty} (1-p)^k p = p \sum_{k=0}^{\infty} (1-p)^k = p \cdot \frac{1}{1-(1-p)} = 1$$

Supongamos ahora que repetimos el experimento hasta que ocurren 2 éxitos.

Consideremos ahora la variable aleatoria X que cuenta el n.º de fracasos antes de obtener 2 éxitos. $X \in \{0, 1, 2, 3, \dots\}$

Se tiene que $P(X=k) = \binom{k+1}{1} (1-p)^k p^2$ pues los casos favorables a que $X=k$ podríamos representarlos así: $\underbrace{FF \dots FF}_{k+1} FE, FE$, donde el primer éxito puede ocurrir entre las posiciones 1 y $k+1$; en total, $k+1$ casos favorables.

¿Y si repetimos el experimento hasta que ocurren 3 éxitos?

Consideremos ahora la variable aleatoria X que cuenta el n.º de fracasos antes de obtener 3 éxitos. $X \in \{0, 1, 2, 3, \dots\}$

$$P(X=k) = \binom{k+2}{2} p^3 (1-p)^k$$

¿Y si repetimos el experimento hasta que ocurren r éxitos?

Consideremos ahora la variable aleatoria X que cuenta el n.º de fracasos antes de obtener r éxitos. $X \in \{0, 1, 2, 3, \dots\}$

$$P(X=k) = \binom{k+r-1}{r-1} p^r (1-p)^k$$

Para comprobar que $\sum_{k=0}^{\infty} P(X=k) = 1$, necesitamos calcular $\sum_{k=0}^{\infty} \binom{k+r-1}{r-1} (1-p)^k$.

La función $f(x) = (1-x)^r$ es infinitamente diferenciable en $|x| < 1$. Por tanto, se puede escribir así: $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(0)}{k!} x^k$, donde $f^{(k)}$ es la derivada k -ésima.

$$f^{(k)}(x) = (-r)(-r-1) \dots (-r-k+1) x^{-r-k} \cdot (-1)^k = \binom{r+k-1}{r-1} x^k$$

(serie de Taylor)

$$\text{Por tanto, } \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} (1-p)^k = (1-(1-p))^r = p^r.$$

FUNCIÓN CARACTERÍSTICA

$$\psi_X(t) = E(e^{itX}) = \sum_{k=0}^{\infty} e^{itk} \binom{k+r-1}{r-1} p^r (1-p)^k =$$

$$= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} [(1-p)e^{it}]^k = p^r (1 + (p-1)e^{it})^{-r}$$

FUNCIÓN GENERATRIZ DE MOMENTOS

$$\psi_X(t) = E(e^{tX}) = \sum_{k=0}^{\infty} e^{tk} \binom{k+r-1}{r-1} p^r (1-p)^k =$$

$$= p^r \sum_{k=0}^{\infty} \binom{k+r-1}{r-1} ((1-p)e^t)^k =$$

$$= p^r (1 + (p-1)e^t)^{-r}$$

$$\uparrow$$

si $|(p-1)e^t| < 1 \Leftrightarrow t < \ln\left(\frac{1}{1-p}\right)$

$$\text{" } (1-p)e^t$$

Puede comprobarse que sus media y varianza son

$$\mu = n \frac{M}{N}, \quad \sigma^2 = \frac{nM(N-n)(N-M)}{N^2(N-1)} \quad (2.11)$$

y se obtienen de manera directa. No se propone una expresión analítica alternativa a la de la propia definición para las funciones generatriz y característica en la tabla 2.1. Por otra parte, podemos reorganizar la expresión (2.10) de la forma

$$P(X = k) = \binom{n}{k} \cdot \frac{M}{N} \cdots \frac{M-k+1}{N-k+1} \cdot \frac{N-M}{N-k} \cdots \frac{N-M-n+k+1}{N-n+1} \quad (2.12)$$

Nótese que, si el tamaño de la población N tiende a infinito junto con el de la categoría A , que vale M , pero de manera que el cociente $M/N \rightarrow p \in (0, 1)$, se sigue de (2.12) que

$$P(X = k) \longrightarrow \binom{n}{k} p^k (1-p)^{n-k}$$

que corresponde a la distribución $B(n, p)$. En ese sentido, podríamos decir que la distribución hipergeométrica es una especie de versión de la binomial para muestreo en poblaciones finitas, o mejor, la que se obtiene en lugar de la binomial en un muestreo con reemplazamiento. En el ejemplo de la figura, $X \sim HG(22, 9, 5)$ da $P(X = 2) = 0.39$, mientras que $X \sim B(5, 9/22)$ da $P(X = 2) = 0.34$. Esa diferencia se debe a que, en poblaciones tan pequeñas ($N = 22$), el efecto del reemplazamiento es notorio.

2.2. Modelos de distribución continuos

Los caracterizaremos inicialmente mediante sus respectivas funciones de densidad y enunciaremos sus propiedades fundamentales.

2.2.1. Distribución normal

Es la distribución central de la Estadística en virtud de TCL. De entrada, el caso particular de $N(0, 1)$ es invariante ante la transformada de Fourier. Utilizaremos dicha transformada para probar TCL. Dados $\mu > 0$, $\sigma > 0$ y $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$, se dice que $X \sim N(\mu, \sigma)$ cuando admite como función de densidad

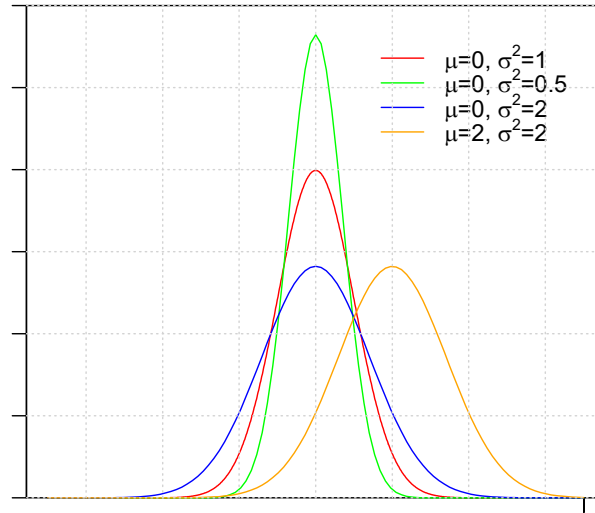
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}} \quad (2.13)$$

No podemos aportar una expresión analítica de la función primitiva de f , por lo que la comprobación de que su integral es uno no es trivial y se basa en que

$$\int_0^\infty e^{-x^2} dx = \sqrt{\frac{\pi}{2}} \quad (2.14)$$

El caso particular $\mu = 0$ y $\sigma = 1$ se denomina normal estándar y se denota usualmente por Z . Aplicando el cambio de variable $x = \mu + \sigma z$, se deduce que cualquier distribución

Figura 2.4: Densidades de normales con diferentes parámetros



normal puede obtenerse como una transformación afín de Z . La función generatriz de Z se calcula así:

$$\begin{aligned} g_Z(t) &= \frac{1}{\sqrt{2\pi}} \int e^{tz} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(z^2 - 2tz + t^2 - t^2)} dz \\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}} \int e^{-\frac{1}{2}(z-t)^2} dz = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int e^{-\frac{1}{2}u^2} du \\ &= e^{t^2/2} \int f(u) du = e^{t^2/2} \end{aligned}$$

Derivando la función generatriz concluimos que la media de x es 0 y su varianza 1. También lo es su desviación típica. Teniendo en cuenta (1.1), la media y desviación típica de X son μ y σ , respectivamente. De 1.7 se deduce que la función generatriz de $X \sim N(\mu, \sigma)$ es $g_X(t) = \exp\{\mu t + (\sigma t)^2/2\}$, siendo su función característica $\varphi_X(t) = \tilde{g}_X(it) = \exp\{it\mu - t^2/2\}$. En particular,

$$\varphi_Z(t) = e^{-t^2/2} \quad (2.15)$$

es la función característica de $N(0, 1)$, que se representa en la figura 1.3 y será la función límite en TCL.

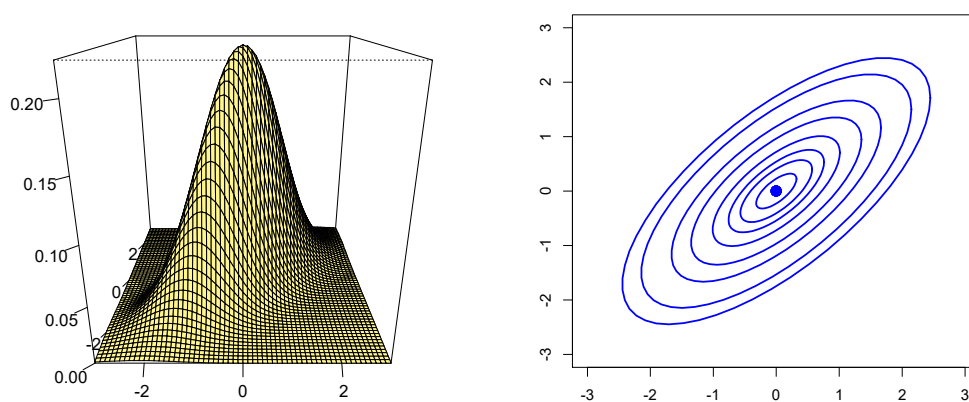
La distribución $N(\mu, \sigma)$ es simétrica respecto a su media, indicando σ el grado de dispersión en relación a la misma, tal y como se ilustra en la figura 2.4. El hecho de que quede caracterizada por estos dos parámetros, media y desviación típica, es de gran trascendencia en lo que se denomina Estadística Paramétrica.

Distribución normal multivariante: Se trata de la extensión multidimensional de la distribución normal. Si $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^n$, $\mu \in \mathbb{R}^n$ y $\Sigma \in \mathcal{M}_{n \times n}$ simétrica y definida positiva, decimos que $X \sim N_n(\mu, \Sigma)$ cuando admite la siguiente función de densidad

$$f_X(x) = \frac{1}{(2\pi|\Sigma|)^{n/2}} \exp\left\{-\frac{1}{2}(x - \mu)' \Sigma^{-1}(x - \mu)\right\}, \quad x \in \mathbb{R}^n \quad (2.16)$$

Las propiedades son por completo análogas a las de la normal unidimensional, aparte de algunas adicionales que enunciaremos a continuación. Podemos hablar también de una distribución normal multivariante estándar $N_n(0, Id)$, a partir de la cual se obtiene el resto de normales en la misma dimensión mediante una transformación afín adecuada. La distribución $N_n(\mu, \Sigma)$ tiene media μ y matriz de covarianza Σ . La parte izquierda de la figura 2.5 ilustra precisamente la función de densidad de una distribución Normal bivalente con alta correlación entre sus dos componentes. Las funciones generatriz y característica se muestran en la tabla 2.1. Además, se verifica:

Figura 2.5: Densidad y contornos de normal bivalente con fuerte correlación



- (a) Si $X \sim N_n(\mu, \Sigma)$, $A \in \mathcal{M}_{k \times n}$ y $b \in \mathbb{R}^k$, entonces

$$AX + b \sim N_k(A\mu + b, A'\Sigma A)$$

Indicación: Calcula la correspondiente función característica.

- (b) $X_i \sim N(\mu_i, \sigma_i^2)$ independientes $i = 1, \dots, n$ si, y sólo si,

$$\begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} \sim N_n \left(\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{pmatrix} \right)$$

Indicación: Aplica al apartado anterior para $A = (1, 0, \dots, 0)$ etc. Razona con la función característica.

- (c) $X \sim N_n(\mu, \Sigma)$ sii $a'X \sim N(a'\mu, \sqrt{a'\Sigma a}) \forall a \in \mathbb{R}^n$.

Indicación: Ten en cuenta que $\varphi_X(t) = \varphi_{\nu_X}(1)$.

- (d) Correlación lineal: si

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{n_1+n_2} \left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \right)$$

se verifica entonces que $P^{X_1|X_2=x_2} \sim N_{n_1}(\alpha + \beta'x_2, \Sigma)$, donde

$$\beta = \Sigma_{22}^{-1}\Sigma_{21}, \quad \alpha = \mu_1 - \beta'\mu_2 \text{ y } \Sigma = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

En particular, si $n_1 = n_2 = 1$, $X_1 = \alpha + \beta X_2 + \varepsilon$, con $\varepsilon \sim N(0, \sigma)$ independiente de X_2 , donde $\beta = \sigma_{12}/\sigma_2^2$, $\alpha = \mu_1 - \beta\mu_2$, $\sigma^2 = \sigma_1^2(1 - \rho_{12}^2)$

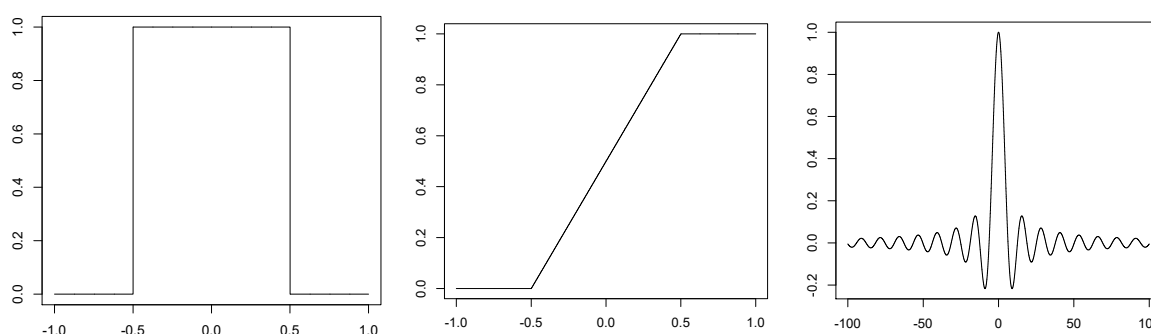
Indicación: Trabaja con las función de densidad condicional.

- (f) Dado $k > 0$, la región de \mathbb{R}^n compuesta por los x tales que $f(x) = k$ es un elipsoide (ver parte derecha de la figura 2.5). Si $\Sigma = \sigma^2 Id$ para algún $\sigma^2 > 0$, entonces es una esfera en \mathbb{R}^n . De ahí que una distribución del tipo $X \sim N_n(\mu, \sigma^2 Id)$ se denomina esférica. Será la distribución de partida del Modelo Lineal Normal.

2.2.2. Distribución uniforme

Es el modelo de distribución que se asocia al azar puro (máxima entropía). Puede definirse en cualquier conjunto A de medida finita $\mu(A)$ en \mathbb{R}^k , siendo $\mu(A)^{-1}I_A$ su función de densidad. A la hora de calcular probabilidades mediante dicha función debemos tener presente (4). No obstante, nos centraremos aquí un intervalo en \mathbb{R} del tipo $[-a, a]$. En ese caso, decimos que $X \sim U[-a, a]$ cuando su función de densidad es $1/(2a) \cdot I_{[-a, a]}$. La generatriz y característica pueden encontrarse en la tabla 2.1. La media es el centro de intervalo, en este caso 0, y la varianza $a^2/3$.

Figura 2.6: Funciones de densidad, distribución y característica de $U[-0.5, -0.5]$



2.2.3. Distribución Gamma

Se trata de una familia de distribuciones entre las que se encuentran, como caso particular, la exponencial y χ^2 , que se relaciona a su vez con la normal multivariante esférica. Dado $\alpha > 0$ puede comprobarse que el parámetro

$$\Gamma[\alpha] = \int_0^\infty x^{\alpha-1} e^{-x} dx$$

es finito y verifica las siguientes propiedades:

- (a) $\Gamma[1] = 1$

$$\text{¿ } \Gamma\left(\frac{1}{2}\right) ? \quad \Gamma\left(\frac{1}{2}\right) = \int_0^{\infty} x^{-1/2} e^{-x} dx =$$

$$\begin{matrix} x = y^2/2 \\ y = \sqrt{2x} \end{matrix} \quad dy = \frac{1}{\sqrt{2x}} dx$$

$$= \sqrt{2} \int_0^{\infty} e^{-y^2/2} dy = \sqrt{2} \sqrt{\frac{\pi}{2}} = \sqrt{\pi}$$

DISTRIBUCIÓN NORMAL

$Z \sim N(0,1)$ si su función de densidad $f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, $x \in \mathbb{R}$

Comprobemos que $\int_{-\infty}^{\infty} f_Z(x) dx = 1$

Para ello, veremos que $\int_{-\infty}^{\infty} f_Z(u) du \cdot \int_{-\infty}^{\infty} f_Z(v) dv = 1$.

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv \stackrel{\text{teorema de Fubini}}{=}$$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-u^2/2} e^{-v^2/2} dv du = \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(u^2+v^2)/2} dv du$$

Cambio a coordenadas polares: $\begin{cases} u = r \cos \theta \\ v = r \sin \theta \end{cases} \Rightarrow u^2 + v^2 = r^2$

$$\text{Jacobiano} = \begin{vmatrix} \partial u / \partial r & \partial u / \partial \theta \\ \partial v / \partial r & \partial v / \partial \theta \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r$$

$$= \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} \cdot r dr d\theta = \frac{1}{2\pi} \int_0^{2\pi} \left(-e^{-r^2/2} \Big|_0^{\infty} \right) d\theta =$$

$$= \frac{1}{2\pi} 2\pi = 1$$

FUNCIÓN GENERATRIZ DE MOMENTOS

(1) $X \sim N(0,1)$

$$g_X(t) = E(e^{tx}) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{tx} e^{-x^2/2} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{2tx-x^2}{2}} dx =$$
$$= e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} dx = e^{t^2/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = e^{t^2/2}$$

cambio $y = x-t$, $dy = dx$

(2) $X \sim N(\mu, \sigma)$, $\mu \in \mathbb{R}$, $\sigma > 0$.

Sabemos que $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$. Así, $X = \sigma Z + \mu$

$$g_X(t) = E(e^{tX}) = E(e^{t(\sigma Z + \mu)}) = e^{t\mu} E(e^{\sigma t Z}) =$$
$$= e^{t\mu} \cdot e^{\frac{\sigma^2 t^2}{2}} = e^{t\mu + \frac{\sigma^2 t^2}{2}}$$

FUNCIÓN CARACTERÍSTICA

De una manera similar obtenemos

(I) si $Z \sim N(0,1) \Rightarrow \psi_Z(t) = e^{-t^2/2}$

(II) si $X \sim N(\mu, \sigma) \Rightarrow \psi_X(t) = e^{it\mu - \frac{\sigma^2 t^2}{2}}$

MOMENTOS

$z \sim N(0,1) \Rightarrow E(z^n) = 0$ si n impar

Si $z \sim N(0,1)$, n par $E(z^n) = \int_{-\infty}^{+\infty} z^n \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz =$

$$= \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^n e^{-\frac{z^2}{2}} dz \stackrel{\substack{\frac{z^2}{2} = y \Rightarrow z dz = dy \\ z = (2y)^{1/2}}}{=} \frac{1}{\sqrt{2\pi}} \int_0^{\infty} (2y)^{\frac{n-1}{2}} e^{-y} dy =$$

$$= \frac{2^{\frac{n-1}{2}} \cdot 2}{\Gamma(1/2) 2^{1/2}} \int_0^{\infty} y^{\frac{n-1}{2}} e^{-y} dy = \frac{2^{\frac{n}{2}}}{\Gamma(1/2)} \Gamma\left(\frac{n+1}{2}\right) =$$

$$= (n-1)(n-3) \dots 3 \cdot 1 = \frac{n!}{(\frac{n}{2})! 2^{\frac{n}{2}}} \text{ pues}$$

1) Se tiene que $\Gamma\left(\frac{n+1}{2}\right) = \frac{n-1}{2} \cdot \Gamma\left(\frac{n-1}{2}\right) \cdot \frac{n-1}{2} \cdot \frac{n-3}{2} \Gamma\left(\frac{n-3}{2}\right) =$
 $= \frac{n-1}{2} \cdot \frac{n-3}{2} \dots \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{(n-1)(n-3) \dots 1}{2^{\frac{n}{2}}}$
 $1 = n - (n-1)$

2) $\frac{n!}{(\frac{n}{2})! 2^{\frac{n}{2}}} = \frac{n(n-1)(n-2)(n-3) \dots 3 \cdot 2 \cdot 1}{\frac{n}{2} \cdot \frac{n-2}{2} \cdot \frac{n-4}{2} \dots 2 \cdot 2^{\frac{n}{2}}} =$
 $= \frac{(n-1)(n-3) \dots 3 \cdot 1}{\frac{2^{n/2}}{2^{n/2}}} = (n-1)(n-3) \dots 3 \cdot 1$

Si $X \sim N(\mu, \sigma) \Rightarrow X = \sigma z + \mu$, $E(X^n) = E((\mu + \sigma z)^n) = E\left(\sum_{k=0}^n \binom{n}{k} \mu^k (\sigma z)^{n-k}\right) =$

$$= \sum_{k=0}^n \binom{n}{k} \mu^k \sigma^{n-k} E(z^{n-k}) = \sum_{\substack{k=0 \\ n-k \text{ par}}}^n \binom{n}{k} \mu^k \sigma^{n-k} \frac{(n-k)!}{(\frac{n-k}{2})! 2^{\frac{n-k}{2}}}$$

USANDO LA FUNCIÓN GENERATRIZ

Si n es impar $\Rightarrow E((X-\mu)^2) = 0$ pues ...

$$g_X(t) = e^{t\mu + \sigma^2 t^2 / 2}$$

$$g'_X(t) = (\mu + \sigma^2 t) e^{t\mu + \sigma^2 t^2 / 2} ; g'_X(0) = \mu$$

$$g''_X(t) = (\mu + \sigma^2 t) \cdot (\mu + \sigma^2 t) e^{t\mu + \sigma^2 t^2 / 2} + \sigma^2 e^{t\mu + \sigma^2 t^2 / 2} ;$$

$$g''(0) = \mu^2 + \sigma^2 (= E(X^2))$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

EJERCICIO Sean X_1, X_2 , variables aleatorias independientes,

$X_k \sim N(\mu_k, \sigma_k^2)$. Entonces:

$$X = a_1 X_1 + a_2 X_2 \sim N(a_1 \mu_1 + a_2 \mu_2, \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2})$$

Usamos la función característica pero antes tenemos en cuenta que:

$$E(a_k X_k) = a_k E(X_k) = a_k \mu_k, \quad k=1,2. \quad \text{Para } E(X_k^2), \text{ calculemos } \varphi_{X_k}''(t)$$

$$\varphi_{X_k}(t) = e^{it\mu_k - \frac{t^2 \sigma_k^2}{2}}$$

$$\varphi_{X_k}'(t) = (i\mu_k - \sigma_k^2 t) e^{it\mu_k - \frac{t^2 \sigma_k^2}{2}}$$

$$\varphi_{X_k}''(t) = -\sigma_k^2 e^{it\mu_k - \frac{t^2 \sigma_k^2}{2}} + (i\mu_k - \sigma_k^2 t)^2 e^{it\mu_k - \frac{t^2 \sigma_k^2}{2}}$$

$$\varphi_{X_k}''(0) = -\sigma_k^2 - \mu_k^2 \Rightarrow E(X_k^2) = \mu_k^2 + \sigma_k^2.$$

Así, $\text{Var}(X_k) = \mu_k^2 + \sigma_k^2 - \mu_k^2 = \sigma_k^2$. Por lo tanto

$$E(a_k^2 X_k^2) = a_k^2 E(X_k^2) = a_k^2 (\mu_k^2 + \sigma_k^2)$$

$$\text{Var}(a_k X_k) = E(a_k^2 X_k^2) - E(a_k X_k)^2 = a_k^2 \sigma_k^2$$

$$\varphi_X(t) = E(e^{it(a_1 X_1 + a_2 X_2)}) = E(e^{it a_1 X_1} e^{it a_2 X_2}) \stackrel{\text{independencia}}{=} E(e^{it a_1 X_1}) E(e^{it a_2 X_2})$$

$$= E(e^{it a_1 X_1}) E(e^{it a_2 X_2}) = e^{it a_1 \mu_1 + \frac{a_1^2 t^2 \sigma_1^2}{2}} e^{it a_2 \mu_2 + \frac{a_2^2 t^2 \sigma_2^2}{2}} =$$

$$= e^{it(a_1 \mu_1 + a_2 \mu_2) + \frac{(a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2) t^2}{2}}$$

que es la función característica de una

$$N(a_1 \mu_1 + a_2 \mu_2, \sqrt{a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2})$$

COROLARIO Si $X_1, \dots, X_n \sim N(\mu, \sigma)$ independientes, entonces $S_n \sim (n\mu, \sqrt{n}\sigma)$ y $\bar{X}_n = \frac{S_n}{n} \sim N(\mu, \sigma/\sqrt{n})$

$$X_1, \dots, X_n \sim N(\mu, \sigma) \Rightarrow \varphi_{S_n}(t) = \prod_{j=1}^n \varphi_{X_j}(t) =$$

$$= \left(e^{it\mu - \frac{\sigma^2 t^2}{2}} \right)^n = e^{itn\mu - \frac{n\sigma^2 t^2}{2}}, \text{ es decir, } S_n \sim N(n\mu, \sqrt{n}\sigma)$$

$$\text{y } \varphi_{\frac{S_n}{n}}(t) = E\left(e^{it \frac{S_n}{n}}\right) = \varphi_{S_n}\left(\frac{t}{n}\right) = e^{it\mu - n\sigma^2 \frac{t^2}{2n^2}} = e^{it\mu - \frac{\sigma^2}{n} \frac{t^2}{2}}, \text{ es decir, } \bar{X}_n = \frac{S_n}{n} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

DISTRIBUCIÓN LOG-NORMAL

Decimos que $X \sim LN(\mu, \sigma)$ si $Y = \ln(X) \sim N(\mu, \sigma)$

¿función de densidad? La función de densidad de $Y \leftarrow$

$$f_Y(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \quad y \in \mathbb{R}$$

Hacemos el cambio $y = \ln(x)$, $dy = \frac{1}{x} dx$ ($\ln(x): (0, \infty) \rightarrow \mathbb{R}$ biyectiva)
 $x = e^y$

$$\text{Así, } f_X(x) = f_Y(\ln(x)) \cdot \left| \frac{1}{x} \right| = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, \text{ si } x > 0$$

EJERCICIOS

1. función característica. Momentos

2. Si $X \sim LN(\mu, \sigma) \Rightarrow aX^b \sim LN(b\mu + \ln(a), b\sigma)$, $a, b > 0$

3. $X_i \sim LN(\mu_i, \sigma_i)$, $i=1, 2 \Rightarrow X_1 \cdot X_2 \sim LN(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2})$

$$E(X) = \int_0^\infty x \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx =$$

$$E(X^2) = \int_0^\infty x^2 \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx =$$

$$y = \frac{\ln(x) - \mu}{\sigma}, \quad x = e^{\sigma y + \mu} \quad dy = \frac{1}{\sigma} \frac{1}{x} dx \quad (x > 0 \Rightarrow y \in \mathbb{R})$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{(\sigma y + \mu)} e^{-y^2/2} dy = \frac{e^{\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2 - 2\sigma y)/2} dy$$

$$= \frac{e^{\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-\sigma)^2/2} dy = e^{\mu} e^{\sigma^2/2} = e^{\mu + \sigma^2/2}$$

$$y = \frac{\ln(x) - \mu}{\sigma}, \quad x = e^{\sigma y + \mu} \quad dy = \frac{1}{\sigma} \frac{1}{x} dx \quad (x > 0 \Rightarrow y \in \mathbb{R})$$

$$= \int_{-\infty}^{\infty} e^{2(\sigma y + \mu)} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy = \frac{e^{2\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2 - 4\sigma y)/2} dy$$

$$= \frac{e^{2\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-2\sigma)^2/2} dy = e^{2(\mu + \sigma^2/2)}$$

$$\text{Var}(X) = E(X^2) - E(X)^2 = e^{2(\mu + \sigma^2)} - e^{2(\mu + \sigma^2/2)} = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$$

$$E(X^n) = \frac{e^{n\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y^2 - 2n\sigma y)/2} dy = \frac{e^{n\mu}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(y-n\sigma)^2/2} dy$$

$$= e^{n\mu + \frac{n^2 \sigma^2}{2}}$$

También lo podríamos haber hecho así:

$$E(X^n) = E((e^Y)^n) = E(e^{nY}) =$$

$$Y = \ln(X) \sim N(\mu, \sigma) \Rightarrow \psi_Y(t) = E(e^{tY}) = e^{t\mu + \frac{\sigma^2}{2} t^2}$$

$$\text{Así que } E(X^n) = \psi_Y(n) = e^{n\mu + \frac{n^2}{2} \sigma^2} \quad (\text{usando la función generatriz de momentos})$$

DISTRIBUCIÓN χ^2 (chi cuadrado, ji cuadrado)

Decimos que $X: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{R})$ abs. continua tiene distribución χ^2 de parámetro k ($k=1, 2, \dots$), $X \sim \chi_k^2$ si su función de densidad es

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(k/2) 2^{k/2}} x^{k/2-1} e^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{en otro caso} \end{cases}$$

Esta es la función de densidad de $\Gamma(k/2, 2)$

TEOREMA Si X_1, \dots, X_n son variables aleatorias independientes, $X_k \sim N(0, 1)$, $k=1, \dots, n$. Entonces

$$\sum_{k=1}^n X_k^2 \sim \chi_n^2$$

Demostación

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{si } x \in \mathbb{R}. \quad \text{Tomamos } y = g(x) = x^2.$$

$$\text{Así, } \begin{cases} \text{si } x \geq 0 \Rightarrow x = \sqrt{y} \\ \text{si } x < 0 \Rightarrow x = -\sqrt{y} \end{cases} \quad y > 0.$$

$$\text{Hacemos el cambio } y = x^2 \begin{cases} x \geq 0, x \rightarrow \sqrt{y} \\ x < 0, x \rightarrow -\sqrt{y} \end{cases}$$

$$y \Rightarrow \begin{cases} x = \sqrt{y} & \text{si } x \geq 0 \Rightarrow dx = \frac{1}{2\sqrt{y}} \\ x = -\sqrt{y} & \text{si } x < 0 \Rightarrow dx = -\frac{1}{2\sqrt{y}} \end{cases}$$

$$f_Y(y) = f_X(-\sqrt{y}) \left| \frac{1}{2\sqrt{y}} \right| + f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} =$$

$$= \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} + \frac{1}{\sqrt{2\pi}} e^{-y/2} \frac{1}{2\sqrt{y}} =$$

$$= \frac{1}{\sqrt{2\pi}} y^{\frac{1}{2}-1} e^{-y/2} = \frac{1}{2^{1/2} \Gamma(1/2)} y^{\frac{1}{2}-1} e^{-y/2}$$

que es la función de densidad de una χ_1^2 .

Finalmente, para la suma, utilizamos el resultado similar demostrado para la distribución gamma.

DISTRIBUCIÓN t-student

Sea $T: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{R})$ absolutamente continua con función de densidad $f_T(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2) \sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$, $t \in \mathbb{R}$.

Decimos que T tiene distribución t-student de parámetro n (con n grados de libertad, se dice). (escribimos $T \sim t(n)$)

PROPOSICIÓN Sea $X \sim N(0, 1)$ e $Y \sim \chi_n^2$. Entonces

$$T = \frac{X}{\sqrt{Y/n}} \sim t(n)$$

$$X \sim N(\mu, \sigma)$$

$$Y \sim \chi_n^2$$

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}, \quad f_Y(y) = \begin{cases} \frac{1}{\Gamma(n/2) 2^{n/2}} y^{\frac{n}{2}-1} e^{-y/2} & y > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Dado que X e Y son independientes, la función de densidad conjunta del vector aleatorio (X, Y) es:

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \frac{1}{\Gamma(n/2) 2^{n/2}} y^{\frac{n}{2}-1} e^{-y/2}$$

Proponemos el siguiente cambio de variables: $T = \frac{X}{\sqrt{Y/n}}$, $U = Y$

$$(\Omega, \mathcal{A}, P) \xrightarrow{(X, Y)} (\mathbb{R}^2, \mathcal{R}^2) \xrightarrow{\gamma} (\mathbb{R}^2, \mathcal{R}^2)$$

$$\downarrow f_{XY} \quad \downarrow f_{TU}(t, u) = f_{XY}(x, y)$$

Siendo $(t, u) = g(x, y) = \left(\frac{x}{\sqrt{y/n}}, y \right)$. Calculamos $g^{-1}(t, u)$

$$\left. \begin{array}{l} t = \frac{x}{\sqrt{y/n}} \\ u = y \end{array} \right\} \left. \begin{array}{l} t^2 = \frac{x^2}{y/n} \\ u = y \end{array} \right\} \left. \begin{array}{l} t^2 \frac{y}{n} = x^2 \\ u = y \end{array} \right\} \left. \begin{array}{l} x = \sqrt{\frac{t^2 \cdot u}{n}} = t \sqrt{\frac{u}{n}} \\ y = u \end{array} \right\} \begin{array}{l} t \in \mathbb{R} \\ u \geq 0 \end{array}$$

Así, $g^{-1}(t, u) = \left(t \sqrt{\frac{u}{n}}, u \right)$, $|J| = \begin{vmatrix} \sqrt{\frac{u}{n}} & \frac{t}{2\sqrt{n}} \\ 0 & 1 \end{vmatrix} = \sqrt{\frac{u}{n}}$.

Entonces $f_{(t,u)}(t, u) = f_{(x,y)}(g^{-1}(t, u)) \cdot |J| = f_{(x,y)}\left(t \sqrt{\frac{u}{n}}, u\right) \cdot \sqrt{\frac{u}{n}} =$

$$= \frac{1}{\sqrt{2\pi}} e^{-t^2 u / n} \frac{u^{\frac{n}{2}-1} e^{-\frac{u}{2}} \sqrt{\frac{u}{n}}}{\Gamma(\frac{n}{2}) \cdot 2^{\frac{n}{2}}} = \frac{1}{\sqrt{2\pi} \Gamma(\frac{n}{2}) 2^{\frac{n}{2}} \sqrt{n}} u^{\frac{n+1}{2}-1} e^{-u \left(\frac{t^2}{2n} + \frac{1}{2} \right)}$$

Calculamos la función de densidad de T , será la marginal de T en f_{TV} .

$$f_T(t) = \int_0^\infty f_{(t,u)}(t, u) du = \frac{1}{\sqrt{2\pi} \Gamma(\frac{n}{2}) 2^{\frac{n}{2}} \sqrt{n}} \int_0^\infty u^{\frac{n+1}{2}-1} e^{-u \left(\frac{t^2}{2n} + \frac{1}{2} \right)} du =$$

Recordamos que $\int_0^\infty e^{-\alpha x} x^{\alpha-1} dx = \Gamma(\alpha) \cdot \alpha^{-\alpha}$

$$= \frac{1}{\sqrt{2\pi} \Gamma(\frac{n}{2}) 2^{\frac{n}{2}} \sqrt{n}} \Gamma\left(\frac{n+1}{2}\right) \cdot \left(\frac{t^2+n}{2n}\right)^{-\frac{n+1}{2}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n} \pi 2^{\frac{n}{2}}} \left(\frac{t^2}{n} + 1\right)^{-\frac{n+1}{2}}$$

$$= \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n} \pi} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad t \in \mathbb{R}$$

que es una distribución $t(n)$.

DISTRIBUCIÓN F-SNEDECOR

Consideremos $X \sim \chi_n^2$ e $Y \sim \chi_m^2$ independientes. Entonces la variable $F = \frac{X/n}{Y/m}$ sigue una distribución de probabilidad que denominaremos F-Snedecor de parámetros n y m (nos referiremos a ellos como grados de libertad)

FUNCIÓN DE DENSIDAD

La de X es: $f_X(x) = \frac{1}{\Gamma(n/2) 2^{n/2}} x^{\frac{n}{2}-1} e^{-x/2}$ si $x > 0$

y la de Y : $f_Y(y) = \frac{1}{\Gamma(m/2) 2^{m/2}} y^{\frac{m}{2}-1} e^{-y/2}$ si $y > 0$

la función de densidad (conjunta) del vector aleatorio (X, Y) es:

$$f_{XY}(x, y) = f_X(x) \cdot f_Y(y).$$

Para calcular la de $X/n/Y/m$, hagamos el siguiente cambio de variables:

$$u = \frac{x/n}{y/m} \quad v = y.$$

Así, $x = \frac{n}{m} uv$, $g^{-1}(u, v) = \left(\frac{n}{m} uv, v \right)$ $J = \begin{vmatrix} \frac{n}{m} v & \frac{n}{m} u \\ 0 & 1 \end{vmatrix} = \frac{n}{m} v$

$$f_{(u,v)}(u, v) = f_{(x,y)}\left(\frac{n}{m} uv, v\right) \frac{n}{m} v = f_X\left(\frac{n}{m} uv\right) f_Y(v) \frac{n}{m} v =$$

$$= \frac{\frac{n}{m} v}{\Gamma(n/2) \Gamma(m/2) 2^{\frac{n+m}{2}}} \left(\frac{n}{m} uv\right)^{\frac{n}{2}-1} v^{\frac{m}{2}-1} e^{-\left(\frac{n}{m} uv\right)/2 - v/2}$$

$$= \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2) 2^{\frac{n+m}{2}}} v^{\frac{n+m}{2}-1} u^{\frac{n}{2}-1} e^{-\left(\frac{n}{m} uv\right)/2 - v/2}$$

Necesitamos, finalmente, calcular la marginal de U :

$$f_U(u) = \int_0^\infty \frac{\left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2) 2^{\frac{n+m}{2}}} v^{\frac{n+m}{2}-1} u^{\frac{n}{2}-1} e^{-\left(\frac{n}{m} uv\right)/2 - v/2} dv =$$

$$= \frac{u^{\frac{n}{2}-1} \left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2) 2^{\frac{n+m}{2}}} \int_0^\infty v^{\frac{n+m}{2}-1} e^{-\left(\frac{\frac{n}{m}u+1}{2}\right)v} dv =$$

$$t = \left(\frac{n}{m}u + 1\right)v/2, \quad dt = \left(\frac{n}{m}u + 1\right)\frac{1}{2} dv$$

$$v = \frac{2t}{\frac{n}{m}u + 1}$$

$$= \frac{u^{\frac{n}{2}-1} \left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2) 2^{\frac{n+m}{2}}} \left(\frac{2}{\frac{n}{m}u + 1}\right)^{\frac{n+m}{2}-1} \frac{2}{\left(\frac{n}{m}u + 1\right)} \int_0^\infty t^{\frac{n+m}{2}-1} e^{-t} dt =$$

$$= \frac{u^{\frac{n}{2}-1} \left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2)} \frac{1}{\left(\frac{n}{m}u + 1\right)^{\frac{n+m}{2}}} \Gamma\left(\frac{n+m}{2}\right) = \frac{\Gamma\left(\frac{n+m}{2}\right) \left(\frac{n}{m}\right)^{\frac{n}{2}}}{\Gamma(n/2) \Gamma(m/2)} \frac{u^{\frac{n}{2}-1}}{\left(\frac{n}{m}u + 1\right)^{\frac{n+m}{2}}}$$

(b) $\Gamma[\alpha + 1] = \alpha\Gamma[\alpha]$

(c) $\Gamma[1/2] = \sqrt{\pi}$. Indicación: Aplica un cambio de variable teniendo en cuenta (2.14).

(d) Si $\beta > 0$ entonces

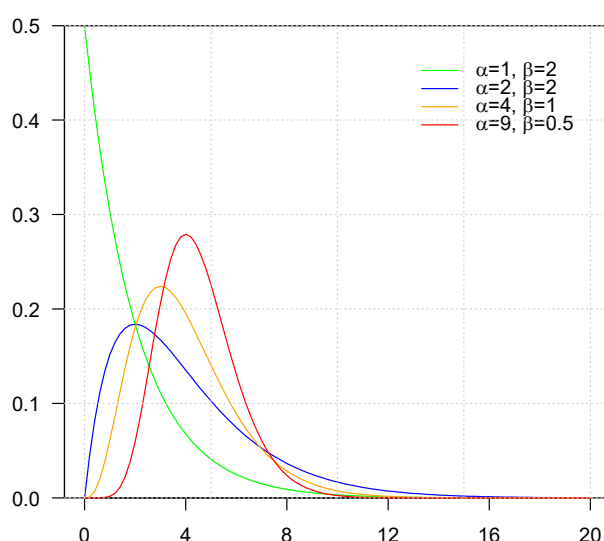
$$\int_0^\infty x^{\alpha-1} e^{-x/\beta} dx = \Gamma[\alpha]\beta^\alpha \quad (2.17)$$

Se dice entonces que $X \sim \Gamma(\alpha, \beta)$, con $\alpha, \beta > 0$, cuando admite como función de densidad

$$f(x) = (\Gamma[\alpha]\beta^\alpha)^{-1} I_{(0,+\infty)} x^{\alpha-1} e^{-x/\beta} \quad (2.18)$$

En la figura 2.7 se ilustran las densidades de algunas distribuciones de la familia. El parámetro α se denomina forma y β^{-1} , tasa. La función generatriz se obtiene sin dificultad en un entorno de 0 como $g_X(t) = (1 - \beta t)^{-\alpha}$, con lo que la función característica se deduce de manera trivial. Derivando g_X obtenemos la como media y varianza $\alpha\beta$ y $\alpha\beta^2$, respectivamente.

Figura 2.7: Algunas distribuciones tipo Gamma



Puede probarse haciendo uso de la función característica que, si $(X_i)_{i=1}^n$ son independientes y tales que $X_i \sim G(\alpha_i, \beta)$, entonces

$$\sum_{i=1}^n X_i \sim G\left(\sum_{i=1}^n \alpha_i, \beta\right) \quad (2.19)$$

2.2.4. Distribución exponencial

Se trata de una subfamilia de la familia Gamma. Concretamente, se dice que X sigue un modelo exponencial de parámetro β cuando $X \sim G(1, \beta)$. Su función de densidad es pues del tipo

$$f(x) = \beta^{-1} e^{-x/\beta} I_{(0,+\infty)} \quad (2.20)$$

Como ejemplo tenemos la distribución verde en la figura 2.7 es $Exp(2)$. La distribución exponencial se considera apropiada para modelizar un proceso de tiempo de espera con pérdida de memoria. Efectivamente, consideremos una variable $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}^+$ que mide el tiempo que hay que esperar hasta que ocurra cierto fenómeno raro. Supongamos que es absolutamente continua con densidad continua, siendo λ el tiempo medio esperado. Supongamos también que se verifica, para cada $s, t > 0$, la propiedad siguiente, denominada “pérdida de memoria”:

$$P(X > s + t | X > s) = P(X > t) \quad (2.21)$$

Eso implica que $P(X > s + t) = P(X > s)P(X > t)$ y, por lo tanto, que existe $\beta > 0$ tal que $\log(1 - F(x)) = -\beta^{-1}x$ para todo $x > 0$. Luego, $F'(x) = \beta^{-1}e^{-x/\beta}I_{\mathbb{R}^+}$ es la densidad de P^X , que pertenece pues a la familia exponencial con $\beta = \lambda$.

El proceso de Poisson: Supongamos que cada vez que el fenómeno ocurre ponemos a 0 el cronómetro y volvemos a medir el tiempo de espera hasta la siguiente ocurrencia, y así sucesivamente. Supongamos por último que los distintos tiempos de espera $(X_i)_i$ son iid. En ese caso, siguen distribuciones $G(1, \lambda)$. Teniendo en cuenta (2.19), la suma de los n primeros tiempos, $S_n = \sum_{i=1}^n X_i$, sigue un modelo $G(n, \lambda)$. Por lo tanto, dado un instante $t > 0$, la probabilidad de que se hayan registrado menos de n ocurrencias equivale a $P(S_n > t)$ que, considerando la función de densidad correspondiente e integrando iterativamente por partes resulta

$$\frac{1}{\Gamma[n]\lambda^n} \int_t^\infty x^{n-1} e^{-x/\lambda} dx = \sum_{k=0}^{n-1} e^{-t/\lambda} \frac{(t/\lambda)^k}{k!} \quad (2.22)$$

Se corresponde pues con la función de distribución de un modelo de Poisson de media t/λ , por lo que el número de fenómenos que ocurren hasta un tiempo determinado t con un tiempo medio de espera λ y con pérdida de memoria sigue un modelo $\mathcal{P}(t/\lambda)$. Puede ser el caso, por ejemplo, de un conteo de fotones que llegan a un fotodiodo en ambientes de baja luminosidad.

2.2.5. Distribución χ^2

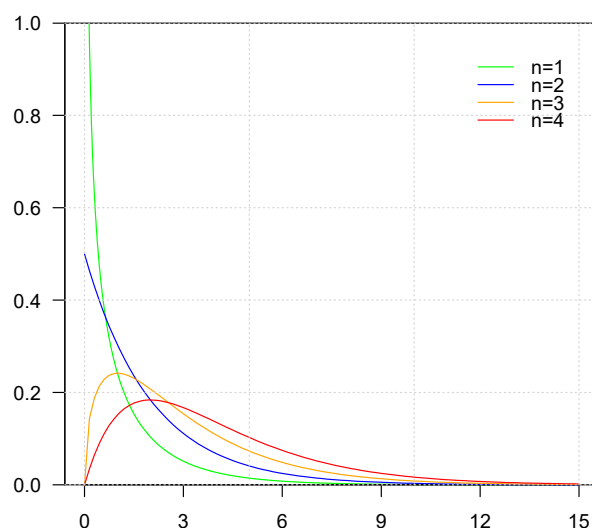
Decimos que Y sigue un modelo de distribución χ_n^2 con $n \geq 1$ cuando $Y \sim G(n/2, 2)$. En consecuencia, tanto sus parámetros principales como sus funciones de densidad, generatriz y característica se obtienen automáticamente, según figuran en la tabla 2.1. La densidad es, concretamente, la siguiente

$$f(x) = (\Gamma[n/2]2^{n/2})^{-1} x^{n/2-1} e^{-x/2} I_{(0,\infty)} \quad (2.23)$$

La media es n y la varianza $2n$. Se ilustran algunos ejemplos de la misma en la figura 2.8. El siguiente resultado aclara el interés de la distribución χ^2 .

Proposición 3. Si $X \sim N_n(0, Id)$ entonces $\|X\|^2 \sim \chi_n^2$.

Demostración. (i) Caso $n = 1$: Sea $B \in \mathcal{B}$ positivo y comprobemos que la función f definida en (2.23) con $n = 1$ es la densidad de la distribución de X^2 . Para ello,

Figura 2.8: Algunas distribuciones tipo χ_n^2


haremos uso del cambio de variable $x = +\sqrt{u}$:

$$\begin{aligned} P(X^2 \in B) &= 2P(X \in \sqrt{B}) = \frac{2}{\sqrt{2\pi}} \int_{\sqrt{B}} e^{-x^2/2} dx \\ &= (\Gamma[1/2]2^{1/2})^{-1} \int_B u^{-1/2} e^{-u/2} du = \int_B f(u) du \end{aligned}$$

- (ii) En general, $\|X\|^2 = \sum_{i=1}^n X_i^2$. Dado que X_1^2, \dots, X_n^2 son iid $G(1/2, 2)$ se sigue de (2.19) que $\sum_i X_i^2 \sim G(n/2, 2)$. □

De manera más general podemos decir que, dado un subespacio vectorial $V \subset \mathbb{R}^n$ y $X \sim N_n(0, Id)$,

$$\|P_V X\|^2 \sim \chi_{\dim V}^2 \quad (2.24)$$

El uso de la distribución χ^2 puede extenderse de la siguiente forma: dado $\sigma > 0$, decimos que $Y \sim \sigma^2 \chi_n^2$ cuando $\sigma^{-2}Y \sim \chi_n^2$. De esta forma, si $X \sim N_n(0, \sigma^2 Id)$ y $V \subset \mathbb{R}^n$, $\|P_V X\|^2 \sim \sigma^2 \chi_{\dim V}^2$. Así pues, la distribución χ^2 modeliza las normas euclídeas al cuadrado (sumas de cuadrados) de proyecciones ortogonales de normales multivariantes esféricas, que son el punto de partida del Modelo Lineal Normal.

Distribución χ^2 no central: La distribución anterior se denomina χ^2 central porque podemos concebir otra extensión más del modelo. Puede probarse que, si $X \sim N_n(\mu, \sigma^2 Id)$, la distribución de $\|X\|^2$ depende del vector (μ, σ^2) a través de $(\|\mu\|^2/\sigma^2, \sigma^2)$, de manera que podemos definir la distribución que induce como $\sigma^2 \chi_n^2(\delta)$, siendo $\delta = \|\mu\|^2/\sigma^2$. Se trata de una composición entre $\mathcal{P}(\delta)$ y la familia $\{\sigma^2 \chi_{2k+1}^2 : k \in \mathbb{N}\}$. Omitimos aquí sus funciones de densidad, generatriz y característica, pero indicamos que su media es $n\sigma^2 + \delta$. Por lo tanto, δ es valor de desplazamiento hacia la derecha, de gran trascendencia en los contrastes de hipótesis del Modelo Lineal. Ese mismo desplazamiento puede contemplarse para las distribuciones que estudiaremos a continuación.

2.2.6. Distribución F -Snedecor

Dadas $X, Y : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ independientes con distribuciones χ_m^2 y χ_n^2 , respectivamente, se define la distribución $F_{m,n}$ como aquélla que induce la transformación

$$F = \frac{X/m}{Y/n} \quad (2.25)$$

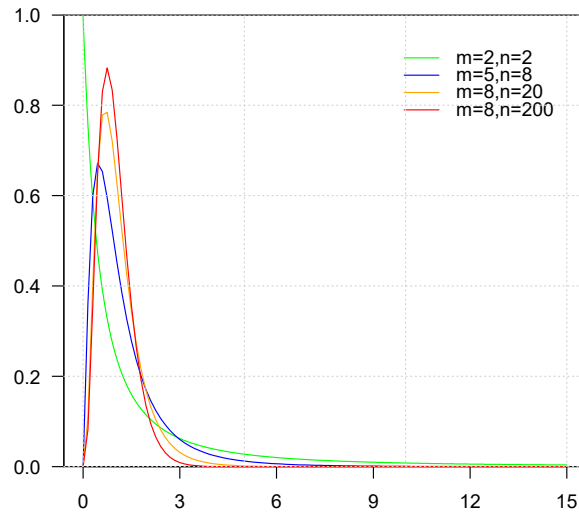
Cuando la distribución de X es no central de parámetro δ , se dice que F es no central con el mismo parámetro. La función de densidad de $F_{m,n}$ es la siguiente:

$$f_F(x) = \frac{\left(\frac{m}{n}\right)^{m/2} \Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right) \Gamma\left(\frac{n}{2}\right)} \cdot x^{\frac{m}{2}-1} \left(\frac{m}{n}x + 1\right)^{-(m+n)/2} I_{(0,\infty)}(x) \quad (2.26)$$

Indicación: Aplica sucesivamente (1.18) y (1.17).

En la figura 2.9 podemos encontrar algunos ejemplos. Es difícil diferenciarla de la distribución χ^2 . De hecho, en el capítulo 4 veremos que se da la convergencia $F_{m,n} \rightarrow \chi_m^2$ cuando $n \rightarrow \infty$.

Figura 2.9: Algunas distribuciones tipo $F_{m,n}$



Puede probarse que la media de una distribución $F_{m,n}$ es $n/(n-2)$ si $n > 2$ y su varianza es, si $n > 4$,

$$\sigma^2 = \frac{n^2}{m} \cdot \frac{2(m+n-2)}{(n-2)^2(n-4)} \quad (2.27)$$

Indicación: : integra $xf_F(x)$ considerando el cambio de variables $u = (1 + mx/n)^{-1}$ y teniendo presente (2.29) para $B(n/2-1, m/2-1)$, así como las propiedades de la función Γ . Con el mismo cambio de variable y teniendo presente de nuevo (2.29) para $B(n/2-2, m/2-2)$, se sigue que

$$E[X^2] = \left(\frac{n}{m}\right)^2 \frac{(m+2)m}{(n-2)(n-4)}, \quad (2.28)$$

a partir de lo cual obtenemos la varianza.

La función generatriz sólo se define en 0 y no podemos dar tampoco una expresión analítica de la función característica.

2.2.7. Distribución Beta

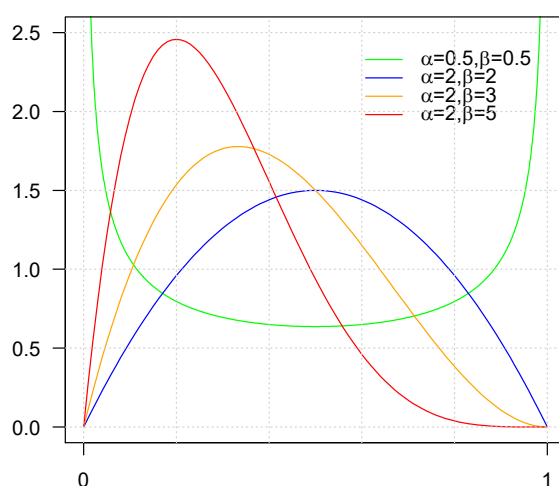
Dados $\alpha, \beta > 0$, se define $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$. Puede probarse que

$$B(\alpha, \beta) = \frac{\Gamma[\alpha]\Gamma[\beta]}{\Gamma[\alpha + \beta]} \quad (2.29)$$

Se dice entonces que $X \sim \text{Beta}(\alpha, \beta)$ cuando admite como función de densidad

$$f(x) = B(\alpha, \beta)^{-1} x^{\alpha-1} (1-x)^{\beta-1} I_{(1,1)}(x) \quad (2.30)$$

Figura 2.10: Algunas distribuciones tipo $\text{Beta}(\alpha, \beta)$



En la figura 2.10 podemos apreciar algunos ejemplos de distribuciones tipo Beta. Su función generatriz de momentos está bien definida en todo \mathbb{R} . Concretamente,

$$g_X(t) = \sum_{k=0}^{\infty} \frac{t^k}{k!} \frac{\Gamma[\alpha + k]\Gamma[\alpha + \beta]}{\Gamma[\alpha + \beta + k]\Gamma[\alpha]}, \quad t \in \mathbb{R} \quad (2.31)$$

Indicación: : desarrolla e^{tx} como serie de potencias en $g_X(t)$.

A partir de (2.31) podemos obtener trivialmente los momentos de orden $k = 1$ y $k = 2$, de los que deducimos la media y varianza, que valen respectivamente

$$\mu = \frac{\alpha}{\alpha + \beta}, \quad \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

Veamos algunas propiedades de la distribución Beta:

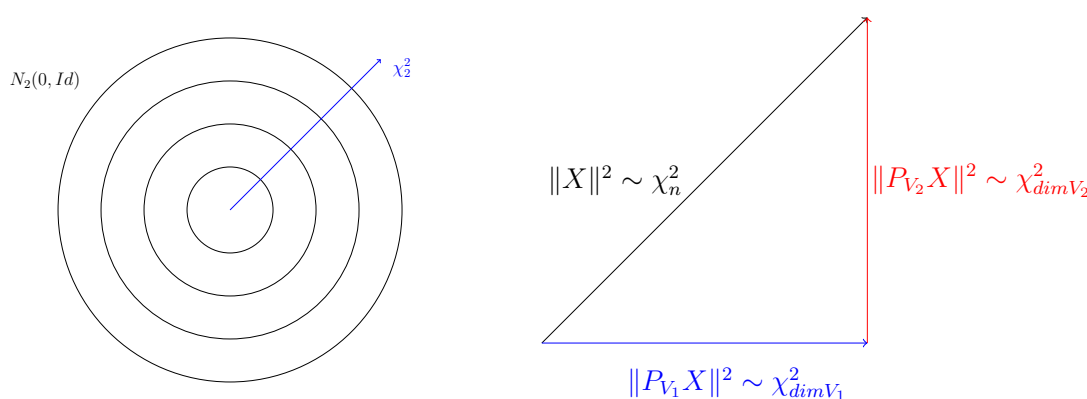
(a) $X \sim \text{Beta}(\alpha, \beta) \Leftrightarrow 1 - X \sim \text{Beta}(\beta, \alpha)$

(b) $Beta(1, 1) = U_{[0,1]}$

(c) $X \sim F(m, n)$ sii $\left(1 + \frac{m}{n}X\right)^{-1} \sim Beta(n/2, m/2)$. Indicación: considera en (2.26) el cambio de variable $u = \frac{n}{mx+n}$.

La propiedad (c) es muy importante pues permite interpretar la distribución Beta, al igual que la F , en términos de cocientes entre χ^2 independientes, como se ilustra en la figura 2.11. Concretamente, mientras que F se obtiene como el cociente entre los cuadrados de los catetos azul y rojo, divididos por las dimensiones de los espacios correspondientes, Beta se obtiene como cociente entre los cuadrados del cateto azul y la hipotenusa. Se trata pues de la distribución del coeficiente de correlación lineal múltiple R^2 bajo las condiciones del Modelo Lineal Normal. No obstante y por esa misma razón, la distribución Beta suele quedar relegada por F .

Figura 2.11: Ilustración de las distribuciones χ^2 , F , Beta y t



2.2.8. Distribución t de Student

Esta distribución fue introducida oficialmente en el año 1908 por William Sealy Gosset, que utilizó el sinónimo “Student”, aunque había sido ya estudiada previamente por Hemert. Por definición, la distribución t_n es la inducida por $Z/\sqrt{Y/n}$, donde Z e Y son independientes y tales que $Z \sim N(0, 1)$ y $Y \sim \chi_n^2$. Aplicando sucesivamente los cambios de variables (1.18), (1.19), (1.17) y, por último $u(x) = 2^{-1}(y^2 + n)x^2$, obtenemos entonces la correspondiente función de densidad

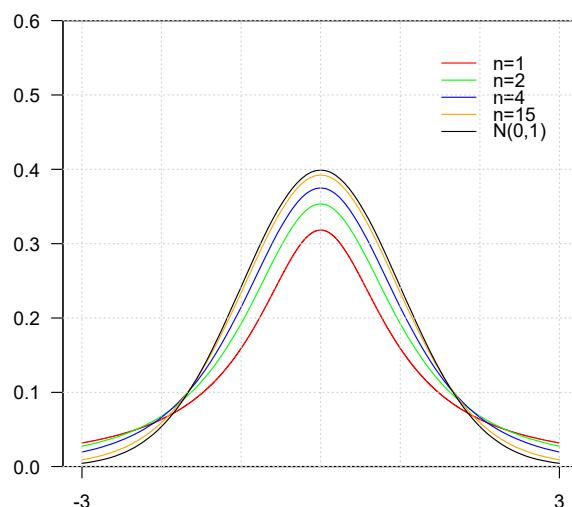
$$f(x) = \frac{\Gamma\left[\frac{n+1}{2}\right]}{\sqrt{n\pi}\Gamma\left[\frac{n}{2}\right]} \cdot \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}, \quad x \in \mathbb{R} \quad (2.32)$$

En la figura 2.12 podemos apreciar algunos ejemplos de distribuciones tipo t , con $N(0, 1)$ en negro. Al igual que la distribución F , carece de función generatriz y no podemos aportar una expresión analítica de su función característica. Si $n > 1$ la media existe y es por lo tanto nula, pues se trata de una distribución simétrica. Si $n > 2$, su momento de orden 2 (varianza en este caso) resulta ser

$$\sigma^2 = n/(n - 2)$$

Indicación: calcula la integral de $x^2 f(x)$ considerando el cambio de variable $x = \sqrt{n(1-u)/u}$ y teniendo presente la función $B[(n-2)/2, 3/2]$.

Figura 2.12: Algunas distribuciones tipo t_n



En la figura 2.12 podemos apreciar una convergencia de t_n a $N(0,1)$ a medida que $n \rightarrow \infty$, que puede considerarse un caso particular de la convergencia de $F_{m,n} \rightarrow \chi^2$. Efectivamente y según las propias definiciones, se verifica que, si $X \sim t_n$, $X^2 \sim F_{1,n}$; igualmente, si $Z \sim N(0,1)$, entonces $Z^2 \sim \chi_1^2$. En ese sentido, la distribución t_n o, mejor dicho, t_n^2 , tiene cabida en la figura 2.11 como distribución tipo F cuando $\dim V_1 = 1$. En definitiva, las distribuciones F , $Beta$ y t pueden entenderse como cocientes entre χ^2 independientes que, a su vez, resultan de calcular la norma euclídea al cuadrado (suma de cuadrados) a partir de una distribución normal multivariante esférica.

2.2.9. Distribución log-normal

Dada $X \sim N(\mu, \sigma)$, $\text{lnorm}(\mu, \sigma)$ es por definición la distribución en \mathbb{R}^+ inducida por e^X , cuya densidad es, según (1.20),

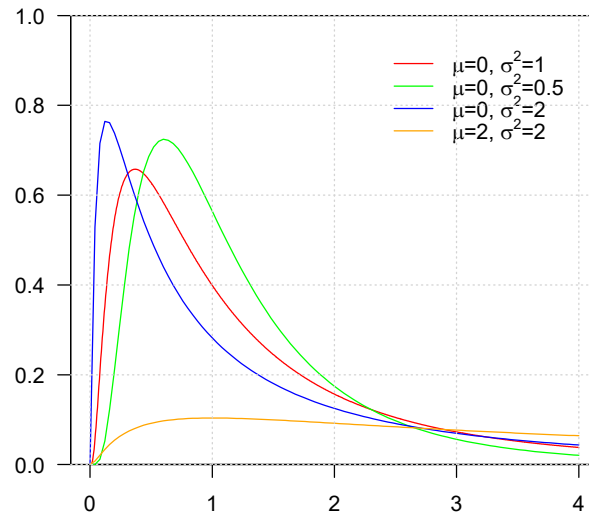
$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\log x - \mu)^2}{2\sigma^2}\right\}, \quad x > 0 \quad (2.33)$$

En la figura 2.13 podemos apreciar algunos ejemplos de distribuciones tipo lnormal. Su función generatriz no está definida en un entorno de 0 y no podemos aportar una expresión analítica de su función característica. Utilizando el cambio de variables $x = \exp\{\mu + \sigma u\}$ obtenemos la media y varianza que figuran en la tabla 2.1

2.2.10. Distribución de Cauchy

Aunque se trata de una familia de distribuciones de interés en Física, destacamos aquí la denominada estándar que se trata realmente de la distribución t_1 y que admite por lo

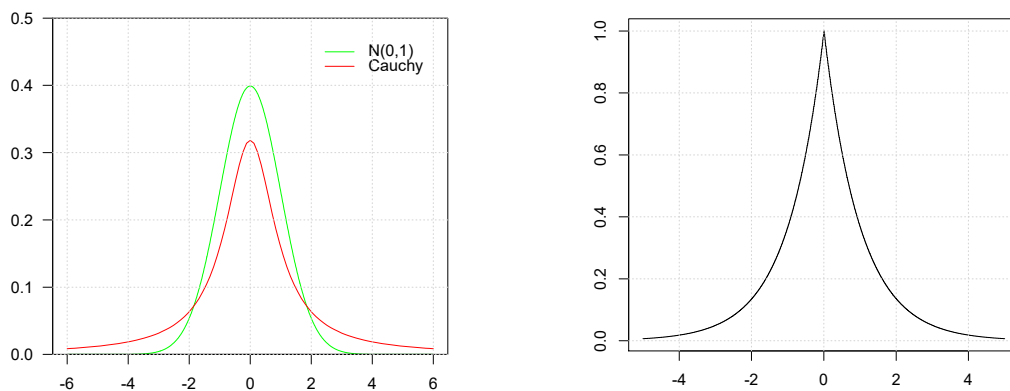
Figura 2.13: Algunas distribuciones tipo Inormal



tanto la función de densidad

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad x \in \mathbb{R} \quad (2.34)$$

Carece de media, en consecuencia, también de varianza y función generatriz definida en un entorno de 0. Su función característica se calcula mediante integración compleja. En la figura 2.14 podemos ver la función de densidad (izquierda), comparada con la de $N(0, 1)$, y característica (derecha).

 Figura 2.14: Funciones de densidad y característica de Cauchy = t_1


DISTRIBUCIÓN BETA

PROPOSICIÓN

Sean $X_j \sim \Gamma(\alpha_j, \beta)$ $j=1 \dots n$ independientes.
Entonces $S_n = \sum_{j=1}^n X_j \sim \Gamma\left(\sum_{j=1}^n \alpha_j, \beta\right)$

Demostación. Si $\varphi_{X_j}(t)$ es la función característica de X_j , $j=1 \dots n$, al ser las X_j independientes, la función característica de S_n , $\varphi_{S_n}(t)$ es $\prod_{j=1}^n \varphi_{X_j}(t)$, que es $\prod_{j=1}^n \frac{1}{(1+it\beta)^{\alpha_j}} = \frac{1}{(1+it\beta)^{\sum_{j=1}^n \alpha_j}}$ que es la función característica de una variable con distribución $\Gamma\left(\sum_{j=1}^n \alpha_j, \beta\right)$. Como la función característica caracteriza a la distribución, $S_n \sim \Gamma\left(\sum_{j=1}^n \alpha_j, \beta\right)$, C.Q.D.

Demostremoslo de otra forma: ($n=2$)

$$(\Omega, \mathcal{A}, P) \xrightarrow{(X_1, X_2)} (\mathbb{R}^2, \mathcal{B}^2) \xrightarrow{g} (\mathbb{R}^2, \mathcal{B}^2)$$

$$(y_1, y_2) = g(x_1, x_2) \quad \text{donde} \quad (y_1, y_2) = g(x_1, x_2) = (x_1 + x_2, x_2)$$

$$\text{Así, } g^{-1}(y_1, y_2) = (y_1 - y_2, y_2)$$

$$J_{g^{-1}}(y_1, y_2) = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$$

La función de densidad conjunta de (Y_1, Y_2) es:

$$f_{(Y_1, Y_2)}(y_1, y_2) = \int_{(x_1, x_2)} (g^{-1}(y_1, y_2)) J_{g^{-1}}(y_1, y_2) = f_{X_1}(y_1 - y_2) f_{X_2}(y_2)$$

Ahora calculamos la densidad marginal de Y_1 y tenemos la de $X_1 + X_2$

Dado que $X_1, X_2 \in (0, \infty) \Rightarrow Y_1 \in (0, \infty)$.

Sea $y_1 \in (0, \infty)$,

$$f_{Y_1}(y_1) = \int_0^{y_1} \frac{1}{\Gamma(\alpha_1)\beta^{\alpha_1}} (y_1 - y_2)^{\alpha_1 - 1} e^{-(y_1 - y_2)/\beta} \frac{1}{\Gamma(\alpha_2)\beta^{\alpha_2}} y_2^{\alpha_2 - 1} e^{-y_2/\beta} dy_2 =$$

$$= \frac{e^{-y_1/\beta}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \int_0^\infty (y_1 - y_2)^{\alpha_1-1} y_2^{\alpha_2-1} dy_2 \stackrel{\substack{\uparrow \\ u = \frac{y_2}{y_1} \quad y_2 = y_1 u \quad dy_2 = y_1 du}}{=}$$

$$= \frac{e^{-y_1/\beta}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \int_0^1 (y_1 - y_1 u)^{\alpha_1-1} (y_1 u)^{\alpha_2-1} y_1 du = \frac{e^{-y_1/\beta} y_1^{\alpha_1-1} y_1^{\alpha_2-1} y_1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \int_0^1 (1-u)^{\alpha_1-1} u^{\alpha_2-1} du$$

$\beta(\alpha_1, \alpha_2)$

$$= \frac{e^{-y_1/\beta} y_1^{\alpha_1+\alpha_2-1}}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \cdot \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} = \frac{y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta}}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}}$$

que es la función de densidad de una $\Gamma(\alpha_1+\alpha_2, \beta)$.

Principales modelos de distribución

Modelo	Probabilidad/Densidad	(μ, σ^2)	Generatriz	Característica
δ_a	$P(X = a) = 1$	$(a, 0)$	e^{at}	$\cos at + i \sin at$
$B(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$	$(np, np(1-p))$	$(1-p + p\mathbf{e}^t)^n$	$(1-p + p\mathbf{e}^{it})^n$
$M(n; (p_j)_{j=1}^k)$	$\frac{n!}{\prod_{j=1}^k i_j!} \prod_{j=1}^k p_j^{i_j}$	$(np_j, np_j(1-p_j)), -np_i p_j$	$(p_k + \sum_{j=1}^{k-1} p_j \mathbf{e}^{t_j})^n$	$(p_k + \sum_{j=1}^{k-1} p_j \mathbf{e}^{it_j})^n$
$\mathcal{P}(\lambda)$	$e^{-\lambda} \frac{\lambda^k}{k!}$	(λ, λ)	$\exp\{\lambda(\mathbf{e}^t - 1)\}$	$\exp\{\lambda(\mathbf{e}^{it} - 1)\}$
$BN(r, p)$	$\binom{k-1}{r-1} p^r (1-p)^{k-r}$	$\left(\frac{r}{p}, \frac{r(1-p)}{p^2}\right)$	$\frac{(p\mathbf{e}^t)^r}{[1-(1-p)\mathbf{e}^t]^r}$	$\frac{(p\mathbf{e}^{it})^r}{[1-(1-p)\mathbf{e}^{it}]^r}$
$HG(N, M, n)$	$\frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}$	$\left(n \frac{M}{N}, \frac{nM(N-n)(N-M)}{N^2(N-1)}\right)$	$g_X(t)$	$\varphi_X(t)$
$N(\mu, \sigma)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2} \frac{(x-\mu)^2}{\sigma^2}}$	(μ, σ^2)	$\mathbf{e}^{\mu t + \sigma^2 t^2 / 2}$	$\mathbf{e}^{i\mu t - \sigma^2 t^2 / 2}$
$N_n(\mu, \Sigma)$	$\frac{1}{(2\pi \Sigma)^{\frac{n}{2}}} e^{-\frac{1}{2} (x-\mu)' \Sigma^{-1} (x-\mu)}$	(μ, Σ)	$\mathbf{e}^{\mu' t + t' \Sigma t / 2}$	$\mathbf{e}^{i\mu' t - t' \Sigma t / 2}$
$U[-a, a]$	$\frac{1}{2a} I_{[-a, a]}$	$\left(0, \frac{a^2}{3}\right)$	$\frac{1}{2at} (\mathbf{e}^{at} - \mathbf{e}^{-at})$	$\frac{1}{at} \sin at$
$\Gamma(\alpha, \beta)$	$(\Gamma[\alpha]\beta^\alpha)^{-1} x^{\alpha-1} \mathbf{e}^{-x/\beta} I_{\mathbb{R}^+}$	$(\alpha\beta, \alpha\beta^2)$	$(1-\beta t)^{-\alpha}$	$(1-i\beta t)^{-\alpha}$
$Exp(\beta)$	$\beta^{-1} \mathbf{e}^{-x/\beta} I_{\mathbb{R}^+}$	(β, β^2)	$(1-\beta t)^{-1}$	$(1-i\beta t)^{-1}$
χ_n^2	$\frac{1}{\Gamma[n/2]2^{n/2}} x^{n/2-1} \mathbf{e}^{-x/2} I_{\mathbb{R}^+}$	$(n, 2n)$	$(1-2t)^{-n/2}$	$(1-2it)^{-n/2}$
$F_{m,n}$	$\kappa x^{\frac{m}{2}-1} \left(\frac{m}{n}x + 1\right)^{-\frac{m+n}{2}} I_{\mathbb{R}^+}$	$\left(\frac{n}{n-2}, \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}\right)$	-	$\varphi_X(t)$
$Beta(\alpha, \beta)$	$\frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}$	$\left(\frac{\alpha}{\alpha+\beta}, \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}\right)$	$\sum_k \frac{t^k}{k!} \frac{\Gamma[\alpha+k]\Gamma[\alpha+\beta]}{\Gamma[\alpha+\beta+k]\Gamma[\alpha]}$	$\sum_k \frac{(it)^k}{k!} \frac{\Gamma[\alpha+k]\Gamma[\alpha+\beta]}{\Gamma[\alpha+\beta+k]\Gamma[\alpha]}$
t_n	$\frac{\Gamma[\frac{n+1}{2}]}{\sqrt{n\pi}\Gamma[\frac{n}{2}]} \cdot \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}$	$\left(0, \frac{n}{n-2}\right)$	-	$\varphi_X(t)$
$\text{lognorm}(\mu, \sigma)$	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}} I_{\mathbb{R}^+}$	$\left(\mathbf{e}^{\mu + \frac{\sigma^2}{2}}, \mathbf{e}^{2(\mu + \sigma^2)} - \mathbf{e}^{2\mu + \sigma^2}\right)$	-	$\varphi_X(t)$
Cauchy	$\frac{1}{\pi(1+x^2)}$	-	-	$\mathbf{e}^{- t }$

Tabla 2.1: Principales distribuciones.

Mediante este [enlace](#) puedes acceder a gráfico interactivos de diferentes distribuciones continuas.

3. CONVERGENCIAS

En esta segunda parte enunciaremos e intentaremos probar los dos resultados fundamentales del Cálculo de Probabilidades: la Ley de los Grandes Números (LGN) y el Teorema Central del Límite (TCL), que justifican la axiomática de Kolmogorov. Pero, dado que se trata de propiedades límite o asintóticas, se precisa un capítulo previo en el que podamos familiarizarnos con elementos básicos de la Teoría Asintótica.

3.1. Resultados básicos

Esta primera sección está dedicada a enunciar y probar dos resultados que ya se adelantaron en el capítulo preliminar: los teoremas de la Convergencia Monótona (TCM) y de la Convergencia Dominada (TCD), partiendo de algunas definiciones y lemas preliminares. En todo momento partimos de un espacio medible compuesto por un conjunto Ω , una σ -álgebra \mathcal{A} de sucesos de $\mathcal{P}(\Omega)$ y una medida μ σ -finita sobre \mathcal{A} .

3.1.1. Límites superior e inferior

Dada una sucesión de sucesos $(A_n)_n$, las sucesiones de sucesos (medibles) $(B_n)_n$ y $(C_n)_n$ definidas mediante $B_n = \cup_{k \geq n} A_k$ y $C_n = \cap_{k \geq n} A_k$ son, respectivamente, decreciente y creciente, convergiendo por tanto a los sucesos (medibles) $\inf_n B_n$ y $\sup_n C_n$. Se definen como límites superior e inferior sus respectivos límites, es decir:

$$\limsup A_n = \bigcap_{n \in \mathbb{N}} \bigcup_{k \geq n} A_k, \quad \liminf A_n = \bigcup_{n \in \mathbb{N}} \bigcap_{k \geq n} A_k \quad (3.1)$$

Es decir, $\limsup A_n$ es el suceso compuesto por los elementos incluidos en infinitos sucesos de $(A_n)_n$ (podría ser \emptyset), mientras que $\liminf A_n$ es el suceso compuesto por los elementos incluidos en todos los sucesos de $(A_n)_n$ salvo, acaso, un número finito de ellos. En consecuencia, $\liminf A_n \subset \limsup A_n$.

Traslademos estas definiciones a una sucesión $(a_n)_n$ de números reales. Las sucesiones $b_n = \sup_{k \geq n} a_k$ y $c_n = \inf_{k \geq n} a_k$ son decreciente y creciente, por lo que tienden a $\inf_n b_n$ y $\sup_n c_n$. Se define entonces

$$\limsup a_n = \inf_n \sup_{k \geq n} a_k, \quad \liminf a_n = \sup_n \inf_{k \geq n} a_k \quad (3.2)$$

El límite superior es por lo tanto el mayor punto de acumulación de la sucesión en $\overline{\mathbb{R}}$. Equivalentemente, lo podemos caracterizar mediante $[x > \sup a_n]$ sii $[a_n > x]$ para a lo sumo finitos términos de la sucesión]. Análogamente, el límite inferior es el menor punto de acumulación de la sucesión en $\overline{\mathbb{R}}$. Equivalentemente, se verifica que $[x < \liminf a_n]$ sii $[a_n < x]$ para a lo sumo finitos términos de la sucesión]. En consecuencia, se verifica que $a_n \rightarrow a$ sii $\limsup a_n = \liminf a_n = a$. Efectivamente, dado $\varepsilon > 0$ se sigue de las caracterizaciones anteriores que $|a_n - a| > \varepsilon$ para, a lo sumo, finitos términos de la sucesión.

Si $(f_n)_n$ es una sucesión de funciones medibles reales podemos extender los conceptos anteriores definiendo puntualmente (3.2) en cada $\omega \in \Omega$. Las funciones resultantes serán medibles dado que tanto el supremo como el ínfimo conservan la medibilidad. De esta forma, se verifica

$$x > \limsup f_n(\omega) \text{ sii } \omega \in \liminf \{f_n < x\} \quad (3.3)$$

Podemos obtener una proposición análoga para el límite inferior.

Lemas de Borel-Cantelli: Los siguientes lemas se enuncian en un espacio probabilístico ($\mu = P$) y son clásicos de esta materia. El primero de ellos se demuestra de manera trivial. La demostración del segundo podemos encontrarla en Rincón (2007).

Lema 5. Si $\sum_{n=1}^{\infty} P(A_n) < \infty$, entonces se verifica que $P(\limsup A_n) = 0$

Lema 6. Si A_1, A_2, \dots son independientes y $\sum_{n=1}^{\infty} P(A_n) = \infty$, entonces se verifica que $P(\limsup A_n) = 1$

3.1.2. Teoremas de convergencia

A continuación, vamos a realizar un pequeño esfuerzo por probar dos teoremas fundamentales que ya se avanzaron en el capítulo preliminar, los de la convergencia dominada y monótona, que establecen condiciones bajo las cuales la integral del límite es el límite de las integrales. Empezamos por un lema de demostración trivial.

Lema 7. Si $s : (\Omega, \mathcal{A}, \mu) \rightarrow \mathbb{R}$ es una función simple y $(B_n)_n \subset \mathcal{A}$ tal que $B_n \nearrow \Omega$, entonces $\int_{B_n} s \, d\mu \rightarrow \int s \, d\mu$

Teorema 8 (TCM). Si $(X_n)_n \geq 0$ es una sucesión monótona creciente ($X_{n+1} \geq X_n$) de variables reales definidas en $(\Omega, \mathcal{A}, \mu)$ y $X = \sup_n X_n$, entonces $\int X_n \, d\mu \rightarrow \int X \, d\mu$.

Demostración. Nótese que el límite X existe y es medible por ser supremo de funciones medibles y su integral existe pudiendo ser ∞ . También existe el límite de $\int X_n \, d\mu$, que es $\sup_n \int X_n \, d\mu$. Se trata pues de probar que, en estas condiciones, la integral del supremo es menor que el supremo de las integrales (la otra desigualdad es obvia).

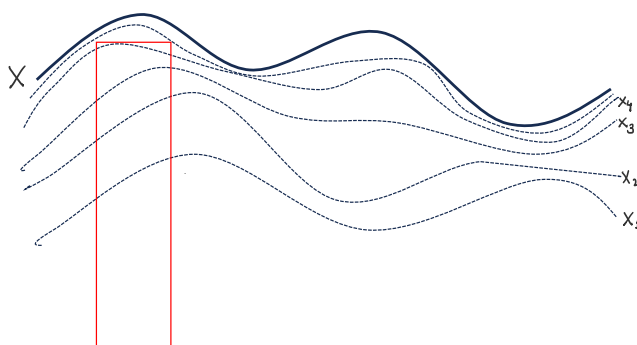
Sea s cualquier función simple tal que $s \leq X$ y $\kappa \in (0, 1)$. Definamos para cada n $B_n = \{\omega \in \Omega : \kappa s(\omega) \leq X_n(\omega)\}$. Dado que $X_n \nearrow X$ y $\kappa s(\omega) < s(\omega)$ para todo ω , se verifica que $B_n \nearrow \Omega$. Por el lema anterior, se verifica para todo $\kappa \in (0, 1)$

$$\kappa \int s \, d\mu = \lim_n \int_{B_n} \kappa s \, d\mu \leq \lim_n \int X_n \, d\mu$$

Por continuidad, la igualdad se verifica también para $\kappa = 1$, i.e., $\int s \, d\mu \leq \lim_n \int X_n \, d\mu$, lo cual ocurre para toda función simple inferior a X y, en consecuencia, para el supremo que es por definición $\int X \, d\mu$. \square

En la figura 3.1 se ilustra la demostración. Veamos dos consecuencias inmediatas. Concretamente, la sucesión $(X_n)_n$ acaba rebasando cualquier función indicador estrictamente inferior a X .

Figura 3.1: Ilustración del TCM

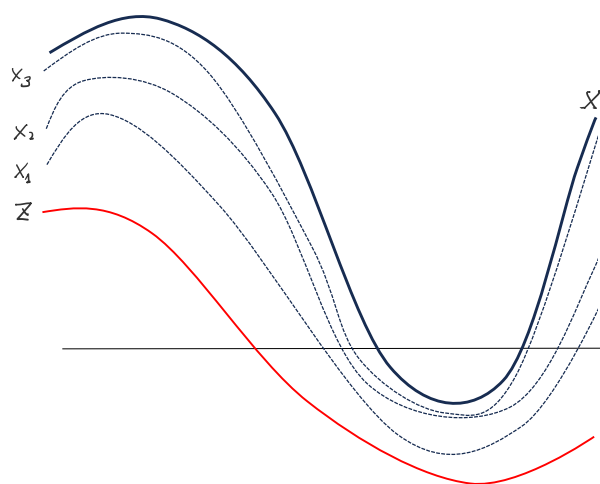


Corolario 9. Si $X_n \nearrow X$ (no necesariamente positivos) y $X_1 \geq Z$ tal que $\int Z \, d\mu > -\infty$, entonces $\int X_n \, d\mu \rightarrow \int X \, d\mu$.

Demostración. Descompongamos $X_n = X_n^+ - X_n^-$ y $X = X^+ - X^-$. En ese caso, $X_n^+ \nearrow X^+$ y $X_n^- \searrow X^-$, i.e., $Z^- - X_n^- \nearrow Z^- - X^-$. Por TCM se verifica que $\int X_n^+ \, d\mu \rightarrow \int X^+ \, d\mu$, por una parte. Por otra, se tiene que $\int (Z^- - X_n^-) \, d\mu \rightarrow \int (Z^- - X^-) \, d\mu$. Como $\int Z^- \, d\mu < +\infty$, podemos descomponer $\int (Z^- - X_n^-) \, d\mu = \int Z^- \, d\mu - \int X_n^- \, d\mu$ y $\int (Z^- - X^-) \, d\mu = \int Z^- \, d\mu - \int X^- \, d\mu$. En consecuencia, $\int X_n^- \, d\mu \rightarrow \int X^- \, d\mu$, lo cual concluye la prueba.

□

Figura 3.2: Ilustración corolario 9 TCM



Corolario 10. Si $X_n \searrow X$ y $X_1 \leq Z$ tal que $\int Z d\mu < \infty$, entonces $\int X_n d\mu \rightarrow \int X d\mu$.

Demostración. Razonando con $(-X_n) \nearrow -X$ con $-X_1 \geq -Z$ y $\int(-Z) d\mu > -\infty$, se deduce que $\int -X_n d\mu \rightarrow \int -X d\mu$, lo cual concluye la prueba. \square

El siguiente lema se precisa para la demostración de TCD.

Lema 11 (Fatou). (i) Si $(X_n)_n \geq Z$ y $\int Z d\mu > -\infty$, $\liminf \int X_n d\mu \geq \int \liminf X_n d\mu$
(ii) Si $(X_n)_n \leq Z$ y $\int Z d\mu < \infty$, $\limsup \int X_n d\mu \leq \int \limsup X_n d\mu$

Demostración. Veamos el apartado (i): si se denota $Z_n = \inf_{k \geq n} X_k \geq Z$ entonces, por el corolario 9, $\int \liminf X_n d\mu = \lim \int Z_n d\mu \leq \lim \int \inf X_n d\mu$. El apartado (ii) se prueba de manera análoga. \square

Teorema 12 (TCD). Si $X_n \rightarrow X$ y $\exists Z > 0$ tal que $\int Z d\mu < +\infty$ y $|X_n| < Z$ para todo n , entonces $\int X_n d\mu \rightarrow \int X d\mu$

Demostración. Aplicando a Z y $-Z$, respectivamente, los apartados (ii) (i) del lema de Fatou, se tiene que

$$\limsup \int X_n d\mu \leq \int \limsup X_n d\mu = \int \liminf X_n d\mu \leq \liminf \int X_n d\mu$$

Como la desigualdad contraria es obvia, estamos ante una igualdad, lo que prueba la tesis. \square

Estos resultados fueron utilizados en capítulos anteriores y serán utilizados con frecuencia en lo que sigue.

3.2. Tipos de convergencias

En lo que sigue nos situaremos en un marco exclusivamente probabilístico, es decir, partiendo de un espacio del tipo (Ω, \mathcal{A}, P) , donde consideraremos una sucesión $(X_n)_n$ de variables aleatorias reales y otra variable X que será el límite en distintos sentidos. De hecho y más allá de la convergencia puntual, que carece de interés en este contexto, definiremos cuatro tipos de convergencia y estableceremos algunas conexiones entre las mismas.

Aunque no es formalmente necesario, conviene entender, tal y como se explicó en el capítulo preliminar, (Ω, \mathcal{A}, P) como $(\Omega_0^{\mathbb{N}}, \mathcal{A}_0^{\mathbb{N}}, P_0^{\mathbb{N}})$. Es decir, cada elemento $\omega \in \Omega$ se entiende como una sucesión infinita de réplicas independientes del fenómeno aleatorio que rige P_0 . Entonces, dada una variable X definida sobre $(\Omega_0, \mathcal{A}_0, P_0)$, podemos considerar, para cada $i \in \mathbb{N}$, la variable X_i definida en (Ω, \mathcal{A}, P) como $X \circ \pi_i$, es decir, la que asigna a la réplica i -ésima su valor según X . Por ejemplo, si el fenómeno aleatorio estudiado es el lanzamiento de un dado y la variable X es la puntuación obtenida, X_i es la puntuación del i -ésimo lanzamiento. La secuencia $(X_i)_{1 \leq i \leq n}$ se denomina, según definimos en el capítulo preliminar, muestra aleatoria simple de tamaño n de P_0^X .

Un ejemplo muy importante de sucesión de variables aleatorias es $(\bar{X}_n)_n$, donde $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, es decir, la puntuación media tras n lanzamientos, denominada media muestral. Realmente, podemos redefinir aquí cualquier parámetro descriptivo como una

variable sobre (Ω, \mathcal{A}, P) para cada valor de n y estamos muy interesados en establecer una conexión entre los mismos y los análogos probabilísticos definidos para X , como la media μ . Nótese que cualquier parámetro probabilístico es un número que puede interpretarse en esta teoría como una variable aleatoria constante.

3.2.1. Convergencia casi seguro

Se dice que $X_n \rightarrow X$ c.s. cuando $P(\lim_{n \rightarrow \infty} X_n = X) = 1$. Nótese que tiene sentido formal pues

$$\{\omega \in \Omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = \{\omega \in \Omega : \liminf_{n \rightarrow \infty} X_n(\omega) = \limsup_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}$$

que es un suceso medible. Equivalentemente, decimos que se da la convergencia c.s. cuando

$$\exists N : P(N) = 0 \text{ y } \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \quad \forall \omega \in \overline{N} \quad (3.4)$$

3.2.2. Convergencia en L^2

Aunque cabe definir la coinvergencia en L^p con $p \geq 1$ nos centramos aquí únicamente en el caso $p = 2$ que nos sitúa en la métrica habitual (Euclídea) basada en el cuadrado de la diferencia. Dados $(X_n)_n$, X de cuadrado integrable decimos que $X_n \rightarrow X$ en L^2 cuando

$$\int (X_n - X)^2 dP \rightarrow 0 \quad (3.5)$$

Veamos primeramente que las convergencias c.s. y L^2 son distintas. Por ejemplo, sea P la distribución uniforme en $[0, 1]$ y $(X_n)_n$ la sucesión de variables definida mediante

$$X_n(\omega) = \begin{cases} \sqrt{n} & \text{si } 0 \leq \omega \leq \frac{1}{n} \\ 1 & \text{si } \frac{1}{n} < \omega \leq 1 \end{cases} \quad (3.6)$$

$(X_n)_n$ converge c.s. a 0 pero no converge en L^2 , pues $\int (X_n - X)^2 dP = 1$ para todo n . Recíprocamente en el ejemplo de la figura 3.3 podemos apreciar una sucesión que no baja de la línea discontinua y que, por lo tanto, no puede ser convergente a 0 c.s, aunque sí lo es en L^2 . No obstante, se sigue del TCD que la convergencia c.s. implica convergencia en L^2 si existe Z de cuadrado integrable tal que $|X_n| < Z$ para todo n , cosa que no sucede en (3.6).

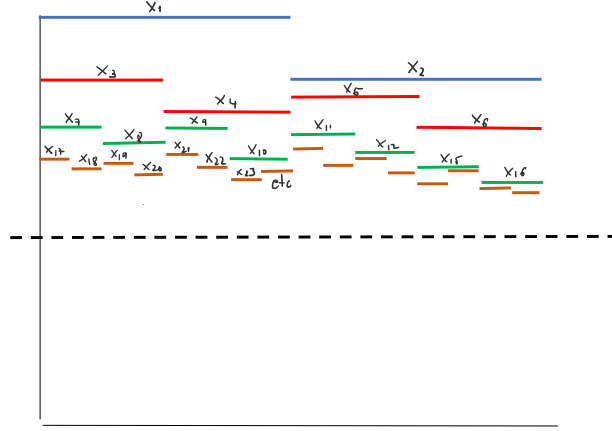
3.2.3. Convergencia en probabilidad

Decimos que $X_n \rightarrow X$ en probabilidad P cuando se verifica que

$$\forall \varepsilon > 0 \quad P(|X_n - X| > \varepsilon) \rightarrow 0 \quad (3.7)$$

El contraejemplo (3.6) muestra una sucesión que converge en probabilidad pero no en L^2 , mientras que el contraejemplo de la figura 3.3 muestra una sucesión que converge en probabilidad pero no c.s. Podemos ir más lejos y afirmar que la convergencia en probabilidad es estrictamente más débil que las convergencias anteriores.

Figura 3.3: Contraejemplo convergencia L^2 no c.s.



Proposición 4. Si $X_n \xrightarrow{c.s.} X$ entonces $X_n \xrightarrow{P} X$

Demostración. Dado $\varepsilon > 0$ y si $X_n \xrightarrow{c.s.} X$, $P(\{|X_n - X| > \varepsilon \text{ para infinitos } n\}) = 0$. Luego, $P(|X_n - X| > \varepsilon) \leq P(\cup_{k \geq n} \{|X_k - X| > \varepsilon\}) \searrow P(\cap_{n \in \mathbb{N}} \cup_{k \geq n} \{|X_k - X| > \varepsilon\}) = 0$. \square

No obstante, puede probarse a partir del primer lema de Borel-Cantelli que, si $X_n \xrightarrow{P} X$, existe una subsección de $(X_n)_n$ que converge c.s. a X . Vamos a probar también que la convergencia en probabilidad es más débil que la convergencia L^2 , en cuyo caso habremos demostrado que la convergencia en L^2 implica la convergencia c.s. de alguna subsección. También demostraremos dos conocidas desigualdades.

Proposición 5 (Desigualdad de Markov). Si Y es de cuadrado integrable y $\varepsilon > 0$,

$$P(|Y| > \varepsilon) \leq \frac{1}{\varepsilon^2} \int Y^2 dP \quad (3.8)$$

Demostración. $\int Y^2 dP = \int_{|Y| < \varepsilon} Y^2 dP + \int_{|Y| > \varepsilon} Y^2 dP \geq \varepsilon^2 P(|Y| > \varepsilon)$ \square

Basta aplicar (3.8) a $X_n - X$ para obtener el resultado buscado.

Corolario 13. Si $X_n \xrightarrow{L^2} X$ entonces $X_n \xrightarrow{P} X$

Además, si aplicamos (3.8) a $Y - E[Y]$ obtenemos la otra desigualdad de interés.

Proposición 6 (Desigualdad de Chebishev). Si $E[Y] = \mu$ y $\text{var}[Y] = \sigma^2$,

$$P(|Y - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2} \quad (3.9)$$

Ley Débil de los Grandes Números (LDGN): Antes de enunciar este resultado fundamental, debemos aclarar una importante propiedad de la media muestral. Téngase en cuenta que, si X_1, X_2 son variables independientes e integrables, se sigue de los teoremas de Fubini y de la Medida Imagen que $E[X_1 X_2] = E[X_1] E[X_2]$. Además, si X_1, \dots, X_n iid según la distribución de X , de media μ , $E[\bar{X}_n] = n^{-1} \sum_{i=1}^n E[X_i] = \mu$. Veamos cómo se comporta la varianza de la media \bar{X} de una muestra aleatoria de tamaño n .

Proposición 7. Si X es de cuadrado integrable siendo $\text{var}[X] = \sigma^2$, entonces

$$\text{var}[\bar{X}_n] = \frac{\sigma^2}{n} \quad (3.10)$$

Demostración. Por lo dicho anteriormente la varianza de \bar{X}_n equivale a $E[(\frac{1}{n} \sum_{i=1}^n X_i - \mu)^2] = \frac{1}{n^2} E[\sum_{i=1}^n (X_i - \mu)^2] + \frac{1}{n^2} \sum_{i \neq j} E[(X_i - \mu)]E[(X_j - \mu)] = \sigma^2/n$ \square

Teorema 14 (LDGN). Si $(X_k)_k$ iid según una distribución de media μ y varianza σ^2 , entonces $\bar{X}_n \xrightarrow{P} \mu$.

Demostración. dado $\varepsilon > 0$ y aplicando (3.9) y (3.10) se tiene que $P(|\bar{X}_n - \mu| > \varepsilon) \leq n^{-1} \varepsilon^{-2} \sigma^2 \rightarrow 0$ \square

En particular, si A es un suceso en \mathcal{A}_0 la función $X = I_A$ toma el valor 1 si A ocurre y 0 en caso contrario. Si $(X_n)_n$ es una muestra aleatoria de X , la media muestral \bar{X}_n equivale a frecuencia relativa $p_n(A)$ tras n réplicas independientes. Estamos afirmando pues que $p_n(A)$ converge a $P(A)$ en probabilidad, lo cual da cierto sentido a la axiomática de Kolmogorov, que será más fuerte si logramos probar la convergencia según el criterio c.s. (LFGN). Es más, LDGN se podrá deducir como corolario tanto de LFGN como de TCL. Lo veremos en el próximo capítulo.

3.2.4. Convergencia en distribución

Dados $(P_n)_n$, P distribuciones en $(\mathbb{R}, \mathcal{B})$, siendo $(F_n)_n$, F sus respectivas funciones de distribución, se dice que P_n converge a P cuando

$$F_n(x) \rightarrow F(x) \quad \forall x \text{ punto de continuidad de } F \quad (3.11)$$

Decimos que X_n converge en distribución (d) a X cuando P^{X_n} converge a P^X en el sentido anterior. La convergencia en distribución no implica convergencia en probabilidad ni, por lo tanto, c.s. ni L^2 . Para entenderlo basta considerar el siguiente contraejemplo definido sobre $[0, 1]$ con la distribución uniforme:

$$\forall n \in \mathbb{N}, X_n(\omega) = \begin{cases} 1 & \text{si } 0 \leq \omega \leq 1/2 \\ 0 & \text{si } 1/2 < \omega \leq 1 \end{cases}, \quad X(\omega) = \begin{cases} 0 & \text{si } 0 \leq \omega \leq 1/2 \\ 1 & \text{si } 1/2 < \omega \leq 1 \end{cases} \quad (3.12)$$

Efectivamente, tanto F_n como F valen 0 si $x < 0$, $1/2$ si $0 < x < 1$ y 1 si $x > 1$. Sin embargo, X_n no converge a X en probabilidad. Veamos que la convergencia en probabilidad es estrictamente más fuerte, por lo que la convergencia en distribución es la más débil de las cuatro estudiadas.

Proposición 8. Si $X_n \xrightarrow{P} X$ entonces $X_n \xrightarrow{d} X$

Demostración. Si x es un punto de continuidad de F y $\varepsilon > 0$, se verifica que

$$\begin{aligned} F_n(x) &= P(X_n \leq x) \leq P(X_n \leq x, |X_n - X| \leq \varepsilon) + P(|X_n - X| > \varepsilon) \\ &\leq P(X \leq x + \varepsilon) + P(|X_n - X| > \varepsilon) \xrightarrow{n \rightarrow \infty} F(x + \varepsilon) \end{aligned}$$

Luego, $\limsup F_n(x) \leq F(x + \varepsilon)$ para todo $\varepsilon > 0$. Como F es continua por la derecha en x , $\limsup F_n(x) \leq F(x)$. Razonando de manera análoga con $1 - F_n(x) = P(X_n > x)$ y teniendo en cuenta que F es por hipótesis también continua por la izquierda, se deduce que $\limsup F_n(x) \leq F(x) \leq \liminf F_n(x)$. Por lo tanto $F_n(x) \rightarrow F(x)$ \square

El recíproco es cierto si X es contante, es decir, si la distribución límite es del tipo δ_a .

Proposición 9. Si $X_n \xrightarrow{d} a$ entonces $X_n \xrightarrow{P} a$

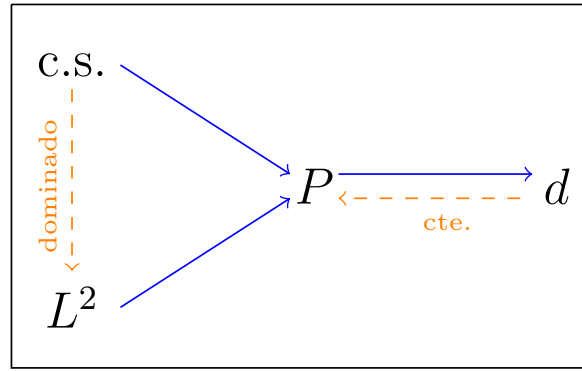
Demostración. Dado $\varepsilon > 0$, F es continua en $a \pm \varepsilon$. Por lo tanto,

$$\begin{aligned} P(|X_n - a| > \varepsilon) &= 1 - P(|X_n - a| \leq \varepsilon) \\ &= 1 - [P(X_n - a \leq \varepsilon) - P(X_n - a < -\varepsilon)] \\ &= 1 - F_n(a + \varepsilon) + F_n(a - \varepsilon) - P(X_n = a - \varepsilon) \\ &\leq 1 - F_n(a + \varepsilon) + F_n(a - \varepsilon) \xrightarrow{n \rightarrow \infty} 1 - F(a + \varepsilon) + F(a - \varepsilon) \\ &= 1 - 1 + 0 \end{aligned}$$

□

El esquema [3.4](#) resume las relaciones estudiadas hasta el momento.

Figura 3.4: Relaciones entre criterios



3.3. Convergencias y operaciones

En esta sección estudiaremos cómo se comportan los diferentes criterios de convergencia cuando realizamos operaciones sencillas o componemos con una función continua, lo cual permitirá una nueva caracterización de la convergencia en distribución.

Lema 15. Si $X_n \rightarrow X$ y $Y_n \rightarrow Y$ según el criterio c.s. o P y $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ es continua, entonces $f(X_n, Y_n) \rightarrow f(X, Y)$ en el criterio correspondiente.

Demostración. La demostración es más fácil para el criterio c.s. que para el criterio P , por lo que desarrollaremos sólo el segundo caso. Dados $\varepsilon > 0$ y $\delta > 0$, se trata de encontrar $n_0 \in \mathbb{N}$ tal que $P(|f(X_n, Y_n) - f(X, Y)| > \varepsilon) < \delta$ si $n \geq n_0$. Dado que $\{|x| > n, |y| > n\} \searrow \emptyset$, $\exists N > 0$ tal que $P(\max\{|X|, |Y|\} > N - 1) < \delta/2$. Como f es uniformemente continua en $[-N, N] \times [-N, N]$, si nos restringimos a dicho conjunto $\exists \gamma \in (0, 1)$ tal que $\max\{|x_1 - x_2|, |y_1 - y_2|\} \leq \gamma$ implica $|f(x_1, y_1) - f(x_2, y_2)| < \varepsilon$. Como X_n, Y_n convergen en probabilidad a X, Y , respectivamente, $\exists n_0 \in \mathbb{N}$ tal que $P(\max\{|X_n - X|, |Y_n - Y|\} > \gamma) < \delta/2$ si $n > n_0$. En ese caso, $P(\max\{|X|, |Y|\} \leq N - 1, \max\{|X_n - X|, |Y_n - Y|\} < \gamma) > 1 - \delta$. En el suceso anterior ocurre que $|X_n|, |Y_n|, |X|, |Y| \leq N$ y $|f(X_n, Y_n) - f(X, Y)| < \varepsilon$. □

Nótese que el lema anterior puede extenderse a una función continua sobre \mathbb{R}^k y es también válido para una función continua sobre \mathbb{R} .

Proposición 10. *Se verifica:*

- (i) Si $X_n \rightarrow X$ c.s. o en probabilidad y f es continua, entonces $f(X_n) \rightarrow f(X)$ según el criterio correspondiente.
- (ii) Si $X_n \rightarrow X$ y $Y_n \rightarrow Y$ según el criterio c.s, L^2 o P y $\alpha, \beta \in \mathbb{R}$, entonces $\alpha X_n + \beta Y_n \rightarrow \alpha X + \beta Y$ según el criterio correspondiente.
- (iii) Si $X_n \rightarrow X$ y $Y_n \rightarrow Y$ según el criterio c.s o P , entonces $X_n Y_n \rightarrow XY$ según el criterio correspondiente.

Indicación: La convergencia en L^2 del apartado (ii) es trivial. El resto es corolario directo del lema 15

El lema 15 justifica afirmaciones como la siguiente. Supongamos que deseamos estimar un parámetro como la varianza σ^2 a partir de una muestra de tamaño n , X_1, \dots, X_n . La varianza puede expresarse como función continua de los momentos de orden 1 y 2 mediante $f(\mu_1, \mu_2) = \mu_2 - \mu_1^2$. Sabemos por LDGN que \bar{X}_n converge en probabilidad a μ ; por el mismo razonamiento y si existe el momento de orden 4, $n^{-1} \sum_{i=1}^n X_i^2 \xrightarrow{P} \mu_2$. En consecuencia, la varianza muestral $\hat{\sigma}_n^2$ verifica

$$\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2 \quad (3.13)$$

De igual forma se puede razonar con otros parámetros, como el coeficiente de correlación, el de asimetría, etc., pues esta afirmación se extiende a funciones continuas de varios momentos. En el siguiente capítulo veremos que también puede extenderse al criterio c.s. Respecto a la convergencia en distribución tenemos este conocido resultado.

Teorema 16 (Slutsky). *Si $X_n \xrightarrow{d} X$ y $Y_n \xrightarrow{P} a$, entonces:*

- (i) $X_n + Y_n \xrightarrow{d} X + a$
- (ii) $X_n Y_n \xrightarrow{d} aX$
- (iii) $X_n / Y_n \xrightarrow{d} X/a$ si $a \neq 0$

Demostración. (i) Nótese que x es un punto de continuidad de F_{X+a} sii $P(X + a = x) = 0$, lo cual equivale a que $x - a$ sea un punto de continuidad de F_X . En ese caso y dado $\varepsilon > 0$ y tal que $x - a - \varepsilon$ es punto de continuidad de X , se verifica

$$\begin{aligned} F_{X_n+Y_n}(x) &= P(X_n \leq x - a - (Y_n - a)) \leq P(X_n \leq x - a - \varepsilon) + P(|Y_n - a| > \varepsilon) \\ &= F_{X_n}(x - a - \varepsilon) + P(|Y_n - a| > \varepsilon) \xrightarrow{n \rightarrow \infty} F_X(x - a - \varepsilon) \end{aligned}$$

Por lo tanto, $\limsup F_{X_n+Y_n}(x) \leq F_X(x - a - \varepsilon)$ si $x - a - \varepsilon$ es punto de continuidad de X . Dado que los puntos de continuidad de F_X son densos, eso ocurre para una sucesión de valores de ε convergiendo a 0. Como F_X es continua por la izquierda en $x - a$, se tiene que $\limsup F_{X_n+Y_n}(x) \leq F_X(x - a)$. Razonando de manera análoga con $1 - F_{X_n+Y_n}(x)$ se obtiene que $\liminf F_{X_n+Y_n}(x) \geq F_X(x - a)$, de lo que se deduce la convergencia.

- (ii) Supongamos que $a = 0$ y probemos que $X_n Y_n$ converge a 0 en probabilidad. Efectivamente, si $\varepsilon > 0$ y $\delta > 0$, existe $M > 0$ tal que $\pm M$ son puntos de continuidad de F_X y $P(|X| > M) \leq \varepsilon/4$. En ese caso, existe $n_1 \in \mathbb{N}$ tal que $n \geq n_1$ implica que $|F_{X_n}(M) - F_X(M)| < \varepsilon/4$ y $|F_{X_n}(-M) - F_X(-M)| < \varepsilon/4$. Entonces,

$$\begin{aligned} P(|X_n| > M) &= 1 - F_{X_n}(M) + F_{X_n}(-M) \\ &\leq P(|X| > M) + |F_{X_n}(M) - F_X(M)| + |F_{X_n}(-M) - F_X(-M)| \\ &< 3\varepsilon/4 \end{aligned}$$

Además, existe $n_2 \in \mathbb{N}$ tal que $n \geq n_2$ implica $P(|Y_n| > \delta M^{-1}) < \varepsilon/4$. Tomado $n_0 = \max\{n_1, n_2\}$ se acaba.

En general, podemos descomponer $X_n Y_n = X_n(Y_n - a) + aX_n$. Así, por la proposición 9 $Y_n - a$ convergería a 0 en probabilidad, por lo que el primer sumando convergería a 0 en probabilidad y por lo tanto en distribución. En virtud de 10(ii), el segundo sumando convergería en probabilidad, y por lo tanto en distribución, a aX . Luego, aplicando el apartado (i), $X_n Y_n$ convergería en distribución a $aX + 0$.

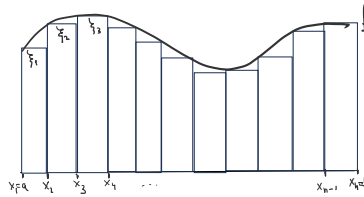
- (iii) Aplicando el lema 15 a la función continua $f(x) = x^{-1}$ se deduce que $1/Y_n \rightarrow 1/a$ en probabilidad. La tesis se sigue entonces del apartado (ii). □

El siguiente resultado es necesario para poder caracterizar la convergencia en distribución de una forma más operativa. Se precisa de un lema previo.

Lema 17. Si $f : [a, b] \rightarrow \mathbb{R}$ es continua y $\mathcal{N} \subset [0, 1]$ es numerable, para todo $\varepsilon_0 > 0$ podemos encontrar una función $g = \sum_{i=1}^{k-1} \xi_i I_{[x_i, x_{i+1}]}$ con $a = x_1 < x_2 < \dots < x_k = b$, todos en $\overline{\mathcal{N}}$, tal que $|f(x) - g(x)| < \varepsilon_0$ para todo $x \in [a, b]$ y $g \leq \max\{f(x) : x \in [a, b]\}$.

Indicación: Basta con tener en cuenta que f es uniformemente continua en $[a, b]$ y que, por lo tanto, existe $\delta > 0$ tal $|f(x) - f(y)| < \varepsilon_0$ si $|x - y| < \delta$.

Figura 3.5: Ilustración del lema 17



Proposición 11. Si $X_n \xrightarrow{d} X$ y f es continua y acotada, entonces $E[f \circ X_n] \rightarrow E[f \circ X]$.

Demostración. Supongamos que $|f(x)| \leq T$ para todo x . Dado $\varepsilon > 0$, sea $M > 0$ tal que $\pm M$ son puntos de continuidad de F_X y $P(|X| > M) < \varepsilon/(20T)$. Sea entonces $g = \sum_{i=1}^{k-1} \xi_i I_{[x_i, x_{i+1}]}$ en las condiciones del lema anterior con, $a = -M$, $b = M$, \mathcal{N} los puntos de discontinuidad de F_X y $\varepsilon_0 = \varepsilon/5$. Como $X_n \rightarrow X$ en distribución, existe $n_0 \in \mathbb{N}$

tal que, para $n \geq n_0$, se verifica que $|F_{X_n}(x_i) - F_X(x_i)| < \varepsilon/(20Tk)$ para $1 \leq i \leq k$, lo cual implica que $P(|X_n| > M) < \varepsilon/(10T)$. En ese caso

$$\begin{aligned}
 |E[f \circ X_n] - E[f \circ X]| &\leq \int_{|X|, |X_n| \leq M} |f \circ X_n - g \circ X_n| dP \\
 &+ \int_{|X|, |X_n| \leq M} |f \circ X - g \circ X| dP \\
 &+ 2T[P(|X| > M) + P(|X_n| > M)] \\
 &+ \int_{|X|, |X_n| \leq M} |g \circ X_n - g \circ X| \\
 &\leq \frac{4\varepsilon}{5} + \int_{|X|, |X_n| \leq M} |g \circ X_n - g \circ X| dP
 \end{aligned}$$

Analicemos el último sumando:

$$\begin{aligned}
 \int_{|X|, |X_n| \leq M} |g dP \circ X_n - g \circ X| dP &\leq \sum_{i=1}^{k-1} |\xi_i| \cdot |P(X_n \in]x_i, x_{i+1}]) - P(X \in]x_i, x_{i+1}])| \\
 &\leq T \sum_{i=1}^{k-1} |F_{X_n}(x_{i+1}) - F_X(x_{i+1}) + F_X(x_i) - F_{X_n}(x_i)| \\
 &\leq 2T \sum_{i=1}^k |F_{X_n}(x_i) - F_X(x_i)| \\
 &\leq 2Tk \cdot \frac{\varepsilon}{10Tk} = \frac{\varepsilon}{5}
 \end{aligned}$$

□

Teorema 18 (Continuidad de Levy). $X_n \xrightarrow{d} X$ sii $\varphi_{X_n}(t) \rightarrow \varphi(t) \forall t \in \mathbb{R}$.

Demostración. \Rightarrow Supongamos que $X_n \xrightarrow{d} X$. Dado $t \in \mathbb{R}$, se sigue de la proposición 11 que $\varphi_{X_n}(t) = E[\cos tX_n] + iE[\sin tX_n] \rightarrow E[\cos tX] + iE[\sin tX] = \varphi_X(t)$.

\Leftarrow La demostración puede encontrarse en Gut(2013).

□

Esta nueva caracterización es la que utilizaremos para probar el TCL en el siguiente capítulo, de ahí la importancia de la función característica en nuestra teoría.

4. PRINCIPALES TEOREMAS LÍMITE

Tal y como avanzamos en la introducción, uno de los dos principales objetivos de esta materia es dar sentido a la axiomática de Kolmogorov. Ahora estamos en condiciones de abordar esa tarea estudiando los dos principales resultados del Cálculo de Probabilidades: la Ley de los Grandes Números y el Teorema Central del Límite en algunas de sus múltiples versiones.

4.1. Leyes de los Grandes Números

En el anterior capítulo demostramos ya sin mucho esfuerzo la denominada Ley Débil de los Grandes Números (LDGN), que garantiza la convergencia en probabilidad de la media muestral a la media probabilística siempre y cuando la varianza sea finita. Estudiaremos aquí un resultado más fuerte y costoso denominado Ley Fuerte de los Grandes Números (LFGN) que garantiza la convergencia c.s. y, además, en una versión más general, admite incluso que la media sea infinita (en cuyo caso la varianza no puede ser finita). Como hemos dicho, se trata de un teorema arduo que precisa de varios resultados previos, los cuales se exponen a continuación.

Lema 19 (Toeplitz). Sean $(x_n)_n \subset \mathbb{R}$ convergente a $x \in \mathbb{R}$, $(\omega_n)_n \geq 0$ tal que $b_n = \sum_{k=1}^n \omega_k \rightarrow \infty$. Entonces $b_n^{-1} \sum_{k=1}^n \omega_k x_k \rightarrow x$. En particular, $n^{-1} \sum_{k=1}^n x_k \rightarrow x$

Demostración. Podemos suponer sin pérdida de generalidad que $x = 0$ pues, en caso contrario, razonaríamos sin problemas con la sucesión $(X_n - x)_n$. Entonces, dado $\varepsilon > 0$, $\exists n_1 \in \mathbb{N}$ tal que $|x_n| < \varepsilon/4$ si $n \geq n_1$, en cuyo caso

$$\begin{aligned} \left| \frac{1}{b_n} \sum_{k=1}^n \omega_k x_k \right| &\leq \left| \frac{1}{b_n} \sum_{k=1}^{n_1} \omega_k x_k \right| + \frac{b_n - b_{n_1}}{b_n} \frac{1}{b_n - b_{n_1}} \sum_{k=n_1+1}^n \omega_k |x_k| \\ &\leq \frac{1}{b_n} \sum_{k=1}^{n_1} \omega_k |x_k| + \frac{b_n - b_{n_1}}{b_n} \frac{1}{b_n - b_{n_1}} (b_n - b_{n_1}) \frac{\varepsilon}{4} \end{aligned}$$

Dado que el primer sumando está acotado y $(b_n) \rightarrow \infty$, $\exists n_2 \in \mathbb{N}$ tal que dicho sumando es inferior a $\varepsilon/2$ si $n \geq n_2$. También existe $n_3 \in \mathbb{N}$ tal que $(b_n - b_{n_1}) < 2$ para $n \geq n_3$. Tomando $n_0 = \max\{n_1, n_2, n_3\}$ acabamos. \square

Lema 20 (Kroonecker). Sean $(x_n)_n \subset \mathbb{R}$, $(a_n)_n > 0$ tal que $a_n \nearrow \infty$ y $(d_n)_n$, definida mediante $d_n = \sum_{k=1}^n x_k/a_k$, converge a d . Entonces, $a_n^{-1} \sum_{k=1}^n x_k \rightarrow 0$.

Demostración. Definamos $a_0 = d_0 = 0$ y $\omega_k = a_{k+1} - a_k > 0$ para todo k . En ese caso, $\sum_{k=1}^n \omega_k = a_n \rightarrow \infty$. Se verifica para todo $k \geq 1$ que $x_k = a_k(d_k - d_{k-1})$. Por lo tanto,

$$\frac{1}{a_n} \sum_{k=1}^n x_k = \frac{1}{a_n} \left(\sum_{k=1}^n a_k d_k - \sum_{k=0}^{n-1} a_{k+1} d_k \right) = d_n - \frac{1}{a_n} \sum_{k=0}^{n-1} (a_{k+1} - a_k) d_k d_n - \frac{1}{a_n} \sum_{k=1}^{n-1} \omega_k d_k$$

El primer sumando converge a d por hipótesis mientras que el segundo también por el lema de Toeplitz, lo que concluye la prueba. \square

Teorema 21 (Desigualdad de Kolmogorov). Si X_1, \dots, X_n independientes con media nula y varianza finita, $S_k = \sum_{i=1}^k X_i$ para $k = 1, \dots, n$ y $\kappa > 0$, entonces

$$P \left(\max_{1 \leq k \leq n} |S_k| > \kappa \right) \leq \frac{1}{\kappa^2} \sum_{k=1}^n \text{var}[X_k] \quad (4.1)$$

En particular, si $(X_k)_k$ iid con varianza σ^2 , $n\kappa^{-2}\sigma^2$ es cota superior.

Demostración. Definamos $A_1 = \{|X_1| > \kappa\}$ y $A_k = \{\max_{1 \leq j \leq k-1} |S_j| \leq \kappa, |S_k| > \kappa\}$ para $k = 2, \dots, n$. En ese caso, A_1, \dots, A_n son disjuntos y tales que $\{\max_{1 \leq k \leq n} |S_k| > \kappa\} = \cup_{k=1}^n A_k$. Dado que X_1, \dots, X_n son independientes y de media nula se verifica:

$$\begin{aligned} \sum_{k=1}^n \text{var}[X_k] &= E[S_n^2] \geq \sum_{k=1}^n E[S_n^2 I_{A_k}] = \sum_{k=1}^n E[(S_n - S_k + S_k)^2 I_{A_k}] \\ &\geq \sum_{k=1}^n E[(S_k^2 + 2S_k(S_n - S_k)) I_{A_k}] \\ &= \sum_{k=1}^n E[S_k^2 I_{A_k}] + 2E[S_k I_{A_k}] \cdot (E[S_n] - E[S_k]) \\ &= \sum_{k=1}^n E[S_k^2 I_{A_k}] \geq \kappa^2 \sum_{k=1}^n P(A_k) = \kappa^2 P(\cup_{k=1}^n A_k) \\ &= \kappa^2 P \left(\max_{1 \leq k \leq n} |S_k| > \kappa \right) \end{aligned}$$

\square

Teorema 22 (Criterio de convergencia de Komogorov). Si $(X_n)_{n \in \mathbb{N}}$ son independientes y tales que $\sum_{n=1}^{\infty} \text{var}[X_n] < \infty$, entonces la sucesión $\sum_{k=1}^n (X_k - E[X_k])$ converge c.s. Si, además, $\sum_{k=1}^n E[X_k]$ es convergente, entonces $\sum_{k=1}^n X_k$ es convergente c.s.

Demostración. Manteniendo las notaciones del teorema anterior vamos a probar que, para todo $\varepsilon > 0$,

$$P \left(\max_{k \geq n} |S_k - E[S_k] - (S_n - E[S_n])| > \varepsilon \right) \xrightarrow{n \rightarrow \infty} 0 \quad (4.2)$$

Efectivamente, dado $p \geq n$, y aplicando la Desigualdad de Kolmogorov a la sucesión $(X_k - E[X_k])_{k \geq n+1}$, se verifica

$$\begin{aligned} P\left(\max_{n \leq k \leq p} |S_k - E[S_k] - (S_n - E[S_n])| > \varepsilon\right) &= P\left(\max_{n \leq k \leq p} \left|\sum_{k=n+1}^p (X_k - E[X_k])\right| > \varepsilon\right) \\ &\leq \frac{1}{\varepsilon^2} \sum_{k=n+1}^p \text{var}[X_k] \end{aligned}$$

Por lo tanto, $P(\max_{k \geq n} |S_k - E[S_k] - (S_n - E[S_n])| > \varepsilon) \leq \varepsilon^{-2} \sum_{k=n+1}^{\infty} \text{var}[X_k]$, que converge a 0 por hipótesis. Ahora bien, si se denota por $A_{n\varepsilon}$ a cada suceso que verifica la condición (4.2), se tiene que, para cada par $n, m \in \mathbb{N}$, $\cap_{j \in \mathbb{N}} A_{j, \frac{1}{m}} \subset A_{n, \frac{1}{m}}$. Luego, la probabilidad de la intersección es nula y, por lo tanto, también lo es $P(\bar{A})$, donde $A = \cup_{m \in \mathbb{N}} \cap_{j \in \mathbb{N}} A_{j, \frac{1}{m}}$. Nótese que si $\omega \in \bar{A}$ y $\delta > 0$, entonces, $\forall m \in \mathbb{N}$ y, en particular, para cualquiera tal que $2/m < \varepsilon$, existe $n_0 \in \mathbb{N}$ tal que, si $j_1 \geq n_0$, $|S_{j_1}(\omega) - E[S_{j_1}] - (S_{n_0}(\omega) - E[S_{n_0}])| < 1/m$. Luego, si $j_1, j_2 \geq n_0$, se verifica que $|S_{j_1}(\omega) - E[S_{j_1}] - (S_{j_2}(\omega) - E[S_{j_2}])|$ es menor o igual que $|S_{j_1}(\omega) - E[S_{j_1}] - (S_{n_0}(\omega) - E[S_{n_0}])| + |S_{j_2}(\omega) - E[S_{j_2}] - (S_{n_0}(\omega) - E[S_{n_0}])| \leq \varepsilon$. Es decir, que $\sum_{k=1}^n (X_k(\omega) - E[X_k])$ es una sucesión de Cauchy y, por lo tanto, convergente si $\omega \in \bar{A}$, como queríamos probar.

Por último, si $\sum_{k=1}^n E[X_k]$ converge entonces $\sum_{k=1}^n X_k$ que descompone en $\sum_{k=1}^n (X_k - E[X_k]) + \sum_{k=1}^n E[X_k]$ es convergente c.s. \square

Teorema 23. Si $(X_n)_{n \in \mathbb{N}}$ son independientes con media nula y varianzas finitas $(\sigma_n^2)_{n \in \mathbb{N}}$, respectivamente, y tales que $\sum_{n=1}^{\infty} \sigma_n^2/n^2 < \infty$, entonces $n^{-1} \sum_{k=1}^n X_k$ converge a 0 c.s.

Demostración. dado que la sucesión de variables aleatorias $(X_n/n)_n$, con varianzas (σ_n^2/n^2) y medias nulas verifica las hipótesis del Criterio de Convergencia de Kolmogorov, se verifica $\sum_{k=1}^n X_k/k$ es convergente c.s., es decir, si $\omega \in \bar{A}$, donde $P(A) = 0$. Aplicando el Lema de Kronecker con $x_n = X_n(\omega)$ y $a_n = n$, concluimos que $n^{-1} \sum_{k=1}^n X_k \rightarrow 0$ c.s. \square

El resultado buscado es consecuencia directa del teorema 23 teniendo en cuenta que $E[X - \mu] = 0$ y que la serie $\sum_{n=1}^{\infty} 1/n^2$ es convergente.

Corolario 24 (LFGN de Kolmogorov). Dada $(X_n)_{n \in \mathbb{N}}$ iid con media μ y varianza finita σ^2 , se verifica que \bar{X}_n converge a μ c.s.

Si lo aplicamos al fenómeno aleatorio en el que se evalúa la ocurrencia de un suceso A con probabilidad $P(A)$, X_n denota la variable I_A correspondiente a la n -ésima réplica del mismo y $p_n(A)$ a la frecuencia relativa tras n réplicas, entonces se tiene que $p_n(A)$ converge c.s. a $P(A)$. Comentamos anteriormente que, en Teoría Asintótica, la probabilidad P que rige las infinitas réplicas debe entenderse como el producto $P_0^{\mathbb{N}}$ de la probabilidad P_0 que rige el fenómeno original. Queda pues demostrada la afirmación (11) que fundamenta la axiomática de Kolmogorov.

Con algo más de esfuerzo, LFGN puede incluso generalizarse, de manera que no sea necesario que la varianza sea finita. Incluso es válido aunque la media tampoco lo sea, siempre que exista (podría ser $\pm\infty$). La demostración podemos encontrarla en Gut(2013).

Teorema 25 (LFGN extendida). Si $(X_n)_n$ iid con media $\mu \in \bar{\mathbb{R}}$, entonces $\bar{X}_n \rightarrow \mu$ c.s.

Para acabar esta sección podemos también mejorar la afirmación (3.13). Concretamente, si existe en \mathbb{R} el momento μ_k de orden k , entonces $n^{-1} \sum_{j=1}^n X_j^k$ converge c.s. a μ_k . Es más, si tenemos en cuenta el lema 15, podemos afirmar:

Teorema 26 (Método de los momentos). *Si $(X_n)_n$ iid con momentos de orden k finito y $g : \mathbb{R}^k \rightarrow \mathbb{R}$ continua, entonces*

$$g\left(\frac{1}{n} \sum_{j=1}^n X_j, \dots, \frac{1}{n} \sum_{j=1}^n X_j^k\right) \xrightarrow{c.s.} g(\mu_1, \dots, \mu_k) \quad (4.3)$$

En consecuencia, todos los parámetros muestrales, univariantes o multivariantes, que puedan expresarse como funciones continuas de momentos de diferentes órdenes (varianza, covarianza, a simetría, etc.) convergen c.s. a sus homólogos probabilísticos.

El siguiente teorema, cuya demostración puede encontrarse en Gut (2013), es consecuencia de LFGN. Se aplica a la función de distribución empírica o muestral que, dada una muestra de tamaño n , X_1, \dots, X_n , se define como $\mathbb{F}_n(x, \omega) = n^{-1} \sum_{i=1}^n I_{(-\infty, x]}(X_i(\omega))$, es decir, la proporción de veces que la variable es inferior a x . en la muestra. El teorema afirma lo siguiente :

Teorema 27 (Glivenko-Cantelli). *Si $(X_n)_n$ iid con función de distribución F entonces existe un suceso N de probabilidad nula tal que*

$$\sup\{x \in \mathbb{R} : |\mathbb{F}_n(x, \omega) - F(x)|\} \xrightarrow[n \rightarrow \infty]{} 0, \quad \forall \omega \in \overline{N} \quad (4.4)$$

Este resultado implica la convergencia c.s. de las frecuencias relativas de sucesos del tipo $(-\infty, x]$ y, por el Teorema de Dynkin, de cualquier suceso en \mathcal{B} , tal y como se deduce de LFGN, pero garantiza además una convergencia uniforme.

4.2. Teorema Central del Límite

Una vez sabido a qué convergen los parámetros muestrales que pueden expresarse como media de potencias, vamos a estudiar cómo converge dicha media, que será siguiendo aproximadamente (en límite) una distribución tipo normal. Estamos hablando de una colección de teoremas de los cuales probaremos sólo los más básicos. Para las versiones extendidas aconsejamos al lector Gut(2013).

Realmente, la versión fundamental, atribuida a Lidemberg y Levy, también denominada clásica o iid, es fácil de demostrar si asumimos el Teorema de Continuidad de Levy (teorema 18) junto con las propiedades de la función característica y del número e . En particular, hemos de tener en cuenta, lo siguiente:

- (a) La convergencia en distribución equivale a la convergencia puntual en función característica.
- (b) Se sigue de (1.12) con $k = 2$ que, si $E[X] = 0$ y $var[X] = 1$, se verifica para todo $t \in \mathbb{R}$ que $\varphi_X(t) = -\frac{1}{2}t^2 + t^2 o(1)$, donde $o(1)$ denota una función medible de t que converge a 0 cuando $t \rightarrow 0$.
- (c) $\varphi_{N(0,1)}(t) = e^{-\frac{1}{2}t^2}$

Teorema 28 (TCL Lindemberg-Levy o clásico). *Si $(X_n)_n$ iid con media μ y varianza finita σ^2 , entonces*

$$\sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \xrightarrow{d} N(0, 1) \quad (4.5)$$

Demostración. Sea, para cada $n \in \mathbb{N}$, $Z_n = \sum_{k=1}^n (X_k - \mu)/\sigma$ y $K_n = Z_n/\sqrt{n}$, que coincide con el término a la izquierda en (4.5). Por lo tanto y teniendo en cuenta que X_1, \dots, X_n son iid, existe una función $o(t)$ tal que, $\forall t \in \mathbb{R}$,

$$\begin{aligned} \varphi_{K_n}(t) &= \varphi_{Z_n}(t/\sqrt{n}) = \varphi_{\sum_{k=1}^n \frac{X_k - \mu}{\sigma}}(t/\sqrt{n}) = \prod_{k=1}^n \varphi_{\frac{X_k - \mu}{\sigma}}(t/\sqrt{n}) = \left(\varphi_{\frac{X_1 - \mu}{\sigma}}(t/\sqrt{n}) \right)^n \\ &= \left(1 - \frac{t^2}{2n} + (t/\sqrt{n})^2 o(t/\sqrt{n}) \right)^n = \left[\left(1 - \frac{1}{\frac{2n}{t^2(1+2o(t/\sqrt{n}))}} \right)^{\frac{2n}{t^2(1+2o(t/\sqrt{n}))}} \right]^{\frac{t^2(1+2o(t/\sqrt{n}))}{2}} \end{aligned}$$

Si $n \rightarrow \infty$, $t/\sqrt{n} \rightarrow 0$ y, en consecuencia, $o(t/\sqrt{n}) \rightarrow 0$. Por lo tanto,

$$a_n(t) = \frac{2n}{t^2(1+2o(t/\sqrt{n}))} \xrightarrow{n \rightarrow \infty} \infty, \quad b_n(t) = \frac{t^2(1+2o(t/\sqrt{n}))}{2} \xrightarrow{n \rightarrow \infty} \frac{t^2}{2}$$

En definitiva, se verifica por (2.6):

$$\varphi_{K_n}(t) = \left[\left(1 - \frac{1}{a_n(t)} \right)^{a_n(t)} \right]^{b_n(t)} \rightarrow e^{-\frac{t^2}{2}}$$

□

Nótese que, si se denota $S_n = \sum_{i=1}^n X_i$, (4.5) equivale a

$$\frac{S_n - E[S_n]}{\sqrt{\text{var}[S_n]}} \xrightarrow{d} N(0, 1) \quad (4.6)$$

Se puede generalizar (4.5) al caso multivariante. La demostración se basa en primera instancia en la versión multivariante del Teorema de Continuidad de Levy, de la cual se deduce la denominada Astucia de Cramer-Wold.

Lema 29 (Astucia de Cramer-Wold). *Si $(X_n)_n$ una sucesión de vectores aleatorios k -dimensionales, $\mu \in \mathbb{R}^k$ y $\Sigma \in \mathcal{M}_{k \times k}$ definida positiva, entonces*

$$\left[X_n \xrightarrow{d} N_k(0, \Sigma) \right] \Leftrightarrow \left[a'X \xrightarrow{d} N(a'\mu, a'\Sigma a) \quad \forall a \in \mathbb{R}^k \right] \quad (4.7)$$

Teorema 30 (TCM multivariante). *Si $(X_n)_n$ en \mathbb{R}^k iid con media μ matriz de covarianzas $\Sigma > 0$, entonces*

$$\sqrt{n}\Sigma^{-1/2}(\bar{X}_n - \mu) \xrightarrow{d} N_k(0, Id) \quad (4.8)$$

Indicación: Comprueba que para cada $a \in \mathbb{R}^k$ $\sqrt{n}(a'\Sigma a)^{-1/2}(\bar{X}_n - \mu)$ converge a $N(0, a'\Sigma a)$ y aplica (4.7).

4.2.1. Aplicaciones del TCL

Hemos utilizado la función característica para probar este teorema, pero conviene volver a la función de distribución para entender sus implicaciones. En particular y en lo sucesivo, Φ denotará la función característica de la distribución límite $N(0, 1)$.

Inferencias sobre la media: Por ejemplo, en un problema de inferencia estadística deseamos contrastar, a partir de una muestra aleatoria de tamaño n , X_1, \dots, X_n , la hipótesis inicial de que la media de cierta variable X , cuya varianza σ^2 es finita, es μ_0 . Si X siguiera una distribución normal, la variable

$$t = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \quad (4.9)$$

seguiría automáticamente un modelo de distribución $Z \sim N(0, 1)$. En ese caso, si el resultado de la muestra es $t = t_{exp}$, se definiría el p -valor bilateral como $P(|t| > |t_{exp}|) = 1 - [\Phi(|t_{exp}|) - \Phi(-|t_{exp}|)] = 2[1 - \Phi(|t_{exp}|)]$, que es el resultado del test de hipótesis. En el contexto asintótico, tenemos realmente una sucesión $(t_n)_n$. El p_n -valor exacto se define de manera análoga para una probabilidad P genérica, pero verifica en virtud de TCL:

$$p_n = P(|t| > |t_{exp}|) = 1 - [F_{t_n}(|t_{exp}|) - F_{t_n}(-|t_{exp}|)] \xrightarrow{n \rightarrow \infty} 1 - [\Phi(|t_{exp}|) - \Phi(-|t_{exp}|)] = p \quad (4.10)$$

Es decir, si n es suficientemente grande, el p -valor calculado a partir de la tabla $N(0, 1)$ es aproximadamente válido. No obstante, la variable (4.9) no puede calcularse en la prác-

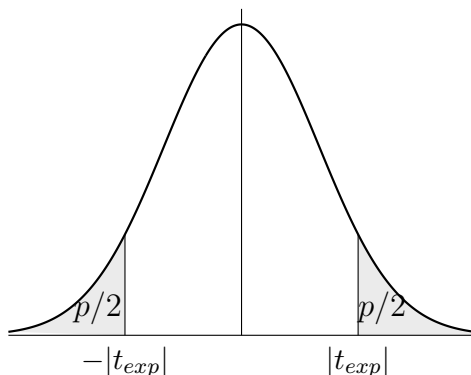


Figura 4.1: Ilustración del p -valor

tica, a menos que se conozca la varianza σ^2 , lo cual resulta enrevesado. Sin embargo, el Teorema 16 (Slutsky) permite aplicar TCL reemplazando σ^2 por su estimador $\hat{\sigma}_n^2$ que, según el Método de los Momentos, converge c.s. (basta que sea en probabilidad) a σ^2 . Efectivamente, se verifica

$$t^* = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\hat{\sigma}_n} = \sqrt{n} \frac{\bar{X}_n - \mu_0}{\sigma} \cdot \frac{\sigma}{\hat{\sigma}_n} \xrightarrow{d} N(0, 1) \quad (4.11)$$

El primer factor converge en distribución a $N(0, 1)$, mientras que el segundo converge en probabilidad a 1, por lo que el producto converge en distribución a $N(0, 1)$. Así pues, t^* es el denominado estadístico de contraste del test de Student, que confrontará su resultado

con la tabla de Φ aunque no se verifique la normalidad, siempre y cuando la muestra sea lo suficientemente grande.

Otra pregunta que podemos plantear es qué entendemos por una muestra suficientemente grande. Mejor dicho: ¿De qué depende la rapidez de la convergencia a Φ en TCL? Podemos encontrar una respuesta parcial en la fórmula de la expansión de Edgeworth, (Gut (2013)).

Teorema 31 (Expansión de Edgeworth). *Si $(X_n)_n$ iid según un modelo de distribución con momento de orden 3 finito y F_n denota la función de distribución de $\sqrt{n}\sigma^{-1}(\bar{X}_n - \mu)$, se verifica*

$$F_n(x) = \Phi(x) + (1 - x^2) \frac{e^{-\frac{x^2}{2}}}{6\sqrt{2n\pi}} \gamma_1 + \frac{1}{\sqrt{n}} o(1) \quad (4.12)$$

En consecuencia, el aspecto más crítico desde el punto de vista asintótico es la asimetría, expresada por el coeficiente γ_1 . Si es baja podemos aceptar la aproximación de la media muestral a la distribución normal para tamaños de muestra pequeños. Si es alta, puede que necesitemos trabajar con una gran muestra para poder utilizar Φ . Un desarrollo de la expresión (4.12) con un grado más de aproximación revela que el coeficiente de curtosis γ_4 también influye, aunque en menor medida, siendo la convergencia más rápida cuando $\gamma_2 \simeq 3$, que es el valor que corresponde a curva normal.

En la figura 4.2 se ilustra la influencia de la asimetría en la distribución de la media de una muestra de tamaño $n = 9$. La distribución de la izquierda corresponde a la media muestral de una variable con $\gamma_1 \simeq 0$ (y $\gamma_2 \simeq 1.5$), mientras que la de la derecha tiene $\gamma_1 \simeq 2$.

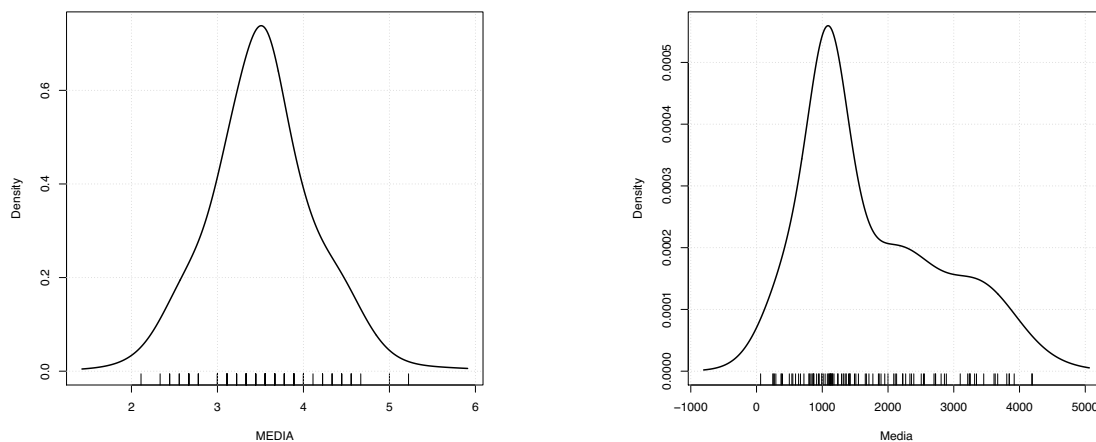


Figura 4.2: Distribución de la media muestral

Aproximaciones de las distribuciones Binomial y Poisson: Como corolario directo de TCL tenemos lo que ya se intuía en las figuras 1.1, 1.2 y 2.1, que es la convergencia en distribución de la binomial a la normal:

$$B(n, p) \longrightarrow N(np, \sqrt{np(1-p)}) \quad (4.13)$$

Con la expresión (4.13) nos hemos permitido la licencia de expresar esto otro: Si $(X_n)_n$ iid $B(p)$ y $S_n = \sum_{i=1}^n X_i$, entonces $\sqrt{n}(S_n - np)/\sqrt{p(1-p)}$ converge en distribución a $N(0, 1)$. Hemos dicho que la validez de la aproximación depende principalmente del grado de asimetría de la variable estudiada, que es $B(p)$. Su asimetría es nula cuando $p = 0.5$ y aumenta conforme p tiende a 0 (equivalentemente si p tiende a 1), es decir, cuanto más raro es el suceso a evaluar (equivalentemente su complementario). Por eso se suele utilizar como criterio de aproximación $np, n(1-p) \geq 5$. Si se verifica, podemos comprobar que $P(S_n = k) \simeq \Phi(k + 0.5) - \Phi(k - 0.5)$ para $k = 0, 1, \dots, n$. En la figura 4.3 podemos comparar las funciones de probabilidad de $B(20, 0.5)$ y $B(20, 0.05)$.

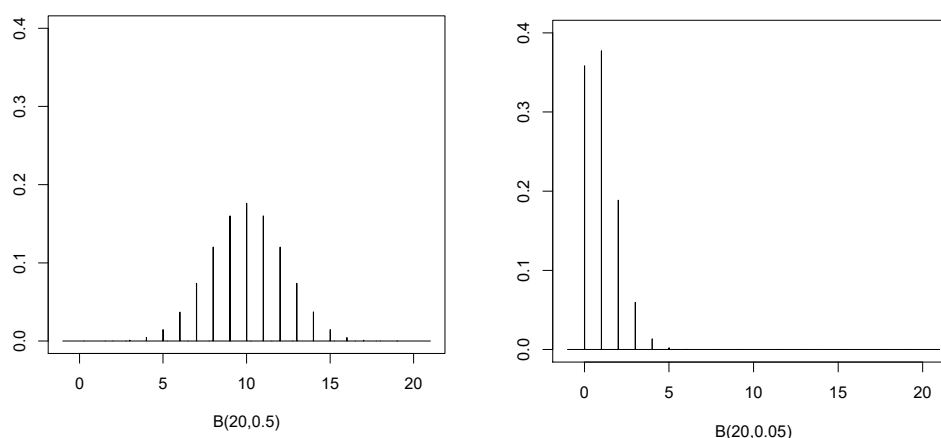


Figura 4.3: Binomiales con $p = 0.5$ y $p = 0.05$

Por otra parte, recordemos que el teorema 4 garantiza una aceptable aproximación de $B(n, p)$ a $\mathcal{P}(np)$ si n es grande y p pequeño. Es decir, que la distribución de Poisson es válida para modelizar el conteo de un suceso raro tras una larga serie de intentos independientes. A la izquierda figura 4.4 podemos apreciar la función de probabilidad de una distribución $\mathcal{P}(\lambda)$ con $\lambda = 1$, que presenta una evidente asimetría. No obstante, la distribución de $\mathcal{P}(\lambda)$ tiende a la normalidad a medida que λ aumenta, tal y como se aprecia en la parte derecha de la misma figura. Esta afirmación constituye otro teorema de los que se agrupan bajo la denominación de TCL y que probaremos a continuación.

Teorema 32. Se verifica que $\frac{\mathcal{P}(\lambda) - \lambda}{\sqrt{\lambda}} \xrightarrow{d} N(0, 1)$ cuando $\lambda \rightarrow \infty$.

Demostración. Si $X \sim \mathcal{P}(\lambda)$ entonces $\varphi_X(t) = \exp\{\lambda(e^{it} - 1)\}$. Al tipificar tenemos que $\varphi_{(X-\lambda)/\sqrt{\lambda}}(t) = \exp\{\lambda(e^{it/\sqrt{\lambda}} - 1)\}e^{-it/\sqrt{\lambda}}$. Si desarrollamos la serie de potencias se verifica

$$e^{it/\sqrt{\lambda}} = 1 + it/\lambda - \frac{1}{2} \frac{t^2}{\lambda} \cdot f(t, \lambda), \text{ donde } \lim_{\lambda \rightarrow \infty} f(t, \lambda) = 1$$

Por lo tanto, $\varphi_{(X-\lambda)/\sqrt{\lambda}}(t) = \exp\{-t^2/2 \cdot f(t, \lambda)\} \xrightarrow{\lambda \rightarrow \infty} e^{-t^2/2}$ □

Lo dicho anteriormente se esquematiza en la figura 4.5

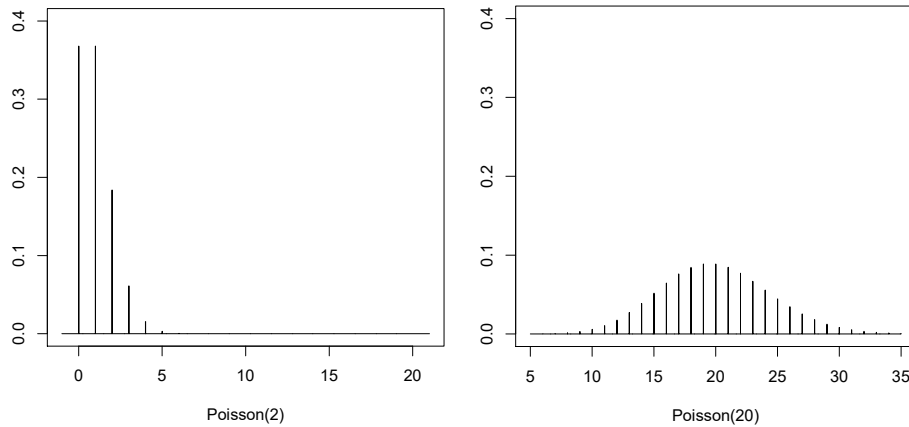
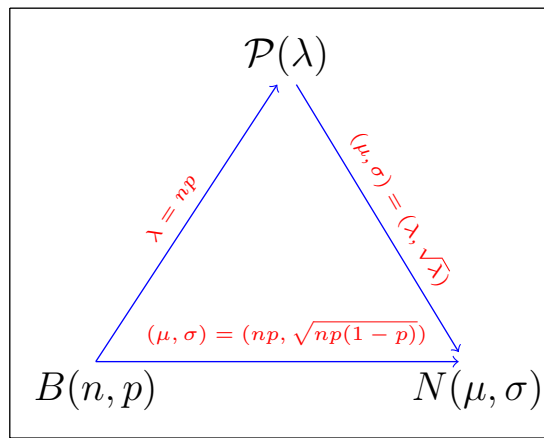

 Figura 4.4: $\mathcal{P}(\lambda)$ con $\lambda = 1$ y $\lambda = 20$


Figura 4.5: Aproximaciones entre Binomial, Poisson y Normal

El método Delta: En ocasiones podemos tener problemas a la hora de poner en práctica TCL por la presencia un parámetro probabilístico en la distribución límite. Eso sucedió, por ejemplo, en (4.9), pues lo que se afirma en ese caso es que, si la media es μ y la varianza σ^2 , $\sqrt{n}(\bar{X}_n - \mu)$ converge en distribución a $N(0, \sigma)$. La desviación típica σ ejerce como un parámetro fantasma que nos vemos obligados a estimar mediante $\hat{\sigma}_n$, de manera que el Teorema de Slutsky soluciona asintóticamente el problema. En otras ocasiones, el parámetro que figura como desviación típica límite puede ser función de la propia media. En esos casos el recurso anterior es menos apropiado que el que explicamos a continuación.

Supongamos que $H_n = \sqrt{n}(T_n - \nu) \xrightarrow{d} N(0, \theta)$ y que g es una función real de derivada continua tal que $g'(\nu) \neq 0$. Teniendo en cuenta el desarrollo de Taylor de grado 1 de g en ν , existe una sucesión (ν_n) de funciones medibles tales quw, para todo $n \in \mathbb{N}$,

$$g(T_n) = g(\nu) + g'(\nu_n)(T_n - \nu), \quad |\nu_n - \nu| \leq |T_n - \nu| \quad (4.14)$$

Del Teorema de Slutsky se sigue también $(T_n - \nu) = n^{-1/2}H_n$ converge a 0 en distribución y, por lo tanto, en probabilidad. Con mayor razón, $\nu_n \xrightarrow{P} \nu$. Al ser g' continua, se deduce de la proposición 10(i) que $g'(\nu_n)$ converge en probabilidad a $g'(\nu)$. Aplicando de nuevo

el Teorema de Slutsky, tenemos que

$$\sqrt{n}[g(T_n) - g(\nu_n)] = \sqrt{n}g'(\nu_n)(T_n - \nu) \xrightarrow{d} N(0, g'(\nu)\theta) \quad (4.15)$$

Luego, si g es tal que $g'(\nu) = \theta^{-1}$, conseguimos que $\sqrt{n}[g(T_n) - g(\nu)]$ converja en distribución a $N(0, 1)$. Por ejemplo, supongamos conocido que, si $(X_n)_n$ iid $N(\mu, \sigma)$, entonces $\sqrt{n}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{d} N(0, \sqrt{2}\sigma^4)$. Tomando $g = (\log x)/\sqrt{2}$ se tiene entonces que

$$\sqrt{\frac{n}{2}}(\log s_n^2 - \log \sigma^2) \xrightarrow{d} N(0, 1) \quad (4.16)$$

4.2.2. Extensiones del TCL

El teorema [28](#) admite diversas generalizaciones que no se probarán aquí. Se aconseja consultar Gut (2013) para un estudio más profundo. No obstante, enunciaremos al menos la más popular, conocida como versión de Lindemberg-Feller. Partimos de $(X_k)_k$ independientes (no se precisa aquí que sean idénticamente distribuidas) con medias y varianzas finitas $(\mu_k)_k$ y $(\sigma_k^2)_k$, respectivamente y siendo $S_n = \sum_{k=1}^n X_k$ y $c_n^2 = \text{var}[S_n]$ para todo $n \in \mathbb{N}$.

Teorema 33 (TCL Lindemberg-Feller). *Consideremos las hipótesis siguientes:*

$$\lim_{n \rightarrow \infty} \frac{1}{c_n^2} \sum_{k=1}^n E[(X_k - \mu_k)^2 I_{\{|X_k - \mu_k| > \varepsilon c_n\}}] = 0 \quad (4.17)$$

$$\frac{S_n - E[S_n]}{c_n} \xrightarrow{d} N(0, 1) \quad (4.18)$$

$$\lim_{n \rightarrow \infty} \max_{1 \leq k \leq n} \frac{\sigma_k^2}{c_n^2} = 0 \quad (4.19)$$

Se verifica entonces que [\(4.17\)](#) \Leftrightarrow [\(4.18\)](#) + [\(4.19\)](#).

La condición [\(4.17\)](#) se atribuye a Lindemberg y es pues suficiente para que se verifique la convergencia [\(4.18\)](#). La demostración de que es necesaria si añadimos [\(4.19\)](#) se atribuye a Feller. En las condiciones del TCL clásico (iid) se verifica [\(4.17\)](#) pues, aplicando TCD a $(X_1 - \mu)^2 I_{\{|X_1 - \mu| > \sqrt{n}\sigma\varepsilon\}}$ tenemos que

$$\frac{1}{n\sigma^2} \sum_{k=1}^n E[(X_k - \mu)^2 I_{\{|X_k - \mu| > \sqrt{n}\sigma\varepsilon\}}] = \frac{1}{\sigma^2} E[(X_1 - \mu)^2 I_{\{|X_1 - \mu| > \sqrt{n}\sigma\varepsilon\}}] \xrightarrow{n \rightarrow \infty} 0$$

La condición de Lindemberg resulta menos cómoda de verificar que esta otra, atribuida a Lyapunov, pero que es estrictamente más fuerte:

$$\exists \delta > 0 \text{ tal que } \frac{1}{c_n^{2+\delta}} \sum_{k=1}^n E[|X_k - \mu_k|^{2+\delta}] \xrightarrow{n \rightarrow \infty} 0 \quad (4.20)$$

El teorema [33](#) puede generalizarse a una versión matricial en la que, para cada $n \in \mathbb{N}$, tanto las variables como las medias y varianzas pueden ser diferentes. Concretamente, para cada n tenemos $(X_{nk})_{1 \leq k \leq n}$ variables independientes con medias y varianzas finitas

$(\mu_{nk})_{1 \leq k \leq n}$ y $(\sigma_{nk}^2)_{1 \leq k \leq n}$, respectivamente y siendo $S_n = \sum_{k=1}^n X_{nk}$ y $c_n^2 = \text{var}[S_n]$ para todo $n \in \mathbb{N}$. La condición de Lindenberg se expresa entonces mediante:

$$\lim_{n \rightarrow \infty} \frac{1}{c_n^2} \sum_{k=1}^n E[(X_{nk} - \mu_{nk})^2 I_{\{|X_{nk} - \mu_{nk}| > \varepsilon c_n\}}] = 0 \quad (4.21)$$

En ese caso se verifica igualmente (4.18). Veamos una situación bastante natural en la que este resultado es de aplicación.

Sean $(Z_n)_n$ iid con media 0 y varianza 1 (en caso contrario basta considerar las variables tipificadas) y $(\omega_n)_n$ una sucesión de vectores donde $\omega_n = (\omega_{n1}, \dots, \omega_{nn})' \in \mathbb{R}^n$ y tal que $\|\omega_n\|^2 = \sum_{k=1}^n \omega_{nk}^2 = 1$. Para cada $n \in \mathbb{N}$ y para $1 \leq k \leq n$ definimos $X_{nk} = \omega_{nk} Z_k$. En ese caso, $c_n = \sum_{k=1}^n \omega_{nk}^2 = 1$. Puede comprobarse fácilmente mediante el TCD que (4.21) se cumple si $\lim_{n \rightarrow \infty} \|\omega_n\|_\infty = 0$, donde $\|\omega_n\|_\infty = \max_{1 \leq k \leq n} \omega_{nk}$. En tal caso y aplicando TCL queda garantizada la convergencia en distribución a $N(0, 1)$. Tenemos pues la versión definitiva, al menos para nosotros, del Teorema Central del Límite, que atribuiremos a Huber. Se enuncia así:

Teorema 34 (TCM Huber).

Si $(Z_n)_n$ iid de media nula y varianza 1 y $(\omega_n)_n$ verifica que $\|\omega_n\| = 1 \ \forall n \in \mathbb{N}$, y $\|\omega_n\|_\infty \xrightarrow{n \rightarrow \infty} 0$ entonces,

$$\omega_{n1}Z_1 + \dots + \omega_{nn}Z_n \xrightarrow[n \rightarrow \infty]{d} N(0, 1) \quad (4.22)$$

Nótese que el TCL clásico puede entenderse como un corolario del teorema 34 considerando $Z_k = \sigma^{-1}(X_k - \mu)$ y, para cada $n \in \mathbb{N}$,

$$\omega_n = \left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right)$$

Hablando en términos heurísticos, esta versión de TCL permite extender la convergencia de la media aritmética a una combinación lineal genérica con la condición de que las ponderaciones ω_{nk} respeten una cierta equidad. Esto puede aplicarse al problema de comparación de dos medias a partir de sendas muestras independientes de tamaños n_1 y n_2 (test de Student para muestras independientes). En ese caso, hay que aplicar de manera independiente el teorema 34 a las secuencias

$$\begin{aligned} \omega_{n_1}^1 &= \left(\frac{1}{\sqrt{n_1}}, \dots, \frac{1}{\sqrt{n_1}}, 0, \dots, 0 \right) \\ \omega_{n_2}^2 &= \left(0, \dots, 0, \frac{1}{\sqrt{n_2}}, \dots, \frac{1}{\sqrt{n_2}} \right) \end{aligned}$$

La validez del test puede probarse entonces teniendo en cuenta (5.2) y el Teorema de Slutsky. De la misma forma puede aplicarse a la comparación de r medias (anova de una factor) y, en general, a cualquier problema que pueda entenderse como una regresión lineal múltiple (Modelo Lineal).

Aplicación al Modelo Lineal: Veamos pues como la versión de TCL de Huber puede aplicarse a un problema genérico de Regresión Lineal Simple, aunque puede generalizarse sin dificultad al Múltiple. El modelo consiste en principio en suponer que existen $\alpha, \beta \in \mathbb{R}$, $\sigma^2 > 0$ y $(\mathbf{X}_i)_{1 \leq i \leq n} \subset \mathbb{R}$ tales que, para todo $i = 1, \dots, n$,

$$Y_i = \alpha + \beta \mathbf{X}_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma) \text{ independientes} \quad (4.23)$$

En tal caso, si se definen $Y = (Y_1, \dots, Y_n)'$, $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_n)'$ y

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n Y_i, \quad \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i, \quad s_y^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{y})^2, \quad s_{\mathbf{x}}^2 = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{x}})^2, \quad (4.24)$$

así como

$$s_{y\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{y})(\mathbf{X}_i - \bar{\mathbf{x}}), \quad \hat{\beta} = \frac{s_{y\mathbf{x}}}{s_{\mathbf{x}}^2}, \quad s_{\hat{\mathbf{e}}}^2 = \frac{1}{n-2} \sum [y_i - \hat{\alpha} + \hat{\beta} \mathbf{x}_i]^2 \quad (4.25)$$

se verifica que

$$t = \sqrt{n} \frac{\hat{\beta} - \beta}{s_{\hat{\mathbf{e}}}/s_{\mathbf{x}}} \sim t_{n-2} \quad (4.26)$$

Este hecho permite realizar inferencias sobre β . Eso es al menos lo que deberíamos saber si conocemos el Modelo Lineal Normal. Veamos qué ocurre cuando $n \rightarrow \infty$ si prescindimos del supuesto de normalidad. En este contexto, redefinimos para cada valor de n los parámetros en (4.24) y (4.25) como $\bar{y}[n]$, $\bar{\mathbf{x}}[n]$, etc., incluyendo el estimador de β que se denotará $\hat{\beta}_n$ y el estadístico t que se denotará t_n .

Veamos una condición suficiente para que $t_n \xrightarrow{d} N(0, 1)$. Para cada $n \in \mathbb{N}$ y cada $i = 1, \dots, n$ definimos la tipificación $z_{ni} = (\mathbf{X}_i - \bar{\mathbf{x}}[n])/s_{\mathbf{x}}[n]$ y $\omega_{ni} = n^{-1/2} z_{ni}$. En ese caso, $\|\omega_n\|^2 = 1$. La condición $\|\omega_n\|_{\infty} \rightarrow 0$ equivale a que

$$\frac{1}{\sqrt{n}} \max_{1 \leq i \leq n} |z_{ni}| \rightarrow 0 \quad (4.27)$$

Si se verifica (4.27), se sigue de (4.22) que

$$\begin{aligned} \sum_{i=1}^n \omega_{ni} \frac{\varepsilon_i}{\sigma} &= \frac{\sum_{i=1}^n (X_i - \bar{\mathbf{x}}[n])[Y_i - (\alpha + \beta \mathbf{X}_i)]}{n\sigma s_{\mathbf{x}}[n]} \\ &= \sqrt{n} \frac{\sum_{i=1}^n (X_i - \bar{\mathbf{x}}[n])(Y_i - \bar{y}[n])}{n\sigma s_{\mathbf{x}}[n]} - \sqrt{n}\beta \frac{\sum_{i=1}^n (X_i - \bar{\mathbf{x}}[n])^2}{n\sigma s_{\mathbf{x}}[n]} \\ &= \frac{s_{\mathbf{x}}[n]}{\sigma} \sqrt{n} \left(\frac{s_{y\mathbf{x}}[n]}{s_{\mathbf{x}}^2[n]} - \beta \right) = \sqrt{n} \frac{\hat{\beta}[n] - \beta}{\sigma/s_{\mathbf{x}}[n]} \xrightarrow{d} N(0, 1) \end{aligned}$$

Como $s_{\hat{\mathbf{e}}}[n] \xrightarrow{P} \sigma$, se sigue del Teorema de Slutsky que

$$t_n = \sqrt{n} \frac{\hat{\beta}[n] - \beta}{s_{\hat{\mathbf{e}}}[n]/s_{\mathbf{x}}[n]} \xrightarrow{d} N(0, 1) \quad \text{si se verifica (4.27)} \quad (4.28)$$

Así pues, no sólo hemos establecido una condición débil (especialmente si los valores \mathbf{X}_i están controlados) para garantizar la validez asintótica de (4.26) ante la violación del supuesto de normalidad, sino que hemos probado la convergencia en distribución de t_n a $N(0, 1)$ y, por lo tanto, de $F_{1,n}$ a χ_1^2 . Este resultado se puede generalizar, de lo cual se deduce que $F_{m,n}$ converge en distribución a χ_m^2 cuando $n \rightarrow \infty$.

5. PROBLEMAS PROPUESTOS

1. Supongamos que (X, Y) sigue una distribución uniforme en el círculo de radio 1. Indica cuál es su función de densidad. Calcula la distribución marginal de X y su función de distribución. Dado $x \in [-1, 1]$, calcula la densidad condicional de Y dado $X = x$, así media y su función de distribución.
2. Repite el ejercicio anterior para la distribución uniforme en el cuadro de vértices $(-1, -1)$, $(-1, 1)$, $(1, 1)$ y $(1, -1)$. ¿Sabrías indicar cuál es la diferencia fundamental entre ambos problemas?
3. Un mono dotado de vida eterna teclea indefinidamente y de forma completamente aleatoria una máquina de escribir con 50 caracteres diferentes. ¿Cuál es la probabilidad de que en algún intervalo de tiempo teclee exactamente El Quijote, carácter por carácter? ¿Cuál es la probabilidad de que teclee el Quijote infinitas veces?
4. Un empleado de un banco sustituye un billete bueno por uno falso en cada fajo de 100 billetes. Si el interventor del banco toma 50 fajos y selecciona al azar un billete en cada uno de ellos, ¿cuál será la probabilidad de que descubra al empleado?
5. En una urna hay cinco bolas blancas y otras cinco negras que se van extrayendo al azar sin reemplazamiento. ¿Cuál es la probabilidad de que en la sexta extracción se saque por tercera vez un bola blanca?
6. Consideremos las tres formas siguientes de juzgar a un presunto delincuente:
 - (a) Un único juez decide. Hay una probabilidad de 0.9 de que tome la decisión correcta.
 - (b) Se decide por mayoría entre las tres personas de un jurado: dos de ellas son responsables y toman la decisión correcta con probabilidad 0.9 y la otra es un irresponsable que absuelve o condena lanzando una moneda al aire.
 - (c) Se decide por mayoría entre las 10 personas de un jurado que actúan independientemente tomando cada una de ellas la decisión correcta con probabilidad 0.8.

¿Con cuál de las tres formas se consigue la mayor probabilidad de tomar la decisión correcta?

7. Consideremos una función aleatoria bidimensional con función de densidad

$$f_{X,Y}(x,y) = \begin{cases} ky(1-x-y) & \text{si } x, y > 0, x+y < 1 \\ 0 & \text{resto} \end{cases}$$

- (a) Calcula k , las distribuciones (marginales) de X e Y y sus momentos de orden 1.
 - (b) Calcula $P(X \leq 3/4 | Y = 1/3)$ y $P(X \leq Y)$
8. Razona por qué podemos afirmar que, si el momento de orden k de una distribución es finito, lo son todos los momentos de orden inferior.
9. Un huevo de cierto insecto da lugar a un nuevo insecto con probabilidad p . Es conocido que el número de huevos puestos por estos insectos en una flor sigue una distribución de Poisson de parámetro λ .
- (a) Determina la distribución de probabilidad del número de insectos que nacen en una flor.
 - (b) Se observa una flor y se ha visto que el número de insectos que han nacido en ella es n . Estudia la distribución de la variable “número de huevos que habrá en dicha flor”.
10. Consideremos una moneda con probabilidad p de salir cara. Se lanza n veces y se introducen en una urna tantas bolas blancas como caras se hayan obtenido y tantas bolas negras como cruces. Después se extraen r bolas de la urna sin reemplazamiento.
- (a) Determina la distribución de probabilidad correspondiente a la variable aleatoria “número de bolas blancas obtenidas”.
 - (b) Si se han obtenido y bolas blancas, determina la probabilidad de que en la urna queden j .
11. Dados $\alpha, \lambda \in \mathbb{R}$, consideremos la variable aleatoria X con función de densidad

$$f(x) = \begin{cases} \lambda \alpha (\lambda x)^{\alpha-1} e^{-(\lambda x)^\alpha} & \text{si } x > 0 \\ 0 & \text{resto} \end{cases}$$

- (a) Comprueba que es una función de densidad.
 - (b) Calcula $E[X]$.
 - (c) Calcula la función generatriz de $\log X$.
12. Sean X e Y iid con densidad

$$f(t) = \begin{cases} 2e^{-2t} & \text{si } t \geq 0 \\ 0 & \text{resto} \end{cases}$$

- (a) Calcula la función de densidad y la media de X^3 .

- (b) Calcula la función de densidad de $X + Y$.
- (c) Calcula las funciones de distribución y densidad de $\max\{X^3, Y^3\}$
13. Considera la variables aleatoria X con función de densidad
- $$f(x) = \begin{cases} \frac{e^x}{2} & \text{si } x \leq 0 \\ \frac{e^{-x}}{2} & \text{si } x > 0 \end{cases}$$
- (a) Comprueba que es una función de densidad.
- (b) Calcula su función de distribución.
- (c) Calcula su media.
- (d) Calcula $P(\{|X| \leq 2\} \cap \{X \geq 1\})$
- (e) Calcula el cuantil de orden 0.75 de la distribución.
14. Obtén, utilizando los cambios de variables adecuados, la función de densidad de la distribución $F_{mn,n}$.
15. Obtén, utilizando los cambios de variables adecuados, la función de densidad de la distribución t_n .
16. Este problema es muy completo: obtén todos los momentos y funciones generatrices de momentos que aparecen en la tabla [2.1](#).
17. Prueba que, si X_1, \dots, X_n son independientes y normalmente distribuidas, el vector aleatorio n -dimensional que componen es normal n -dimensional.
18. Prueba que, si X es un vector aleatorio n -dimensional con media μ y matriz de covarianzas Σ tal que $a'X \sim N(a'\mu, \sqrt{a'\Sigma a})$ para todo $a \in \mathbb{R}^n$, entonces $X \sim N_n(\mu, \Sigma)$.
19. Determina las funciones de densidad, característica, media y matriz de covarianzas de la distribución [\(5.1\)](#). Determina la distribución marginal de X y la condicional de Y dado $X = x$, indicando sus respectivas densidades.
- $$\begin{pmatrix} Y \\ X \end{pmatrix} \sim N_2 \left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0.7 \\ 0.7 & 1 \end{pmatrix} \right) \quad (5.1)$$
20. Prueba que, bajo los supuestos del Modelo Lineal Normal, el estadístico de contraste del test de Student sigue un modelo de distribución $t_{n_1+n_2-2}$.
21. Prueba que, bajo los supuestos del Modelo Lineal Normal, el estadístico de contraste del test F para el anova de un factor con k niveles sigue un modelo de distribución $F_{k-1, n-k}$, siendo $n = n_1 + \dots + n_k$.
22. Prueba que, bajo los supuestos del Modelo Lineal Normal, el estadístico de contraste de la hipótesis $H_0 : \beta_1 = 0$ sigue un modelo de distribución t_{n-2} . ¿Qué modelo de distribución sigue según la hipótesis H_0 el coeficiente de determinación r^2 ?

23. Razona por qué la varianza muestral de una muestra aleatoria de tamaño n si converge en probabilidad a la varianza probabilística cuando $n \rightarrow \infty$.
24. Razona cómo podemos construir un intervalo al 95 % de confianza para la probabilidad π de un suceso a partir de las proporciones de resultados favorables, p_n , tras n réplicas independientes del fenómeno aleatorio, suponiendo que n sea lo suficientemente grande.
25. ¿Cómo puede utilizarse el resultado del apartado anterior en el estudio de una proporción poblacional, como puede ser la prevalencia de la diabetes II?
26. ¿Qué tipo de fenómenos pueden modelizarse mediante un modelo de distribución lognormal?
27. ¿Es cierto que si X_n y Y_n son independientes entre sí para todo n y las sucesiones $(X_n)_n$ y $(Y_n)_n$ convergen en distribución a X y Y , respectivamente, entonces X y Y son también independientes?
28. Considera dos sucesiones variables aleatorias independientes entre sí $(X_n)_n$ y $(Y_m)_m$, tales que

$$\sqrt{n} \frac{X_n - \mu_1}{\sigma} \xrightarrow[n \rightarrow \infty]{d} N(0, 1), \quad \sqrt{m} \frac{Y_m - \mu_2}{\sigma} \xrightarrow[m \rightarrow \infty]{d} N(0, 1)$$

Prueba que

$$\frac{X_1 - Y_m - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \xrightarrow[n, m \rightarrow \infty]{d} N(0, 1) \quad (5.2)$$

29. Calcula el coeficiente de asimetría de una distribución Bernoulli de parámetro p . Relaciónalo con la velocidad de convergencia de $B(n, p)$ a $N(np, \sqrt{np(1-p)})$.
30. La aplicación del TCL otorga validez asintótica al test de Student para muestras independientes si los tamaños de muestra verifican cierta propiedad. ¿Cuál? ¿Afecta este hecho a la probabilidad de error de tipo I, de tipo II o a ambas?
31. Razona en qué condiciones el resultado (p -valor) obtenido mediante el anova de un factor puede considerarse aproximadamente válido si se viola el supuesto de normalidad.
32. Supongamos conocido que, si r_n denota el coeficiente de correlación de Pearson, correspondiente a una muestra aleatoria de tamaño n de una distribución normal bidimensional con coeficiente de correlación ρ , se verifica que

$$\sqrt{n}(r_n - \rho) \xrightarrow{d} N(0, 1 - \rho^2)$$

Utiliza la denominada transformación de Fisher g para lograr una convergencia a $N(0, 1)$.

$$g(x) = \frac{1}{2} \log \frac{1+x}{1-x}$$

33. Sea $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ con media μ y varianza finita $\sigma^2 > 0$, $(X_n)_{n \in \mathbb{N}}$ iid según P^X y, para cada $n \in \mathbb{N}$, $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Razona formalmente si se verifica o no que

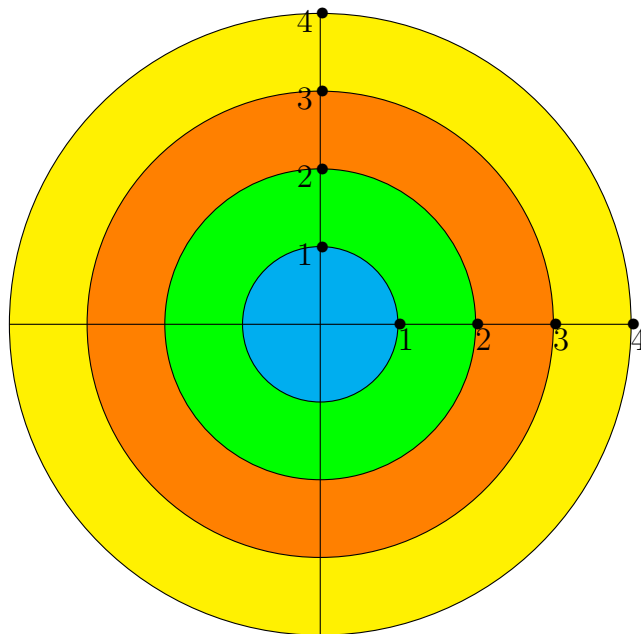
$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{P} 0$$

34. Sea $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ con media μ y varianza finita $\sigma^2 > 0$, $(X_n)_{n \in \mathbb{N}}$ iid según P^X y, para cada $n \in \mathbb{N}$, $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Razona formalmente si se verifica o no que

$$\frac{(\bar{X}_n - \mu)^2}{\sigma/\sqrt{n}} \xrightarrow{P} 0$$

35. Sea $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ con varianza finita $\sigma^2 > 0$ y coeficiente de curtosis $\tau_2 = 2$. Si $(X_n)_{n \in \mathbb{N}}$ iid según P^X , construye de manera razonada un intervalo de confianza para $\log \sigma^2$ que sea válido asintóticamente. **Indicación:** : utiliza TCL para $(X - \mu)^2$ junto con el método delta.

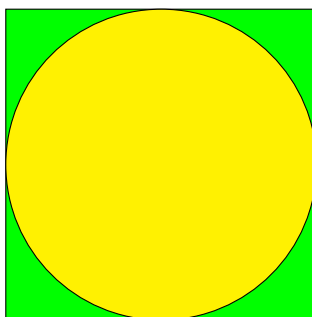
36. Se produce un bombardeo de n partículas que impactan con total seguridad en la diana de manera independiente e idénticamente distribuida según un modelo uniforme. Supongamos que cada impacto puntúa 420 si se produce en la zona azul, 140 en la zona verde, 84 en la naranja y 60 en la amarilla.



Calcula la probabilidad de que, de $n = 16$ impactos, al menos uno de ellos se localice en el círculo azul. Puedes aproximar dicha probabilidad en virtud de resultados conocidos pero razonándolo adecuadamente.

37. En las condiciones del ejercicio anterior, calcula la probabilidad de que, tras $n = 100$ impactos, la puntuación media obtenida sea superior a 100. Puedes aproximar dicha probabilidad en virtud de resultados conocidos pero razonándolo adecuadamente.

38. Una placa está sometida a un bombardeo de partículas, de tal forma que tras un minuto se registran 132 impactos; el siguiente minuto se registran 98 impactos; el tercer minuto 124 y el cuarto, 167. Suponiendo que estamos ante réplicas en idénticas condiciones de un mismo fenómeno aleatorio, razona cuál sería un valor aproximado de la probabilidad de que en el quinto minuto se registren menos de 100 impactos.
39. Imagina que necesitas conocer el cuantil 0.95 de la distribución χ^2_{100} pero sólo dispone de las tablas de χ^2_n con $n < 20$. Razona cómo podrías averiguarlo.
40. **Método de Montecarlo.** Imaginemos que desconocemos el valor del número π y pretendemos estimarlo partiendo de tres premisas:
- Podemos bombardear con $n = 10000$ partículas el cuadrado de la figura, de manera que las localizaciones de los impactos sean independientes e idénticamente e distribuidas según un modelo de distribución uniforme en el cuadrado.
 - Podemos discriminar si cada partícula impacta o no en el círculo inscrito. De hecho, resulta que 7850 impactan y el resto, no.
 - Conocemos la función de distribución Φ de $N(0, 1)$.



Construye de manera razonada y a partir de los resultados obtenidos tras los 10000 impactos un intervalo aproximado de confianza al 95 % para el valor de π .

41. El examen de cierta asignatura consiste en 100 preguntas tipo test. Con cada pregunta se acompañan 5 posibles respuestas (sólo una de las 5 respuestas es correcta). La puntuación del examen se realiza de la siguiente forma: por cada pregunta correcta se suma un punto, por cada pregunta incorrecta se resta 0.25 puntos y por cada pregunta no contestada ni se suman ni se restan puntos. Si la puntuación total es positiva, entonces la calificación asignada al examen es dicha puntuación dividida entre 10 y, si es negativa o cero, se califica con un 0. Un estudiante que se presenta a dicho examen decide seguir la siguiente regla de actuación: Ante cada una de las 100 preguntas, lanza una moneda correcta y, si sale cara, contesta la pregunta; en caso contrario no la contesta. Cuando ha de contestar la pregunta lo hace eligiendo al azar una entre las 5 posibles respuestas. Determina de forma aproximada la probabilidad de que dicho estudiante obtenga calificación 0 en el examen.