Resumen de Proba

Leopoldo Lerena

Resumen

Este es un resumen muy abreviado de la cursada de Probabilidad y Estadística de la licenciatura en Cs. Matemáticas de UBA.

Índice

1. Definiciones basicas

De ahora en más X es un conjunto cualesquiera. En esta parte están las definiciones más básicas del objeto de estudio de la materia: los espacios de probabilidad. Aparecen también algunos resultados que en sí no son propios de la probabilidad sino que valen para todos los espacios de medida pero resultan muy útiles en subsequentes demostraciones.

Definición. Una σ - álgebra Ω es un subconjunto de $\mathcal{P}(X)$ que cumple las siguientes propiedades.

- Sea $(A_n)_{n\in\mathbb{N}} \in \Omega$ luego $\cup_{n\in\mathbb{N}} A_n \in \Omega$
- Si $A \in \Omega \implies X \setminus A \in \Omega$

Definición. Una función de probabilidad es una función $\mathbb{P} : \Omega \to \mathbb{R}_{\geq 0}$ que cumple las siguientes condiciones.

- $\mathbb{P}(X) = 1$
- si $A \subset B \implies \mathbb{P}(A) \leq \mathbb{P}(B)$
- si $A \cap B = \emptyset \implies \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$

Definición. Definimos un espacio de probabilidad como un triple $(X, \Omega, \mathbb{P})$.

Teorema 1. *Sea $(A_n)_{n\in\mathbb{N}}$ una sucesión creciente de eventos.*

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cup_{n\in\mathbb{N}} A_n).$$

Análogamente si $(B_n)_{n\in\mathbb{N}}$ una sucesión decreciente de eventos.

$$\lim_{n\to\infty} \mathbb{P}(B_n) = \mathbb{P}(\cap_{n\in\mathbb{N}} B_n).$$

Teorema 2 (Fórmula de inclusión-exclusión)**.** *Sean $(X, \Omega, \mathbb{P})$, $A_1 \dots A_n$ eventos.*

$$\mathbb{P}\left(\cup_{i=1\dots n} A_i\right) = \sum_{k=1}^{n} \sum_{\substack{\mathcal{J}\in\{1,\dots,n\},\\ \#((\mathcal{J})=k)}} (-1)^{k+1}\, \mathbb{P}\left(\cap_{i\in\mathcal{J}}\right)$$

Sean las sig sumas,

$$S_k = \sum_{1\leq i_1,\dots,i_k\leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k})$$

Una manera de generalizar estos resultados anteriores es considerar las siguientes desigualdades:

Teorema 3 (Desigualdad de Bonferroni). *En el caso que k es impar,*

$$\mathbb{P}(\bigcup A_i) \leq \sum_{j=1}^{k} (-1)^{j-1} S_j.$$

Mientras que en el caso que k es par,

$$\mathbb{P}(\bigcup A_i) \geq \sum_{j=1}^{k} (-1)^{j-1} S_j.$$

2. Independencia de variables aleatorias

Es la primer idea básica de la materia. Es de vital importancia para la resolución de problemas más aún que para demostrar otros resultados más avanzados. Definiciones y teoremas fáciles pero de gran importancia.

Definición. Calculamos la probabilidad condicional de un evento A dado B de la siguiente manera.

$$\mathbb{P}(A|B) = \left(\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \right)$$

Teorema 4 (Regla de la probabilidad total). *Sea $\{B_i\}$ una partición en conjuntos disjuntos de X. Podemos recuperar la probabilidad de un evento A cualesquiera por la siguiente formula*

$$\mathbb{P}(A) = \sum_{i} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

Teorema 5 (Formula de Bayes). *Es una formula para obtener la probabilidad de B_i condicionada con A si tenemos la otra, es decir la de A condicionada con B_i. Consideramos los $\{B_i\}$ de 4.*

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(B_i|A)}{\sum_i \mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Demostración. Es una consecuencia rápida de 4. □

3. Funciones de distribución

Sea una función $X : \Omega \to \mathbb{R}$ tal que a cada evento nos asigna un número. Nos gustaría entender como es que estas funciones tienen distribuidas la probabilidad, es decir como se comporta en cierta manera más cualitativa.

Es un nuevo enfoque para ver las variables aleatorias , nos permite *clasificarlas* con respecto a su distribución de la proba. Hay muchos tipos de distribuciones fundamentales pero no voy a escribir todas sino algunos resultados de caracterización que resultan importantes.

Definición. Una funcion de distribucion de una v.a X, es una función $F : \mathbb{R} \to [0, 1]$ definida de la siguiente manera.

$$F(x) = \mathbb{P}\left(X^{-1}(-\infty, x)\right)$$

Acá es fundamental notar que para que la probabilidad tenga sentido es fundamental que $X^{-1}(-\infty, x)$ sea un evento en la σ-álgebra. Esto es lo único que le pedimos a X para poder ser llamada una variable aleatoria .

Propiedades. *La función de distribución F cumple lo siguiente.*

1. *F es monótona no decreciente*
2. $\lim_{x\to\infty} F(x) = 1$ *y* $\lim_{x\to-\infty} F(x) = 0$
3. *F es continua a derecha.*

A su vez toda función que cumpla estas propiedades resulta ser una función de distribución.

Demostración. Estas propiedades son bastante visibles si uno recuerda los gráficos de estas funciones.

1. Inmediata dado que la probabilidad es una función aditiva de conjuntos.
2. Esto se sigue de que $X^{-1}(-\infty, x) \nearrow \Omega$ y dado que tenemos el resultado 1.
3. Acá de nuevo volvemos a usar el resultado 1 pero en este caso para conjuntos decrecientes. El claro contraejemplo para ver que no puede resultar continua está dado por la distribución de una variable aleatoria discreta en algún punto que tenga probabilidad positiva.

$\square$

Obs 1. F es continua $\iff \mathbb{P}(X = x) = 0 \qquad \forall x \in \mathbb{R}$

3.1. Distribuciónes importantes

Definición. A partir de una distribución F podemos armarnos una variable aleatoria que tenga esa misma distribución. Para eso tomamos una $U \sim \mathcal{U}[0,1]$ y decimos que

$$X := F^{-1}(U)$$

es la inversa generalizada de F

Definición. Decimos que una variable aleatoria exhibe la propiedad de pérdida de memoria cuando el tiempo de espera de un evento no le afecta cuánto tiempo se haya esperado. Esto es si $i, j \in \mathbb{R}$

$$\mathbb{P}(X > i + j | X > i) = \mathbb{P}(X > j)$$

Sea X una variable aleatoria continua. Vale la siguiente caracterización de las variables aleatorias que echiben la propiedad de pérdida de memoria.

Teorema 6 (Pérdida de memoria)**.**

$$X \textit{ tiene pérdida de memoria} \iff X \sim \mathcal{E}(\lambda)$$

Demostración. La vuelta requiere expandir la cuenta de la pérdida de memoria y todo se sigue. Hay que tener cuidado en la justificación con las indicadoras.

La ida consiste en un argumento más delicado con respecto a las propiedades que definen a la función e^x.

Si definimos $g(t) := \mathbb{P}(X > t)$. Como tiene pérdida de memoria sabemos que en particular vale la siguiente igualdad.

$$g(t + s) = g(t)g(s)$$

A partir de esa igualdad, lo que podemos hacer es demostrar para los $\frac{p}{q} \in \mathbb{Q}$ (por medio de inducción en q) que valen las siguientes igualdades.

$$g\left(\frac{p}{q}\right)^q = g(p)$$
$$g\left(\frac{p}{q}\right) = g(1)^{\frac{p}{q}}$$

Veamos cuanto vale la función en un racional cualesquiera, tomando logaritmo y después elevandolo, nos queda lo siguiente

$$\begin{aligned} g\left(\frac{p}{q}\right) &= e^{\log\left(g\left(\frac{p}{q}\right)\right)} \\ &= e^{\frac{p}{q}\log(g(1))} \end{aligned}$$

Entonces ya estamos. LLamamos $\lambda := g(1)$ y como la $g(t) = e^{\lambda t} \forall t \in \mathbb{Q}$ esto implica que por ser una función continua la igualdad vale para todos los números reales.

$\square$

Obs 2. Este mismo resultado vale para variables aleatorias discretas si en vez de considerar las exponenciales tomamos su contraparte, las geométricas. La cuenta es análoga.

4. Vectores aleatorios

Definición. Un vector aleatorio es una variable aleatoria que va a parar a $\mathbb{R}^d$. Muchas definiciones y propiedades son las extensiones naturales de lo visto para una variable aleatoria .

Teorema 7 (Independencia de vectores aleatorios)**.** *Las siguientes condiciones son equivalentes para que un vector aleatorio tenga componentes independientes. En este caso lo escribo para un vector puramente discreto pero vale para vectores puramente continuos. Si* $\overline{X} = (X_1, \dots, X_d)$

1. *Las componentes del vector* $\overline{X}$ *son independientes.*
2. $p_{\overline{X}}(x_1, \dots, x_d) = p_{X_1}(x_1) \dots p_{X_d}(x_d)$.
3. *Se factoriza la funcion de distribución en funciones que dependen de una única variable.*
4. *Se factoriza la función de probabilidad puntual en funciones que dependen de una única variable.*

Demostración. El teorema tiene bastantes implicaciones directas, solo hay una que requiere un poco más de trabajo y es la siguiente.
$4 \implies 2$.
Para esta demostración si sabemos que

$$p_{\overline{X}}(x_1, \dots, x_d) = q_1(x_1) \dots q_d(x_d)$$

Primero podemos ver que $p_{X_1}(x) = c_1 q_1(x_1)$. Esto sale de que si tomamos $x \in R_{x_1}$ y nos queda libre en el rango de las otras coordenadas obtenemos una seria convergente y estrictamente positiva.

$$c_i := \sum_{(x_2, \dots, x_d) \in R_{x_1} \times \dots \times R_{x_d}} q_2(x_2) \dots q_d(x_d)$$

Y finalmente para chequear que $\frac{1}{c_1 \dots c_d} q_1(x_1) \dots q_d(x_d) = p_{X_1}(x_1) \dots p_{X_d}(x_d)$ nos quedaría ver que $\frac{1}{c_1 \dots c_d} = 1$ □

El siguiente teorema es el mismo de toda la vida. Las hipótesis son las mismas y la demostración es idéntica si uno no se quiere poner tan riguroso con las integrales.

Teorema 8 (Cambio de variables)**.** *Sea $\overline{X} = (X_1, \dots, X_d)$ un vector aleatorio absolutamente continuo con densidad $f_{\overline{X}}$, $G \subseteq \mathbb{R}^d$ abierto tal que concentra la proba de $\overline{X}$ y $g : G \to U \subseteq \mathbb{R}^d$ que cumple:*

- $G = \dot{\bigcup} G_i$
- $g_i : G_i \to U$ *biyectivas*

Entonces lo que vale es lo siguiente, la variable aleatoria $Y \coloneqq g(\overline{X})$ es abs. cont. con función de densidad

$$f_Y = \sum f_{\overline{X}} \circ g_i^{-1} |Dg_i|^{-1} \chi_U$$

5. Esperanza

La *esperanza* es un operador lineal que va a los $\mathbb{R}$. Básicamente nos da el valor medio de una variable aleatoria .

El siguiente teoremita es un resultado que nos permite calcular la esperanza por medio de una fórmula alternativa.

Teorema 9. $\mathbb{E}X < \infty$ *entonces vale la siguiente fórmula*

$$\mathbb{E}X = \int_0^{+\infty} (1 - F_X(x))\, dx - \int_{-\infty}^0 F_X(x) dx$$

Demostración. La idea es bastante directa. Dividimos la integral de la esperanza en las dos partes del enunciado

$$\mathbb{E}X = \int_0^{+\infty} x f_X(x) dx + \int_{-\infty}^0 x f_X(x) dx$$

y usamos integración por partes en ambos sumandos para obtener lo que queríamos. □

5.1. Covarianza

La esperanza se puede usar para otorgar una norma al espacio vectorial de variables aleatorias . Este espacio se lo denomina $\mathcal{L}^2$. Obviamente hay muchísimas propiedades importantes de este espacio pero nosotros no las vamos a desarrollar en completo sino que vamos a ver cuáles son las analogías con el especio euclídeo. El siguiente concepto nos da una idea similar a la que el producto interno nos da en el espacio euclídeo entre 2 vectores. La idea es darnos que tan distantes dos variables aleatorias X e Y son, este número lo vamos a llamar la covarianza

$$\mathrm{cov}(X,Y) = \mathbb{E}\left[(X-\mathbb{E}X)\,(Y-\mathbb{E}Y)\right]$$

A partir de este concepto definimos el análogo para los ángulos que vamos a llamar el coeficiente de correlación y que nos dice que tan similares son, se define de la siguiente manera

$$\rho(X,Y) = \frac{\mathrm{cov}(X,Y)}{(\mathbb{V}X)(\mathbb{V}Y)}$$

se puede chequear que $|\rho(X,Y)| \leq 1$.

5.2. Desigualdades clásicas

En esta sección voy a introducir las desigualdades más importantes que usan la esperanza y que son de suma importancia (aunque las cotas sean medio bruscas la mayoría de las veces, la utilidad radica en que siempre valen).

Teorema 10. *[Desigualdad de Markov] X variable aleatoria positiva entonces vale lo siguiente*

$$\mathbb{P}\left(X > \epsilon\right) < \frac{\mathbb{E}X}{\epsilon}$$

Obs 3. La desigualdad de arriba sigue valiendo si componemos a X con una función creciente como se mantienen las desigualdades y también vale si la componemos con una función convexa usando la desigualdad de Jensen.

Teorema 11. *[Desigualdad de Chebyshev] Sea X tal que $\mu \coloneqq \mathbb{E}X < +\infty$ entonces vale la siguiente desigualdad*

$$\mathbb{P}\left(|X-\mu| > \epsilon\right) < \frac{\mathbb{V}X}{\epsilon^2}$$

5.3. Esperanza condicional

El tema de la esperanza condicional es difícil de desarrollar en su generalidad por lo tanto arrancaremos viendolo en el caso de variables aleatorias discretas. Llegaremos a una caracterización menos constructiva de la esperanza condicional y por medio de un resultado que no demostraremos, veremos que siempre existe una esperanza condicional y así lo generalizamos para variables aleatorias cualesquiera.

Definición. Dadas dos variables aleatorias X e Y discretas definidas en un mismo espacio de probabilidad decimos que la densidad de Y condicionada con X=x está dada por la siguiente fórmula

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

que es un función de y.

Definición. Definimos la esperanza condicional como la variable aleatoria dada por la siguiente fórmula

$$g(x) = \mathbb{E}(Y|X=x)$$

El siguiente teorema si bien es fácil de demostrar en el caso de las discretas es de suma importancia en las aplicaciones.

Teorema 12. *Si* $\mathbb{E}Y < \infty$ *luego para toda* $h : \mathbb{R} \to \mathbb{R}$ *acotada y medible vale que*

$$\mathbb{E}(h(X)\mathbb{E}(Y|X=x)) = \mathbb{E}(Yh(X))$$

Demostración. Solamente hay que ampliar y reacomodar la serie y fácilmente se obtiene el resultado. □

Obs 4. Este resultado nos dice que en particular vale lo siguiente

$$\mathbb{E}(\mathbb{E}(Y|X=x)) = \mathbb{E}(Y)$$

Con esta caracterización vamos a basarnos la nueva definición, equivalente, de una esperanza condicional pero en este caso va a ser más general.

Definición. Una variable aleatoria Z es la esperanza condicional de $Y|X$ si cumple

1. Existe g medible tal que $Z = g(X)$
2. Vale la siguiente igualdad para toda h medible y acotada, $\mathbb{E}(h(X)Z) = \mathbb{E}(Yh(X))$

Obs 5. Aunque no lo veremos, siempre existe esta variable aleatoria definida como la esperenza condicional y es rápido de chequear que tiene que ser única.

6. Convergencias de v.a

En esta parte hay muchos resultados importantes de convergencia de variables aleatorias . Los resultados se corresponden con las implicaciones de los distintos tipos de convergencia, con resultados que lo relacionan con convergencia de series que pueden ser más fáciles de manejar y con los grandes teoremas de proba como la ley de los grandes números fuerte.

Primero veamos unas condiciones equivalentes a las distintas convergencias. En el caso de la convergencia c.s podemos llevarla a una de límites superiores que parece horrible pero tiene la ventaja de ser muy útil dado que podemos usar Borel Cantelli.

Teorema 13. *Tenemos la siguiente equivalencia.* $\forall \epsilon > 0$ *vale que* $\limsup_n \mathbb{P}(|X_n - X| > \epsilon) = 0 \iff X_n \overset{c.s}{\to} X$

Demostración. Para la ida consideremos al siguiente evento

$$B_k = \left\{ \limsup A_n(\frac{1}{k}) \right\}^C$$

tal que $\mathbb{P}(B_k) = 1$. Podemos verificar que $\mathbb{P}(\cap B_k) = 1$. Entonces ahora nos basta tomar para cada ω un k lo suficientemente grande para que

$$\frac{1}{k} < \epsilon$$

y así confirmar que $X_n(\omega) \to X(\omega)$.

La vuelta es simplemente notar que la definición de límite superior es lo mismo que aparezca infinitas veces. Pero como converge sobre todo evento debe ser que no es posible que aparezca infinitas veces.

□

Teorema 14. *[Convergencia acotada] Sean* $X_n \overset{P}{\to} X$ *y* $|X_n|, |X| < M$ *entonces vale que*

$$\mathbb{E}(X_n) \to \mathbb{E}(X)$$

Demostración. Queremos ver que $\mathbb{E}(|X - X_n|) \to 0$. La idea simplemente es dividirla en dos partes a esta esperanza.

$$\mathbb{E}(|X - X_n|) = \mathbb{E}(|X - X_n|)\chi_{\{|X - X_n| < \epsilon\}} + \mathbb{E}(|X - X_n|)\chi_{\{|X - X_n| > \epsilon\}}$$

Sabiendo que tenemos convergencia en proba podemos acotar a $\mathbb{P}(|X - X_n| > \epsilon)$ por un δ adecuado. Entonces el primer sumando lo acotamos por arriba por ϵ y al segundo sumando por $2M\delta$. □

Queremos demostrar una equivalencia de convergencia en distribución. Para eso requerimos de un lema técnico de demostrar pero sin dudas un resultado muy fuerte.

Lema 15. *[Skorohod] Sean* F_n, F *funciones de distribución tales que tenemos convergencia c.t.p* $F_n(t) \to F(t)$ *luego existen variables aleatorias* X_n, X *tales que* $X_n \sim F_n, X \sim F$ *y que* $X_n \overset{c.s}{\to} X$

Teorema 16.

$$X_n \overset{D}{\to} X \iff \mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \textit{para toda } f \textit{ continua y acotada}$$

Demostración. La ida es un poco más fácil que la vuelta que requiere más manipulaciones.

$\Longrightarrow$ Para esto usamos el lema de Skohorold 15, que nos garantiza la existencia de Y_n, Y tales que $X_n \sim Y_n, X \sim Y$ y $Y_n \stackrel{c.s}{\to} Y$. Luego al componerla con una función continua obtengo nuevas variables aleatorias y se preserva la convergencia casi segura. Esto es que $f(Y_n) \stackrel{c.s}{\to} f(Y)$. Como f es acotada podemos usar el teorema de convergencia acotada 14 y obtenemos como conclusión que $\mathbb{E}f(Y_n) \to \mathbb{E}f(Y)$. Solo queda verificar el detalle que $\mathbb{E}f(Y_n) = \mathbb{E}f(X_n)$.

$\Longleftarrow$ Sería ideal que la función $f_0(y) = \chi_{\{y \leq x\}}$ sea continua para luego usar que

$$\mathbb{E}f_0(X_n) = \mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x)$$

pero esto no va a poder ser tan directo. Para eso vamos a ensanguchar a f_0 con dos funciones continuas que la aproximen [1]. Finalmente tomando límites verificamos que se cumple $\mathbb{E}f_0(X_n) \to \mathbb{P}(X \leq x)$ como queríamos ver.

□

6.1. Implicaciones de convergencias

Propiedades. *Estas implicaciones valen siempre sin pedir más.*

$$X_n \stackrel{c.s}{\to} X \implies X_n \stackrel{P}{\to} X \implies X_n \stackrel{D}{\to} X$$

El siguiente resultado nos da una pequeña vuelta a la primer implicación.

Teorema 17.

$$X_n \stackrel{P}{\to} X \implies X_{n_k} \stackrel{c.s}{\to} X$$

Demostración. La idea es tomar una sucesión que esté acotada por algo sumable ya que tenemos el resultado más fuerte 13. Tomamos la siguiente sucesión indexada por k.

$$\mathbb{P}\left(\left\{\omega \in \Omega : |X_n - X| > \frac{1}{k}\right\}\right) < \frac{1}{2^k}.$$

□

El siguiente resultado aunque no lo parezca es un resultado muy importante ya que sirve como resultado previo a la ley de los grandes números débil. Es una vuelta de cuándo vale que conv. en dist. implica conv. en proba.

Teorema 18. *[Slutsky] $X_n \stackrel{D}{\to} X, Y_n \stackrel{D}{\to} c$. Entonces vale lo siguiente.*

1. $Y_n \stackrel{P}{\to} c$

2. $X_n + Y_n \stackrel{D}{\to} X + Y$

[1] Acá hace falta que sean f_ϵ idénticas salvo en un intervalito de tamaño ϵ donde crece de manera continua.

3. $X_n Y_n \xrightarrow{D} XY$
4. $\dfrac{X_n}{Y_n} \xrightarrow{D} \dfrac{X}{c}$

Demostración. Solo escribo las demostraciones de la primera y la segunda dado que las otras se parecen en esencia a la segunda.

1. Queremos escribir la convergencia en proba como una de distribución entonces basta con controlar

$$\mathbb{P}(|Y_n - c| > \epsilon) = F_n(c - \epsilon) + 1 - F_n(c + \epsilon)$$

donde ambos sumandos tienden a 0 debido a que $X_n \xrightarrow{D} c$.

2. La demostración es un poco más rebuscada. Vamos a usar fuertemente el resultado 16 y acotar con cuidado.

Sea g continua y acotada, lo que queremos acotar es lo siguiente

$$|\mathbb{E}g(X_n + Y_n) - \mathbb{E}g(X + c)|$$

entonces intercalando y usando la desigualdad triangular obtenemos

$$\leq |\mathbb{E}g(X_n + Y_n) - \mathbb{E}g(X_n + c)| + |\mathbb{E}g(X_n + c) - \mathbb{E}g(X + c)|$$

Para el segundo sumando, usamos el resultado 14. Chequeamos que la función $g(\cdot + c)$ es continua y acotada, y que $X_n \xrightarrow{D} X$ por lo tanto vemos que este término tiende a 0.

Para el primer sumando vamos a usar un truco muy interesante de indicadoras de conjuntos elegidos con especial cuidado. Como $Y_n \xrightarrow{P} c$ entonces existe $M(\epsilon)$ tal que $\mathbb{P}(|X| > M(\epsilon)) < \epsilon$. Lo que vamos a hacer es reescribir al sumando de la siguiente manera

$$|\mathbb{E}g(X_n + Y_n) - \mathbb{E}g(X_n + c)| \left(\chi_{\{X_n < M(\epsilon)\}} + \chi_{\{X_n \geq M(\epsilon)\}}\right) \left(\chi_{\{|Y_n - c| < \delta\}} + \chi_{\{|Y_n - c| \geq \delta\}}\right)$$

y esto nos da la ventaja de expandir los productos para obtener diferentes sumandos que se pueden acotar con la continuidad uniforme si restringimos el dominio (en el caso en especial que $X_n < M(\epsilon)$ y $|Y_n - c| < \delta$) y sino acotando por arriba a la función y tomando límites en ϵ, δ.

$\square$

El siguiente teorema es muy útil en las demostraciones y no es muy complicado de demostrar.

Teorema 19 (Borel-Cantelli)**.** *Sea A_n una flia de eventos, $A := \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_n$*

1. *Si $\sum \mathbb{P}(A_n) < +\infty \implies \mathbb{P}(A) = 0$*
2. *Si A_n son independientes y $\sum \mathbb{P}(A_n) < +\infty \implies \mathbb{P}(A) = 1$*

Demostración. Ambas demostraciones son directas salvo la segunda que requiere un poco de ingenio al final.

Sabemos que

$$\mathbb{P}(A) = \mathbb{P}(\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} A_n) \leq \mathbb{P}(\bigcup_{k\geq n} A_n) \leq \sum_{k\geq n} \mathbb{P}(A_n)$$

Y esta es la cola de una serie conv. entonces se va a 0.

2. Quiero ver que $\mathbb{P}(\Omega \setminus A) = 0$.

Para eso vemos que $\Omega \setminus A = \bigcup_{k\in n} \bigcap_{n\geq k} \Omega \setminus A_n$. Como es una unión creciente tenemos continuidad del límite de las probabilidades 1.

$$\mathbb{P}(\Omega \setminus A) = \lim_{k\to\infty} \mathbb{P}(\bigcap_{n\geq k} \Omega \setminus A_n)$$

Y ahora usamos la hipótesis fundamental de la independencia para escribir todo como un producto y con un poco de magia de análisis encontrar lo que buscabamos [2].

$$\lim_{k\to\infty} \mathbb{P}(\bigcap_{n\geq k} \Omega \setminus A_n) = \prod_{n\geq k} 1 - \mathbb{P}(A_n) \leq e^{-\sum_{n\geq k} \mathbb{P}(A_n)} = 0$$

Y como vimos que $\mathbb{P}(\Omega \setminus A) = 0 \implies \mathbb{P}(A) = 1$

□

El siguiente resultado nos da una condición suficiente para que las esperanzas converjan.

[2] Usamos la siguiente desigualdad, término a término, que sale del polinomio de Taylor:

$$1 - x \leq e^{-x}$$

6.2. Leyes de los grandes números

Teorema 20 (Ley débil de los grandes números)**.** *Sean $X_1, X_2, \ldots$ variables aleatorias iid con varianza σ^2 finita. Entonces vale que $\overline{X_n} \xrightarrow{p} \mu$.*

Demostración. La idea es usar Chebyshev ya que tenemos varianza finita por hipótesis y la varianza del promedio nos va a dar algo que depende inversamente en n por lo que tomando límite tiende a 0. □

Hay dos versiones del teorema. La fuerte nos da la convergencia casi segura y la demostración es un poco más difícil ya que no se usa funciones características.

Teorema 21 (Ley fuerte de los grandes números)**.** $\overline{X}_n \xrightarrow{c.s} \mu$

Demostración. Vamos a usar fuertemente el siguiente resultado 13.

Miramos la sucesión $Y_n = X_n - \mu$ tal que $\mathbb{E}(Y_n) = 0$. Lo que vamos a hacer es ver qué pasa con el término n-ésimo de la serie. Este término acotandolo con Markov nos da lo siguiente. $\mathbb{P}(\overline{Y}_n > \epsilon) \leq \dfrac{\sigma}{n\epsilon^2}$. Pero esta cota no es buena porque la seria harmónica diverge. Para eso miramos cuando tomamos los índices como n^2 y en tal caso sí converge, porque nos queda $\mathbb{P}(\overline{Y}_{n^2} > \epsilon) \leq \dfrac{\sigma}{n^2\epsilon^2}$. Entonces vamos a intercalar esto que sabemos que converge (viejo truco de análisis) y hacer algunas manipulaciones cuidadosas para llegar a lo que queríamos.

Recapitulando, queremos acotar $|\overline{Y}_k|$. Para eso notamos lo siguiente

$$\begin{aligned} |\overline{Y}_k| &= \left| \frac{S_k}{k} \right| \\ &= \left| \frac{S_k - S_{n(k)^2} + S_{n(k)^2}}{k} \right| \\ &\leq \left| \frac{S_k - S_{n(k)^2}}{k} \right| + \left| \frac{S_{n(k)^2}}{n(k)^2} \right| \end{aligned}$$

Lo que sabemos es que el segundo sumando nos da una serie convergente, por lo tanto nos tenemos que encargar del primer sumando únicamente.

Para el primer sumando miramos al conjunto

$$A_\epsilon = \left\{ \omega \in \Omega : \left| \frac{S_k(\omega) - S_{n(k)^2}(\omega)}{k} \right| \geq \epsilon \right\}$$

y usamos la desigualdad de Chebyshev para encontrar una cota que va a ser de $o(\frac{1}{n^2})$. Una vez encontrada la cota hay que terminar la demostración verificando que la serie es convergente pero estas son cuentas usuales.

□

7. Funciones caracteristicas

Las funciones características son como una funciones generatrices de las variables aleatorias . La idea es codificar bastante información en estos objetos y que sean más fáciles de manejar que las variables aleatorias. Transforman el problema de la convergencia en proba y distribución que suelen ser difíciles de manejar en un problema de convergencia de funciones en c.t.p. Todo esto se debe al gran teorema de **Paul Levy** 22 que permite dar lindas demostraciones de teoremas fundamentales de convergencia como el *TCL*. Lamentablemente este teorema no es nada trivial de demostrar, requeriría varios lemas e introducir incluso un tema nuevo. Por lo tanto si confíamos en que es cierto podemos gozar de cortas demostraciones para teoremas importantes.

Definición. La función caracteristica de una variable aleatoria X es la siguiente función.

$$\phi_X(t) \coloneqq \mathbb{E}(e^{itX})$$

Teorema 22. *[Paul Levy] Las siguientes afirmaciones son equivalentes.*

1. $\phi_n(t) \to \phi(t)$ *c.t.p.*
2. $X_n \xrightarrow{D} X$.

7.1. Consecuencias del teorema de Lévy

Una vez demostrado o asumiendo que vale el resultado de Lévy podemos pasar a las demostraciones de los posibles 2 teoremas más importantes de convergencia de variables aleatorias. Existen otras posibles demostraciones pero éstas requieren un uso mínimo de variables complejas y polinomio de Taylor.

Teorema 23 (Ley de los grandes números débil)**.** *(X_i) variables aleatorias iid tales que $\mathbb{E}(X_i) = \mu < \infty$ entonces $\overline{X_i} \xrightarrow{P} \mu$*

Demostración. La idea es escribir la función característica de $\overline{X_i}$ y usando la independencia de las X_i y el teorema de Taylor de orden 1 [3] para llegar a

$$\left(1 + \frac{i\mu t + nR(t/n)}{n}\right)^n \to e^{i\mu t}$$

esto quiere decir que las funciones características convergen a la función característica de la variable aleatoria μ y por lo tanto por 22 y 18, llegamos a lo que queríamos. $\square$

[3] Para usar el teorema de Taylor estamos usando que es derivable en 0 y eso es equivalente a que tenga esperanza finita.

Obs 6. Notemos que en el enunciado del teorema anterior solo pedimos que la esperanza fuera finita y no necesitamos ninguna hipótesis extra sobre la varianza. Esto nos dice que esta versión del teorema es más fuerte aún que la anterior.

Obs 7. Si $Z \sim N(0,1)$ entonces $\phi_Z(t) = e^{-t^2/2}$.

Teorema 24 (Teorema central del límite). *X_n i.i.d, $\mathbb{V}X_n, \mathbb{E}X_n < \infty$, luego vale lo siguiente*

$$Z_n = \frac{\overline{X_n} - \mathbb{E}\overline{X_n}}{\sqrt{\mathbb{V}\overline{X_n}}} \xrightarrow{D} Z \sim N(0,1)$$

Demostración. Análogamente que para el teorema anterior vamos a desarrollar las funciones características usando la independencia y Taylor de orden 2 en este caso dado que al tener varianza finita implica que su segundo momento también lo es. Vemos que en el límite nos queda la característica de una normal y luego usando el teorema de Lévy concluimos la demostración. □

8. Estadística

En esta sección voy a definir varios conceptos básicos de la estadística. No va a haber demostraciones pero sí descripciones de los métodos y de las definiciones.

Decimos que una muestra se corresponde con un vector aleatorio $\overline{X} = (X_1, \ldots, \overline{X}_n)$. La idea es a partir de la muestra obtener información de la variable aleatoria X. Por ejemplo si queremos saber cual es su valor medio o cualquier otro parámetro θ, lo que hacemos es mirar $\overline{\theta}(\overline{X})$ que es una función que va a aproximar a θ. A esta función la llamamos el estimador. Decimos que un estimador es insesgado si su esperanza es θ

$$\mathbb{E}\left(\overline{\theta}\left(\overline{X}\right)\right) = \theta$$

y otra propiedad que nos va a interesar es cuando converge en probabilidad al parámetro y decimos que es consistente

$$\overline{\theta}\left(\overline{X}\right) \to \theta$$

Existen dos métodos distintos para poder encontrar buenos estimadores. Uno es el llamado método de los momentos que consiste en despejar el estimador del cálculo de los momentos. El otro es el método de máxima verosimilitud que consiste en despejar el valor que maximiza la probabilidad de una observación dada.

9. Procesos

Para empezar esta sección vamos a introducir la idea central de un proceso. En pocas palabras un proceso es una extensión de las variables aleatorias . En un proceso miramos la suma de variables aleatorias i.i.d tal que su suma resulta tener otra distribución. El caso más sencillo es la suma de Bernouilli que resulta ser un proceso binomial. El caso más usado que si bien nosotros no llegamos a abarcar pero resulta sumamente importante en el desarrollo teórico se corresponde al proceso de Poisson. Para esto vamos a tener cuidado dado que el tiempo ya no va a ser más discreto sino que va a pasar a ser continuo. En este caso si definimos $\tau_i \sim \mathcal{E}(\lambda)$. Si ahora definimos

$$Y_n = \sum_{i=1}^{n} \tau_i$$

entonces decimos que la variable aleatoria nos dice el tiempo hasta el n-ésimo éxito. Observamos que $Y_n \sim \Gamma(n, \lambda)$. Ya con esto definimos el contador de éxitos que va a depender del tiempo continuo

$$N_t = \text{máx}\{k : Y_k \leq t\}$$

entonces decimos que N_t es el proceso de Poisson de parámetro λ. El resultado que habría que corroborar es que $N_t \sim \mathcal{P}(\lambda t)$.

Propiedades. *Las siguientes son algunas propiedades que definen a los procesos de cualquier tipo.*

1. *Los incrementos son independientes.*

2. *Los incrementos son estacionarios.*

Demostración. Antes que todo definamos lo que aún no definimos. Que los incrementos sean independientes quiere decir que si tomamos intervalos disjuntos del tiempo entonces los procesos van de tomando valores en esos intervalos van a ser independientes.

$$N_{t+s} - N_s \text{ es un procesos de Poisson de parámetro } \lambda \text{ independiente de } \mathbb{N}_r : r \leq s$$

Por otro lado que sea estacionario siginifica que la cantidad de llegadas en un intervalo tiene distribución a un proceso de Poisson de la medidad del intervalo. Es decir que lo que determina la cantidad de llegadas es la longitud del intervalo y no cuál intervalo es.

$$N_{t_i} - N_{t_{i-1}} \sim N_{t_i - t_{i-1}}$$

.

□

10. Cadenas de Markov

Las cadenas de Markov son procesos estocásticos (X_n) que toman valores en un conjunto numerable que denotaremos S. Si $X_n = x$ la cadena está en el estado x a tiempo n. El paso de un estado a otro no depende en lo absoluto del tiempo en qué ocurra sino del estado previo. Si juntamos todos los posibles cambios de estado en una matriz Q, obtenemos una matriz llamada la matriz de transición. La propiedad que define a estas cadenas es que no importa el pasado a la hora de condicionar. [4]

$$\mathbb{P}(X_{k+1} = y | X_k = x_k, \ldots, X_0 = x_0) = \mathbb{P}(X_{k+1} = y | X_k = x_k)$$

A partir de la definición podemos deducir las ecuaciones de Kolmogorov-Chapman que nos caracterizan el pasaje de estados como un producto de potencias de la matriz de transición.

$$Q^{m+n}(x,y) = \sum_{z \in S} Q^m(x,z) Q^n(z,y)$$

Nuestro foco ahora va a girar a los vectores de probabilidades que son vectores positivos tales que sus coordenadas suman 1. Si miramos la matriz de transición y a sus autovectores, vamos a denominar al autovector de autovalor 1 una medida invariante π. Siempre existe una medida invariante por cómo está armada la matriz de transición y éste es el vector $(1, \ldots, 1)$.

El teorema más importante de esta parte es un teorema del álgebra lineal que nos garantiza la unicidad de la medida invariante.

Teorema 25. *[Perron-Frobenius] Sea Q una matriz de transición de $n \times n$. Luego Q tiene como autovalor dominante a 1 y se cumple que*

1. *Los otros autovalores λ son tales que $|\lambda| < 1$.*
2. *El autovector asociado a 1 tiene todas sus coordenadas positivas.*
3. *Los otros autovectores son tales que sus coordenadas suman 0.*
4. *1 es un autovalor simple.*

Demostración. El más difícil de demostrar es último ítem que requiere un lema que relaciona determinantes con derivadas y cómo no se demostró en esta cursada no voy anotar su demostración, aunque se la puede leer en estas notas [5]. Para los otros ítems basta con expandir $|Q(x,y)v|$.

1 Miramos $v_k = \max_i |v_i|$, con su autovalor correspondiente λ y consideramos $|\lambda v_k|$ que con unas simples cuentas podemos acotar por arriba por $|v_k|$, por lo tanto $|\lambda| < 1$

[4] Pensar en el random walk de un borracho que no importan los primeros k pasos sino el último

[5] `http://www.math.uchicago.edu/~may/VIGRE/VIGRE2011/REUPapers/WuB.pdf`

2 y 3 Como vale la siguiente igualdad $Q^t v = \lambda v$ si consideramos la suma de las coordenadas de ambos vectores y las comparamos podemos obtener que $\sum_i v_i = \lambda \sum_i v_i$ y de esto deducimos que $\lambda = 1$ o bien que $\sum_i v_i = 0$

$\square$