Teoria de la probabilitat

Xavier Povill Clarós

24 d'octubre de 2020

Índex

1
Espais de probabilitat

La teoria de la probabilitat neix de la necessitat de trobar un model que representi fenòmens que depenen de l'atzar (o fenòmens dels quals per alguna raó no se'n pot predir el resultat a priori). A aquest tipus de fenòmens els anomenarem *experiments*.

Cada experiment té un *resultat*, pero sovint no ens interessarà el resultat en sí, sinó una pregunta concreta que ens fem sobre el resultat (per exemple, ens pot interessar saber si el dau ha tret un nombre parell, en lloc del nombre concret que ha sortit). Aquestes preguntes és el que es coneix com a *successos* o *observables*.

1.1 Experiments i probabilitat

Definició 1.1.1 (Experiment)**.** Un experiment és un parell $(\Omega, \mathcal{A})$ on Ω és un conjunt i $\mathcal{A}$ és un subconjunt de Ω tal que

- $\emptyset \in \mathcal{A}$
- Si $A \in \mathcal{A}$, aleshores $\bar{A} \in \mathcal{A}$
- Si $\{A_n\}_{n\geq 1}$ és una col·lecció numerable d'elements de $\mathcal{A}$, aleshores $\bigcup_{n\geq 1} A_n \in \mathcal{A}$

Aquestes condicions equivalen a demanar que $\mathcal{A}$ sigui una σ-àlgebra.

A Ω li direm *espai mostral*, un element $\omega \in \Omega$ serà un *resultat* i un $A \in \mathcal{A}$ serà un *succés* o *esdeveniment*.

Exemple 1.1.1.

- Tirar una moneda: $\Omega = \{0, 1\}$ (cara i creu) i $\mathcal{A} = \mathcal{P}(\Omega)$.
- Tirar una moneda fins que surti cara: $\Omega = \{1, 2, \dots\}$ i $\mathcal{A} = \mathcal{P}(\Omega)$.

- Tirar dos daus: $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$ i $\mathcal{A} = \mathcal{P}(\Omega)$. Si no sabem quin és quin, aleshores $\Omega = \{(i,j) : 1 \leq i \leq j \leq 6\}$.

A continuació definirem una funció probabilitat que a cada esdeveniment (cada element de $\mathcal{A}$) li asigni una probabilitat d'ocórrer. Per tal que aquesta sigui compatible amb la noció prèvia intuitiva de probabilitat, demanarem les següents condicions:

Definició 1.1.2 (Espai de probabilitat)**.** Un *espai de probabilitat* és una terna $(\Omega, \mathcal{A}, p)$ on $(\Omega, \mathcal{A})$ és un experiment i $p : \mathcal{A} \longrightarrow \mathbb{R}$ és una funció tal que

- $p(\emptyset) = 0$
- $p(A) \geq 0$ per qualsevol $A \in \mathcal{A}$
- Si $\{A_n\}_{n\geq 1}$ és una col·lecció d'esdeveniments dos a dos disjunts, aleshores $p\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} p(A_n)$
- $p(\Omega) = 1$

És a dir, un espai de probabilitat és un espai de mesura normalitzat, on $p(\Omega) = 1$. Una funció p que satisfà aquestes condicions s'anomena *funció de probabilitat.*

Exemple 1.1.2.

- **Espai discret**. Si Ω és numerable, aleshores ens podem agafar $\mathcal{A} = \mathcal{P}(\Omega)$. Sigui $\Omega = \{\omega_1, \omega_2, \dots\}$, aleshores podem prendre les seves probabilitats com els elements d'una sèrie que sumi 1. Habitualment escriurem $p(\{\omega_i\})$ com $p(\omega_i)$, per alleugerir la notació.
- L'**espai clàssic** és un espai discret amb $|\Omega| = N$ i $p_i := p(\omega_i) = 1/N$. En aquests espais, la probabilitat segueix la regla clàssica de *casos favorables entre casos possibles.*
- Tirar dos daus. Recordem que teníem $\Omega = \{(i,j) : 1 \leq i,j \leq 6\}$. Si considerem que tots els successos individuals són equiprobables, aleshores tenim $p(\{(i,j)\}) = 1/36$.

  Si els daus són indistingibles, aleshores $\Omega = \{(i,j) : 1 \leq i \leq j \leq 6\}$, i

  $$p(\{(i,j)\}) = \frac{2 - \delta_{ij}}{36}$$

- Tirar una moneda fins que surti una cara. Teníem $\Omega = \{1, 2, \dots\}$, $\mathcal{A} = \mathcal{P}(\Omega)$. Definim $p(\{i\}) = 1/2^i$, ja que per tal que s'hagi de tirar la moneda i vegades, les $i-1$ primeres han de ser creu i la i-èssima ha de ser cara (és a dir, fixem i resultats).
- Durada bateria mòbil. Mesurem quant triga a esgotar-se la bateria d'un mòbil. L'espai mostral és $\Omega = [0, \infty) \subset \mathbb{R}$. A la pràctica, podem considerar un límit superior per la durada de la bateria L, de manera que $\Omega = [0, L]$.

  Observem que en aquest cas no podem agafar $\mathcal{A} = \mathcal{P}(\Omega)$, ja que aleshores no existeix cap mesura compatible. Habitualment en un cas com aquest ens interessaria mesurar la probabilitat d'un interval (és a dir, que la bateria duri entre a i b hores).

Per tant, agafem com a σ-àlgebra $\mathcal{B} = \sigma(I)$, la σ-àlgebra més petita que contingui tots els intervals oberts. Aquesta σ-àlgebra existeix i és única, i s'anomena *σ-àlgebra de Borel.* Es pot construir una mesura μ compatible amb aquesta σ-àlgebra, que s'anomena *mesura de Lebesgue.* Per tal de que sigui una funció de probabilitat, la multipliquem per $1/L$, ja que $\mu([0, L]) = L$, i aleshores

$$p(\Omega) = \frac{1}{L}\mu([0, L]) = 1$$

Aquesta funció de probabilitat no és gaire útil, ja que és bastant uniforme i no modela bé la durada de la bateria d'un mòbil. Al capítol 4 veurem com construir funcions de probabilitat més interessants.

A continuació donarem una sèrie de propietats dels espais de probabilitat que es poden provar com a exercici.

Proposició 1.1.1. *Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat,*

1. *Per $r \geq 2$, siguin $A_1, \ldots, A_r \in \mathcal{A}$ dos a dos disjunts, aleshores $p\left(\bigcup_{i=1}^r A_i\right) = \sum_{i=1}^r p(A_i)$*
2. *Si $A, B \in \mathcal{A}$ i $A \subseteq B$, aleshores $p(B \setminus A) = p(B) - p(A)$ (i per tant $p(A) \leq p(B)$)*
3. *$p(\bar{A}) = 1 - p(A)$ per tot $A \in \mathcal{A}$*
4. *(Desigualtat de Boole) Si $A_1, \ldots, A_r \in \mathcal{A}$, aleshores $p\left(\bigcup_{i=1}^r A_i\right) \leq p(A_1) + \cdots + p(A_r)$*
5. *Si $\{A_i\}_{i\geq 1} \in \mathcal{A}$ i $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, aleshores $p\left(\bigcup_{i\geq 1} A_i\right) = \lim_{i\to\infty} p(A_i)$. Similarment, si $A_1 \supseteq A_2 \supseteq \ldots$, aleshores $p\left(\bigcap_{i\geq 1} A_i\right) = \lim_{i\to\infty} p(A_i)$.*

Demostració. Provarem només la propietat 5 (la resta són un exercici). Siguin $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $B_3 = A_3 \setminus A_2, \ldots$, que són dos a dos disjunts, ja que cada A_i conté a tots els A_j amb $j < i$. Per tant,

$$p\Big(\bigcup_{i\geq 1} A_i\Big) = p\Big(\bigcup_{i\geq 1} B_i\Big) = \sum_{i\geq 1} p(B_i) = \lim_{N\to\infty} \underbrace{\sum_{i=1}^{N} p(B_i)}_{p(A_N)} = \lim_{N\to\infty} p(A_N)$$

La segona proposició es demostra equivalentment passant a complementaris. □

A continuació donem una proposició que és l'equivalent en espais de probabilitat del principi de inclusió-exclusió.

Proposició 1.1.2. *Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat. Siguin $A_1, \ldots, A_r \in \mathcal{A}$. Per tot $I \subseteq [r]$, definim $A_I = \bigcap_{i\in I} A_i$ i per $k \in [r]$ definim $S_k = \sum_{|I|=k} p(A_I)$. Aleshores,*

$$p\Big(\bigcup_{i=1}^{r} A_i\Big) = \sum_{k=1}^{r} (-1)^{k+1} S_k$$

Demostració. Hi ha moltes maneres de demostrar-ho, es recomana fer-ho per inducció com a exercici. □

Si trunquem la suma en conjunts de no més de $k < r$ elements, aleshores obtindrem una desigualtat, el signe de la qual dependrà de la paritat de k:

Proposició 1.1.3 (Desigualtats de Bonferroni)**.** *Sigui* $M_T = \sum_{k=1}^{T}(-1)^{k+1}S_k$. *Aleshores,*

$$\begin{cases} p\left(\bigcup_{i=1}^{r} A_i\right) \leq M_T, & \text{si } T \text{ és senar} \\ p\left(\bigcup_{i=1}^{r} A_i\right) \geq M_T, & \text{si } T \text{ és parell} \end{cases}$$

Demostració. Novament, la demostració es pot fer per inducció. □

1.2 Probabilitat condicionada

Si tirem dos daus, la probabilitat de que almenys un dels daus hagi tret un 1 és $11/36$. Observem però que, si ens donen la suma dels daus, aleshores la probabilitat de que almenys un dels daus hagi tret un 1 es veurà influenciada pel fet que sabem la suma total. Si la suma és petita, la probabilitat serà més gran, mentra que si la suma és gran, la probabilitat serà més petita. Per quantificar aquesta intuició, s'uutilitza el concepte de *probabilitat condicionada.*

Expressat de manera més formal, sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat. Sigui $B \in \mathcal{A}$ amb $p(B) > 0$. Volem trobar una manera de calcular probabilitats si sabem del cert que B ha ocorregut.

Definició 1.2.1 (Probabilitat condicionada)**.** Sigui $B \in \mathcal{A}$ amb $p(B) > 0$, i $A \in \mathcal{A}$. Definim la *probabilitat de A condicionada a B* com

$$p(A|B) := \frac{p(A \cap B)}{p(B)}$$

Observació.

- La probabilitat de A condicionada a B ($p(A|B)$) pot ser més gran, igual o més petita que $p(A)$.
- Fixat B, podem definir una funció $A \mapsto p_B(A) := p(A|B)$ que compleix totes les condicions per ser una funció de probabilitat, tant en $(\Omega, \mathcal{A})$ com en $(\Omega, \mathcal{A}_B := \{A \cap B : A \in \mathcal{A}\})$.

Quan tenim una partició de l'espai mostral, la probabilitat condicionada ens serveix per calcular la probabilitat total:

Proposició 1.2.1 (Fórmula de probabilitat total)**.** *Sigui $B_1, \ldots, B_n$ una partició de Ω (amb $B_i \in \mathcal{A}$ i $p(B_i) > 0$ per tot i). Aleshores, si $A \in \mathcal{A}$,*

$$p(A) = \sum_{i=1}^{r} p(A|B_i)p(B_i)$$

Demostració. Donat que $B_1, \ldots, B_r$ és una partició, $A \cap B_1, \ldots A \cap B_r$ són dos a dos disjunts i la seva unió és A. Per tant, per additivitat finita,

$$p(A) = p\Big(\bigcup_{i=1}^{r} A \cap B_i\Big) = \sum_{i=1}^{r} p(A \cap B_i) = \sum_{i=1}^{r} p(B_i)p(A|B_i)$$

on la última igualtat és terme a terme a partir de la definició de probabilitat condicionada. □

Exemple 1.2.1 (Problema de la ruina del jugador (Huygens))**.** Suposem que tenim un jugador de cartes amb un capital inicial de $k \geq 1$ i que vol arribar a un objectiu de $N \geq k$. A cada torn, o bé guanya 1, o bé perd 1, equiprobablement. Suposant que el jugador no para de jugar fins que perd tots els diners o arriba al seu objectiu, quina és la probabilitat de que el jugador s'arruini?

Observem que els torns són independents entre sí. Per tant, si guanya la 1a ronda, la probabilitat d'arruinar-se serà la mateixa que si hagués començat amb un capital de $k+1$, mentre que si perd la 1a ronda, la probabilitat d'arruinar-se serà la mateixa que si hagués començat amb $k-1$.

Sigui B l'esdeveniment "a la 1a jugada 1, guanya 1" i $R_k =$ "s'ha arruinat amb capital inicial k". Aleshores,

$$p(R_k) = p(R_k|B)p(B) + p(R_k|\bar{B})p(\bar{B}) = \frac{1}{2}(R_{k+1} + R_{k-1})$$

Per tant, si definim $p_k := p(R_k)$, tenim l'equació recurrent

$$p_k = \frac{1}{2}p_{k+1} + \frac{1}{2}p_{k-1}$$

amb condicions inicials $p_0 = 1$, $p_N = 0$. Per resoldre-ho, observem que podem reescriure la recurrència com

$$\frac{1}{2}(p_k - p_{k-1}) = \frac{1}{2}(p_{k+1} - p_k)$$

i que per tant la diferència entre dos termes consecutius és constant. Aleshores, $p_k = 1 - k/N$.

Un cop hem resolt el problema, ens podríem preguntar quin és l'espai de probabilitat en el que estem. Com que el jugador podria no parar mai, considerem $\Omega = \{0,1\}^{\mathbb{N}}$ (seqüències infinites de zeros i uns). Observem que a cada real a $[0,1]$ li podem associar la seqüència de zeros i uns corresponent a la seva forma en binari de manera injectiva (no bijectiva perquè 0.110000... i 0.1011111... es corresponen les dues a $3/4$), de manera que Ω és no numerable.

Ens definim la funció probabilitat en aquest espai com $p(A) = \mu(\phi(A))$, on $\phi : \{0,1\}^{\mathbb{N}} \longrightarrow [0,1] \subset \mathbb{R}$ és l'aplicació que hem descrit abans i μ és la mesura de Lebesgue. Aleshores, prenem $\mathcal{A}$ com la antiimatge per ϕ dels borelians de $[0,1]$.

Per tant, individualment cada successió tindrà probabilitat zero, però successos com "el primer dígit és un 1" no.

Proposició 1.2.2 (Teorema de Bayes)**.** *Siguin $A, B \in \mathcal{A}$, aleshores*

$$P(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

Demostració. De la definició de probabilitat condicionada tenim que

$$p(B|A)p(A) = p(A \cap B) = p(A|B)p(B)$$

Reordenant obtenim la fórmula de Bayes. □

1.3 Independència

El concepte d'independència és clau per desenvolupar la teoria de la probabilitat i, de fet, R. Durrett diu en un dels seus llibres que la independència és allà on acaba la teoria de mesura i comença la probabilitat.

Definició 1.3.1 (Independència)**.** Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat. Direm que dos esdeveniments $A, B \in \mathcal{A}$ són independents si

$$p(A \cap B) = p(A)p(B)$$

Observació. Observem que aquesta definició quadra amb la idea de que dos successos independents no s'influeixen l'un en l'altre. Si calculem la probabilitat condicionada d'un respecte l'altre, tenim

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = p(A)$$

de manera que la probabilitat de A condicionada a B és simplement la probabilitat de A (recordem que la probabilitat condicionada només està definida si $p(B) > 0$).

A continuació es donen un parell de proposicions fàcils la demostració de les quals es deixa com a exercici:

Proposició 1.3.1.

- $\emptyset$ *i* Ω *són independents amb qualsevol* $A \in \mathcal{A}$.
- *Si* A *i* B *són independents, aleshores* $\bar{A}$ *i* B *també.*

Exemple 1.3.1. Considerem un experiment en el que tirem una moneda dos cops. Observem que els successos $A_1 =$"la 1a tirada és cara", $A_2 =$"la 2a és cara", $A_3 =$"les dues tirades són iguals" són independents dos a dos.

En canvi, $p(A_1 \cap A_2 \cap A_3) = 1/4$ però $p(A_1)p(A_2)p(A_3) = 1/8$. Per tant, no podem generalitzar directament la noció d'independència a més de dos conjunts a partir de la definició d'independència entre dos conjunts. És a dir, necessitem una noció alternativa d'independència per a conjunts de més de dos successos:

Definició 1.3.2 (Independència)**.** Sigui $\{A_i\}_{i \in I}$ una col·lecció de successos. Direm que aquests són *independents* si $\forall J \subseteq I$ amb $|J| < \infty$,

$$p(\bigcap_{j \in J} A_j) = \prod_{j \in J} p(A_j)$$

En particular, aquesta definició inclou la que havíem donat anteriorment pel cas de dos conjunts.

1.4 Espais producte

Suposem que tenim dos espais de probabilitat $(\Omega_1, \mathcal{A}_1, p_1)$ i $(\Omega_2, \mathcal{A}_2, p_2)$. Volem trobar un espai de probabilitat que ens modeli la realització successiva dels dos experiments (suposant que el 1r no influeix en el 2n). Aquest espai tindria un espai mostral $\Omega_1 \times \Omega_2$, i compliria que

$$p(A_1 \times A_2) = p_1(A_1)p_2(A_2)$$

per tot $A_1 \in \mathcal{A}_1$ i $A_2 \in \mathcal{A}_2$.

Per aconseguir-ho, no podem agafar

$$\mathcal{A}_1 \times \mathcal{A}_2 := \{A_1 \times A_2 : A_i \in \mathcal{A}_i\}$$

ja que aquest conjunt no és una σ-àlgebra, degut a que el complementari d'un element qualsevol és

$$\overline{A_1 \times A_2} = \left(\overline{A_1} \times \overline{A_2}\right) \cup \left(\overline{A_1} \times A_2\right) \cup \left(A_1 \times \overline{A_2}\right)$$

i aquest no té perquè estar contingut en $\mathcal{A}_1 \times \mathcal{A}_2$. Per tant, agafarem la σ-àlgebra generada per $\mathcal{A}_1 \times \mathcal{A}_2$, que escriurem com $\sigma(\mathcal{A}_1 \times \mathcal{A}_2)$, i que es defineix com la intersecció de totes les σ-àlgebres que contenen $\mathcal{A}_1 \times \mathcal{A}_2$.

Per construir una probabilitat, utilitzarem el següent teorema:

Proposició 1.4.1 (Tma d'extensió)**.** *Sigui p_0 una premesura en una àlgebra $\mathcal{A}_0$. Aleshores existeixen una σ-àlgebra $\mathcal{A}^* \supseteq \sigma(\mathcal{A}_0)$ i una mesura p^* tal que p^* coincideix amb p_0 en $\mathcal{A}_0$. A més, si p_0 és finita, p^* és única.*

Recordem que una àlgebra és una σ-àlgebra però tancada només per unions finites, i una premesura és com una mesura però en la que es demana que sigui numerablement additiva només per les unions numerables que pertanyen a l'àlgebra.

Per aplicar aquest teorema, agafem $\mathcal{A}_0 = \{A \subseteq \Omega_1 \times \Omega_2 : A \text{ és unió finita d'elements de } \mathcal{A}_1 \times \mathcal{A}_2\}$. Com a exercici, es pot comprovar que $\mathcal{A}_0$ és una àlgebra i que la unió finita amb que es defineix A es pot prendre a més disjunta. A més, també s'haurien de comprovar algunes propietats més per veure que es comporta bé, però no entrarem en els detalls en aquest curs.

També es pot definir un espai de probabilitat a partir d'un producte numerable d'espais de probabilitat, però això tampoc entra dins els continguts del curs.

Exemple 1.4.1 (Problema de l'agulla de Buffon)**.** Considerem un conjunt de línies paral·leles a distància L les unes de les altres. Tirem una agulla de longitud $l \leq L$. Quina és la probabilitat que l'agulla talli alguna de les línies?

Parametritzem la posició de l'agulla a partir de la menor distància del seu punt mig a alguna de les rectes (que anomenarem d), i l'angle que forma amb la direcció paral·lela a les rectes (que anomenarem α). Observem que només amb aquestes dades ja podem veure si talla o no.

Aleshores, l'agulla tallarà una recta si

$$\frac{l}{2}\sin(\pi - \alpha) = \frac{l}{2}\sin\alpha \geq d$$

Agafem $\Omega = [0, \pi] \times [0, \frac{L}{2}]$. Tant α com l estan uniformement distribuïts en els seus sengles intervals (?), de manera que

$$p(\text{talla}) = \frac{\text{àrea regió tq } d \leq (l/2)\sin\alpha}{\text{area rectangle } [0, \pi] \times [0, L/2]} = \cdots = \frac{2l}{L\pi}$$

1.5 Lemes de Borel-Cantelli

Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat, i sigui $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ un conjunt numerable d'esdeveniments.

Definició 1.5.1 (Límit superior, límit inferior)**.** Definim el *límit superior* i el *límit inferior* d'una successió d'esdeveniments $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ com

$$\limsup A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k$$
$$\liminf A_n = \bigcup_{n\geq 1} \bigcap_{k\geq n} A_k$$

Observem que ambdós formen part de $\mathcal{A}$, ja que es construeixen a partir d'interseccions i unions numerables de conjunts de $\mathcal{A}$.

Observació. Observem que per un cert $\omega \in \Omega$, tindrem que $\omega \in \limsup A_n \iff \forall n \geq 1$, $\omega \in \bigcup_{k\geq n} A_k$. Per tant, els elements de $\limsup A_n$ són els elements $\omega \in \Omega$ que pertanyen a infinits A_n.

De la mateixa manera, a partir de la definició veiem que $\omega \in \Omega$ pertany a $\liminf A_n \iff \omega \in A_n$ per tot n més gran que un cert n_0.

A partir d'aquestes caracteritzacions, observem que es segueix que $\liminf A_n \subseteq \limsup A_n$.

Un cop definida la noció de límits a $\mathcal{A}$, podem enunciar els dos lemes de Borel-Cantelli:

Lema 1.5.1 (Borel-Cantelli 1)**.** *Si* $\sum_{n\geq 1} p(A_n) < \infty$*, aleshores* $p(\limsup A_n) = 0$.

Demostració. Observem que $\bigcup_{k\geq n} A_k$ és una successió d'esdeveniments encaixats decreixents, de manera que

$$p(\limsup A_n) = p\Big(\bigcap_{n\geq 1} \bigcup_{k\geq n} A_k\Big) = \lim_{n\to\infty} p\Big(\bigcup_{k\geq n} A_k\Big) \leq \lim_{n\to\infty} \sum_{k\geq n} p(A_n) = 0$$

on la última igualtat es deu a que el límit de la suma de les cues d'una sèrie convergent ha de valdre zero. $\square$

Observem que la condició de que la sèrie convergeixi és necessària ja que, per exemple, si prenem tots els A_n iguals, aleshores $p(\limsup A_n) = p(A_n)$, que potser no val zero.

En el cas que la sèrie sigui divergent, podrem donar un resultat del mateix estil si els successos $\{A_n\}_{n\geq 1}$ són independents:

Lema 1.5.2. *Si* $\sum_{n\geq 1} p(A_n) = \infty$ *i els* $\{A_n\}_{n\geq 1}$ *són independents, aleshores* $p(\limsup A_n) = 1$.

Demostració. Per utilitzar la hipòtesi d'independència, necessitem tenir interseccions de A_n's. Per tant, passem al complementari:

$$p(\overline{\limsup A_n}) = p(\bigcup_{n\geq 1}\bigcap_{k\geq n}\overline{A_k}) = p(\liminf \overline{A_n})$$

Observem que $\bigcap_{k\geq n} A_k$ és una successió en n de successos encaixats creixents. Per tant,

$$p(\bigcup_{n\geq 1}\bigcap_{k\geq n}\overline{A_k}) = \lim_{n\to\infty} p(\bigcap_{k\geq n}\overline{A_k})$$

Només hem definit la independència per a un nombre de conjunts finits, de manera que expressem la intersecció infinita com un límit d'interseccions finites:

$$p(\bigcap_{k\geq n}\overline{A_k}) = \lim_{r\to\infty} p(\bigcap_{k=n}^{r}\overline{A_k})$$

Donat que els A_k són independents, els $\overline{A_k}$ també. Per tant,

$$\lim_{n\to\infty}\lim_{r\to\infty} p(\bigcap_{k=n}^{r}\overline{A_k}) = \lim_{n\to\infty}\lim_{r\to\infty}\prod_{k=n}^{r}(1-p(A_k))$$

Utilitzant la desigualtat $1-x\leq e^{-x}$ per $x\geq 0$, podem passar el producte a suma:

$$\lim_{n\to\infty}\lim_{r\to\infty}\prod_{k=n}^{r}(1-p(A_k)) \leq \lim_{n\to\infty}\lim_{r\to\infty}\prod_{k=n}^{r}e^{-p(A_k)} = \lim_{n\to\infty}\lim_{r\to\infty} e^{-\sum_{k=n}^{r}p(A_k)}$$

I donat que la sèrie divergeix,

$$\lim_{n\to\infty}\underbrace{\lim_{r\to\infty} e^{-\sum_{k=n}^{r}p(A_k)}}_{=0} = 0$$

i aleshores

$$p(\limsup A_n) = 1 - p(\overline{\limsup A_n}) = 1.$$

□

Exemple 1.5.1. Suposem que tirem una moneda repetidament, de manera independent, i definim p_n com la probabilitat de que hagi sortit cara a l'n-èssima tirada. Si la moneda té $p_n = 1/2$, aleshores la sèrie $\sum_{n\geq 1} p_n$ és divergent i, pel segon lema de Borel-Cantelli, surt un nombre infinit de cares amb probabilitat 1.

De la mateixa manera, si $p_n = 1/n$, la sèrie també és divergent i, aplicant un altre cop el segon lema de Borel-Cantelli, sortirà un nombre infinit de cares amb probabilitat 1. En canvi, si $p_n = 1/n^2$, la sèrie convergeix i, aplicant el primer lema, sortirà un nombre finit de cares amb probabilitat 1.

2

Variables aleatòries

2.1 Definició i distribució

Definició 2.1.1 (Variable aleatòria)**.** Una variable aleatòria és una aplicació $X : \Omega \longrightarrow \mathbb{R}$ tal que $\forall B \in \mathcal{B}(\mathbb{R})$, es té que $X^{-1}(B) \in \mathcal{A}$

Observem que les variables aleatòries associen a cada resultat d'un espai de probabilitat $(\Omega, \mathcal{A}, p)$ un nombre real. Això és útil per exemple si tenim un dau amb espai mostral $\{1, \ldots, 6\}$ i volem estudiar la suma al tirar-lo dues vegades, ja que en lloc de treballar amb parelles d'enters podrem treballar directament als enters.

Observació.

- La definició de v.a. es pot expressar com que X és v.a. $\iff$ X és mesurable respecte $(\mathbb{R}, \mathcal{B})$.
- No fa falta veure que X es comporta bé amb tots els borelians, sinó que tenim que X v.a. $\iff \forall\, a \in \mathbb{R}$, $X^{-1}((-\infty, a]) \in \mathcal{A}$
- Qualsevol "expressió raonable" escrita a partir de variables aleatòries és també variable aleatòria (suma, resta, valor absolut, etc.)

Comentari. Habitualment escriurem $p(X^{-1}(B))$ com $p(X \in B)$, tot i que formalment aquesta expressió no tingui sentit. Igualment, utilitzarem expressions com $p(X \leq x)$, $p(X = x)$ o $p(X > x)$.

Un avantatge d'utilitzar variables aleatòries és que observarem que podem classificar les variables per tipus, de manera que trobarem que variables aleatòries d'experiments diferents es comporten igual, i es poden estudiar utilitzant les mateixes eines. Per veure si dues variables aleatòries són del mateix tipus, utilitzarem les seves funcions de distribució:

Definició 2.1.2 (Funció de distribució)**.** La *funció de distribució* (acumulada) d'una v.a. X és

$$\begin{aligned} F_X :& \mathbb{R} \longrightarrow [0,1] \\ & x \mapsto p(X \leq x) \end{aligned}$$

Comentari. Recordem que $p(X \leq x)$ és notació per $p(\{\omega \in \Omega : X(\omega) \leq x\})$.

Exemple 2.1.1.

- La *variable indicadora* d'un succés $A \in \mathcal{A}$ és la variable aleatòria

$$\begin{aligned} I_A : \Omega &\longrightarrow \mathbb{R} \\ \omega &\mapsto \begin{cases} 1, & \text{si } \omega \in A \\ 0, & \text{si } \omega \notin A \end{cases} \end{aligned}$$

  Observem que la funció de distribució d'una variable indicadora pren la forma següent:

$$F_{I_A}(x) = p(I_A \leq x) = \begin{cases} 0, & \text{si } x < 0 \\ 1 - p(A), & \text{si } 0 \leq x < 1 \\ 1, & \text{si } x \geq 1 \end{cases}$$

- Sigui Ω el conjunt de resultats de tirar dos daus i $X(d_1, d_2) = |d_1 - d_2|$ la variable aleatòria que ens dona la diferència absoluta entre els dos valors. Aleshores, la funció de distribució de X és

$$F(x) = \begin{cases} 0, & \text{si } x < 0 \\ 1/6, & \text{si } 0 \leq x < 1 \\ 16/36, & \text{si } 1 \leq x < 2 \\ 24/36, & \text{si } 2 \leq x < 3 \\ (...) & \end{cases}$$

Proposició 2.1.1. *Si X és v.a., la seva funció indicadora F_X satisfà*

1. *F_X és creixent.*
2. $\lim_{x\to\infty} F_X(x) = 1$ *i* $\lim_{x\to-\infty} F_X(x) = 0$
3. *F_X és contínua per la dreta* ($\lim_{h\to 0^+} F_X(x+h) = F_X(x)$).

Demostració.

1. Si $x_1 \leq x_2$, aleshores

$$F_X(x_1) = p(\{\omega \in \Omega : X(\omega) \leq x_1\}) \leq p(\{\omega \in \Omega : X(\omega) \leq x_2\}) = F_X(x_2)$$

   on hem utilitzat que el conjunt de l'esquerra està contingut en el de la dreta.

2. Exercici (es fa similar al 3).

3. Sigui $\{h_n\}_{n\geq 1}$ una successió de reals decreixent que tendeix a 0 per la dreta. Aleshores,

$$\lim_{h\to 0^+} F_X(x+h) = \lim_{n\to\infty} F_X(x+h_n) = \lim_{n\to\infty} p(\{\omega\in\Omega : X(\omega \leq x+h_n\})$$

Observem que els successos $A_n = \{\omega \in \Omega : X(\omega \leq x + h_n\}$ estan encaixats, ja que $\{h_n\}$ és decreixent i per tant $A_i \supseteq A_j$ per $i \leq j$. Per tant, el límit anterior és

$$\lim_{h\to 0^+} F_X(x+h) = p\big(\bigcap_{n\geq 1}\{\omega\in\Omega : X(\omega)\leq x+h_n\}\big) = p(\{\omega\in\Omega : X(\omega)\leq x\}) = F_X(x)$$

□

Proposició 2.1.2. *Si tenim una* $F : \mathbb{R} \longrightarrow [0,1]$ *tal que satisfaci les 3 propietats de 2.1.1, aleshores existeix una variable aleatòria* X *tal que* $F_X = F$.

Demostració. La demostració és molt tècnica i no la farem, ja que no ho utilitzarem gaire. □

Proposició 2.1.3. *Sigui* X *una v.a. amb funció indicadora* F_X. *Aleshores,*

1. $\lim_{y\to x^-} F_X(y) = p(X < x) = F_X(x) - p(X = x)$
2. $p(X > x) = 1 - F_X(x)$
3. $p(x_1 < X \leq x_2) = F_X(x_2) - F_X(x_1)$

Demostració. Exercici. □

Si tenim una variable aleatòria X, aquesta indueix una funció de probabilitat p_X en $(\mathbb{R}, \mathcal{B})$ donada per

$$p_X(B) = p(X\in B), \qquad B\in\mathcal{B}$$

A aquesta funció de probabilitat se la coneix com a *llei* de X o com a *mesura de probabilitat induïda per* X.

Aquest concepte és útil per comparar variables aleatòries, ja que ens podem oblidar de l'experiment del que provenen i classificar la variable aleatòria segons la seva llei.

2.2 Moments d'una variable aleatòria

Si tenim una variable aleatòria X que només pren un conjunt finit de valors $\{a_1, \dots, a_n\}$ aleshores podríem calcular la seva "mitjana" com

$$\sum_{i=1}^{r} a_i\, p(X = a_i)$$

Per generalitzar aquest concepte a variables contínues, que prenen un nombre infinit (i potser no numerable) de valors, es defineix l'esperança de X:

Definició 2.2.1 (Esperança)**.** L'*esperança* d'una variable aleatòria X es defineix com

$$\mathbb{E}[X] = \int_\Omega X\, dp \qquad \text{(sempre que existeixi i sigui finita)}$$

on integrem en l'espai de mesura $(\Omega, \mathcal{A}, p)$.

Proposició 2.2.1 (Propietats de l'esperança)**.** *Siguin X, Y dues variables aleatòries en el mateix espai de probabilitat,*

1. *Si $X = c$ (constant), aleshores $\mathbb{E}[X] = c$.*
2. *Si $X = \mathbb{I}_A$, aleshores $\mathbb{E}[X] = p(A)$.*
3. *Si $X \leq Y$ punt a punt, aleshores $\mathbb{E}[X] \leq \mathbb{E}[Y]$ (sempre que existeixin).*
4. *L'esperança és lineal, és a dir, $\forall a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$*

Observació. Donada una funció $f : \mathbb{R} \longrightarrow \mathbb{R}$, si $f(X)$ és una variable aleatòria (mesurable), aleshores

$$\mathbb{E}[f(X)] := \int_\Omega f(X)\, dp \qquad \text{(sempre que la integral sigui absolutament convergent)}$$

Sovint ens interessa no només conèixer l'esperança d'una variable aleatòria X, sinó també com es distribueix. Per obtenir una mesura de com es desvia la variable del seu punt mig, sembla raonable calcular $\mathbb{E}[X - \mathbb{E}[X]]$. El problema és que aquesta expressió sempre val zero, ja que

$$\mathbb{E}[X - \mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] - \mathbb{E}[X] = 0$$

Observem que els valors negatius de quan X cau a l'esquerra de l'esperança es cancel·len amb els valors positius de quan X cau a la dreta de la l'esperança. Per tant, prendrem quadrats per tal que tots siguin positius:

Definició 2.2.2 (Variància)**.** La *variància* de X és

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

Per corregir el quadrat que hem afegit, sovint es treballa amb l'arrel quadrada de la variància:

Definició 2.2.3 (Desviació típica o estàndar)**.** La *desviació típica o estàndar* de la variable aleatòria X es defineix com

$$\sigma = \sqrt{\mathrm{Var}[X]}$$

Proposició 2.2.2 (Propietats de la variància)**.** *Sigui X una variable aleatòria i $c \in \mathbb{R}$,*

1. $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$
2. $\mathrm{Var}[c] = 0$

3. $\mathrm{Var}[c + X] = \mathrm{Var}[X]$

4. $\mathrm{Var}[cX] = c^2\mathrm{Var}[X]$

Demostració.

1. $\mathrm{Var}[X] = \mathbb{E}[X^2 + \mathbb{E}[X]^2 - 2X\mathbb{E}[X]] = \mathbb{E}[X^2] + \mathbb{E}[X]^2 - 2\mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$.
2. Falta.
3. Falta.
4. Falta.

□

L'esperança i la variància són casos particulars d'un concepte més general que són els moments d'una variable aleatòria.

Definició 2.2.4 (moment k-èssim)**.** El *moment k-èssim* d'una variable aleatòria X és $\mathbb{E}[X^k]$ (només si $\mathbb{E}[|X|^k] < \infty$).
Similarment, definim el *moment k-èssim centrat* de X com $\mathbb{E}[(X - \mathbb{E}[X])^k]$ (novament, només està definit si és absolutament convergent).

Observació. Observem que $\left(\int_\Omega |X|^k\right)^{1/k}$ és una norma (el que es coneix a Anàlisi Real com a norma L_p). Per tant, podrem aplicar desigualtats d'Anàlisi Real en un context probabilístic:

Proposició 2.2.3.

1. ***Desigualtat de Cauchy-Schwartz:*** *si* $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$*, aleshores*

$$\mathbb{E}[XY]^2 \leq \mathbb{E}[X^2]\mathbb{E}[Y^2]$$

2. ***Desigualtat de Hölder:*** *si* $\mathbb{E}[|X|^p], \mathbb{E}[|Y|^q] < \infty$*, amb* $p > 1$ *i* $\frac{1}{p} + \frac{1}{q} = 1$*, aleshores*

$$\mathbb{E}[|XY|] \leq \mathbb{E}[|X|^p]^{(1/p)}\,\mathbb{E}[|Y|^q]^{(1/q)}$$

3. ***Desigualtat de Minkowski:*** *si* $\mathbb{E}[|X|^p], \mathbb{E}[|Y|^p] < \infty$ *i* $p \geq 1$*, aleshores*

$$\mathbb{E}[|X + Y|^p]^{1/p} \leq \mathbb{E}[|X|^p]^{(1/p)} + \mathbb{E}[|Y|^p]^{(1/p)}$$

Demostració. 1. Exercici. S'ha de calcular $\mathbb{E}[(X - \lambda Y)^2]$.

□

En aquest curs utilitzarem més sovint les següents desigualtats, que són pròpies de la teoria de la probabilitat:

Proposició 2.2.4 (Desigualtat de Markov)**.** *Sigui X una variable aleatòria no negativa (és a dir, tal que $X(\omega) \geq 0$ per tot ω). Sigui $a \in \mathbb{R}$ amb $a > 0$. Aleshores,*

$$p(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Demostració. Sigui $A = \{\omega \in \Omega : X(\omega) \geq a\}$. Aleshores, $X \geq a\mathbb{I}_A$ (observem que aquí és on fem servir la hipòtesi que $X \geq 0$). Per tant, per la monotonia de l'esperança,

$$\mathbb{E}[X] \geq \mathbb{E}[a\mathbb{I}_A] = a \cdot p(A) = a \cdot p(X \geq a)$$

□

Proposició 2.2.5 (Desigualtat de Txebixev)**.** *Sigui X una variable aleatòria amb $\mathbb{E}[X] = \mu < \infty$ i $\mathrm{Var}[X] = \sigma^2 < \infty$. Aleshores, per qualsevol $b > 0$, la probabilitat que la variable s'allunyi de l'esperança com a mínim b està fitada per*

$$p(|X - \mu| \geq b) \leq \frac{\sigma^2}{b^2}$$

Demostració. Observem que $b^2 > 0$ i $(X - \mu)^2 \geq 0$, de manera que podem aplicar Markov i obtenim

$$p(|X - \mu| \geq b) = p((X - \mu)^2 \geq b^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{b^2} = \frac{\sigma^2}{b^2}$$

□

2.3 Vectors de variables aleatòries

Suposem que en un únic espai de probabilitat $(\Omega, \mathcal{A}, p)$ hi tenim diverses variables aleatòries $X_1, \ldots, X_n$ que volem estudiar conjuntament.

Definició 2.3.1 (Vector aleatori)**.** Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat. Definim un *vector aleatori*, o *vector de variables aleatòries* com una aplicació

$$\bar{X} = (X_1, \ldots, X_n) : \Omega \longrightarrow \mathbb{R}^n$$

tal que $\forall B \in \mathcal{B}(\mathbb{R}^n)$, $\bar{X}^{-1}(B) \in \mathcal{A}$ (és a dir, $\bar{X}$ és mesurable).

Comentari. Els borelians a $\mathbb{R}^n$ es poden definir igual que a $\mathbb{R}$, com la menor σ-àlgebra que conté els oberts de la topologia euclidiana.

Fet. Per tal que $\bar{X}$ sigui un vector aleatori, n'hi ha prou amb que

$$\bar{X}^{-1}\left((-\infty, a_1] \times \cdots \times (-\infty, a_n]\right) \in \mathcal{A}$$

i donat que l'expressió anterior és equivalent a

$$\bigcap_{i=1}^{n} X_i^{-1}(-\infty, a_i] \in \mathcal{A}$$

tenim que $\bar{X} = (X_1, \ldots, X_n)$ és vector aleatori si $X_1, \ldots, X_n$ són variables aleatòries.

Observació. De fet, treballant amb els Borelians es pot veure que $\bar{X}$ és una vector aleatori sii cada X_i és una variable aleatòria.

Definició 2.3.2 (Vector d'esperances)**.** Donat un vector aleatori $\bar{X}$, definim el seu *vector d'esperances* com el vector format per l'esperança de cada una de les variables aleatòries que formen $\bar{X}$:

$$\mathbb{E}[\bar{X}] = \begin{pmatrix} \mathbb{E}[X_1] \\ \vdots \\ \mathbb{E}[X_n] \end{pmatrix}$$

Definició 2.3.3 (Funció de distribució conjunta)**.** La *funció de distribució conjunta* de $\bar{X} = (X_1, \dots, X_n)$ és l'aplicació

$$\begin{aligned} F_{\bar{X}} : \mathbb{R}^n &\longrightarrow [0,1] \\ (x_1, \dots, x_n) &\mapsto p(X_1 \le x_1, \dots, X_n \le x_n) \end{aligned}$$

Exemple 2.3.1. Sigui $(\Omega, \mathcal{A}, p)$ l'espai de probabilitat donat per l'experiment de tirar dos daus. Siguin $X = \mathbb{I}_{\text{1r dau parell}}$ i $Y = \mathbb{I}_{\text{ els dos daus iguals}}$. Aleshores, la funció de distribució conjunta de $\bar{X} = (X, Y)$ és

$$F_{(X,Y)} = \begin{cases} 0, & X < 0 \text{ o } Y < 0 \\ 5/12, & 0 \le X < 1 \text{ i } 0 \le Y < 1 \\ 5/6, & X \ge 1 \text{ i } 0 \le Y < 1 \\ 1/2, & 0 \le X < 1 \text{ i } Y \ge 1 \\ 1, & X \ge 1 \text{ i } Y \ge 1 \end{cases}$$

Proposició 2.3.1 (Propietats de la funció de distribució conjunta)**.** *Sigui $F_{(X,Y)}(x,y)$ una funció de distribució de dues variables aleatòries. Tenim que*

1. *Si $x_1 \le x_2$ i $y_1 \le y_2$, aleshores $F_{(X,Y)}(x_1, y_1) \le F_{(X,Y)}(x_2, y_2)$.*
2. $\lim_{x,y \to -\infty} F_{(X,Y)}(x,y) = 0, \quad \lim_{x,y \to \infty} F_{(X,Y)}(x,y) = 1.$
3. $\lim_{y \to \infty} F_{(X,Y)}(x,y) = F_X(x)$ *(funció de distribució marginal)*
4. $\lim_{h_1, h_2 \to 0^+} F_{(X,Y)}(x + h_1, y + h_2) = F_{(X,Y)}(x,y).$
5. (Propietat dels rectangles) $F_{(X,Y)}(b,d) - F_{(X,Y)}(b,c) - F_{(X,Y)}(a,d) + F_{(X,Y)}(a,c) \ge 0$ *per qualssevol $a \le b, c \le d \in \mathbb{R}$.*

Comentari. Observem que la propietat dels rectangles de la proposició anterior ens dona pel principi d'inclusió-exclusió $p(a < X \le b, c < Y \le d)$, que ha de ser més gran o igual que zero pel fet de ser una probabilitat.

Observació. No ho demostrarem, però les propietats 1, 2, 3 i 5 caracteritzen F com una funció de distribució de dues variables (és a dir, tota F que ho compleixi serà la funció de distribució conjunta d'un cert parell de variables aleatòries).

2.3.1 Independència de variables aleatòries

Suposem que tenim una funció de distribució conjunta $F_{\bar{X}}$ d'un vector aleatori $\bar{X}$. Si els successos $X_1 \leq x_1, \ldots, X_n \leq x_n$ són independents per uns certs $x_1, \ldots, x_n \in \mathbb{R}$, aleshores tenim que

$$F_{\bar{X}}(x_1, \ldots, x_n) = p(X_1 \leq x_1, \ldots, X_n \leq x_n) = \prod_{i=1}^{n} p(X_i \leq x_i)$$

Si la condició anterior es compleix per tot punt de $\mathbb{R}^n$, aleshores direm que les variables $X_1, \ldots, X_n$ són independents.

Definició 2.3.4 (Independència de variables aleatòries)**.** Direm que $X_1, \ldots, X_n$ són *variables aleatòries independents* si

$$F_{\bar{X}}(x_1, \ldots, x_n) = F_{X_1}(x_1) \ldots F_{X_n}(x_n)$$

per qualssevol $x_1, \ldots, x_n \in \mathbb{R}$.

Fet. Si $X_1, \ldots X_n$ són independents, aleshores $\forall\, B_1, \ldots, B_n \in \mathcal{B}$ es té que $p(X_1 \in B_1, \ldots, X_n \in B_n) = \prod p(X_i \in B_i)$. Per tant, podem extendre la noció d'independència a un conjunt de borelians qualssevol.

2.4 Covariància i (una) Llei dels Grans Nombres

Sovint ens interessa estudiar com estan relacionades les distribucions de dues variables aleatòries. Per tal de quantificar-ho, utilitzem la covariància:

Definició 2.4.1 (Covariància)**.** Siguin X i Y dues variables aleatòries en el mateix espai de probabilitat. Definim la *covariància* de X i Y com

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))]$$

Proposició 2.4.1 (Propietats de la covariància)**.**

1. $\mathrm{Cov}(X, X) = \mathbb{E}[(X - E(X))^2] = \mathrm{Var}(X)$
2. $\mathrm{Cov}(c, X) = 0, \quad \text{on } c \in \mathbb{R}.$
3. $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$
4. $\mathrm{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
5. $\mathrm{Cov}(aX + bY, Z) = a\,\mathrm{Cov}(X, Z) + b\,\mathrm{Cov}(Y, Z)$

Demostració. Surten directament de la definició de covariància i de les propietats de l'esperança. □

Definició 2.4.2 (Coeficient de correlació de Pearson)**.** Donades dues variables aleatòries X i Y en el mateix espai de probabilitat, calcularem el seu *coeficient de correlació* com

$$\rho_{X,Y} = \frac{\mathrm{Cov}(X,Y)}{\sigma_x \sigma_y}$$

Observació. Observem que, aplicant la desigualtat de Cauchy-Schwarz a la definició de covariància, obtenim

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{\mathbb{E}[(X-\mathbb{E}[X)^2]\mathbb{E}[(Y-\mathbb{E}[Y])^2]} = \sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)} = \sigma_X \sigma_Y$$

Per tant, tindrem sempre que $|\rho_{XY}| \leq 1$.

El coeficient de correlació ens indicarà la relació que existeix entre les dues variables.

Definició 2.4.3. Direm que X i Y són

- *incorrelades*, si $\rho_{XY} = 0 \iff \mathrm{Cov}(X,Y) = 0$
- *positivament correlades*, si $\rho_{XY} = 1$ (si una puja, l'altra també)
- *negativament correlades*, si $\rho_{XY} = -1$ (si una puja, l'altra baixa)

Observació. Quan diem que dues variables són incorrelades, estem dient intuitivament que el valor d'una no influeix en el valor de l'altra. Hem de tenir en compte però que el coeficient de Pearson només ens calcula la "correlació lineal", de manera que podria ser que $\rho_{XY} = 0$ per una certa parella de variables aleatòries X i Y però que X i Y tinguessin alguna mena de relació no lineal entre elles.

Fet: Si X i Y són variables aleatòries independents, aleshores són incorrelades. El recíproc no és cert (veure per ex. els apunts del curs de Juanjo Rué o un dels problemes del tema 3).

Observació. Si X i Y són incorrelades, aleshores

$$\mathrm{Var}(X,Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

De la mateixa manera es pot comprovar que, si $X_1, \dots, X_n$ són dos a dos incorrelades, aleshores

$$\mathrm{Var}(X_1 + \cdots + X_n) = \mathrm{Var}(X_1) + \cdots + \mathrm{Var}(X_n)$$

Quan tenim un vector de variables aleatòries, podem expressar les seves covariàncies dos a dos en forma de matriu:

Definició 2.4.4 (Matriu de covariàncies)**.** Sigui $\bar{X} = (X_1, \dots, X_n)$ un vector de variables aleatòries. Definim la seva *matriu de covariància* com

$$\mathrm{Cov}(\bar{X}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1,X_2) & \dots & \mathrm{Cov}(X_1,X_n) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \mathrm{Cov}(X_n,X_1) & \dots & \dots & \mathrm{Var}(X_n) \end{pmatrix}$$

Llei feble dels grans nombres

Siguin $X_1, \dots, X_n$ variables aleatòries amb $\mathbb{E}[X_i] = \mu \in \mathbb{R}$ i $\mathrm{Var}(X_i) = \sigma^2 \in \mathbb{R}$ per $i \in [n]$. Suposem a més que són dos a dos incorrelades. (Per exemple, podríem prendre com a experiment tirar tres daus, on $n = 3$ i X_i representa el resultat de la tirada i-èssima. En aquest cas tindríem que $\mu = 7/2$ i $\sigma^2 \cong 27.4$.)

Definim la variable aleatòria $\bar{X}_n$ com la mitjana de $X_1, \dots, X_n$:

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$$

L'esperança i la variància d'aquesta variable seran

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{n\mu}{n} = \mu$$

$$\mathrm{Var}(\bar{X}_n) = \frac{1}{n^2}\mathrm{Var}(\underbrace{X_1 + \cdots + X_n}_{\text{incorrelades 2 a 2}}) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Sigui $\varepsilon > 0$. Aplicant la desigualtat de Txebyxev, tenim que la probabilitat que $\bar{X}_n$ s'allunyi més de ε de la seva esperança està fitada per

$$p(\left|\bar{X}_n - \mu\right| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon}$$

Observem que, si fem tendir $n \to \infty$, tindrem que

$$\lim_{n\to\infty} p(\left|\bar{X}_n - \mu\right| \geq \varepsilon) = 0$$

És a dir, que si repetim un experiment moltes vegades, la mitjana d'una certa variable aleatòria al llarg de totes les iteracions de l'experiment acabarà tendint a la seva esperança amb probabilitat 1.

Aquesta és una versió més feble d'un resultat que es coneix com a *Llei dels Grans Nombres*, que veurem al capítol 6.

Exercici. Comprovar que aquest resultat també és vàlid si les X_i són 2 a 2 negativament correlades ($\rho < 0$).

Indicació: Proveu que, en aquest cas,

$$\mathrm{Var}(\sum_i X_i) \leq \sum_i \mathrm{Var}(X_i)$$

3

Variables aleatòries discretes

Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat.

Definició 3.0.1 (Variable aleatòria discreta)**.** Direm que la variable aleatòria $X : \Omega \to \mathbb{R}$ és *discreta* si $\operatorname{Im} X$ és numerable.

Observem que en la definició no ens importa si Ω és numerable o no. Lo important és que la imatge de la variable aleatòria (el conjunt de successos amb els que volem treballar) sigui numerable.

En aquest capítol suposarem que totes les variables aleatòries que es mencionin són discretes, tot i que no es digui explícitament.

3.1 Funció de probabilitat, esperança, independència

En primer lloc, veurem com els conceptes que hem definit en el capítol anterior es particularitzen pel cas de variables aleatòries discretes.

Sigui $\operatorname{Im} X = \{a_1, a_2, \dots\}$. Observem que aleshores podem escriure l'espai mostral com

$$\Omega = \bigcup_{n \geq 1} \{\omega \in \Omega : X(\omega) = a_n\}$$

Per tant, donant que tots aquests conjunts són disjunts i en tenim un nombre numerable, tenim que $\forall\, B \in \mathcal{B}$,

$$p(X \in B) = \sum_{a_i \in B} p(X = a_i)$$

Aleshores, el que coneixem com *llei* de X, $p_X(B) := p(X \in B)$ (que recordem que definia una probabilitat en $(\mathbb{R}, \mathcal{B})$), queda totalment determinada per $\{a_1, a_2, \dots\}$ i $\{p(X = a_i)\}_{i \geq 1}$.

A la funció $p(X = a_i) : \mathbb{N} \to \mathbb{R}$ l'anomenarem *funció de probabilitat* de X. Observem que no tenim un anàleg de la funció de probabilitat per variables aleatòries contínues, ja que cada $a \in \operatorname{Im} X$ té probabilitat zero.

Funció de distribució:

En una variable aleatòria discreta, la funció de distribució pren la forma

$$F_X(x) = p(X \leq x) = \sum_{a_i \leq x} p(X = a_i)$$

On s'ordenen les imatges de manera que $a_1 < a_2 < a_3 < \dots$

Esperança:

Aplicant la definició d'esperança,

$$\mathbb{E}[X] = \sum_{n \geq 1} a_n p(X = a_n) = \sum_{x \in \operatorname{Im} X} x p(X = x) \qquad \text{(sempre que convergeixi absolutament)}$$

de manera que en lloc de treballar amb integrals, podem treballar únicament amb sèries.

Similarment, l'esperança d'una funció de X queda

$$\mathbb{E}[g(X)] = \sum_{x \in \operatorname{Im} X} g(x) p(X = x)$$

Independència:

Siguin X, Y variables aleatòries. Per definició, X i Y són independents si $F_{(X,Y)}(x, y) = F_X(x)F_Y(y)$. Quan les variables aleatòries són discretes, tenim una manera més fàcil de comprobar-ho:

$$F_{(X,Y)}(x, y) = F_X(x)F_Y(y) \quad \forall\, x, y \in \mathbb{R} \iff p(X = x, Y = y) = p(X = x)p(Y = y) \quad \forall\, x, y \in \mathbb{R}$$

Per convèncer-nos d'això, anem a comprovar que dues variables discretes independents són incorrelades:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_{\substack{u \in \operatorname{Im} X \\ u \neq 0}} u \cdot p(XY = u) = \sum_{\substack{x \in \operatorname{Im} X \\ x \neq 0}} \sum_{\frac{u}{x} \in \operatorname{Im} Y} u \cdot p(X = x, Y = \frac{u}{x}) = \\
&= \sum_{\substack{x \in \operatorname{Im} X \\ x \neq 0}} \sum_{\frac{u}{x} \in \operatorname{Im} Y} u \cdot p(X = x) p(Y = \frac{u}{x}) = \sum_{\substack{x \in \operatorname{Im} X \\ x \neq 0}} x \cdot p(X = x) \sum_{\frac{u}{x} \in \operatorname{Im} Y} \frac{u}{x} \cdot p(Y = \frac{u}{x}) = \\
&= \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

Observem que hem utilitzat que XY és una variable aleatòria discreta si X, Y són variables aleatòries discretes.

3.2 Funcions generadores de probabilitat

Les funcions generadores de probabilitat són una eina molt útil per treballar amb variables aleatòries discretes.

Només podrem calcular la funció generadora de variables aleatòries que comptin alguna cosa (per exemple, el nombre de persones amb samarretes vermelles a la classe, o el nombre de cotxes que passen per un carrer en un minut). Aquestes variables les anomenarem variables de recompte:

Definició 3.2.1 (Variable de recompte)**.** Una variable aleatòria és *de recompte* si $\operatorname{Im} X \subseteq \{0, 1, 2, \ldots, \}$.

Definició 3.2.2 (Funció generadora de probabilitat)**.** La *funció generadora de probabilitat* (*fgp*) d'una variable aleatòria de recompte X és la sèrie formal de potències

$$G_X(z) = \sum_{n \geq 0} p(X = n) z^n$$

Quan diem que aquesta sèrie de potències és formal, ens referim a que la tractem com un objecte amb el que podem operar sense preocupar-nos de si la sèrie convergeix o no.

Exemple 3.2.1. A l'experiment de tirar un dau, li associem la funció generadora de probabilitat

$$G_X(z) = z + z^2 + \cdots + z^6$$

Per qualsevol funció generadora de probabilitat, tenim les propietats següents:

Proposició 3.2.1.

1. $G_X(0) = 0$
2. $G_X(1) = \sum_{n \geq 0} p(X = n) = 1$
3. $G_X(z)$ *convergeix absolutament* $\forall\, z : |z| \leq 1$
4. $G_X(z)$ *té radi de convergència* $R \geq 1$
5. $G_X(z)$ *defineix una funció analítica* $\forall\, z : |z| < R$

Per tant, almenys en l'interior del disc unitat, tenim que la funció generadora és una funció analítica i no només una expressió formal.

Observació. Observem que si derivem formalment la sèrie de potències (és a dir, derivant terme a terme) tenim que

$$G'_X(z) = \sum_{n \geq 1} n \cdot p(X = n) z^{n-1} \implies G'_X(1) = \mathbb{E}[X]$$

Però aquesta expressió només coincidirà amb la derivada analítica si el radi de convergència és estrictament més gran que 1. (En cas contrari ens pot sortir esperança infinita.)

Si tornem a derivar la funció generadora, obtenim

$$G''_X(z) = \sum_{n\geq 1} n(n-1)\cdot p(X=n)z^{n-2} \implies G''_X(1) = \mathbb{E}[X(X-1)] = \mathbb{E}[X^2] - \mathbb{E}[X]$$

Per tant, podem expressar la variància com

$$\mathrm{Var}[X] = G''_X(1) + G'_X(1) - (G'_X(1))^2$$

En general, si coneixem la funció generadora de probabilitat, aquest mètode serà el més ràpid per calcular l'esperança o la variància.

Observació. Donat que $G_X(z) = \sum_{n\geq 0} p(X=n)z^n$, podem entendre la funció de probabilitat com l'esperança de z^X:

$$G_X(z) = \mathbb{E}[z^X]$$

Es pot comprovar que si X i Y són independents, aleshores z^X i z^Y també ho són. Per tant, la funció generadora de probabilitat es comportarà bé per la suma de variables independents:

$$G_{X+Y}(z) = \mathbb{E}[z^{X+Y}] = \mathbb{E}[z^X z^Y] \underset{\text{(independents)}}{=} \mathbb{E}[z^X]\mathbb{E}[z^Y] = G_X(z)G_Y(z)$$

A continuació ho reenunciarem en forma de proposició:

Proposició 3.2.2. *Siguin X i Y dues variables de recompte independents. Aleshores,*

$$G_{X+Y}(z) = G_X(z)G_Y(z)$$

Demostració. En la observació anterior ho hem provat veient que z^X i z^Y eren independents, però també es pot fer directament.

$$\begin{aligned} G_{X+Y}(z) &= \sum_{n\geq 0} p(X+Y=n)z^n = \sum_{n\geq 0}\sum_{i=0}^{n} p(X+Y=n,\ X=i)z^n = \\ &= \sum_{n\geq 0}\sum_{i=0}^{n} p(X=i, Y=n-i)z^n = \\ &= \sum_{n\geq 0}\sum_{i=0}^{n} p(X=i)p(Y=n-i)z^i z^{n-i} = G_X(z)G_Y(z) \end{aligned}$$

□

3.3 Distribucions i esperances condicionades

Sigui $(\Omega, \mathcal{A}, p)$ un espai de probabilitat i siguin X, Y variables aleatòries discretes. Sigui $A \in \mathcal{A}$ un esdeveniment qualsevol amb probabilitat no nul·la. Ens podem preguntar quina és la probabilitat que Y valgui y sabent que ha passat A, és a dir, $p(Y = y \mid A)$. Per la definició de probabilitat condicionada, aquesta ve donada per la fórmula

$$p(Y = y \mid A) = \frac{p(\{Y = y\} \cap A)}{p(A)}$$

Si fixem el A i iterem per totes les $y \in \mathbb{R}$, podem definir una variable aleatòria $Y|A$, que tindrà funció de probabilitat

$$p(Y|A = y) = \frac{p(\{Y = y\} \cap A)}{p(A)}$$

Si prenem $A = \{X = x\}$ (sempre que $p(X = x) > 0$), obtenim la *funció de probabilitat de Y condicionada a X*:

$$p_{Y|X}(y|x) = p(Y = y|X = x)$$

La seva funció de distribució (*funció de distribució condicionada*) serà

$$F_{Y|X}(y|x) = p(Y \leq y \mid X = x)$$

Observació. Si X i Y són independents, aleshores $F_{Y|X}(y|x) = F_Y(y)$.

Exemple 3.3.1. Considerem l'espai de probabilitat donat per tirar dos daus. Sigui X el valor del primer dau i Y la suma dels valors dels dos daus. Aleshores,

$$p(Y = 2 \mid X = 1) = \frac{1}{6} \neq p(Y = 2) = \frac{1}{36}$$

Tornem un altre cop al cas genèric, on $A \in \mathcal{A}$ és un esdeveniment qualsevol. Aleshores, podem calcular la esperança de Y condicionada a A com

$$\mathbb{E}[Y|A] = \sum_{y \in \operatorname{Im} Y} y \, p(Y = y \mid A)$$

que observem que es correspon a $\mathbb{E}[Y]$ a l'espai $(\Omega, \mathcal{A}, p_A)$, on p_A és la probabilitat condicionada a A.

Exemple 3.3.2. Reprenent l'exemple dels daus, tenim que

$$\mathbb{E}[Y \mid X = 1] = 4.5, \qquad \mathbb{E}[Y \mid X = 6] = 9.5$$

Quan agafem $A = \{X = x\}$, observem que per cada x obtenim un valor de $\mathbb{E}[Y \mid X = x]$. Per tant, podem entendre l'esperança condicionada a $\{X = x\}$ com una variable aleatòria tal que

$$\mathbb{E}[Y \mid X](\omega) = \mathbb{E}[Y \mid X = X(\omega)]$$

Exemple 3.3.3. A l'exemple dels daus, la variable aleatòria $\mathbb{E}[Y \mid X]$ pren valors $4.5, 5.5, \ldots,$ 9.5 amb probabilitat $1/6$.

Com que és una variable aleatòria, en podem calcular l'esperança:

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = 7 = \mathbb{E}[Y]$$

Aquest resultat no és casualitat, sinó que tenim la proposició següent:

Proposició 3.3.1. *L'esperança de l'esperança de Y condicionada a X és l'esperança de Y, és a dir,*

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \mathbb{E}[Y]$$

Demostració. Estem suposant implícitament que X i Y són variables aleatòries discretes en un mateix espai de probabilitat. Per tant,

$$\begin{aligned}\mathbb{E}[\mathbb{E}[Y \mid X] &= \sum_{x \in \operatorname{Im} X} \mathbb{E}[Y \mid X = x] \cdot p(X = x) = \sum_{x \in \operatorname{Im} X} \left(\sum_{y \in \operatorname{Im} y} y \cdot p(Y = y \mid X = x) \right) \cdot p(X = x) \\ &= \sum_{x \in \operatorname{Im} X} \sum_{y \in \operatorname{Im} y} y \cdot p(Y = y \mid X = x)\, p(X = x) \\ &= \sum_{y \in \operatorname{Im} Y} y \underbrace{\sum_{x \in \operatorname{Im} X} p(Y = y \mid X = x)\, p(X = x)}_{p(Y=y)} = \mathbb{E}[Y]\end{aligned}$$

□

Aquesta proposició ens serà útil quan sigui difícil calcular $\mathbb{E}[Y]$ directament però sigui fàcil calcular $\mathbb{E}[Y \mid X]$.

Sigui $\psi(x) = \mathbb{E}[Y \mid X = x]$. Ens podríem preguntar, donat un cert $\alpha \in \operatorname{Im} \psi$, quant val $p(\mathbb{E}[Y \mid X] = \alpha)$. Observem que

$$\begin{aligned}p(\mathbb{E}[Y \mid X] = \alpha) &= p(\{\omega : \mathbb{E}[Y \mid X](\omega) = \alpha\}) = p(\{\omega : \psi(X(\omega)) = \alpha\}) = \\ &= p(\bigcup_{x \in \psi^{-1}(\alpha)} \{\omega : X(\omega) = x\}) = \sum_{x \in \psi^{-1}(\alpha)} p(X = x)\end{aligned}$$

Per tant, el que veiem és que la probabilitat que l'esperança de Y condicionada a X valgui α és la suma de $p(X = x)$ per tots els x tals que l'esperança de Y condicionada a $X = x$ és α.

Exemple 3.3.4. Tornem a l'exemple dels daus anterior en el qual tirem dos daus i definim X com el valor del primer i Y com la suma. Observem que

$$\mathbb{E}[Y \mid X = i] = \frac{2i + 7}{2} = i + \frac{7}{2}$$

Habitualment ens interessa expressar la variable esperança condicionada en funció de X. En aquest cas,

$$\mathbb{E}[Y \mid X] = X + \frac{7}{2}$$

En general, tindrem que l'esperança condicionada serà la "millor" manera d'expressar una variable en funció d'una altra (en l'exemple anterior, d'escriure la suma de dos daus en funció del valor de un).

No està clar a que ens referim quan parlem de "millor", però a continuació considerarem uns quants casos particulars per veure què volem dir quan diem que $\mathbb{E}[Y \mid X]$ és la millor aproximació de Y a través de X.

- $Y = f(X)$: Aleshores,

$$\mathbb{E}[f(X) \mid X = x] = \sum_{x' \in \operatorname{Im} X} f(x') \cdot p(X = x' \mid X = x) = \sum_{x' \in \operatorname{Im} X} f(x')\delta_{xx'} = f(x)$$

  Per tant, $\mathbb{E}[f(X) \mid X] = f(X)$. És a dir, si Y ja és funció de X, aleshores $\mathbb{E}[Y \mid X]$ ens dona $f(X)$, que ja és la millor aproximació que teníem de Y en funció de X.

  Com a exercici, es pot veure que $\mathbb{E}[Yf(x) \mid X] = f(X)\mathbb{E}[Y \mid X]$, i utilitzar això per demostrar també el resultat anterior.

- X i Y són independents:

$$\begin{aligned}\mathbb{E}[Y \mid X = x] &= \sum_{y \in \operatorname{Im} Y} y\, p(Y = y \mid X = x) = \sum_{y \in \operatorname{Im} Y} y \frac{p(Y = y,\, X = x)}{p(X = x)} \\ &= \sum_{y \in \operatorname{Im} Y} y\, p(Y = y) = \mathbb{E}[Y]\end{aligned}$$

  Per tant, si X i Y són independents, el resultat de Y no ve influenciat pel resultat de X, i la millor aproximació de Y en funció de X és simplement $\mathbb{E}[Y]$.

3.4 Models de variables aleatòries discretes

Comentari. Si X i Y tenen la mateixa llei, aleshores escrivim $X \sim Y$.

En aquest apartat estudiarem els models de variables aleatòries discretes que s'utilitzen més sovint.

- **Uniforme**: $X \sim U(a_1, \dots, a_n)$ amb $a_1, \dots, a_n \in \mathbb{R}$
  - $\operatorname{Im} X = \{a_1, \dots, a_n\}$
  - $p(X = a_i) = \frac{1}{n}$
  - $\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^{n} a_i$
- **Bernoulli**: $X \sim B(p)$ amb $p \in [0, 1]$
  - $\operatorname{Im} X = \{0, 1\}$
  - $p(X = 1) = p$ i $p(X = 0) = 1 - p$
  - $G_X(z) = (1 - p) + pz$
  - $\mathbb{E}[X] = p$
  - $\operatorname{Var}[X] = p(1 - p)$
- **Binomial**: $X \sim \operatorname{Bin}(n, p)$ amb $n \in \mathbb{Z}$, $n \geq 1$ i $p \in [0, 1]$

  Diem que $X \sim \operatorname{Bin}(n, p)$ si existeixen $X_1, \dots, X_n$ variables aleatòries independents, $X_i \sim B(p)$ tals que $X = X_1 + \cdots + X_n$. És a dir, una variable $\operatorname{Bin}(n, p)$ compta el nombre d'èxits al realitzar un experiment n vegades amb probabilitat d'èxit p.

- $\operatorname{Im} X = \{0, 1, \ldots, n\}$
- $p(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$
- $G_X(z) = G_{X_1}(z)^n = (1 - p + pz)^n$
- $\mathbb{E}[X] = np$
- $\operatorname{Var}[X] = np(1-p)$

Observem que la suma de variables binomials independents de la mateixa p és la suma de variables de Bernoulli de la mateixa p, que és una binomial. Per tant,

$$\operatorname{Bin}(n_1, p) + \operatorname{Bin}(n_2, p) = \operatorname{Bin}(n_1 + n_2, p)$$

- **Geomètrica**: $X \sim \operatorname{Geom}(p)$ amb $p \in [0, 1]$

  Una variable aleatòria geomètrica $X \sim \operatorname{Geom}(p)$ compta el nombre de repeticions d'una Bernoulli de probabilitat p fins que s'obtingui el primer èxit.

  - $\operatorname{Im}(X) = \{1, 2, \ldots\}$
  - $p(X = k) = (1-p)^{k-1} p$
  - $G_X(z) = \sum_{n \geq 1} p(1-p)^{n-1} z^n = \dfrac{pz}{1-(1-p)z}$
  - $\mathbb{E}[X] = 1/p$
  - $\operatorname{Var}[X] = \dfrac{1-p}{p^2}$

  Si no volem incloure el primer èxit en el nombre de repeticions, denotarem la variable com $X \sim \operatorname{Geom}_0(p)$ (veure exercici 13).

- **Binomial negativa**: $X \sim \operatorname{BinN}(p, r)$ amb $p \in [0, 1]$, $r \in \mathbb{Z}$, $r \geq 1$

  Diem que $X \sim \operatorname{BinN}(p, r)$ si existeixen $Y_1, \ldots, Y_r$ independents tals que $Y_i \sim \operatorname{Geom}(p)$ i $X = Y_1 + \cdots + Y_r$. Així doncs, les variables binomials negatives són una generalització de les geomètriques, i compten el nombre de repeticions d'una $B(p)$ fins a obtenir r èxits.

  - $\operatorname{Im}(X) = \{r, r+1, \ldots\}$
  - $p(X = k) = \underbrace{\binom{k-1}{r-1}}_{\text{maneres de distribuir els primers } r-1 \text{ èxits}} p^r (1-p)^{k-r}$
  - $G_X(z) = \left(\dfrac{pz}{1-(1-p)z}\right)^r$
  - $\mathbb{E}[X] = r/p$
  - $\operatorname{Var}[X] = \dfrac{r(1-p)}{p^2}$

- **Poisson**: $X \sim \operatorname{Po}(\lambda)$ amb $\lambda \in \mathbb{R}$, $\lambda > 0$

  Una variable de Poisson és el límit d'una binomial $\operatorname{Bin}(n, p)$ quan $n \to \infty$ i $\lambda := np$ es manté constant.

- $\mathrm{Im}\,(X) = \{0, 1, 2, \ldots, \}$
- $p(X = k) = \frac{1}{k!} e^{-\lambda} \lambda^k$
- $G_X(z) = \sum_{k \geq 0} \frac{1}{k!} e^{-\lambda} \lambda^i z^i = e^{-\lambda} e^{\lambda z} = e^{\lambda(z-1)}$
- $\mathbb{E}[X] = \lambda$
- $\mathrm{Var}[X] = \lambda$

3.5 Altres aplicacions de les funcions generadores de probabilitat

3.5.1 Sumes amb nombre aleatori de summands

Suposem que tenim un conjunt de variables aleatòries de recompte independents $X_1, X_2, \ldots,$ tals que $X_i \sim X$ per qualsevol i.

Sabem que la funció generadora de la suma és el producte de funcions generadores:

$$G_{X_1+\cdots+X_n}(z) = G_{X_1}(z) \ldots G_{X_n}(z) = G_X(z)^n$$

Ara suposem que tenim una altre variable de recompte N, independent de les $\{X_i\}$. Volem condicionar el nombre de X_i que sumem al valor de la variable N (és a dir, volem estudiar l'esperança de $X_1 + \cdots + X_N$, on N és una altra variable aleatòria de recompte independent).

Tenim que

$$\begin{aligned} G_{X_1+\cdots+X_N}(z) &= \mathbb{E}[z^{X_1+\cdots+X_N}] = \mathbb{E}[\mathbb{E}[z^{X_1+\cdots+X_N} \mid N]] \\ &= \sum_{n \in \mathrm{Im}\, N} \mathbb{E}[z^{X_1+\cdots+X_N} \mid N = n]\, p(N = n) = \sum_{n \in \mathrm{Im}\, N} \mathbb{E}[z^{X_1+\cdots+X_n}]\, p(N = n) \\ &= \sum_{n \in \mathrm{Im}\, N} p(N = n)\, G_{X_1+\cdots+X_n}(z) = \sum_{n \in \mathrm{Im}\, N} p(N = n)\, G_X(z)^n = G_N(G_X(z)) \end{aligned}$$

Això ho podem resumir amb el següent lema:

Lema 3.5.1. *Amb les hipòtesis anteriors, tenim que* $G_{X_1+\cdots+X_N}(z) = G_N(G_X(z))$.

Observem que el truc d'utilitzar l'esperança condicionada ens ha permès particularitzar l'expressió per n's concretes, fent més fàcil calcular-ho.

A continuació veurem un exemple d'aplicació d'aquesta fórmula.

3.5.2 Arbres de Galton-Watson

Els arbres de Galton-Watson són un model que es va proposar al segle XIX per estudiar l'extinció dels cognoms entre l'aristocràcia anglesa.

Segons els costums anglesos, els cognoms passen de pares a fills (ja que se suposa que les filles agafaran el cognom del seu marit). Podem representar l'evolució d'un cognom com un arbre, on cada node té un pare (del qual hereda el cognom), i una sèrie de fills (als quals transmet el cognom).

Ens volem preguntar quina és la probabilitat que el cognom s'extingueixi, és a dir, que existeixi un nivell de l'arbre que sigui buit.

Sigui Z_i el nombre d'individus a la generació $i+1$-èssima. Suposarem que a la primera generació només hi ha un únic individu amb el cognom ($Z_0 = 1$), i que cada node té la mateixa distribució de probabilitat de nombre de fills (que seguirà una llei X). Suposarem a més que tots els individus es reprodueixen independentment de tota la resta.

Aleshores, la probabilitat d'extinció és

$$p\Big(\bigcup_{n\geq 1}\{Z_n = 0\}\Big) = \lim_{n\to\infty} p(Z_n = 0) = \lim_{n\to\infty} G_{Z_n}(0)$$

on hem utilitzat que els $\{Z_n = 0\}$ són successos encaixats per passar de la probabilitat de la unió al límit de probabilitats.

Proposició 3.5.1. *Sigui $G_n(z) = G_{Z_n}(z)$. Aleshores, es satisfà la recurrència*

$$G_{n+m}(z) = G_n(G_m(z))$$

per qualssevol $n, m \geq 0$.

Demostració. Cada individu de la generació $n+m$ té un únic antecedent de la generació m. Per tant, el nombre d'individus a la generació $n+m$ és

$$Z_{n+m} = Y_1 + \cdots + Y_{Z_m}$$

on Y_i és el nombre de descendents a la generació $n+m$ de l'i-èssim membre de la generació m.

Aleshores observem que, per simetria, $Y_i \sim Z_n$, de manera que tenim una suma de variables aleatòries amb la mateixa distribució i tals que el nombre de summands ve regulat per una altra variable aleatòria. Aplicant l'apartat anterior,

$$G_{m+n}(z) = G_m(G_n(z))$$

□

Exercici: Trobeu l'esperança i la variància de Z_n en funció de l'esperança i la variància de X.

Proposició 3.5.2. *Sigui η la solució positiva més petita de $s = G_X(s)$. Aleshores, la probabilitat d'extinció del cognom és η. A més, si $\mu = \mathbb{E}[X]$, tenim que*

$$\begin{cases} \eta = 1, & \text{si } \mu < 1 \\ \eta < 1, & \text{si } \mu > 1 \\ \eta = 1, & \text{si } \mu = 1 \text{ i } \mathrm{Var}(X) > 0 \end{cases}$$

Demostració. Sigui $\eta_n = p(Z_n = 0) = G_n(0)$. Teníem que

$$G_n(z) = (\underbrace{G_X \circ \cdots \circ G_X}_{n \text{ cops}})(z)$$

de manera que

$$\eta_n = G_X(G_{n-1}(0)) = G_X(\eta_{n-1})$$

Quan $n \to \infty$, si volem que existeixi $\eta = \lim_n \eta_n$, tindrem que aquesta η serà la solució de

$$\eta = G_X(\eta)$$

Observem que la funció $y(x) = x$ té pendent 1 arreu. Per tant, si $\mu = G'_X(1) < 1$, tindrem que la primera solució de $G_X(x) = x$ serà a $x = 1$, mentre que si $G'_X(1) > 1$, tindrem que hi ha una solució abans.

En el cas degenerat $G'_X(x) = 1$ també tindrem que el primer punt de tall entre $G(x)$ i x és a $x = 1$ (excepte en el cas especial en el que $\mathrm{Var}(X) = 0$ i $G_X(x) = x$). $\square$