Universitat de Barcelona

Facultat de Matemàtiques i Enginyeria Informàtica

Apunts de classe

Grau en Matemàtiques

Curs 2023-2024 | Vuitè Semestre

Estadística (EST)

Autor:
Mario Vilar

Professor/a:
Dr. David Márquez

Presentació de l'assignatura

Conèixer les propietats bàsiques dels estimadors puntuals i d'interval. Saber plantejar el mètode de Neyman-Pearson. Comprendre el mètode de la raó de versemblança per construir contrastos d'hipòtesis. Entendre els contrastos d'ajustament i independència. Conèixer el model lineal. **Incompletes**.

Classificació AMS (2020): 03B48, 60-00, 60-01, 60A05, 60B12, 60F05.

Índex

Introducció

We'll all be planning that route
We're gonna take real soon
We're waxing down our surfboards
We can't wait for June
We'll all be gone for the summer
We're on surfari to stay
Tell the teacher we're surfin'.

THE BEACH BOYS, *Surfin' USA* (1963)

Primer de tot, trobareu que hi ha un índex, on hi distingim els diferents apartats ordenats seguint el meu propi criteri i, de tant en tant, seguint l'ordre cronològic del curs. Hi ha capítols, seccions, subseccions (i fins i tot subsubseccions). Us faig cinc cèntims de com he organitzat els encapçalaments de cada pàgina:

1. el número de l'últim capítol/secció/subsecció, depèn de la profunditat que hi hagi definida en aquell moment, figurarà en cada cantonada superior de pàgina parella (per exemple, *1.2*);
2. el nom del capítol es trobarà a la part dreta de la capçalera de les pàgines parelles (per exemple, «Divisibilitat i nombres primers»);
3. el nom de l'última secció/subsecció de la pàgina, a la cantonada dreta superior de les pàgines parelles (per exemple, «Polinomis: algorisme d'Euclides»);
4. el número de l'últim teorema, definició... de la pàgina en qüestió es trobarà a les pàgines senars, a la cantonada superior dreta, *destacat en el color de la seva capçalera corresponent* (per exemple, **1.2.3**).

A més, hi ha una taula, la taula de continguts. En aquest sentit, tal com acabem de dir, es veu fàcilment que s'ha seguit una mena de *sorting-by-color* per poder treballar de manera més eficient amb els diferents tipus d'enunciats matemàtics. D'aquesta manera, si busqueu una definició, un teorema... podreu distingir que estan destacats amb colors diferents (ara els introduïm) i trobar-los molt ràpidament:

1. Teoremes, proposicions, lemes, corol·laris, propietats, conjectures, processos i exercicis tindran aquest format (capçalera destacada amb color gris fosc):

   Teorema. *Compte! L'enunciat del teorema també serà en cursiva! Jove xef, porti whisky amb quinze glaçons d'hidrogen, coi!*

2. Les definicions i notacions tindran aquest format (capçalera de color gris clar):

   Definició. Aqueix betzol, Jan, comprava whisky de figa.

3. Les remarques i exemples tindran aquest format:

   Observació. Zel de grum: quetxup, whisky, cafè, bon vi; ja!

Després de molts anys fent-ho malament, ara sol quedaran numerades aquelles equacions a les quals em referiré més endavant. És la *filosofia dominant* en la majoria de textos matemàtics (té nom i tot, es diu la *regla d'Occam*).

Per últim, m'estalviaré de comentar l'índex terminològic perquè el seu propòsit és clar i, en efecte, paral·lel al de l'organització d'aquest document: poder facilitar-vos al màxim la feina per localitzar qualsevol concepte que desitgeu. Espero que us serveixin d'alguna cosa aquests apunts, els he fet amb tot l'amor del món. Sort!

Mario VILAR
Sitges, Barcelona
4 de juny de 2024

Taula de continguts

IV Capítol 4 IV

B | Capítol B | B

I

Model estadístic

1.1

Introducció

Definició 1.1.1 (Model estadístic). És una terna $(\Omega, \mathscr{A}, \mathscr{P})$ en què $(\Omega, \mathscr{A})$ és un espai mesurable i $\mathcal{P}$ és una família de probabilitats en $(\Omega, \mathscr{A})$.

Definició 1.1.2 (Model paramètric). Un model estadístic $(\Omega, \mathscr{A}, \mathscr{P})$ serà paramètric si podem escriure la família de probabilitats com $\mathscr{P} = \{P_\Theta \bigm| \Theta \in \Theta\}$, on Θ és un subconjunt de $\mathbb{R}^d$ amb $d \geq 1$. El conjunt Θ l'anomenem espai de paràmetres.

Observació 1.1.3. Una família finita $\mathscr{P}$ és sempre parametritzable.

Exemple 1.1.4 (Control de qualitat). Imaginem-nos que tenim una cadena de producció d'una certa peça d'una maquinària i estem interessats en estudiar si aquest procés produeix un excés de peces defectuoses. Acceptem que el nombre de peces de la cadena és molt gran i no podem estudiar-les totes, per tant haurem d'obtenir una mostra i treballar amb ella.

Per prendre una decisió al respecte el que podem fer, simplificant la modelització al màxim, és observar si la proporció de peces defectuoses de la nostra mostra no supera una quantitat $\theta_0 \in (0, 1)$ fixada a priori. En el cas de ser superada aquesta quantitat direm que la cadena produeix masses peces en mal estat i la rebutjarem. Seguint amb la modelització, suposarem que la mostra de peces és escollida pel mètode del mostreig aleatori simple i té mida n, és a dir, cadascuna de les n peces és triada a l'atzar i amb reemplaçament.

A cada peça li podem associar dos possibles valors, un 0 si la peça està en bon estat i un 1 si és defectuosa. Així, cadascuna d'aquestes n observacions es pot interpretar com el valor d'una variable aletòria amb llei Bernoulli de paràmetre $\theta \in (0, 1)$ desconegut. Les variables que generen aquestes observacions són independents i idènticament distribuïdes. El model estadístic associat és

$$(\Omega, \mathcal{A}, \{P_\theta, \theta \in (0, 1)\}),$$

on $\Omega = \{0, 1\}^n$, $\mathcal{A}$ és la família de tots els subconjunts de Ω i, per $x = (x_1, \ldots, x_n) \in \{0, 1\}^n$,

$$P_\theta(\{x\}) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{n\bar{x}_n}(1-\theta)^{n(1-\bar{x}_n)},$$

amb $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Exemple 1.1.5 (Tria de nombres reals). Tenim un interval $[0, \theta]$, amb $\theta > 0$ desconegut, d'on s'han triat a l'atzar i de manera independent n nombres reals, $x_1, \ldots, x_n$. Es pretén analitzar la longitud θ d'aquest interval.

S'assumeix que la variable aleatòria que modelitza aquesta tria d'un nombre a l'atzar entre 0 i θ segueix una uniforme en $[0, \theta]$, és a dir, té una densitat

$$\frac{1}{\theta}\mathbb{1}_{[0,\theta]}(y) = \begin{cases} \frac{1}{\theta}, & \text{si } y \in [0, \theta], \\ 0, & \text{altrament.} \end{cases}$$

El problema de saber que la llei d'aquesta variable és una uniforme el resol l'estadística no paramètrica. Per tant, aquest exemple es pot modelitzar fent servir el model estadístic següent:

$$\left([0, +\infty)^n, \mathcal{B}\left([0, +\infty)^n\right), \{P_\theta, \theta > 0\}\right),$$

on per $\mathcal{B}\left([0, +\infty)^n\right)$ entenem els Borelians de $[0, +\infty)^n$ i P_θ és la probabilitat absolutament contínua en $[0, +\infty)^n$ amb densitat

$$f_\theta(x) = \prod_{i=1}^{n} \left[\frac{1}{\theta}\mathbb{1}_{[0,\theta]}\left(x_i\right)\right] = \frac{1}{\theta^n}\mathbb{1}_{[0,\theta]}\left(x_{(n)}\right),$$

on $x_{(n)} = \max\left(x_1, \ldots, x_n\right)$.

1.2 Funció de versemblança

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$ i un vector aleatori $X = (X_1, \ldots, X_n) : \tilde{\Omega} \longrightarrow \Omega$, on $\Omega \subset \mathbb{R}^n$ és l'espai de les observacions. Recordem que P_θ és la llei en $(\Omega, \mathscr{A})$ del vector aleatori X. El concepte de funció de versemblança va ser introduït per Fisher i, encara que es pot presentar de forma molt general, aquí donarem només la definició per als dos casos més usuals.

Definició 1.2.1 (Funció de versemblança). És l'aplicació $\mathscr{L} : \Omega \times \Theta \longrightarrow \mathbb{R}^+$ definida diferenciant els dos casos següents:

1. En el cas que P_θ sigui una llei discreta, $\mathscr{L}(x, \theta) = P_\theta(\{x\})$.
2. En el cas que P_θ sigui una llei absolutament contínua amb densitat f_θ, $\mathscr{L}(x, \theta) = f_\theta(x)$.

Observació 1.2.2.

1. Aquesta definició no engloba tots els casos possibles perquè el vector aleatori pot tenir moltes més expressions. En realitat, la funció de versemblança pot ser definida de manera única, sense haver-se de separar en dos casos, però per a això es necessiten uns coneixements més amplis de teoria de la mesura dels que no assumim.

2. Fixat $\theta \in \Theta$, la funció de versemblança, com a funció de x, caracteritza la llei del vector X.

Sovint treballarem amb mostres aleatòries simples de mida n (sobretot en els exemples), i això vol dir que les variables aleatòries $X_1, \ldots, X_n$ són independents i idènticament distribuïdes. Per a aquest tipus de mostres, per a tot $x = (x_1, \ldots, x_n) \in \Omega$, la funció de versemblança és

$$\mathscr{L}(x,\theta) = \begin{cases} \prod_{i=1}^n P_\theta\left(\{x_i\}\right), & \text{en el cas discret} \\ \prod_{i=1}^n f_\theta\left(x_i\right), & \text{en el cas absolutament continu} \end{cases}$$

La funció de versemblança té una dependència clara de la dimensió de l'espai de les observacions Ω i de l'espai de paràmetres Θ.

Exemple 1.2.3 (Control de qualitat)**.** En aquest cas tenim:

$$\mathscr{L}(x,\theta) = \theta^{n\bar{x}_n}(1-\theta)^{n(1-\bar{x}_n)} = \exp\left\{\log\left(\theta^{n\bar{x}_n}(1-\theta)^{n(1-\bar{x}_n)}\right)\right\} = \exp\left\{n\bar{x}_n \log\left(\frac{\theta}{1-\theta}\right) + n\log(1-\theta)\right\},$$

d'on:

$$C(\theta) = n\log\left(\frac{\theta}{1-\theta}\right),\ \Phi(x) = \bar{x}_n,\ D(\theta) = n\log(1-\theta),\ \Psi(x) = 0.$$

Per tant, és un model exponencial (cf 1.2.4). A classe hem escrit que el model estadístic és $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), \mathcal{P})$, on $\mathcal{P} = \{P_\theta \mid \theta \in (0,1)\}$, i

$$\mathscr{L}(x,\theta) = \mathscr{L}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i}.$$

Definició 1.2.4 (Model exponencial)**.** Direm que un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$ és exponencial si existeixen funcions reals mesurables $\Psi, \Phi_1, \ldots, \Phi_r : \Omega \longrightarrow \mathbb{R}$ i funcions reals $D, C_1, \ldots, C_r : \Theta \longrightarrow \mathbb{R}$ tals que podem expressar la funció de versemblança de la manera següent:

$$\mathscr{L}(x,\theta) = \exp\left\{\sum_{k=1}^r C_k(\theta)\Phi_k(x) + D(\theta) + \Psi(x)\right\} \equiv \tilde{D}(\theta)\tilde{\Psi}(x)\exp\left\{\sum_{k=1}^r C_k(\theta)\Phi_k(x)\right\},$$

per a $\tilde{D}(\theta), \tilde{\Psi}(x) > 0$. És important fer constar que les funcions que apareixen en un model exponencial no són úniques. En el cas particular que l'espai de paràmetres Θ sigui unidimensional la funció de versemblança es podrà escriure com:

$$\mathscr{L}(x,\theta) = \exp\left\{C(\theta)\Phi(x) + D(\theta) + \Psi(x)\right\}.$$

Trobarem una relació entre models exponencials i suficiència a 1.4.10, que permetrà donar un criteri senzill per a establir un model regular.

1.3

Suficiència i regularitat

Com a la secció anterior, considerem el model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$ i un vector $X = (X_1, \dots, X_n) : \tilde{\Omega} \longrightarrow \Omega \subset \mathbb{R}^n$. Estudiarem transformacions de la mostra que conservin tota la informació.

Definició 1.3.1 (Estadístic)**.** Anomenem estadístic tota aplicació mesurable $T : (\Omega, \mathscr{A}) \longrightarrow (\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ on $m \geq 1$ és un enter.

Exemple 1.3.2. La mitjana i la variància mostral, el màxim i el mínim en són alguns exemples unidimensionals:

$$\begin{aligned}
T_1(x) &= \bar{x}_n := \frac{1}{n}\sum_{i=1}^n x_i. \\
T_2(x) &= s_n^2 := \frac{1}{n}\sum_{i=1}^n (x_i - \bar{x}_n)^2. \\
T_3(x) &= x_{(n)} := \max(x_1, \dots, x_n). \\
T_4(x) &= x_{(1)} := \min(x_1, \dots, x_n).
\end{aligned}$$

Definició 1.3.3 (Estadístic suficient)**.** Direm que un estadístic $T : \Omega \longrightarrow \mathbb{R}^n$ és *suficient* si la llei del vector X condicionada per l'estadístic T no depèn de θ. Intuïtivament, direm que un estadístic serà suficient si la coneixença de la mostra no aporta cap informació addicional de θ que la donada per T.

Teorema 1.3.4 (de factorització de Neyman-Fisher)**.** *Sigui $T : \Omega \longrightarrow \mathbb{R}^m$ un estadístic. Aleshores, T és suficient si, i només si, existeixen dues funcions mesurables $\psi : \mathbb{R}^m \times \Theta \longrightarrow \mathbb{R}$ i $h : \Omega \longrightarrow \mathbb{R}$ tals que, per a tot $x \in \Omega$ i $\theta \in \Theta$, es compleix que $\mathscr{L}(x, \theta) = \psi(T(x), \theta)h(x)$.*

Propietat 1.3.5.

1. *En un model exponencial, si T és un estadístic i φ mesurable, $T = (\varphi_1(x), \dots, \varphi_r(x))$ és suficient.*
2. *Si T és suficient i φ mesurable i bijectiva, $\varphi(T)$ és suficient.*
3. *Si T és suficient i existeix H mesurable tal que $T = H(S)$, aleshores S és suficient.*
4. *T és un estadístic suficient i minimal si per a qualsevol altre estadístic suficient S existeix H mesurable tal que $T = H(S)$.*

Definició 1.3.6 (Model regular)**.** Un model estadístic és regular i ho denotarem per a (R) si es compleix que:

1. El subconjunt $\{x \in \Omega \mid \mathscr{L}(x, \theta) > 0\} \subset \Omega$ no depèn de θ.
2. Exisiteixen les derivades $\partial_\theta \ln \mathscr{L}(x, \theta)$ i $\partial_\theta^2 \ln \mathscr{L}(x, \theta)$. A més a més, per a cada $\tilde{\theta} \in \Theta$ existeix també una funció positiva $h : \Omega \longrightarrow \mathbb{R}$ amb $\sum_{x \in \Omega} h(x) < \infty$ (cas discret) o $\int_\Omega h(x) < \infty$ (cas continu). Per a tot θ en un entorn de $\tilde{\theta}$ tenim:

$$|\partial_\theta \mathscr{L}(x, \theta)| \vee |\partial_\theta^2 \mathscr{L}(x, \theta)| \leq h(x).$$

3. Per a tot $\theta \in \Theta$, $0 < \mathbb{E}_\theta \left(|\partial_\theta^2 \ln \mathscr{L}(x, \theta)|^2\right) < \infty$.

En la majoria de problemes i exemples la primera condició ens dirà si el model és regular o bé no ho és.

1.4 Informació de Fisher

Ara volem donar una eina que mesuri la informació que tenim del paràmetre. Una bona mesura serà la sensibilitat que la mostra reflecteix davant les variacions del paràmetre.

Definició 1.4.1 (Informació de Fisher). En un model estadístic paramètric regular (R), la informació de Fisher és la funció del paràmetre següent:

$$I_F(\theta) := \mathbb{E}_\theta \left(|\partial_\theta \ln \mathscr{L}(x, \theta)|^2\right) = \mathbb{E}_\theta(\mathrm{sc}(\theta)^2).$$

1. En el cas discret:

$$\mathbb{E}_\theta \left(|\partial_\theta \ln \mathscr{L}(x, \theta)|^2\right) = \sum_{x \in \Omega} |\partial_\theta \ln \mathscr{L}(x, \theta)|^2 \mathscr{L}(x, \theta);$$

2. En el cas absolutament continu:

$$\mathbb{E}_\theta \left(|\partial_\theta \ln \mathscr{L}(x, \theta)|^2\right) = \int_\Omega |\partial_\theta \ln \mathscr{L}(x, \theta)|^2 \mathscr{L}(x, \theta)\, dx\,.$$

Proposició 1.4.2. *La informació de Fisher val zero (és a dir, $I_F(\theta) = 0$) si, i només si, $\mathrm{sc}(\theta) = \partial_\theta \ln \mathscr{L}(x, \theta) = 0$ per a tot $\theta \in \Theta$. Equivalentment, $\mathscr{L}(x, \theta)$ no depèn de θ (la versemblança no depèn del paràmetre).*

Demostració. Una variable no negativa té esperança zero si, i només si, és zero quasi segurament. Aplicant la definició, ja hem acabat. ∎

Proposició 1.4.3. *En un model estadístic paramètric regular (R), es compleix que $\mathbb{E}_\theta(\partial_\theta \ln \mathscr{L}(x, \theta)) = 0$.*

Demostració. Farem la demostració en el cas absolutament continu. Per a cada $\theta \in \Theta$, $\mathscr{L}(x,\theta)$ és la funció de densitat de X i, per tant, gràcies a la segona condició d'un model regular (que tenim $\int_\Omega h(x) < \infty$ per a una funció positiva h), tenim que

$$\int_\Omega \partial_\theta \mathscr{L}(x,\theta)\,dx = \partial_\theta \left(\int_\Omega \mathscr{L}(x,\theta)\,dx \right) = \partial_\theta 1 = 0.$$

Aleshores,

$$\begin{aligned} \mathbb{E}_\theta\left(\partial_\theta \ln \mathscr{L}(x,\theta)\right) &= \int_\Omega \partial_\theta \ln \mathscr{L}(x,\theta) \cdot \mathscr{L}(x,\theta)\,dx \\ &= \int_\Omega \frac{1}{\mathscr{L}(x,\theta)} (\partial_\theta \mathscr{L}(x,\theta)) \cdot \mathscr{L}(x,\theta)\,dx = \int_\Omega \partial_\theta \mathscr{L}(x,\theta)\,dx = 0. \end{aligned}$$

El cas discret funciona exactament igual intercanviant la integral sobre Ω pel sumatori sobre Ω:

$$\begin{aligned} E_\theta(\partial_\theta \ln \mathscr{L}(x,\theta)) &= \sum_{x\in\Omega} \partial_\theta \ln \mathscr{L}(x,\theta) \cdot \mathscr{L}(x,\theta) \\ &= \sum_{x\in\Omega} \partial_\theta \mathscr{L}(x,\theta) = \partial_\theta \sum_{x\in\Omega} \mathscr{L}(x,\theta) = \partial_\theta \sum_{x\in\Omega} P_\theta(x) = 0. \end{aligned}$$

A causa d'aquesta proposició és immediat el resultat següent. ■

Corol·lari 1.4.4. *En un model estadístic paramètric regular* (R) *tenim que*

$$I_F(\theta) = \mathrm{Var}_\theta\left(\partial_\theta \ln \mathscr{L}(x,\theta)\right) = \mathrm{Var}_\theta(\mathrm{sc}(\theta)).$$

Proposició 1.4.5. *Assumint* (R) *es compleix*

$$I_F(\theta) = -\mathbb{E}_\theta\left(\partial_\theta^2 \ln \mathscr{L}(x,\theta)\right).$$

Demostració. Primerament, en el cas absolutament continu:

$$\begin{aligned} E_\theta(\partial_\theta^2 \ln \mathscr{L}(x,\theta)) &= \int_\Omega \partial_\theta^2 \ln \mathscr{L}(x,\theta) \cdot \mathscr{L}(x,\theta)\,dx = \int_\Omega \partial_\theta \left(\frac{1}{\mathscr{L}(x,\theta)} \cdot \partial_\theta \mathscr{L}(x,\theta) \right) \cdot \mathscr{L}(x,\theta)\,dx \\ &= \int_\Omega \frac{-1}{\mathscr{L}(x,\theta)^2} \cdot (\partial_\theta \mathscr{L}(x,\theta))^2 \cdot \mathscr{L}(x,\theta)\,dx + \int_\Omega \frac{1}{\mathscr{L}(x,\theta)} \partial_\theta^2 \mathscr{L}(x,\theta) \cdot \mathscr{L}(x,\theta)\,dx \\ &= -\int_\Omega \left(\frac{1}{\mathscr{L}(x,\theta)} \partial_\theta \mathscr{L}(x,\theta) \right)^2 \cdot \mathscr{L}(x,\theta)\,dx + \underbrace{\int_\Omega \partial_\theta^2 \mathscr{L}(x,\theta)\,dx}_{\searrow 0} \\ &= -\mathbb{E}_\theta(|\partial_\theta \ln \mathscr{L}(x,\theta)|^2) = I_F(\theta). \end{aligned}$$

Per tant, $I_F(\theta) = -\mathbb{E}_\theta\left(\partial_\theta^2 \ln \mathscr{L}(x,\theta)\right)$, com volíem. Per últim, la farem en el cas discret. Així:

$$\begin{aligned}
\mathbb{E}_\theta\left(\partial_\theta^2 \ln \mathscr{L}(x,\theta)\right) &= \mathbb{E}_\theta\left[\partial_\theta\left(\frac{1}{\mathscr{L}(x,\theta)}\partial_\theta \mathscr{L}(x,\theta)\right)\right] \\
&= \mathbb{E}_\theta\left[\frac{1}{\mathscr{L}(x,\theta)}\partial_\theta^2 \mathscr{L}(x,\theta) - \frac{1}{\mathscr{L}(x,\theta)^2}\left(\partial_\theta \mathscr{L}(x,\theta)\right)^2\right] \\
&= \sum_{x\in\Omega} \mathscr{L}(x,\theta)\cdot\frac{1}{\mathscr{L}(x,\theta)}\cdot\partial_\theta^2 \mathscr{L}(x,\theta) - \sum_{x\in\Omega}\mathscr{L}(x,\theta)\cdot\frac{1}{\mathscr{L}(x,\theta)^2}(\partial_\theta \mathscr{L}(x,\theta))^2 \\
&= \sum_{x\in\Omega}\partial_\theta^2 \mathscr{L}(x,\theta) - \mathbb{E}_\theta(|\partial_\theta \ln \mathscr{L}(x,\theta)|^2).
\end{aligned}$$

Finalitzem la prova per mitjà del fet que

$$\sum_{x\in\Omega}\partial_\theta^2 \mathscr{L}(x,\theta) = \partial_\theta^2\left(\sum_{x\in\Omega}\mathscr{L}(x,\theta)\right) = 0.$$ [1]

I amb això tenim $I_F(\theta) = -\mathbb{E}_\theta\left(\partial_\theta^2 \ln \mathscr{L}(x,\theta)\right)$, com volíem. ■

Exemple 1.4.6 (de control de qualitat)**.** En aquest cas $\mathscr{L}(x,\theta) > 0, \forall x \in \Omega = \{0,\ldots,n\}$. A classe hem posat:

$$\mathscr{L}(x,\theta) = \theta^{\sum_i x_i}(1-\theta)^{n-\sum_i x_i} = \theta^{n\bar{x}_n}(1-\theta)^{n-n\bar{x}_n}.$$

Sigui com sigui, el logaritme de la funció de versemblança val

$$\log \mathscr{L}(x,\theta) = n\bar{x}_n \log\theta + n\left(1-\bar{x}_n\right)\log(1-\theta)$$

i derivant dues vegades

$$\partial_\theta^2 \log \mathscr{L}(x,\theta) = -\frac{n\bar{x}_n}{\theta^2} - \frac{n\left(1-\bar{x}_n\right)}{(1-\theta)^2}.$$

Per tant,

$$I_F(\theta) = -\mathbb{E}_\theta\left(\partial_\theta^2 \log \mathscr{L}(x,\theta)\right) = \frac{n\mathbb{E}_\theta\left(\bar{x}_n\right)}{\theta^2} + \frac{n\left(1-\mathbb{E}_\theta\left(\bar{x}_n\right)\right)}{(1-\theta)^2} = \frac{n}{\theta(1-\theta)},$$

ja que $\mathbb{E}\left(\bar{x}_n\right) = \theta$.

Exemple 1.4.7 (Tria de nombres reals)**.** Aquest model no és regular. No es compleix la primera condició de regularitat perquè l'indicador que apareix a la funció de versemblança depèn del paràmetre.

Observació 1.4.8. Considerem una família de probablitats $\{P_\theta \mid \theta \in \Theta\}$ sobre un subconjunt I de la recta real (un interval, vaja) i assumim que $(I, \mathscr{B}(I), \{P_\theta \mid \theta\in\Theta\})$ és un model regular amb informació de Fisher $I_{F,1}$. Sigui $X_1,\ldots,X_n$ una mostra aleatòria simple de mida n de

[1] Val a dir que el model regular (R) ens permet afitar per coses sumables o integrables. La principal conseqüència pràctica és que podem intercanviar sumatori i derivada parcial, indistintament.

la família $\{P_\theta \mid \theta \in \Theta\}$. El model estadístic $(I^n, \mathscr{B}(I^n), \{P_\theta^n \mid \theta \in \Theta\})$ (el tercer perquè les probabilitats P_θ són independents) associat a la mostra $X_1, \ldots, X_n$ també és regular i, a causa de la independència, té per funció de versemblança $\mathscr{L}(x_1, \ldots, x_n; \theta) = \prod_{i=1}^n \mathscr{L}(x_i, \theta)$. Com que $I_F(\theta) = \mathrm{Var}_\theta(\partial_\theta \ln \mathscr{L}(x, \theta)) = \mathrm{Var}_\theta(\mathrm{sc}(\theta))$, resulta que:

$$I_{F,n}(\theta) = \mathrm{Var}_\theta(\ln \mathscr{L}(x, \theta)) = \mathrm{Var}_\theta\left(\sum_{i=1}^n \partial_\theta \ln \mathscr{L}(x_i, \theta)\right) = \sum_{i=1}^n \mathrm{Var}_\theta(\partial_\theta \ln \mathscr{L}(x_i, \theta)) = nI_{F,1}(\theta).$$

Definició 1.4.9 (Informació de Fisher multidimensional)**.** Quan Θ és un subconjunt de $\mathbb{R}^d$, el concepte d'informació de Fisher té una generalització multiparamètrica. La condició de regularitat és l'extensió natural al cas multidimensional i aleshores la informació de Fisher es defineix com la matriu següent:

$$\begin{aligned} I_F(\theta) &:= \left(\mathbb{E}_\theta\left[\partial_{\theta_i} \ln \mathscr{L}(x, \theta) \cdot \partial_{\theta_j} \ln \mathscr{L}(x, \theta)\right]\right)_{i,j \in \{1,\ldots,d\}} \\ &= \left(-\mathbb{E}_\theta\left[\partial^2_{\theta_i,\theta_j} \ln \mathscr{L}(x, \theta)\right]\right)_{i,j \in \{1,\ldots,d\}}. \end{aligned}$$

L'última igualtat s'obté aplicant arguments similars a la proposició 1.4.5.

El resultat següent ens permet assegurar la regularitat dels models exponencials que compleixen certes condicions.

Proposició 1.4.10. *Sigui $(\Omega, \mathscr{A}, \{P_\theta, \theta \in \Theta\})$ un model estadístic uniparamètric exponencial amb funció de versemblança*

$$\mathscr{L}(x, \theta) = \exp\{C(\theta)\Phi(x) + D(\theta) + \Psi(x)\}.$$

Si es compleix que Θ és un interval de $\mathbb{R}$, que C i D són dues funcions dues vegades contínuament diferenciables amb $C'(\theta) \neq 0$ per a tot θ i Φ una funció tal que, per a tot $\theta \in \Theta, \mathbb{E}_\theta\left(\Phi(x)^2\right) < +\infty$, aleshores el model és regular.

A partir d'aquí no entra a examen.

Si en un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta, \theta \in \Theta\})$ suposem que el veritable valor de θ és θ_1, un es pot demanar en quina mesura el model permet discernir θ_1 d'un altre valor del paràmetre θ_2. Per resoldre aquesta quiestió Kullback va definir, per a cada resultat $x \in \Omega$, el que ell anomenava el poder discriminant entre el veritable valor de θ_1 i un altre valor θ_2

$$\ln \frac{\mathscr{L}(x, \theta_1)}{\mathscr{L}(x, \theta_2)}$$

Kullback va proposar emprar la mitjana del poder discriminant com a mesura de la informació.

Definició 1.4.11 (Informació de Kullback)**.** Per a $\theta_1, \theta_2 \in \Theta$, la informació de Kullback de θ_1 contra θ_2 es defineix per

$$I_K(\theta_1 \mid \theta_2) := \mathbb{E}_{\theta_1}\left[\ln \frac{\mathscr{L}(x, \theta_1)}{\mathscr{L}(x, \theta_2)}\right] = \mathbb{E}_{\theta_1}[\ln \mathscr{L}(x, \theta_1)] - \mathbb{E}_{\theta_1}[\ln \mathscr{L}(x, \theta_2)].$$

Observació 1.4.12. Aquesta informació mesura de manera global la diferència entre la funció de versemblança avaluada en θ_1 i θ_2 i, com a conseqüència del fet que la funció logaritme és còncava, es pot comprovar que la informació de Kullback sempre existeix.

Proposició 1.4.13. *La informació de Kullback sempre és positiva.*

II

Estimació puntual

2.1

ESTIMADORS

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta, \theta \in \Theta\}), \Theta \subseteq \mathbb{R}^d$, on el paràmetre θ és desconegut, i n observacions $x_1, \ldots, x_n$ d'un vector aleatori $X = (X_1, \ldots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ. L'objectiu fonamental de l'estimació puntual consisteix a trobar a partir d'aquestes observacions un valor aproximat del paràmetre θ, o d'alguna funció de $g : \Theta \longmapsto \mathbb{R}^k$.

Definició (Estadístic, recordatori). En el model estadístic paramètric anterior, anomenem *estadístic* tota aplicació mesurable $T : (\Omega, \mathscr{A}) \longrightarrow (\mathbb{R}^m, \mathscr{B}(\mathbb{R}^m))$ on $m \geq 1$ és un enter.

Definició 2.1.1 (Estimador). En aquest context anomenarem estimador de $g(\theta)$ qualsevol estadístic $T : \Omega \longmapsto g(\Theta)$ que ens serveixi per a trobar un valor aproximat de $g(\theta)$.

Estudiarem les propietats dels estimadors per poder elegir en cada cas el més convenient.

Definició 2.1.2 (Funció de pèrdua). Considerem una funció mesurable

$$\begin{array}{rccc} W: & \mathbb{R}^k \times \Theta & \longrightarrow & \mathbb{R}^+ \\ & (t, \theta) & \longmapsto & W(t, \theta) \end{array}$$

que representa l'error que cometem en proposar t com a valor estimat de $g(\theta)$. Aquesta funció és anomenada funció de pèrdua i ha de complir $W(t, \theta) = 0$ per a $t = g(\theta)$.

Definició 2.1.3 (Funció de risc). Donats una funció de pèrdua W i un estadístic T definirem com a funció de risc associada a T la funció que amitjana les pèrdues

$$\begin{array}{rccc} R_T: & \Theta & \longrightarrow & \mathbb{R} \\ & \theta & \longrightarrow & \mathbb{E}_\theta[W(T, \theta)]. \end{array}$$

Aquesta funció està definida només per aquells $\theta \in \Theta$ tals que $R_T(\theta) < \infty$.

Definició 2.1.4 (Uniformement millor). Fixada una funció de pèrdua W i dos estadístics T i S amb funció de risc finita per a qualsevol $\theta \in \Theta$, direm que T és uniformement millor que S si

$$R_T(\theta) \leq R_S(\theta), \quad \forall \theta \in \Theta.$$

Definició 2.1.5 (Òptim). Fixada una funció de pèrdua W i una classe $\mathscr{E}$ d'estadístics que tenen tots funció de risc finita per a qualsevol $\theta \in \Theta$, direm que T és òptim dins la classe $\mathscr{E}$ si

$$R_T(\theta) \leq R_S(\theta), \quad \forall \theta \in \Theta, \forall S \in \mathscr{E}$$

L'objectiu de l'estimació puntual és trobar estimadors òptims dins d'$\mathscr{E}$. Ara bé, aquest pot ser un problema molt difícil de resoldre o fins i tot irresoluble si la classe d'estadístics $\mathscr{E}$ és massa general.

A partir d'ara i per simplificar l'exposició i fer-la més entenedora, ens restringirem essencialment al cas uniparamètric i **utilitzarem com a funció de pèrdua l'error quadràtic**, és a dir:

$$W(\theta, t) = |g(\theta) - t|^2$$

Així, doncs, el que pretenem és trobar un estadístic T, amb funció de risc finita per a qualsevol $\theta \in \Theta$, que minimitzi

$$R_T(\theta) = \mathbb{E}_\theta\left[(T - g(\theta))^2\right]$$

Usant que $(T - g(\theta))^2 = (T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - g(\theta))^2$, no resulta difícil veure que:

$$R_T(\theta) = \mathbb{E}_\theta\left[(T - \mathbb{E}_\theta(T))^2\right] + \mathbb{E}_\theta\left[(\mathbb{E}_\theta(T) - g(\theta))^2\right] + 2\underbrace{\mathbb{E}_\theta\left[(T - \mathbb{E}_\theta(T))(\mathbb{E}_\theta(T) - g(\theta))\right]}_{0}$$

És clar que $\mathrm{Var}_\theta(T) = \mathbb{E}_\theta\left[(T - \mathbb{E}_\theta(T))^2\right]$ i $\mathbb{E}_\theta\left[(\mathbb{E}_\theta(T) - g(\theta))^2\right] = (\mathbb{E}_\theta(T) - g(\theta))^2$ (perquè, per exemple, $E_\theta(T), g(\theta) \in \mathbb{R}$). En definitiva:

$$R_T(\theta) = \mathrm{Var}_\theta(T) + (\mathbb{E}_\theta(T) - g(\theta))^2$$

Sembla natural estudiar estadístics que compleixin $\mathbb{E}_\theta(T) = g(\theta)$ i/o amb variància petita. Aquesta primera condició és la que tractarem tot seguit.

Definició 2.1.6 (Estadístic integrable). Un estadístic T és integrable si té esperança finita per a tot θ; és a dir, $\mathbb{E}_\theta(|T|) < \infty$, $\forall \theta \in \Theta$.

Definició 2.1.7 (Biaix). El biaix d'un estimador integrable T respecte a $g(\theta)$ és la funció $b_T(\theta) = \mathbb{E}_\theta(T) - g(\theta)$.

Definició 2.1.8 (Sense biaix). Direm que un estadístic integrable T és un estimador sense biaix de $g(\theta)$ si $\mathbb{E}_\theta(T) = g(\theta)$, per a tot $\theta \in \Theta$.

Observació 2.1.9. Els estimadors sense biaixos de $g(\theta)$ no tenen per què ser únics i hi ha models en què directament no existeixen models sense biaixos.

Proposició 2.1.10. *Si T_1, T_2 són dos estimadors sense biaix de $g(\theta)$, aleshores $\lambda T_1 + (1-\lambda)T_2$, per a tot $\lambda \in [0,1]$ també és un estimador sense biaix de $g(\theta)$.*

Exemple 2.1.11. En el cas d'una mostra de mida 1 d'una variable aleatòria amb llei de Bernoulli de paràmetre $\theta \in (0,1)$ desconegut, es pot comprovar que no existeix cap estimador sense biaix de θ^2.

Proposició 2.1.12. *Suposem que les variables aleatòries $X_1, \dots, X_n$ són idènticament distribuïdes (i no necessàriament independents), amb $\mathbb{E}_\theta(|x_1|^r) < +\infty$, per a tot $\theta \in \Theta$, $r \geq 1$. Aleshores, es compleix que $T = \frac{1}{n}\sum_{i=1}^n x_i^r$ és un estimador sense biaix de la funció $g(\theta) = \mathbb{E}_\theta(x_1^r)$.*

Demostració. En efecte, la linealitat d'esperança implica que:

$$\mathbb{E}_\theta(T) = \frac{1}{n}\mathbb{E}_\theta\left(\sum_{i=1}^n x_i^r\right) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta(x_i^r) = \mathbb{E}_\theta(x_1^r) = g(\theta).$$ ■

Corol·lari 2.1.13. *Si $X_1, \dots, X_n$ són variables aleatòries idènticament distribuïdes no necessàriament independents, la mitjana mostral $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ és un estimador sense biaix de l'esperança teòrica o mitjana poblacional $\mathbb{E}_\theta(x_1)$; és a dir, $\mathbb{E}_\theta(\bar{X}_n) = \mathbb{E}_\theta(X_1) = g(\theta)$.*

Proposició 2.1.14. *Assumim que les variables $X_1, \dots, X_n$ tenen moment de segon ordre i, a més, són VAIID. La variància mostral és un estimador esbiaixat de la variància poblacional.*

Demostració. Definim $g(\theta) = \mathrm{Var}_\theta(X_1)$ i l'intentem estimar amb $T = \frac{1}{n}\sum_i (X_i - \bar{X}_n)^2$. D'una banda, podem comprovar fàcilment que:

$$\begin{aligned}
\mathbb{E}_\theta(T) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}_\theta\left[(X_i - \bar{X}_n)^2\right] = \frac{1}{n}\sum_{i=1}^n E_\theta\left[(X_i - \mathbb{E}_\theta(X_i) + \mathbb{E}_\theta(X_i) - \bar{X}_n)^2\right] \\
&\overset{2.1.13}{=} \frac{1}{n}\left[\sum_{i=1}^n \mathbb{E}_\theta\left[(X_i - \mathbb{E}_\theta(X_i))^2\right] - 2\sum_{i=1}^n \underbrace{\mathbb{E}_\theta\left[(X_i - \mathbb{E}_\theta(X_i)) \cdot (\bar{X}_n - \mathbb{E}_\theta(\bar{X}_n))\right]}_{\mathrm{Cov}(X_i, \bar{X}_n) = 0,\ \forall i} \right. \\
&\left. + \sum_{i=1}^n \mathbb{E}_\theta\left[(\bar{X}_n - \mathbb{E}_\theta(X_i))^2\right]\right]
\end{aligned}$$

La independència de les variables i 2.1.13 impliquen:

$$\mathbb{E}_\theta\left[(X_i - \mathbb{E}_\theta(X_i))^2\right] = \mathrm{Var}_\theta(X_i) = \mathrm{Var}_\theta(X_1),\ \forall i.$$

$$\mathbb{E}_\theta\left[\left(\bar{X}_n - \mathbb{E}_\theta(X_1)\right)^2\right] = \mathbb{E}_\theta\left[\left(\bar{X}_n - \mathbb{E}_\theta(\bar{X}_n)\right)^2\right] = \mathrm{Var}_\theta(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}_\theta(X_1) = \frac{\mathrm{Var}_\theta(X_1)}{n},$$

i aleshores

$$\mathbb{E}_\theta(T) = \frac{1}{n} \cdot \left[n\,\mathrm{Var}_\theta(X_1) - n \cdot \frac{\mathrm{Var}_\theta(X_1)}{n}\right] = \frac{n-1}{n}\mathrm{Var}_\theta(X_1).$$ ■

Corol·lari 2.1.15. *Sota les mateixes hipòtesis de la proposició anterior, la variància mostral corregida no té biaix respecte a la variància poblacional.*

Demostració. En efecte, $\hat{T} = \frac{n}{n-1} \cdot T$, pel que:

$$\mathbb{E}_\theta(\hat{T}) = \frac{n}{n-1}\mathbb{E}_\theta(T) = \mathrm{Var}_\theta(X_1).$$ ■

Exemple 2.1.16 (Control de qualitat). Considerem una mostra aleatòria simple de mida n d'una variable aleatòria amb llei de Bernoulli de paràmetre desconegut $\theta \in (0,1)$. Com a conseqüència d'un corol·lari i del fet que $\mathbb{E}_\theta(x_1) = \theta$, la mitjana mostral $\bar{x}_n$ és un estimador sense biaix del paràmetre θ.

Exemple 2.1.17 (Succés poc freqüent). En aquest cas tenim una mostra aleatòria simple de mida n d'una variable aleatòria amb llei de Poisson de paràmetre desconegut $\lambda > 0$. Pel mateix argument, la mitjana mostral és un estimador sense biaix de λ. A més a més, com que $\mathrm{Var}_\lambda(x_1) = \lambda$, gràcies a una proposició, la variància mostral corregida també és un estimador sense biaix de λ.

Exemple 2.1.18 (Temps de vida). Tenim una mostra aleatòria simple de mida n d'una variable aleatòria amb llei exponencial $\mathrm{Exp}(\alpha), \alpha > 0$. La mitjana mostral és un estimador sense biaix de la funció del paràmetre $g(\alpha) = \frac{1}{\alpha}$.

Exemple 2.1.19 (Mesurament d'errors). En el cas de tenir una mostra aleatòria simple de mida n d'una variable aleatòria amb llei normal $N(\mu, \sigma^2)$, amb els dos paràmetres $\mu \in \mathbb{R}$ i $\sigma > 0$ desconeguts, la mitjana mostral $\bar{x}_n$ és òbviament un estimador sense biaix del paràmetre μ. D'altra banda, la proposició implica que la variància mostral s_n^2 té biaix respecte la funció $g(\sigma) = \sigma^2$ i, en canvi, la variància mostral corregida $\hat{s}_n^2$ no en té. També es pot comprovar que l'estadístic anomenat Von Neuman

$$T = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

no té biaix respecte σ^2. Finalment, en el cas particular en que el paràmetre μ sigui conegut, resulta evident que l'estadístic $T = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2$ és no esbiaixat respecte σ^2.

2.2 Estimadors uniformement de mínima variància

Definim la classe d'**estimadors sense biaix** de $g(\theta)$ i de quadrat integrable:

$$\mathscr{E}_g = \{T : \Omega \longrightarrow \mathbb{R} \mid \mathbb{E}_\theta(T^2) < \infty \text{ i } \mathbb{E}_\theta(T) = g(\theta),\ \forall \theta \in \Theta\}.$$

Dins d'aquesta classe d'estimadors $\mathscr{E}_g$ la funció del risc per a T val:

$$R_T(\theta) = \mathbb{E}_\theta(|T - g(\theta)|^2) = \mathrm{Var}_\theta(T).$$

Per tant, T serà òptim si $\forall S \in \mathscr{E}_g$, $\mathrm{Var}_\theta(T) \leq \mathrm{Var}_\theta(S)$, $\forall \theta \in \Theta$. En altres paraules, el que tindrà el risc més baix. I d'aquí, la definició següent.

Definició 2.2.1 (UMV). Tot estimador òptim dins la classe d'estimadors $\mathscr{E}_g$ s'anomena *estimador uniformement de mínima variància*, abreujadament UMV.

Dos preguntes naturals que hom es podria plantejar són les següents:

1. *Aquests UMV són únics?* La resposta és que sí.
2. *Existència dels UMV* (procés de construcció)? La resposta a aquesta pregunta es dona en dos teoremes que queden fora de l'abast d'aquesta assignatura: Rao-Blackwell i Lehman-Schefe.

Definició 2.2.2 (Estadístic complet). Direm que un estadístic S és complet si per a tota funció $f : g(\Theta) \longrightarrow \mathbb{R}$ mesurable, tal que $f(S)$ és integrable, es compleix:

$$\mathbb{E}_\theta(f(S)) = 0, \forall\theta \in \Theta \implies P_\theta(f(S) = 0) = 1, \forall\theta \in \Theta.$$

Teorema 2.2.3 (Rao-Blackwell). *Sigui $T \in \mathcal{E}_g$ i S un estadístic suficient. Aleshores, $h(S) = \mathbb{E}(T|S)$ és un estimador sense biaix de $g(\theta)$ i $R_{h(S)}(\theta) \leq R_T(\theta)$, per a tot $\theta \in \Theta$. A més a més:*

$$R_{h(S)}(\theta) = R_T(\theta),\ \forall\theta \in \Theta \iff T = \mathbb{E}(T|S),\ qs,\ \forall\theta \in \Theta.$$

Teorema 2.2.4 (Lehamnn-Sheffé). *Assumim $\mathcal{E}_g \neq \emptyset$ i que existeix un estadístic integrable suficient i complet S. Llavors, per qualsevol $T \in \mathcal{E}_g$, $\mathbb{E}(T|S)$ és un estimador* UMV *de $g(\theta)$.*

Proposició 2.2.5. *Siguin T_1, T_2 dos estimadors UMV de $g(\theta)$. Aleshores, $P_\theta(T_1 = T_2) = 1$, per a tot $\theta \in \Theta$.*

Demostració. Definim un nou estimador $T = \frac{1}{2}(T_1 + T_2)$, que evidentment pertany a $\mathscr{E}_g$. D'una banda, com que $R_{T_1}(\theta) = R_{T_2}(\theta) = \operatorname{Var}_\theta(T_1) = \operatorname{Var}_\theta(T_2)$, tenim que:

$$\begin{aligned} R_{T_1}(\theta) \leq R_T(\theta) &= \operatorname{Var}_\theta\left(\frac{1}{2}(T_1 + T_2)\right) \\ &= \frac{1}{4}\left(\operatorname{Var}_\theta(T_1) + \operatorname{Var}_\theta(T_2) + 2\operatorname{Cov}(T_1, T_2)\right) = \frac{1}{2}(R_{T_1}(\theta) + \operatorname{Cov}_\theta(T_1, T_2)). \end{aligned}$$

Com que:

$$\operatorname{Cov}_\theta(T_1, T_2) = \mathbb{E}_\theta\left[((T_1 - \mathbb{E}_\theta(T_1))^2)^{\frac{1}{2}}((T_2 - E_\theta(T_2))^2)^{\frac{1}{2}}\right] \leq R_{T_1}(\theta)^{\frac{1}{2}} \cdot R_{T_2}(\theta)^{\frac{1}{2}} = R_{T_1}(\theta).$$

Per tant,

$$\frac{1}{2}(R_{T_1}(\theta) + \operatorname{Cov}_\theta(T_1, T_2)) \leq \frac{1}{2}(R_{T_1}(\theta) + R_{T_1}(\theta)) = R_{T_1}(\theta).$$

En particular, totes les desigualtats són igualtats i $\operatorname{Cov}(T_1, T_2) = R_{T_1}(\theta)$. Per tant:

$$\operatorname{Var}_\theta(T_1 - T_2) = \operatorname{Var}_\theta(T_1) + \operatorname{Var}_\theta(T_2) - \operatorname{Cov}_\theta(T_1, T_2) = 0.$$

Per tant, $P_\theta(T_1 - T_2 = C) = 1$, i com que els estimadors T_1, T_2 no tenen biaix, es té $C = 0$. ■

2.3

Estimadors eficients

Sigui ara un model estadístic paramètric regular $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, Θ un interval obert, un vector aleatori $X = (X_1, \ldots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ desconeguda i $\mathscr{L}(x, \theta)$, la funció de versemblança associada.

Definició 2.3.1 (Classe d'estadístics)**.** Designem per $\mathscr{E}_F$ la classe d'estadístic T de quadrat integrable per als quals per a cada $\theta_0 \in \Theta$ existeix una funció $h : \Omega \longrightarrow \mathbb{R}$ amb $\sum_{x \in \Omega} h(x) \ll \infty$ o $\int_\Omega h(x)\, dx \ll \infty$ (segons el cardinal de Ω) tal que $|T(x) \cdot \partial_\theta \mathscr{L}(x, \theta)| \leq h(x)$, per a tot θ en un entorn de θ_0.

Teorema 2.3.2 (Cota de Cramer-Rao)**.** *Suposem que tenim un model estadístic paramètric regular com el d'abans. Sigui $T \in \mathscr{E}_F$ i considerem la funció $g(\theta) = \mathbb{E}_\theta(T)$. Aleshores,* $\operatorname{Var}_\theta(T) \geq \frac{g'(\theta)^2}{I_F(\theta)}$.

Demostració. Apliquem la desigualtat de Cauchy-Schwarz a les variables aleatòries $T - g(\theta)$ i $\partial_\theta \ln \mathscr{L}(x, \theta)$. Tenim que:

$$\left[E_\theta\left([\partial_\theta \ln \mathscr{L}(x, \theta)] \cdot [T - g(\theta)]\right)\right]^2 \leq E_\theta\left((\partial_\theta \ln \mathscr{L}(x, \theta))^2\right) \cdot E_\theta\left((T - g(\theta))^2\right) = I_F(\theta) \cdot \operatorname{Var}_\theta(T).$$

D'altra banda, en el cas absolutament continu i utilitzant les condicions de regularitat,

$$\begin{aligned} E_\theta\left[(\partial_\theta \ln \mathscr{L}(x, \theta))(T - g(\theta))\right] &= E_\theta[T \cdot \partial_\theta \ln \mathscr{L}(x, \theta)] - \underbrace{E_\theta[g(\theta) \cdot \partial_\theta \ln \mathscr{L}(x, \theta)]}_{\searrow 0} \\ &= \int_\Omega T(x) \cdot \mathscr{L}(x, \theta) \cdot \partial_\theta \ln \mathscr{L}(x, \theta)\, dx = \int_\Omega T(x) \cdot \partial_\theta \mathscr{L}(x, \theta)\, dx \\ &= \partial_\theta \left(\int_\Omega T(x) \cdot \mathscr{L}(x, \theta)\, dx \right) = \partial_\theta E_\theta(T) = \partial_\theta g(\theta) = g'(\theta). \end{aligned}$$

El cas discret es fa de manera similar. ■

Definició 2.3.3 (Estimador eficient)**.** Suposem un model estadístic paramètric regular un altre cop. Sigui T un estimador sense biaix de $g(\theta)$ que pertany a $\mathscr{E}_F$. Direm que T és eficient si per a tot $\theta \in \Theta$ es compleix que $\operatorname{Var}_\theta(T) = \frac{g'(\theta)^2}{I_F(\theta)}$. És a dir, un estimador $T \in \mathscr{E}_F$ sense biaix és eficient *si la seva funció de risc coincideix amb la cota de Cramer-Rao.*

Observació 2.3.4.

1. Via 2.3.2, T és estimador eficient de $g(\theta)$ si, i només si, existeix una funció $\lambda(\theta)$ tal que per a tot $\theta \in \Theta$, $\lambda(\theta)\, (\partial_\theta \log \mathscr{L}(x, \theta)) = T(x) - g(\theta)$. Ho formalitzem a la proposició següent, 2.3.5.

2. Si T és eficient, aleshores T és uniformement de mínima variància (UMV) (si existeix un estimador eficient aquest és únic). El recíproc, però, no és pas cert.

Proposició 2.3.5. *Sigui un model estadístic paramètric regular i $T \in \mathscr{E}_F$ un estadístic. Són equivalents:*

1. *T és eficient respecte a $g(\theta)$.*
2. *Existeix una funció $\lambda(\theta)$ tal que per a tot $\theta \in \Theta$:*

$$\lambda(\theta) \cdot \partial_\theta \ln \mathscr{L}(x, \theta) = T(x) - g(\theta), \; P_\theta \; qs.$$

Observació 2.3.6. A partir d'aquí, podem trobar fàcilment la informació de Fisher i la variància de l'estimador T. En efecte, suposem que $\lambda(\theta)$ és derivable; aleshores, derivant respecte de la θ obtenim:

$$\lambda'(\theta)\partial_\theta \ln \mathscr{L}(x, \theta) + \lambda(\theta)\partial_\theta^2 \ln \mathscr{L}(x, \theta) = -g'(\theta).$$

Prenent esperances:

$$\lambda'(\theta)\mathbb{E}_\theta(\partial_\theta \ln \mathscr{L}(x, \theta)) + \lambda(\theta)\mathbb{E}_\theta(\partial_\theta^2 \ln \mathscr{L}(x, \theta)) = -g'(\theta).$$

Llavors, com que $\mathbb{E}_\theta(\partial_\theta \ln \mathscr{L}(x, \theta)) = 0$ i $\mathbb{E}_\theta(\partial_\theta^2 \ln \mathscr{L}(x, \theta)) = -I_F(\theta)$,

$$-\lambda(\theta)I_F(\theta) = -g'(\theta) \iff \mathrm{Var}_\theta(T) = \frac{(g'(\theta))^2}{I_F(\theta)}.$$

Proposició 2.3.7. *Sigui $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$ un model estadístic uniparamètric exponencial amb funció de versemblança:*

$$\mathscr{L}(x, \theta) = \exp(C(\theta)\Phi(x) + D(\theta) + \Psi(x)).$$

Suposem que Θ és un interval de $\mathbb{R}$, C, D dues funcions dues vegades contínuament diferenciables amb $C'(\theta) \neq 0$ i que Φ és una funció tal que, per a tot $\theta \in \Theta$, $E_\theta(\Phi(x)^2) < \infty$. Aleshores, l'estadístic $T = \Phi(x)$ és un estimador sense biaix i eficient de $g(\theta) = -\frac{D'(\theta)}{C'(\theta)}$.

Demostració. Aquest model és regular. Aplicant logaritmes i derivant respecte de θ tenim:

$$\partial_\theta \log(x, \theta) = C'(\theta)\Phi(x) + D'(\theta) = C'(\theta)\left(\Phi(x) + \frac{D'(\theta)}{C'(\theta)}\right).$$

La proposició 2.3.5 implica que $\Phi(x)$ és un estimador sense biaix i eficient de $g(\theta) = -\frac{D'(\theta)}{C'(\theta)}$. ■

Observació 2.3.8. Es pot demostrar que la cota de Cramer-Rao només s'assoleix en models exponencials i, per tant, els estimadors eficients només existeixen en aquest tipus de models.

2.4

Estimadors de màxima versemblança

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^d$, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ i funció de versemblança $\mathscr{L}(x, \theta)$.

Definició 2.4.1 (EMV)**.** Un estadístic $\hat{\theta} = \hat{\theta}(x)$ a valors en Θ s'anomena *estimador de màxima versemblança* del paràmetre θ si maximitza la funció de versemblança $\mathscr{L}(x, \theta)$, és a dir, si compleix $\mathscr{L}(x, \hat{\theta}(x)) = \sup_{\theta \in \Theta} \mathscr{L}(x, \theta)$, per a tot $x \notin N$ amb $P_\theta(N) = 0$ per a tot $\theta \in \Theta$.

Observació 2.4.2 (Equacions de versemblança)**.** Suposem que la funció de versemblança és estrictament positiva per a tot $x \in \Omega$ i per a tot $\theta \in \Theta$ i, a més a més, és una funció $\mathcal{C}^2$ respecte de θ per a tot $x \in \Omega$. Aleshores, un EMV $\hat{\theta}(x)$ haurà de complir les anomenades *equacions de versemblança*:

$$\partial_{\theta_i} \ln \mathscr{L}(x, \theta)|_{\theta = \hat{\theta}(x)} = 0, \quad i = 1, \dots, d.$$

I la matriu:

$$\left(-\partial^2_{\theta_i, \theta_j} \ln \mathscr{L}(x, \theta)|_{\theta = \hat{\theta}(x)} \right)_{i,j=1,\dots,d}$$

haurà de ser definida positiva. Amb aquestes condicions obtenim $\hat{\theta}(x)$ és un màxim local; a la pràctica, però ens faltarà confirmar que és global.

Exemple 2.4.3 (Tria de nombres reals)**.** Suposem ara que la mostra aleatòria simple de mida n té llei uniforme $U(\theta - 1, \theta + 1)$, $\theta \in \mathbb{R}$. La funció de versemblança és:

$$\begin{aligned} \mathscr{L}(x, \theta) = \frac{1}{2^n} \prod_{i=1}^{n} \mathbb{1}_{[\theta-1, \theta+1]}(x_i) &= \frac{1}{2^n} \cdot \mathbb{1}_{[\theta-1, \theta+1]}(x_{(1)}) \cdot \mathbb{1}_{[\theta-1, \theta+1]}(x_{(n)}) \\ &= \begin{cases} \frac{1}{2^n}, & \text{si } x_{(n)} - 1 \leq \theta \leq x_{(1)} + 1, \\ 0, & \text{altrament.} \end{cases} \end{aligned}$$

L'EMV no és únic perquè qualsevol estimador que compleixi:

$$x_{(n)} - 1 \leq \hat{\theta}(x) \leq x_{(1)} + 1, \ \forall x \in \Omega$$

és un EMV de θ.

Exemple 2.4.4 (Mesurament d'errors)**.** Recordem que la funció de versemblança d'una mostra aleatòria simple de mida n d'una variable aleatòria amb llei normal $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$, és

$$\begin{array}{rccc} \mathscr{L}(x, \cdot): & \mathbb{R} \times (0, \infty) & \longrightarrow & (0, \infty) \\ & (\mu, \sigma^2) & \longrightarrow & \mathscr{L}(x, \mu, \sigma^2) \end{array}$$

amb

$$\mathscr{L}(x, \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}.$$

Aleshores, resolent les equacions:

$$\begin{cases} \partial_\mu \log \mathscr{L}(x, \mu, \sigma^2) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu) = 0, \\ \partial_{\sigma^2} \log \mathscr{L}(x, \mu, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - \mu)^2 = 0, \end{cases}$$

i utilitzant la matriu de derivades segones és pot comprovar que els EMV de μ i σ^2 són:

$$\hat{\mu} = \bar{x}_n \text{ i } \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2,$$

respectivament. Per veure que és un màxim global també es pot utilitzar l'argument següent:

1. Fixat σ^2 és fàcil veure que el màxim per a μ es troba en $\bar{x}_n$, per tant,

$$\mathscr{L}\left(x, \mu, \sigma^2\right) \leq \mathscr{L}\left(x, \bar{x}_n, \sigma^2\right), \quad \forall \mu.$$

2. Fixat μ, també es pot veure fàcilment que el màxim per σ^2 es troba en $\frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$, d'on:

$$\mathscr{L}(x, \mu, \sigma^2) \leq \mathscr{L}\left(x, \mu, \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2\right), \ \forall \sigma^2.$$

3. Combinant les dues desigualtats, tenim que per a tot (μ, σ^2) es compleix:

$$\mathscr{L}(x, \mu, \sigma^2) \leq \mathscr{L}(x, \bar{x}_n, \hat{\sigma}^2) \leq \mathscr{L}\left(x, \bar{x}_n, \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2\right).$$

Una propietat molt important i útil és el principi d'invariància funcional.

Proposició 2.4.5 (Principi d'invariància funcional). *Sigui $g : \Theta \longrightarrow \bar{\Theta}$ una aplicació mesurable i bijectiva entre dos oberts de $\mathbb{R}^d$. Si $\hat{\theta}$ és un EMV de θ en el model $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, aleshores $g(\hat{\theta})$ és un EMV de $\bar{\theta} = g(\theta)$ en el model $(\Omega, \mathscr{A}, \{P_{g^{-1}(\bar{\theta})} \mid \bar{\theta} \in \bar{\Theta}\})$.*

Demostració. La funció de versemblança del segon model és $\bar{\mathscr{L}}(x, \bar{\theta}) = \mathscr{L}(x, g^{-1}(\bar{\theta}))$. Així,

$$\bar{\mathscr{L}}(x, g(\hat{\theta}(x))) = \mathscr{L}(x, \hat{\theta(x)}) = \sup_{\theta \in \Theta} \mathscr{L}(x, \theta) = \sup_{\bar{\theta} \in \bar{\Theta}} \bar{\mathscr{L}}(x, \bar{\theta}).$$ ■

Exemple 2.4.6 (Temps de vida). Recordem que aquest era un model estadístic associat al temps que trigava una certa partícula radioactiva a desintegrar-se i que les durades de diferents partícules eren independents amb llei exponencial de paràmetre $\alpha > 0$ desconegut. Suposem que volem fer una estimació de la probabilitat que una partícula visqui més de 100 unitats de temps. La funció de versemblança:

$$\mathscr{L}(x,\alpha) = \alpha^n e^{-\alpha \sum_i x_i} \iff \log \mathscr{L}(x,\alpha) = n \log \alpha - \alpha \sum_i x_i.$$

$$\partial_\alpha \log \mathscr{L}(x,\alpha) = \frac{n}{\alpha} - \sum_i x_i \iff \frac{n}{\alpha} = \sum_i x_i.$$

$$\partial_\alpha^2 \log \mathscr{L}(x,\alpha) = -\frac{n}{\alpha^2}.$$

Tenim que:

$$g(\alpha) = P_\alpha(x \in (100, +\infty)) = \int_{100}^{+\infty} \alpha e^{-\alpha x}\, dx = e^{-100\alpha}.$$

Com que $\hat{\alpha}(x) = \frac{1}{\bar{x}_n}$ pel principi d'invariància funcional tenim que l'estimador de la màxima versemblança d'aquesta probabilitat $g(\alpha)$ és $e^{-\frac{100}{\bar{x}_n}}$.

Proposició 2.4.7. *Assumim que tenim un model estadístic paramètric regular i que $T(x) \in \mathscr{E}_F$ és un estimador sense biaix i eficient de θ. Aleshores, les equacions de versemblança tenen solució única $\hat{\theta}(x) = T(x)$.*

Demostració. T és un estimador eficient de $g(\theta)$ si tenim una relació d'aquest estil, amb una funció $\lambda(\theta)$:

$$\lambda(\theta)(\partial_\theta \log \mathscr{L}(x,\theta)) = T(x) - g(\theta) \xrightarrow[\text{EMV}]{g(\theta)=\theta} \lambda(\theta)(\underbrace{\partial_\theta \log \mathscr{L}(x,\theta)}_{0}) = T(x) - \theta.$$

Obtenim que $\hat{\theta}(x) = T(x)$ és una solució de l'equació de versemblança. La solució de l'equació de versemblança és, doncs, un màxim absolut; en efecte, $\partial_\theta \ln \mathscr{L}(x,\theta)$ és positiu si $\theta < T(x)$ i negatiu si $\theta > T(x)$. ■

2.5 Mètode dels moments

És un mètode molt senzill que es basa en una aplicació de la llei dels grans nombres i en el fet d'expressar els paràmetres mitjançant els moments.

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta, \theta \in \Theta\})$, amb $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$, i n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ. Assumirem que la mostra aleatòria és simple i que la llei P_θ té moment d'ordre $k \in \mathbb{N} \setminus \{0\}$, és a dir, $\mathbb{E}_\theta(|x_1|^k) < +\infty$ (això ens assegura que existeixen tots els moments d'ordre $j \leq k$). Finalment, també suposarem que el vector de paràmetres $\theta = (\theta_1, \dots, \theta_d)$ pot expressar-se en funció dels k moments poblacionals

$$\begin{cases} \theta_1 = h_1\left(m_1, \ldots, m_k\right), \\ \vdots \\ \theta_d = h_d\left(m_1, \ldots, m_k\right), \end{cases}$$

on $m_j = \mathbb{E}_\theta(x_1^j)$, $j = 1, \ldots, k$, i $h_1, \ldots, h_d : \mathbb{R}^k \longrightarrow \mathbb{R}$ són funcions contínues.

La llei dels grans nombres ens diu que els moments empírics convergeixen cap als moments poblacionals:

$$\hat{m}_j := \frac{1}{n}\sum_{i=1}^{n} X_i^j \xrightarrow[n\to\infty]{\text{qs}} m_j = \mathbb{E}_\theta\left(X_1^j\right), \quad j = 1, \ldots, k.$$

Aleshores, tenim que:

$$\begin{cases} h_1(\hat{m}_1, \ldots, \hat{m}_k) \xrightarrow[n\to\infty]{\text{qs}} h_1(m_1, \ldots, m_k) = \theta_1, \\ \vdots \\ h_d(\hat{m}_1, \ldots, \hat{m}_k) \xrightarrow[n\to\infty]{\text{qs}} h_d(m_1, \ldots, m_k) = \theta_d. \end{cases}$$

Definició 2.5.1 (Estimadors per moments)**.** Als estadístics $\hat{\theta}_i(x_1, \ldots, x_n) = h_i(\hat{m}_1, \ldots, \hat{m}_k)$, $i = 1, \ldots, d$, se'ls anomena *estimadors pel mètode dels moments.*

Definició 2.5.2 (Consistència)**.** Si tenim $(T_n)_n$ una successió d'estimadors de $g(\theta)$, direm que és consistent si $T_n \xrightarrow{P_\theta} g(\theta)$ quan $n \to \infty$ (és a dir, en probabilitat) i és fortament consistent si $T_n \xrightarrow{n\to\infty} g(\theta)$ (quasi segurament en P_θ).

Propietat 2.5.3 (Propietats asimptòtiques del mètode de moments)**.** *Podem destacar-ne els aspectes següents:*

1. *Òbviament $k \geq d$, però pot ocórrer que $k > d$.*
2. *Aquests estimadors són fortament consistents.*

Exemple 2.5.4 (Tria de nombres reals)**.** Per una mostra aleatòria simple de mida n d'una variable amb llei uniforme $U(0, \theta)$, $\theta > 0$ desconegut, un estimador pel mètode dels moments és $\hat{\theta} = 2\hat{x}_n$, ja que $\mathbb{E}_\theta(x_1) = 2\theta$.

Exemple 2.5.5 (Temps de vida)**.** Considerem un cas més general, una mostra aleatòria simple de mida n d'una variable aleatòria amb llei gamma $G(\alpha, \beta)$, $\alpha, \beta > 0$. En aquerst cas, sabem que $\mathbb{E}_\theta(x_1) = \frac{\alpha}{\beta}$ i $\mathrm{Var}_\theta(x_1) = \frac{\alpha}{\beta^2}$, $\theta = (\alpha, \beta)$. Aleshores:

$$\begin{cases} m_1 &= \frac{\alpha}{\beta} \\ m_2 &= \frac{\alpha(1+\alpha)}{\beta^2} \end{cases} \Longrightarrow \begin{cases} \alpha &= \frac{m_1^2}{m_2 - m_1^2}, \\ \beta &= \frac{m_1}{m_2 - m_1^2}. \end{cases}$$

Substituint els moments poblacionals pels moments empírics obtenim:

$$\begin{cases} \hat{\alpha} &= \frac{\hat{m}_1^2}{\hat{m}_2 - \hat{m}_1^2} = \frac{\bar{x}_n^2}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}_n^2} = \frac{\bar{x}_n^2}{s_n^2}, \\ \hat{\beta} &= \frac{\hat{m}_1}{\hat{m}_2 - \hat{m}_1^2} = \frac{\bar{x}_n}{\frac{1}{n}\sum_{i=1}^n x_i^2 - \bar{x}_n^2} = \frac{\bar{x}_n}{s_n^2}. \end{cases}$$

2.6

Propietats assimptòtiques dels estimadors

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^d$ i $X_1, \ldots, X_n$ n variables independents i idènticament distribuïdes amb llei P_θ. Sigui $(x_1, \ldots, x_n)$ una observació de $(X_1, \ldots, X_n) : \tilde{\Omega} \longrightarrow \Omega$. Suposem que per a tot n un estimador $T_n : \Omega \longrightarrow g(\Theta)$ que tindrà una forma general en funció de n.

Definició 2.6.1 (Consistent)**.** Direm que una successió d'estimadors $(T_n)_n$ és consistent si per a tot $\varepsilon > 0$ i per a tot $\theta \in \Theta$ tenim que:

$$\lim_{n\to\infty} P_\theta(|T_n - g(\theta)| > \varepsilon) = 0.$$

Aquesta definició és equivalent a dir que la successió d'estimadors $(T_n)_n$ convergeix en probabilitat a $g(\theta)$; és a dir:

$$T_n \xrightarrow[n\to\infty]{P_\theta} g(\theta), \forall \theta \in \Theta.$$

Definició 2.6.2 (Fortament consistent)**.** Direm que una successió d'estimadors $(T_n)_n$ és consistent si convergeix quasi segurament a $g(\theta)$; és a dir:

$$T_n \xrightarrow[n\to\infty]{\text{qs}} g(\theta) \text{ o } \lim_{n\to\infty} \mathbb{E}_\theta(|T_n - g(\theta)|^2) = 0,\ \forall \theta \in \Theta.$$

Definició 2.6.3 (Asimptòticament sense biaix)**.** Direm que una successió d'estimadors integrables $\{T_n, n \geq 1\}$ és asimptòticament sense biaix de $g(\theta)$ si

$$b_{T_n}(\theta) = \mathbb{E}_\theta\left(T_n\right) - g(\theta) \xrightarrow[n\to\infty]{} 0.$$

La condició

$$R_{T_n}(\theta) \xrightarrow[n\to\infty]{} 0$$

implica que la successió $\{T_n, n \geq 1\}$ és asimptòticament sense biaix.

Definició 2.6.4 (Asimptòticament normal)**.** Sigui $\{T_n, n \geq 1\}$ una successió d'estimadors de quadrat integrable sense biaix de $g(\theta)$ i on l'espai de paràmetres Θ és unidimensional. Direm que $\{T_n, n \geq 1\}$ és asimptòticament normal si

$$\sqrt{n}\left(T_n - g(\theta)\right) \xrightarrow[n\to\infty]{\mathcal{L}} N\left(0, \sigma^2(\theta)\right)$$

amb $\sigma(\theta) > 0$. Si el model és regular, direm que una successió asimptòticament normal és asimptòticament eficient si a més:

$$\sigma^2(\theta) = \frac{g'(\theta)^2}{I_F(\theta)}$$

III

Confiança

3.1 Regions de confiança

L'estimació puntual ens proporciona un estimador T que interpretem com una aproximació al valor de la funció $g(\theta)$. Ara desenvoluparem un mètode que consisteix en trobar un subconjunt de $g(\Theta)$ depenent de les observacions x, i que conté el valor del paràmetre amb una probabilitat fixada a priori. Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^d$, amb paràmetre θ desconegut, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ i una funció $g : \Theta \longrightarrow \mathbb{R}^k$. Fixem un valor $\alpha \in (0,1)$.

Definició 3.1.1 (Regió de confiança). Una regió de confiança per a $g(\theta)$ al nivell de confiança α és una família de parts de $g(\Theta)$, $\{S(x) \subset \mathbb{R}^k \mid x \in \Omega\}$, que compleix les dues propietats següents:

1. Per a tot $\theta \in \Theta$, $\{x \in \Omega \mid g(\theta) \in S(x)\} \subset \mathscr{A}$ (aquest conjunt és mesurable i per tant sobre aquest conjunt podrem calcular probabilitats, etc.).
2. Per a tot $\theta \in \Theta$, $P_\theta(x \in \Omega \mid g(\theta) \in S(x)) \geq \alpha$ (permet d'interpretar α com la probabilitat que $S(x)$ inclogui el veritable valor des paràmetre)[1].

Quan g prengui valors reals ($k = 1$) i les regions $S(x)$ siguin intervals, les regions de confiança s'anomenaran *intervals de confiança*.

Definició 3.1.2 (Funció pivotant). Una funció $\pi : \Omega \times g(\Theta) \longrightarrow \mathbb{R}^k$ s'anomena *funció pivotant* per a $g(\theta) \in g(\Theta)$ si compleix:

1. Per a tot $\theta \in \Theta$, $\pi(\cdot, g(\theta))$ és mesurable.
2. La llei de $\pi(\cdot, g(\theta))$ no depèn de θ.

Exemple 3.1.3. Prenem X una variable aleatòria que segueix una uniforme $U(0,\theta)$ amb x una única observació. Aleshores, $\frac{X}{\theta} \sim U(0,1)$; és a dir, $T_\theta := \frac{X}{\theta}$ segueix una llei que no depèn de θ i per tant tenim una funció pivotant.

Observació 3.1.4. Designem per ν la llei de probabilitat en $\mathbb{R}^k$ comuna de tots els vectors aleatoris $\pi(\cdot, g(\theta)) : \Omega \times g(\Theta) \longrightarrow \mathbb{R}^k$ i $\theta \in \Theta$. Sigui $B \in \mathscr{B}(\mathbb{R}^k)$[2] que compleix $\nu(B) \geq \alpha$. Si

[1] Si repetíssim moltes vegades l'experiència aleatòria, la regió $S(x)$ construïda a partir de les observacions x contindria el veritable valor de la funció del paràmetre $g(\theta)$ aproximadament en més d'un $100\alpha\%$ dels casos.
[2] Recordem que $\mathscr{B}(\mathbb{R}^k)$ és el conjunt de borelians de $\mathbb{R}^k$, pel que B és un borelià de $\mathbb{R}^k$.

definim:

$$S(x) = \{y \in g(\Theta) \mid \pi(x,y) \in B\} = \pi(x,\cdot)^{-1}(B) \subset g(\Theta),$$

aleshores $\{S(x) \mid x \in \Omega\}$ és una regió de confiança per a $g(\theta)$ al nivell de confiança α. En efecte, d'una banda tenim que gràcies a 3.1.2.*1.* es compleix que per a tot $\theta \in \Theta$:

$$\{x \in \Omega \mid g(\theta) \in S(x)\} = \{x \in \Omega \mid \pi(x, g(\theta)) \in B\} \in \mathscr{A},$$
$$P_\theta(x \in \Omega \mid g(\theta) \in S(x)) = P_\theta(x \in \Omega \mid \pi(x, g(\theta)) \in B) = \nu(B) \geq \alpha.$$

Exemple 3.1.5. Considerem una mostra aleatòria simple de mida n d'una variable aleatòria amb llei normal $N(\mu, \sigma^2)$ amb μ desconeguda i σ coneguda. Tenim que:

$$\nu \sim \pi(x,\mu) = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$$

és una funció pivotant perquè la llei de $\pi(\cdot, \mu)$ és una normal $N(0,1)$. Podem triar dos valors reals ζ_γ i η_γ tals que $\nu([\zeta_\gamma, \eta_\gamma]) = \gamma$ i, aleshores,

$$P_\mu\left(x \in \mathbb{R}^n \mid \zeta_\gamma \leq \sqrt{n} \cdot \frac{\bar{X}_n - \mu}{\sigma} \leq \eta_\gamma\right) = \gamma.$$

Per tant,

$$P_\mu\left(x \in \mathbb{R}^n \mid \bar{X}_n - \eta_\gamma \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + \zeta_\gamma \frac{\sigma}{\sqrt{n}}\right).$$

Per a qualsevol γ podem triar infinites parelles $\zeta_\gamma, \eta_\gamma$. En aquests casos, s'agafa l'interval més petit i, en el cas de la normal $N(0,1)$ sempre és interval centrat en el zero. Si suposem $\gamma = 0.95$, tindrem $\eta_\gamma = -\zeta_\gamma = 1.96$ i, així, l'interval de confiança per a μ és:

$$S(x) = \left[\bar{x}_n - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x}_n + 1.96\frac{\sigma}{\sqrt{n}}\right].$$

Exemple 3.1.6. Considerem una mostra aleatòria simple de mida n d'una variable aleatòria amb llei absolutament contínua i densitat $f_\theta(x) = \frac{3x^2}{\theta^3}\mathbb{1}_{[0,\theta]}(x)$, $\theta > 0$. Veiem com per mitjà de $x_{(n)}$ (no depèn de θ) podem trobar un interval de confiança per a θ amb nivell de confiança α. És fàcil comprovar que la funció de distribució de $x_{(n)}$ val:

$$F_{x_{(n)}}(t) = P_\theta(x_{(n)} \leq t) = \begin{cases} 0, & \text{si } t < 0 \\ \left(\frac{t}{\theta}\right)^{3n}, & \text{si } t \in [0,\theta) \\ 1, & \text{si } t \geq \theta \end{cases} \quad f_{x_{(n)}}(t) = 3n\left(\frac{t}{\theta}\right)^{3n-1}\mathbb{1}_{(0,\theta)}(t).$$

Pel que $\pi(x,\theta) = \frac{x_{(n)}}{\theta}$ és una funció pivotant perquè la llei de $\pi(\cdot,\theta)$ no depèn de θ. Concretament, la funció de distribució de $\frac{x_{(n)}}{\theta}$ és:

$$F_T(t) = P_\theta(x_{(n)} \leq \theta t) = \begin{cases} 0, & \text{si } t < 0 \\ t^{3n}, & \text{si } t \in [0,1) \\ 1, & \text{si } t \geq 1 \end{cases}$$

Resolem el cas $\alpha = 0,9$. S'han de trobar $\zeta_\alpha, \eta_\alpha \in [0,1]$ tal que:

$$P_\theta\left(x \in \mathbb{R}_+^n \mid \zeta_\alpha \leq \frac{x_{(n)}}{\theta} \leq \eta_\alpha\right) = F_T(\eta_\alpha) - F_T(\zeta_\alpha) = \eta_\alpha^{3n} - \zeta_\alpha^{3n} = 0,9.$$

Equivalentment:

$$P_\theta\left(\frac{x_{(n)}}{\eta_\alpha} \leq \theta \leq \frac{x_{(n)}}{\zeta_\alpha}\right).$$

Com en l'exemple anterior, existeixen infinites parelles possibles. Si volem l'interval de confiança per a θ més petit que compleix aquesta condició tindrem $S(x) = [x_{(n)}, \frac{x_{(n)}}{\sqrt[3n]{0,1}}]$ i, si volem simetria ($\eta_\alpha^{3n} = 1 - \zeta_\alpha^{3n}$) trobarem $S(x) = [\frac{x_{(n)}}{\sqrt[3n]{0,95}}, \frac{x_{(n)}}{\sqrt[3n]{0,95}}]$.

Teorema 3.1.7 (Teorema de Fisher)**.** *Sigui $(x_1, \dots, x_n)$ una mostra aleatòria simple de mida n d'una normal $N(0,1)$. Aleshores, la mitjana $\bar{X}_n$ i la variància mostral ns_n^2 són independents i, a més:*

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^n x_i \sim N\left(0, \frac{1}{n}\right) \text{ i } ns_n^2 = \sum_{i=1}^n (x_i - \bar{x}_n)^2 \sim \chi^2_{(n-1)}.$$

Corol·lari 3.1.8. *Sigui $(x_1, \dots, x_n)$ una mostra aleatòria simple de mida n d'una normal $N(\mu, \sigma^2)$. Aleshores, la mitjana i la variància mostral són independents i:*

$$\bar{x}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right), \ \frac{ns_n^2}{\sigma^2} \sim \chi^2_{(n-1)}, \ \sqrt{n-1} \cdot \frac{\bar{x}_n - \mu}{s_n} \sim t_{(n-1)}.$$

Corol·lari 3.1.9. *Sigui $(x_1, \dots, x_{n_1})$ una mostra aleatòria simple de mida n_1 d'una normal $N(\mu_1, \sigma_2^2)$ i $(y_1, \dots, y_{n_2})$ una mostra aleatòria simple de mida n_2 d'una normal $N(\mu_2, \sigma_2^2)$ independent de l'anterior. Aleshores, les dues variàncies mostrals $s_{n_1}^2$ i $s_{n_2}^2$ associades a cada mostra són independents i:*

$$\frac{n_1 s_{n_1}^2 (n_2 - 1)\sigma_2^2}{n_2 s_{n_2}^2 (n_1 - 1)\sigma_1^2} \sim F_{n_1 - 1, n_2 - 1}.$$

3.2 Regions de confiança en poblacions normals

Exemple 3.2.1 (Interval de confiança per a la mitjana amb variància coneguda)**.** En un exemple ja hem obtingut:

$$S(x) = \left[\bar{x}_n - \eta_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \bar{x}_n + \eta_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right],$$

amb $\nu([-\eta_\alpha, \eta_\alpha]) = \alpha$, on ν és la llei $N(0,1)$.

Exemple 3.2.2 (Interval de confiança per a la mitjana amb variància desconeguda)**.** Considerem una mostra aleatòria simple de mida n d'una variable aleatòria amb llei normal $N(\mu, \sigma^2)$

amb paràmetres μ, σ desconeguts, de la qual volem donar un interval de confiança per a la mitjana μ. Tenim la funció pivotant següent:

$$\nu \sim \pi(x, \mu) = \sqrt{n-1} \cdot \frac{\bar{x}_n - \mu}{s_n} \sim t_{(n-1)}.$$

Com que una t de Student és simètrica al voltant del zero, escollim un $\eta_\alpha > 0$ tal que $\nu([-\eta_\alpha, \eta_\alpha]) = \alpha$. Aleshores:

$$P_\theta\left(x \in \mathbb{R}^n,\ -\nu_\alpha \leq \sqrt{n-1} \cdot \frac{\bar{x}_n - \mu}{s_n} \leq \nu_\alpha\right) = \alpha.$$

Per tant, l'interval de confiança per μ és:

$$S(x) = \left[\bar{x}_n - \eta_\alpha \cdot \frac{s_n}{\sqrt{n-1}}, \bar{x}_n + \eta_\alpha \cdot \frac{s_n}{\sqrt{n-1}}\right].$$

Exemple 3.2.3 (Interval de confiança per a la variància)**.** Considerem la funció pivotant:

$$\nu \sim \pi(x, \sigma^2) = \frac{ns_n^2}{\sigma^2} \sim \chi^2_{(n-1)}.$$

Com que una χ^2 de Pearson no és simètrica, escollim dos reals positius $\zeta_\alpha < \eta_\alpha$ tals que:

$$P_\theta\left(x \in \mathbb{R}^n \mid \frac{ns_n^2}{\sigma^2} \leq \zeta_\alpha\right) = P_\theta\left(x \in \mathbb{R}^n \mid \eta_\alpha \leq \frac{ns_n^2}{\sigma^2}\right) = \frac{1-\alpha}{2}.$$

Així l'interval de confiança per a la variància σ^2 és $S(x) = [\frac{ns_n^2}{\eta_\alpha}, \frac{ns_n^2}{\zeta_\alpha}]$.

Exemple 3.2.4 (Radi de confiança per a la mitjana i la variància)**.** Tenim que:

$$\nu \sim \pi(x, \theta) = \left(\sqrt{n}\frac{\bar{x}_n - \mu}{\sigma}, \frac{ns_n^2}{\sigma^2}\right)$$

és una funció pivotant amb lleis normal $N(0,1)$ i χ^2_{n-1}, respectivament i, a més, independents. S'han de trobar $\zeta_\alpha, \eta_\alpha, \kappa_\alpha > 0$ tals que $\eta_\alpha < \kappa_\alpha$ i:

$$P_\theta\left(x \in \mathbb{R}^n \mid -\zeta_\alpha \leq \sqrt{n}\frac{\bar{x}_n - \mu}{\sigma} \leq \zeta_\alpha,\ \eta_\alpha \leq \frac{ns_n^2}{\sigma^2} \leq \kappa_\alpha\right) = \alpha.$$

Existeixen moltes possibilitats per determinar aquests tres valors, per exemple, denotant $\nu_1 =\sim N(0,1)$ i $\nu_2 \sim \chi^2_{n-1}$, els podem escollir de manera que es compleixin:

$$\nu_1([-\zeta_\alpha, \zeta_\alpha]) = \sqrt{\alpha} \text{ i } \nu_2([0, \eta_\alpha]) = \nu_2([\kappa_\alpha, \infty)) = \frac{1-\sqrt{\alpha}}{2}.$$

Aleshores, la regió de confiança per (μ, σ^2) serà:

$$S(x) = \left[\bar{x}_n - \zeta_\alpha\frac{\sigma}{\sqrt{n}}, \bar{x}_n + \zeta_\alpha\frac{\sigma}{\sqrt{n}}\right] \times \left[\frac{ns_n^2}{\kappa_\alpha}, \frac{ns_n^2}{\eta_\alpha}\right].$$

Exemple 3.2.5 (Raó de variàncies)**.** Un resultat previ ens dona una funció pivotant:

$$\nu \sim \pi\left(x, y, \frac{\sigma_1^2}{\sigma_2^2}\right) = \frac{n_1 s_{n_1}^2 (n_2-1)\sigma_2^2}{n_2 s_{n_2}^2 (n_1-1)\sigma_1^2} = \frac{\tilde{s}_{n_1}^2 \sigma_2^2}{\tilde{s}_{n_2}^2 \sigma_1^2} \sim F_{n_1-1, n_2-1}.$$

Com que una F de Fisher no és simètrica, per trobar l'interval de confiança buscarem dos nombres reals positius $\zeta_\alpha, \eta_\alpha$ tals que:

$$\nu([0, \zeta_\alpha]) = \frac{1-\alpha}{2} \text{ i } \nu([\eta_\alpha, \infty)) = \frac{1-\alpha}{2}.$$

Per tant, l'interval per a la raó és $S(x) = \left[\frac{\tilde{s}_{n_1}^2}{\eta_\alpha \tilde{s}_{n_2}^2}, \frac{\tilde{s}_{n_1}^2}{\zeta_\alpha \tilde{s}_{n_2}^2}\right]$.

Exemple 3.2.6 (IC per a la diferència de mitjanes amb la mateixa variància desconeguda)**.** Per diferents resultats, per exemple, perquè

$$\bar{X}_n - \bar{Y}_n \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1}, \frac{\sigma_2^2}{n_2}\right),$$

tenim que

$$\frac{\bar{x}_{n_1} - \bar{y}_{n_2} - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0,1)$$
$$\frac{n_1 s_{n_1}^2}{\sigma^2} + \frac{n_2 s_{n_2}^2}{\sigma^2} = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_{n_1})^2}{\sigma^2} + \frac{\sum_{i=1}^{n_2} (y_i - \bar{y}_{n_2})^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$$

són independents. Aleshores, per les propietats de la t-Student la funció següent és pivotant:

$$\begin{aligned}\nu &\sim \pi(x, y, \mu_1 - \mu_2) \\ &= \frac{\bar{x}_{n_1} - \bar{y}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{n_1 s_{n_1}^2 + n_2 s_{n_2}^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \sim t_{(n_1+n_2-2)}.\end{aligned}$$

Podem escollir un real positiu ζ_α tal que $\nu([-\zeta_\alpha, \zeta_\alpha]) = \alpha$ i, aleshores, obtenim l'interval de confiança per a $\mu_1 - \mu_2$ següent

$$\begin{aligned}S(x) = \Bigg[&\bar{x}_{n_1} - \bar{y}_{n_2} - \zeta_\alpha \sqrt{\frac{(n_1+n_2)\left(n_1 s_{n_1}^2 + n_2 s_{n_2}^2\right)}{n_1 n_2 (n_1+n_2-2)}}, \\ &\bar{x}_{n_1} - \bar{y}_{n_2} + \zeta_\alpha \sqrt{\frac{(n_1+n_2)\left(n_1 s_{n_1}^2 + n_2 s_{n_2}^2\right)}{n_1 n_2 (n_1+n_2-2)}} \Bigg].\end{aligned}$$

Exemple 3.2.7 (IC per a la diferència de mitjanes amb diferents variàncies desconegudes)**.** En aquest cas no tenim una resolució exacta del problema i el que es fa és donar solucions aproximades. Nosaltres donarem alguns breus comentaris extrets del llibre de Vélez i García. Si la mida de les mostres no és gaire petita ($n_1, n_2 \geq 15$) es substitueix σ_1^2 i σ_2^2 per les variàncies mostrals corregides respectives, obtenint que

$$\frac{\bar{x}_{n_1} - \bar{y}_{n_2} - (\mu_1 - \mu_2)}{\sqrt{\frac{\tilde{s}^2_{n_1}}{n_1} + \frac{\tilde{s}^2_{n_2}}{n_2}}}$$

es comportarà aproximadament com una normal $N(0,1)$. En canvi, en el cas que la mida d'una mostra sigui petita, s'empra que la quantitat anterior es comporta aproximadament com una $t_{(n)}$ on n és l'enter més pròxim a:

$$\frac{\left(\frac{\tilde{s}^2_{n_1}}{n_1} + \frac{\tilde{s}^2_{n_2}}{n_2}\right)^2}{\frac{1}{n_1+1}\left(\frac{\tilde{s}_{n_1}}{n_1}\right)^2 + \frac{1}{n_2+1}\left(\frac{\tilde{s}_{n_2}}{n_2}\right)^2}.$$

3.3 Intervals de confiança asimptòtics

En aquests casos, si la mida de la mostra és prou gran, podem buscar un interval aproximat mitjançant el comportament asimptòtic d'algun estadístic associat al model. Sigui $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}$ (ja que considerem Θ unidimensional per simplificar la presentació) i $(x_n)_n$ una successió d'observacions, cadascuna amb llei P_θ. Suposem:

1. una successió d'estadístics T_n amb moment de segon ordre finit ($\operatorname{Var}_\theta(T_n) \ll \infty$),
2. complint que $\mathbb{E}_\theta(T_n) = g(\theta)$
3. i amb el comportament asimptòtic següent:

$$\frac{T_n - g(\theta)}{\sqrt{\operatorname{Var}_\theta(T_n)}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1).$$

Si la mida de la mostra n és suficientment gran, podrem donar un interval de confiança de $g(\theta)$ utilitzant l'aproximació d'aquest quocient per la distribució normal. És a dir:

$$P_\theta\left(-\eta_\alpha \leq \frac{T_n - g(\theta)}{\sqrt{\operatorname{Var}_\theta(T_n)}} \leq \eta_\alpha\right) = \alpha.$$

Pel que:

$$T_n - \eta_\alpha\sqrt{\operatorname{Var}_\theta(T_n)} \leq g(\theta) \leq T_n + \eta_\alpha\sqrt{\operatorname{Var}_\theta(T_n)}.$$

Exemple 3.3.1. Considerem el model estadístic associat al *succés poc freqüent*, en què tenim una mostra aleatòria simple de mida n d'una variable amb llei Poisson de paràmetre λ desconegut i del qual donarem un interval de confiança asimptòtic al nivell $\alpha = 0.95$. Com que $\mathbb{E}_\lambda(x_1) = \mathrm{Var}_\lambda(x_1) = \lambda$, en aquest cas tenim que:

$$\sqrt{n} \cdot \frac{\bar{x}_n - \lambda}{\sqrt{\lambda}} \xrightarrow[n\to\infty]{\mathcal{L}} Z,\ Z \sim N(0,1).$$

Per tant, per a una mostra prou gran:

$$P_\lambda\left(x \in \mathbb{R}^n \;\Big|\; -1.96 \leq \sqrt{n} \cdot \frac{\bar{x}_n - \lambda}{\sqrt{\lambda}} \leq 1.96\right) \simeq 0.95.$$

L'interval de confiança asimptòtic de λ surt de les desigualtats respecte de λ:

$$-1.96 \leq \sqrt{n} \cdot \frac{\bar{x}_n - \lambda}{\sqrt{\lambda}} \leq 1.96 \iff \bar{x}_n - 1.96\sqrt{\frac{\lambda}{n}} \leq \lambda\bar{x}_n + 1.96\sqrt{\frac{\lambda}{n}}.$$

Observació 3.3.2. Aquest mètode té dos inconvenients: el primer, i més important, és que l'interval que trobarem no serà exacte: serà aproximat. El segon, és que pot resultar complicat aïllar $g(\theta)$ del quocient perquè la variància del denominador també depèn del paràmetre.

Exemple 3.3.3. Tenim que la mitjana i la variància de les alçadaes en centímetres dels 17 nois que cursen l'assignatura d'Estadística són $\bar{x}_{17} = 175.65$ i $s_{17}^2 \simeq 5.24$. Assumint que les observacions provenen d'una mostra aleatòria simple d'una normal $N(\mu, \sigma^2)$ amb μ i σ^2 desconegudes, pretenem donar els intervals de confiança per a μ i σ^2 al nivell de confiança $\alpha = 0.95$. Busquem primer l'interval de confiança per a μ. Si ν és la probabilitat d'una variable aleatòria amb llei t d'Student amb 16 graus de llibertat, tenim que:

$$\mu([-\eta_\alpha, \eta_\alpha]) = 0.95 \implies \eta_\alpha = 2.12 \implies \left[\bar{x}_{17} - 2.12\frac{s_{17}}{\sqrt{16}}, \bar{x}_{17} + 2.12\frac{s_{17}}{\sqrt{16}}\right] = [174.43, 176.86].$$

Per a calcular l'interval de confiança per a σ^2, en primer lloc, si ν és la probabilitat associada a una variable aleatòria amb llei $\chi^2_{(16)}$, hem de torbar $\zeta_\alpha, \eta_\alpha$, $0 < \zeta_\alpha < \eta_\alpha$, tal que $\nu([\zeta_\alpha, \eta_\alpha]) = 0.95$, buscant-los tenim que $\zeta_\alpha = 6.908$ i $\eta_\alpha = 28.845$; aleshores, obtenim:

$$\left[\frac{17s_{17}^2}{28.845}, \frac{17s_{17}^2}{6.908}\right] = [3.09, 12.9].$$

Exemple 3.3.4. Suposem ara que també coneixem la mitjana i la variància de les alçades dels 20 nois de l'assignatura Probabilitat

$$\bar{y}_{20} \simeq 175,55 \quad \text{i} \quad s_{20}^2 \simeq 6,81,$$

i que les dades també provenen d'una mostra aleatòria simple d'una normal $N(\mu, \sigma^2)$ amb paràmetres desconeguts. Ens proposem trobar l'interval de confiança per a la difèrencia de mitjanes al nivell de confiança $\alpha = 0,95$. En primer lloc haurem de veure com són les seves variàncies i en funció d'aquesta dada buscar l'interval per a la diferència de mitjanes. La funció pivotant que hem proposat per a la raó de variàncies és una F de Fisher-Snedecor $F_{16,19}$ i, per tant, l'interval per a la raó de variàncies que trobem és

$$\left[\frac{1}{2,59}\frac{\tilde{s}_{17}^2}{\tilde{s}_{20}^2}, \frac{1}{0,37}\frac{\tilde{s}_{17}^2}{\tilde{s}_{20}^2}\right] = [0,29; 2,10].$$

Com que l'1 pertany a l'interval que hem obtingut per a la raó de variàncies, podem assumir que les variàncies són desconegudes però iguals. En aquest cas, la funció pivotant que usem és una t de Student $t_{(35)}$ i els extrems de l'interval per a la diferència de mitjanes són:

$$\bar{x}_{17} - \bar{y}_{20} \pm 2,03\sqrt{\frac{37\,(17s_{17}^2 + 20s_{20}^2)}{17 \times 20 \times 35}},$$

i, per tant, l'interval és $[-1,60; 1,80]$.

Considerem ara el model del control de qualitat, en què tenim una mostra aleatòria simple de mida n d'una variable aleatòria amb llei de Bernoulli de paràmetre desconegut, $\theta \in (0,1)$. En aquest cas, l'estimador de màxima versemblança per a θ és $\hat{\theta}(x) = \bar{x}_n$. Com que $\mathbb{E}_\theta(\bar{x}_n) = \theta$ i $\mathrm{Var}_\theta(\bar{x}_n) = \frac{\mathrm{Var}_\theta(X)}{n} = \frac{\theta(1-\theta)}{n}$, el teorema central del límit implica que:

$$\sqrt{n}\frac{\bar{x}_n - \theta}{\sqrt{\theta(1-\theta)}} \xrightarrow[n\to\infty]{\mathcal{L}} N(0,1).$$

L'interval de confiança asimptòtic per a θ al nivell de confiança per a θ al nivell α que obtenim és:

$$\bar{x}_n - \eta_\alpha\sqrt{\frac{\theta(1-\theta)}{n}} \le \theta \le \bar{x}_n + \eta_\alpha\sqrt{\frac{\theta(1-\theta)}{n}}.$$

Pel que l'interval resulta:

$$\left[\bar{x}_n - \eta_\alpha\sqrt{\frac{\theta(1-\theta)}{n}}, \bar{x}_n + \eta_\alpha\sqrt{\frac{\theta(1-\theta)}{n}}\right].$$

Hem usat que:

$$-\eta_\alpha \le \frac{\sqrt{n}(\bar{x}_n - \theta)}{\sqrt{\theta(1-\theta)}} \le \eta_\alpha.$$

Prenent η_α tal que, per a $\nu \sim N(0,1)$, $\eta([-\eta_\alpha, \eta_\alpha]) = \alpha$.

Considerem ara que tenim dues mostres aleatòries simples independents de mides n_1 i n_2, respectivament. Cadascuna d'aquestes mostres prové d'una variable aleatòria amb llei de Bernoulli de paràmetre desconegut, $\theta_1, \theta_2 \in (0,1)$. Fent servir els mateixos arguments que en la secció anterior, tenim que

$$\frac{\bar{x}_{n_1} - \bar{y}_{n_2} - (\theta_1 - \theta_2)}{\sqrt{\frac{\theta_1(1-\theta_1)}{n_1} + \frac{\theta_2(1-\theta_2)}{n_2}}} \xrightarrow[n_1 \cdot n_2 \to \infty]{\mathscr{L}} N(0,1).$$

Aleshores, substituint de manera similar el denominador, obtenim que l'interval de confiança asimptòtic al nivell α per a $\theta_1 - \theta_2$ és:

$$\left[\bar{x}_{n_1} - \bar{y}_{n_2} - \eta_\alpha \sqrt{\frac{\bar{x}_{n_1}\left(1-\bar{x}_{n_1}\right)}{n_1} + \frac{\bar{y}_{n_2}\left(1-\bar{y}_{n_2}\right)}{n_2}},\right.$$
$$\left.\bar{x}_{n_1} - \bar{y}_{n_2} + \eta_\alpha \sqrt{\frac{\bar{x}_{n_1}\left(1-\bar{x}_{n_1}\right)}{n_1} + \frac{\bar{y}_{n_2}\left(1-\bar{y}_{n_2}\right)}{n_2}}\right],$$

on η_α és el mateix de la secció anterior.

3.4 Txebixev

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta, \theta \in \Theta\}), \Theta \subseteq \mathbb{R}$, i suposem que tenim un estimador T sense biaix de θ, de quadrat integrable i amb funció de risc $R_T(\theta)$. Per a $\rho > 0$ fixat, la desigualtat de Txebixev aplicada a la funció x^2 implica que

$$P_\theta\left(\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right| > \rho\right) \leq \frac{1}{\rho^2}\mathbb{E}_\theta\left(\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right|^2\right) = \frac{1}{\rho^2}.$$

Ja que:

$$\mathbb{E}_\theta\left(\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right|^2\right) = \frac{1}{R_T(\theta)}\underbrace{\mathbb{E}_\theta(|T-\theta|^2)}_{R_T(\theta)} = 1.$$

Aleshores, si volem un interval de confiança al nivell α, tenim que per a $\rho_\alpha = \frac{1}{\sqrt{1-\alpha}}$.

$$P_\theta\left(\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right| \leq \rho\right) = 1 - P\left(\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right| > \rho\right) \geq 1 - \frac{1}{\rho^2} = 1 - \frac{1}{\frac{1}{1-\alpha}} = 1 - 1 + \alpha = \alpha.$$

Pel que:

$$\begin{aligned}\left|\frac{T-\theta}{\sqrt{R_T(\theta)}}\right| \leq \rho_\alpha &\iff -\rho_\alpha\sqrt{R_T(\theta)} \leq T - \theta \leq \rho_\alpha\sqrt{R_T(\theta)} \\ &\iff T - \rho_\alpha\sqrt{R_T(\alpha)} \leq \theta \leq T + \rho_\alpha\sqrt{R_T(\theta)} \\ &\iff P_\theta\left(T - \rho_\alpha\sqrt{R_T(\theta)} \leq \theta \leq T + \rho_\alpha\sqrt{R_T(\theta)}\right) \geq \alpha.\end{aligned}$$

O bé aïllem θ, o, en el seu defecte, si T és un bon estimador de θ, donem com a interval de confiança

$$\left[T - \rho_\alpha\sqrt{R_T(T)}, T + \rho_\alpha\sqrt{R_T(T)}\right].$$

Aquest mètode no és gaire recomanable perquè la desigualtat de Txebixev no és del tot òptima i els intervals que s'obtenen són molt grans. En altres paraules, la fita és molt grollera, però les hipòtesis que necessitem per aplicar-la són mínimes.

IV

Testos

4.1 Tests d'hipòtesis

Els tests d'hipòtesis permeten contestar la pregunta de si la mostra o conjunt d'observacions corrobora de manera significativa aquesta teoria. Precisarem matemàticament què s'entén per *científicament demostrat*. En altres paraules, ens preguntem com podem formalitzar matemàticament la validació o rebuig d'una hipòtesi de treball.

Exemple 4.1.1. Suposem que tenim una moneda que hem llençat 10 vegades i ens preguntem si realment la moneda és perfecta. El model estadístic associat a aquest fenomen és $(\Omega, \mathscr{A}, \{P_\theta^{10} \mid \theta \in [0,1]\})$, amb $\Omega = \{\text{cara, creu}\}$, $\mathscr{A} = \mathscr{P})(\Omega)$ i $P_\theta(\{\text{cara}\}) = \theta$ i $P_\theta(\{\text{creu}\}) = 1 - \theta$. Si la moneda no és perfecta, és equivalent a formular-se la pregunta que el valor de θ és diferent de $\frac{1}{2}$? Per resoldre aquest problema haurem de proposar una regla de decisió que, a partir de les observacions ens permeti descartar, si és el cas, que $\theta = \frac{1}{2}$.

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$. Qüestionar-se alguna cosa del paràmetre és equivalent a fer una *partició* de l'espai de paràmetres $\Theta = \Theta_0 \cup \Theta_1$ i $\Theta_0 \cap \Theta_1 = \emptyset$, i demanar-se si θ pertany o no a Θ_0.

Definició 4.1.2 (Hipòtesi nul·la)**.** L'afirmació que volem confirmar o rebutjar, és a dir, la qüestió que ens estem plantejant, serà la hipòtesi nul·la: $H_0 \mid \theta \in \Theta_0$.

Definició 4.1.3 (Hipòtesi alternativa)**.** L'altre cas, la hipòtesi diferent de la hipòtesi nul·la pren el nom d'hipòtesi alternativa: $H_1 \mid \theta \in \Theta_1 = \Theta \setminus \Theta_0$.

Definició 4.1.4 (Hipòtesi simple o composta)**.** Quan un dels dos conjunts, Θ_0 o Θ_1, consta d'un sol element, direm que la hipòtesi corresponent és simple. Altrament, direm que és composta.

Observació 4.1.5. En el model estadístic $(\Omega, \mathscr{A}, \mathscr{P})$ podem considerar una partició $\mathscr{P} = \mathscr{P}_0 \cup \mathscr{P}_1$. La gran majoria de nocions presentades a partir d'ara poden ser generalitzades al cas no paramètric.

Exemple 4.1.6 (de la moneda)**.** En aquest exemple, la hipòtesi nul·la correspon al fet que la moneda sigui perfecta, i l'alternativa al fet que no ho sigui. L'espai de paràmetres el model escriure com $\Theta = \{\theta \mid \theta \in [0,1]\}$ i aleshores $H_0 : \theta = \frac{1}{2}$ contra $H_1 : \theta \neq \frac{1}{2}$. En aquest cas, la hipòtesi nul·la $\Theta_0 = \{\frac{1}{2}\}$ és simple i l'alternativa $\Theta_1 = [0,1] \setminus \{\frac{1}{2}\}$ és composta.

Exemple 4.1.7 (del dau). Imaginem que hem llançat 100 vegades un dau i ens preguntem si aquest dau és perfecte o està trucat. El model estadístic associat és:

$$\Omega = \{1,2,3,4,5,6\}^{100}, \mathscr{F} = \mathscr{P}(\Omega), \left\{P_\theta^{100}, \theta \in \Theta\right\},$$

amb $\Theta = \left\{(\theta_1,\ldots,\theta_6) \in [0,1]^6, \sum_{i=1}^6 \theta_i = 1\right\}$, i essent $P_\theta(\{i\}) = \theta_i, \forall i \in \{1,\ldots,6\}$. Aquí, qüestionar-se si el dau és perfecte és plantejar-se si $\theta_i = \frac{1}{6}, \forall i \in \{1,\ldots,6\}$, i ho podem formalitzar de la manera següent:

$$H_0 : \theta_i = \frac{1}{6} \forall i \in \{1,\ldots,6\} \quad \text{contra} \quad H_1 : \theta_i \neq \frac{1}{6} \text{ per algun } i \in \{1,\ldots,6\}.$$

La hipòtesi nul·la també és simple i l'alternativa és composta.

Exemple 4.1.8 (Control de qualitat). Considerem el model estadístic associat a l'exemple del control de qualitat, és a dir:

1. El model: $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n), \{P_\theta^n \mid \theta \in [0,1]\}$.
2. $P_\theta(\{1\}) = \theta$, θ la probabilitat que la peça sigui defectuosa.

En tot procés de fabricació de qualsevol objecte es produeixen una certa proporció de peces defectuoses. Una pregunta lògica és demanar-se si la proporció $\theta \in [0,1]$ de peces defectuoses d'un lot és inferior a una quantitat θ_0, ja que en el cas que la proporció fos més gran que θ_0 rebutjaríem el lot. Per saber-ho podem comprovar-les totes, o bé prendre'n una mostra de mida n i plantejar un test per a les hipòtesis següents:

$$H_0 : \theta \leq \theta_0 \quad \text{contra} \quad H_1 : \theta > \theta_0,$$

que correspon a la partició de l'espai de paràmetres $\Theta_0 = [0,\theta_0]$ contra $\Theta_1 = (\theta_0, 1]$. Les dues hipòtesis són compostes.

Les hipòtesis nul·la i alternativa defineixen un problema de test o contrast d'hipòtesis. Ara s'ha de fixar una regla de decisió tal que a partir de les observacions puguem decidir si acceptem o rebutgem la hipòtesi nul·la. Així, un test d'hipòtesis φ és donat per una partició del conjunt Ω,

$$\Omega = A_0 \cup A_1, \quad A_0 \cap A_1 = \emptyset, \quad \text{amb } A_0 \in \mathscr{A},$$

tal que

$$\begin{aligned} \varphi : \quad \Omega &\longrightarrow \{0,1\} \\ x &\longmapsto \mathbb{1}_{A_1}(x) \end{aligned}$$

La regla de decisió que en resulta és la següent:

1. Si les observacions x pertanyen al conjunt A_0, és a dir, $\varphi(x) = 0$, acceptarem la hipòtesi nul·la.

2. Si les observacions x pertanyen al conjunt A_1, és a dir, si $\varphi(x) = 1$, rebutjarem la hipòtesi nul·la.

D'acord amb aquesta regla de decisió, anomenarem *regió d'acceptació* A_0 i *regió de crítica o rebuig* A_1. Hem de resoldre el problema de com determinar la regió d'acceptació, és a dir, necessitem algun tipus d'eina que ens permeti escollir una regió d'acceptació i no altres. Per això, treballarem amb les probabilitats de prendre decisions incorrectes que comporta cada regla de decisió. Quan prenem una decisió podem cometre dos tipus d'errors:

1. Podem rebutjar la hipòtesi nul·la malgrat que sigui certa.
2. Podem acceptar la hipòtesi nul·la malgrat que sigui falsa.

Definició 4.1.9 (Errors de primera i segona espècie)**.**

1. L'error de *primera espècie* és la probabilitat de rebutjar la hipòtesi nul·la tot i ésser certa (en altres paraules, un *fals positiu*) i que definirem com: $\alpha_1(\theta) = P_\theta(A_1) = \mathbb{E}_\theta(\varphi(x)) = \Phi(\theta)$, per a tot $\theta \in \Theta_0$.
2. L'error de *segona espècie*, que consisteix en la probabilitat d'acceptar la hipòtesi nul·la tot i ser falsa (en altres paraules, un *fals negatiu*), i que definirem com $\alpha_2(\theta) = P_\theta(A_0) = 1 - \mathbb{E}_\theta(\varphi(x)) = 1 - \Phi(\theta)$, per a tot $\theta \in \Theta_1$.

Aquesta $\Phi : \Theta \longrightarrow [0,1]$ és la funció de potència definida per $\mathbb{E}(\varphi(x)) = P_\theta(A_1)$.

	Acceptar H_0	Rebutjar H_0
H_0 certa	Decisió correcta	Error de primera espècie
H_0 falsa	Error de segona espècie	Decisió correcta

Definició 4.1.10 (Potència del test)**.** La restricció de Φ sobre $\theta \in \Theta_1$ s'anomena *potència* del test φ contra l'alternativa θ, i representa la probabilitat de rebutjar la hipòtesi nul·la quan és falsa, ja que el valor vertader del paràmetre és θ.

Definició 4.1.11 (Nivell de significació)**.** L'error màxim de primera espècie s'anomena *nivell de significació*. Sigui $\alpha \in (0,1)$, direm que el test φ té nivell de significació α si $\sup_{\theta\in\Theta_0} \alpha_1(\theta) \leq \alpha$.

Ara plantejarem alguns mètodes per a minimitzar els errors. Idealment, hauríem de minimitzar els dos.

Observació 4.1.12. Considerem el model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$ i sigui $\theta_0 \in \Theta$. A partir de les regions de confiança podem construir tests d'hipòtesis a un nivell de significació α donat per contrastar $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. En efecte, suposem que $[T_1, T_2]$ és un interval de confiança al nivell de confiança γ per al paràmetre θ és a dir,

$$P_\theta(\theta \in [T_1, T_2]) = P_\theta(\omega \in \Omega \mid \theta \in [T_1(\omega), T_2(\omega)] \geq \gamma, \ \forall \theta \in \Theta.$$

Aleshores, $A_0 = \{\omega \in \Omega \mid T_1(\omega) \le \theta_0 \le T_2(\omega)\}$ és la regió d'acceptació d'un test φ al *nivell de significació* $\alpha = 1 - \gamma$, ja que $P_{\theta_0}(A_1) = 1 - P_{\theta_0}(A_0) \le 1 - \gamma = \alpha$. Aquest mètode té l'inconvenient de no controlar l'error de segona espècie, o el que és el mateix, la potència del test.

Hem de fixar una estratègia per determinar un test que controli els dos errors. Dels dos errors, el de primera espècie serà el que tindrà pitjors conseqüències i, entre les decisions possibles, considerarem únicament les que tinguin aquest error afitat per una quantitat fixada a priori; després, dins d'aquesta classe de tests amb error de primera espècie afitat, provarem de trobar el que minimitzi l'error de segona espècie i maximitzi, doncs, la potència del test.

Definició 4.1.13 (UMP). Designarem $\mathscr{D}_\alpha$ el conjunt de tots els tests de nivell de significació α. Aleshores, donats dos testos φ i φ' de $\mathscr{D}_\alpha$ direm que φ és més potent que φ' si $\Phi_{\varphi'}(\theta) \le \Phi_\varphi(\theta)$, per a tot $\theta \in \Theta_1$. Fixat un nivell de significació α, direm que un test de la classe $\mathscr{D}_\alpha$ *uniformement de màxima potència* (UMP) si és més potent que qualsevol altre test de $\mathscr{D}_\alpha$.

Observació 4.1.14.

1. En general pot ocórrer que no existeixi un test UMP o, en cas de no existir, que no sigui únic. Més endavant donarem resultats en aquesta direcció.
2. En aquesta estratègia les dues hipòtesis no són tractades de la mateixa manera i, per tant, el tractament no és pas simètric. Les dues hipòtesis no són intercanviables.

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\}$ amb funció de versemblança associada $L(x, \theta)$.

Lema 4.1.15 (de Neyman-Pearson). *Fixat $\alpha \in (0,1)$ i assumim que existeix una constant positiva c_α tal que $A_0 = \{x \mid \mathscr{L}(x,\theta_1) \le c_\alpha \mathscr{L}(x,\theta_0)\}$ satisfà $P_{\theta_0}(A_0^c) = \alpha$. Aleshores, el test $\varphi(x) = \mathbb{1}_{A_0^c}(x)$ és UMP al nivell de significació α per contrastar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$. Quan $\mathscr{L}(x,\theta_0)$ no s'anul·la (quan no depèn de θ), aleshores posem $A_0 = \{\frac{\mathscr{L}(x,\theta_1)}{\mathscr{L}(x,\theta_0)} \le c_\alpha\}$.*

Observació 4.1.16. *Què succeeix si no podem assolir $P_{\theta_0}(A_0^c) = \alpha$?* Si per exemple P_{θ_0} és una llei discreta, molt probablement no podrem obtenir una regió tal que $P_{\theta_0}(A_0^c) = \alpha$ perquè la funció de distribució associada, F_{θ_0}, tindrà salts. Com hem d'exigir que tingui nivell de significació α buscarem una regió tal que $\alpha' = P_{\theta_0}(A_0^c) \le \alpha$, $\alpha' < \alpha$. El lema de Neyman-Pearson ens assegura que A_0 és una regió d'un test UMP al nivell de significació α'; també, una regió de nivell de significació α, però en aquest darrer cas no sabrem si és UMP.

Ho veiem en el següent exemple.

Exemple 4.1.17 (Control de qualitat). Considerem el model estadístic associat a l'exemple del control de qualitat. Fixem dos valors del paràmetre $\theta_0 < \theta_1$. Aleshores, contrastarem les hipòtesis:

$$H_0 : \theta = \theta_0 \quad \text{contra} \quad H_1 : \theta = \theta_1.$$

Per tractar-lo des de la perspectiva del teorema, analitzarem el quocient de les funcions de versemblança

$$\frac{\mathscr{L}(x,\theta_1)}{\mathscr{L}(x,\theta_0)} = \frac{\theta_1^{S_n(x)}\,(1-\theta_1)^{n-S_n(x)}}{\theta_0^{S_n(x)}\,(1-\theta_0)^{n-S_n(x)}} = \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{S_n(x)},$$

on $S_n(x) = \sum_{i=1}^n x_i$. Ja que $\theta_1\,(1-\theta_0) > \theta_0\,(1-\theta_1)$ tenim que per a tot c_α:

$$\begin{aligned} A_0 &= \left\{ x \mid \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{S_n(x)} \le c_\alpha \right\} = \left\{ x \mid \left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{S_n(x)} \le c'_\alpha \right\} \\ &= \left\{ x \mid S_n(x)\cdot \ln\left(\frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right) \le k_\alpha \right\} = \left\{ x \mid S(x) \le k'_\alpha \right\} \\ &= \left\{ x \mid S_n(x) \le \left(\ln \frac{\theta_1(1-\theta_0)}{\theta_0(1-\theta_1)}\right)^{-1} \left[\ln c_\alpha - n \ln \frac{1-\theta_1}{1-\theta_0}\right] \right\}. \end{aligned}$$

És a dir, podem expressar-la de manera equivalent com $A_0 = \{x \mid S_n(x) \le N_\alpha\}$. El problema apareix en haver de buscar una constant N_α tal que $P_{\theta_0}\left(A_0^c\right) = \alpha$. En efecte, la variable $S_n(x)$ sota la hipòtesi nul·la és una binomial $B\left(n,\theta_0\right)$ i, per tant, la variable és discreta i la seva funció de distribució F té salts, pel que la probabilitat $P_{\theta_0}\left(A_0^c\right) = 1 - F\left(N_\alpha\right)$ pot no assolir el valor α. Aleshores, elegirem N_α el menor nombre real tal que $P_{\theta_0}\left(A_0^c\right) = \alpha' \le \alpha$. El lema de Neyman i Pearson ens assegura que A_0 és un test UMP al nivell α' per contrastar H_0 contra H_1. Observem que si haguéssim triat $\theta_0 > \theta_1$ obtindríem la regió d'acceptació del test amb la desigualtat contrària, $A_0 = \{x \mid S_n \ge N_\alpha\}$.

Demostració de 4.1.15. Òbviament $\varphi = \mathbb{1}_{A_0^c}$ és un test de significació α. Haurem de comprovar que és un test UMP. Sigui $\varphi' \in \mathscr{D}_\alpha$ qualsevol, això vol dir que si B_0 és la regió d'acceptació de φ', $P_{\theta_0}(B_0^c) \le \alpha$. Volem veure que $P_{\theta_1}(B_0^c) \le P_{\theta_1}(A_0^c)$ (és a dir, que un és més potent que l'altre). Podem descompondre A_0 i B_0 en unions disjuntes de la forma:

$$\begin{aligned} A_0 &= (A_0 \cap B_0^c) \cup (A_0 \cap B_0) = A \sqcup B, \\ B_0 &= (B_0 \cap A_0^c) \cup (B_0 \cap A_0) = A' \sqcup B, \end{aligned}$$

amb $A = A_0 \cap B_0^c$, $B = A_0 \cap B_0$ i $A' = B_0 \cap A_0^c$. Aleshores, per provar que $P_{\theta_1}(B_0^c) \le P_{\theta_1}(A_0^c)$, és equivalent a provar $P_{\theta_1}(A') \le P_{\theta_1}(A)$. Per construcció d'$A_0$, sabem que $P_{\theta_0}(A_0) = 1 - \alpha$ i com B_0 és una regió d'acceptació de $\varphi' \in \mathscr{D}_\alpha$, tenim que $P_{\theta_0}(B_0) \ge 1 - \alpha$. Observem que:

$$P_{\theta_0}(A') + P_{\theta_0}(B) = P_{\theta_0}(B_0) \ge P_{\theta_0}(A_0) = P_{\theta_0}(A) + P_{\theta_0}(B) \implies P_{\theta_0}(A') \ge P_{\theta_0}(A).$$

D'una banda, en el conjunt A' tenim que $\mathscr{L}(x,\theta_1) > c_\alpha \cdot \mathscr{L}(x,\theta_0)$ i, suposant per exemple que P_{θ_0} i P_{θ_1} són absolutament contínues, tenim:

$$P_{\theta_1}(A') = \int_{A'} \mathscr{L}(x,\theta_1)\,dx \geq c_\alpha \int_{A'} \mathscr{L}(x,\theta_0)\,dx = c_\alpha \cdot P_{\theta_0}(A'),$$

aquesta desigualtat és vàlida també per a probabilitats discretes ja que només cal canviar les integrals per sumes. D'altra banda, en A es compleix $\mathscr{L}(x,\theta_1) \leq c_\alpha \cdot \mathscr{L}(x,\theta_0)$; per tant $P_{\theta_1}(A) \leq c_\alpha P_{\theta_0}(A)$. Aleshores,

$$P_{\theta_1}(A') \geq c_\alpha P_{\theta_0}(A) \geq P_{\theta_1}(A). \qquad \blacksquare$$

Observació 4.1.18. En el cas senzill de contrastar hipòtesis simples, el lema de Neyman-Pearson ens permet donar una regió d'acceptació A_0 d'un test UMP al nivell α, sempre que existeixi una constant c_α que satisfaci $P_{\theta_0}(A_0^c) = \alpha$. Però això no sempre serà possible. Si ens trobem en un cas d'aquests el que podem fer és determinar la constant c_α de manera que $\alpha' = P_{\theta_0}(A_0^c) \leq \alpha$. Aleshores, A_0 serà la regió d'acceptació d'un test UMP al nivell α'.

Observació 4.1.19 (Extensió a constrastos d'hipòtesis unilaterals). Suposem que tenim un model estadístic paramètric $(\Omega, \mathcal{A}, \{P_\theta \mid \theta \in \Theta\})$, i que mitjançant el lema de Neyman-Pearson hem obtingut un test UMP al nivell de significació α amb regió A_0 per a contrastar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$, assumint $\theta_0 < \theta_1$. Aleshores, si la regió A_0 no depèn de θ_1 i la funció de potència $\Phi(\theta) = P_\theta(A_0^c)$ és creixent[1], la regió d'acceptació A_0 també serà la regió d'acceptació d'un test UMP al nivell de significació α per a contrastar $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$.

4.2 Test de la raó de versemblança

Considerem un model estadístic paramètric $(\Omega, \mathscr{A}, \{P_\theta \mid \theta \in \Theta\})$, $\Theta \subset \mathbb{R}^d$, amb paràmetre θ desconegut, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n) : \tilde{\Omega} \longrightarrow \Omega$ amb llei P_θ i funció de versemblança $\mathscr{L}(x,\theta)$.

Definició 4.2.1 (Raó de versemblança). Donada una partició de l'espai de paràmetres $\Theta = \Theta_0 \cup \Theta_1$; anomenarem *raó de versemblança* el quocient:

$$\lambda(x) = \frac{\mathscr{L}(x,\Theta_0)}{\mathscr{L}(x,\Theta)}; \ \mathscr{L}(x,\Theta_0) = \sup_{\theta\in\Theta_0} \mathscr{L}(x,\theta), \ \mathscr{L}(x,\Theta) = \sup_{\theta\in\Theta} \mathscr{L}(x,\theta)$$

Fixem-nos que $\lambda(x) \in [0,1]$, i si $\hat{\theta}(x)$ és EMV, $\mathscr{L}(x,\Theta) = \mathscr{L}(x,\hat{\theta}(x))$.

[1] Això vol dir que la funció de versemblança creix respecte el paràmetre.

Definició 4.2.2 (Test de la raó de versemblança). Anomenarem *test de la raó de versemblança* el nivell de significació $\alpha \in (0,1)$ per contrastar la hipòtesi $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \in \Theta_1$, el que té per regió d'acceptació $A_0 = \{x \mid \lambda(x) \geq c_\alpha\}$, on c_α ($c_\alpha \leq 1$) es determina de manera que $P_\theta(A_0^c) \leq \alpha$, per a tot $\theta \in \Theta_0$.

Òbviament ens interessa els que tinguin la regió A_0 tan petita com sigui possible.

Observació 4.2.3. Si en aquest model estadístic existeix l'estimador de màxima versemblança $\hat\theta(x)$, tindrem que $\mathscr{L}(x, \hat\theta(x)) = \sup_{\theta\in\Theta} L(x,\theta)$, i si en el model corresponent a la hipòtesi nul·la també podrem obtenir l'estimador de màxima versemblança, $\hat\theta_0(x)$, tindrem:

$$\mathscr{L}(x, \hat\theta_0(x)) = \sup_{\theta\in\Theta_0} \mathscr{L}(x,\theta) \implies \lambda(x) = \frac{\mathscr{L}(x,\hat\theta_0(x))}{\mathscr{L}(x,\hat\theta(x))}.$$

Si les premisses d'aquesta observació són vàlides, la justificació intuïtiva del per què d'aquesta regió d'acceptació seria la següent. Si la hipòtesi nul·la és certa, el valor de $\mathscr{L}(x,\hat\theta_0(x))$ serà proper al de $\mathscr{L}(x,\hat\theta(x))$ i, per tant, el valor de $\lambda(x) \to 1$. En altres paraules, sota hipòtesi simple coincideixen i tenen bons comportament asimptòtics.

Proposició 4.2.4. *Suposem que volem contrastar $H_0 : \theta = \theta_0$ i $H_1 : \theta = \theta_1$, $\Theta = \{\theta_0\} \cup \{\theta_1\}$. Si existeix el test de Neyman-Pearson i existeix el test de la raó de versemblança per contrastar hipòtesis simples, aleshores coincideixen.*

Demostració. Sigui A_0 la regió d'acceptació d'un test de la raó de versemblança al nivell de significació α per contrastar $H_0 : \theta = \theta_0$ contra $H_1 : \theta = \theta_1$. Si definim:

$$U = \left\{x \in \Omega \mid \frac{\mathscr{L}(x,\theta_1)}{\mathscr{L}(x,\theta_0)} \leq \frac{1}{c_\alpha}\right\} \text{ i } V = \{x \in \Omega \mid \mathscr{L}(x,\theta_1) \geq \mathscr{L}(x,\theta_0)\},$$

tenim que

$$\begin{aligned} A_0 &= \left\{x \mid \frac{\mathscr{L}(x,\theta_0)}{\mathscr{L}(x,\theta_0) \vee \mathscr{L}(x,\theta_1)} \geq c_\alpha\right\} \\ &= \left\{x \mid \frac{\mathscr{L}(x,\theta_0)}{\mathscr{L}(x,\theta_1)} \geq c_\alpha, \mathscr{L}(x,\theta_1) \geq \mathscr{L}(x,\theta_0)\right\} \cup \left\{x \mid \frac{\mathscr{L}(x,\theta_0)}{\mathscr{L}(x,\theta_0)} \geq c_\alpha, \mathscr{L}(x,\theta_1) < \mathscr{L}(x,\theta_0)\right\} \\ &\overset{2}{=} \left\{x \mid \frac{\mathscr{L}(x,\theta_0)}{\mathscr{L}(x,\theta_1)} \geq c_\alpha, \mathscr{L}(x,\theta_1) \geq \mathscr{L}(x,\theta_0)\right\} \cup \{x \mid \mathscr{L}(x,\theta_1) < \mathscr{L}(x,\theta_0)\} = (U \cap V) \sqcup V^c. \end{aligned}$$

I com $V^c \subset U$, $A_0 = U$. A l'última igualtat hem usat que $V^c \subset U$. Per tant, A_0 coincideix amb la regió d'acceptació que proporciona el lema de Neyman-Pearson. ■

[2] És clar que $\frac{\mathscr{L}(x,\theta_0)}{\mathscr{L}(x,\theta_0)} = 1$, pel que $1 \geq c_\alpha$, però això és cert per hipòtesi; podem prescindir d'aquesta condició.

Proposició 4.2.5. *Si existeix un estadístic suficient per al paràmetre θ, el test de la raó de versemblança és funció d'aquest estadístic.*

Demostració. Pel criteri de factorització de Neymann i Fisher, si $T(x)$ és suficient es compleix que $L(x, \theta) = \psi(T(x), \theta)h(x)$ i per tant:

$$\lambda(x) = \frac{\sup_{p \in \Theta_0} \mathscr{L}(x, \theta)}{\sup_{p \in \Theta} \mathscr{L}(x, \theta)} = \frac{\sup_{p \in \Theta_0} \psi(T(x), \theta)}{\sup_{p \in \Theta} \psi(T(x), \theta)}.$$ ■

En molts casos serà difícil conèixer la distribució de $\lambda(x)$ sota la hipòtesi nul·la i trobar, per tant, el valor de la constant c_α perquè el test proposat tingui nivell de significació α. El resultat següent ens dona una propietat asimptòtica del test de la raó de versemblança.

Teorema 4.2.6 (Quan no coneixem la llei de la raó de versemblança). *Suposem $\Omega \subset \mathbb{R}$ i que l'espai de paràmetres Θ és un obert d $\mathbb{R}^k$. Assumim que el model estadístic associat és regular i que l'EMV és únic (i.e. per a cada $n \geq 1$, la solució a les equacions de versemblança ho és). Aleshores, quan contrastem $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$ tenim que sota la hipòtesi nul·la:*

$$-2 \ln \lambda_n(x) \xrightarrow[n \to \infty]{\mathcal{L}} \chi^2_{(k)},$$

on $\lambda_n(x)$ és la raó de versemblança per a una mostra de mida n.

Corol·lari 4.2.7. *En condicions suficients de regularitat, si contrastem $H_0 : \theta \in \Theta_0$ contra $H_1 : \theta \notin \Theta_0$, podem utilitzar que:*

$$-2 \ln \lambda_n(x) \xrightarrow[n \to \infty]{\mathcal{L}} \chi^2_{(r)},$$

on r és la diferència entre el nombre de paràmetres lliures de l'espai de paràmetres i el nombre de paràmetres lliures de Θ_0.

Exemple 4.2.8 (del dau perfecte). Suposem que hem llançat n vegades un dau i hem obtingut com a freqüències $n_1, \ldots, n_6$, n_i amb i-èsima cara i $n = n_1 + \cdots + n_6$. Pretenem verificar si el dau és perfecte i, per això, haurem de contrastar les hipòtesis:

$$H_0 : p_1 = \cdots = p_6 = \frac{1}{6} \text{ contra } H_1 : \exists i \mid p_i \neq \frac{1}{6},\ i \in \{1, \ldots, 6\}.$$

La funció de versemblança d'aquest model és:

$$L(x, p) = \frac{n!}{n_1! \cdots n_6!} \cdot p_1^{n_1} \cdots p_6^{n_6},\ p_1 + \cdots + p_6 = 1.$$

Per trobar l'EMV haurem d'usar el mètode conegut com a *multiplicadors de Lagrange*. Considerem així la funció:

$$F(p_1, \ldots, p_6; \zeta) = \sum_{i=1}^{6} n_i \ln p_i + \zeta(p_1 + \cdots + p_6 - 1),$$

on derivant tenim que:

$$\frac{\partial F}{\partial p_i} = \frac{n_i}{p_i} + \zeta = 0,\ \forall i \in \{1, \dots, 6\}.$$
$$\frac{\partial F}{\partial \zeta} = p_1 + \dots + p_6 - 1 = 0.$$

Fent els passos corresponents podem obtenir que el màxim s'assoleix en $\zeta = -n$ i $p_i = \frac{n_i}{n}$, $\forall i \in \{1, \dots, 6\}$ i així:

$$\sup_{p\in\Theta} \mathscr{L}(x,p) = \frac{n!}{n_1! \cdots n_6!} \cdot \left(\frac{n_1}{n}\right)^{n_1} \cdots \left(\frac{n_6}{n}\right)^{n_6},$$

on $\Theta = \{(p_1, \dots, p_6) \in [0,1]^6 \mid p_1 + \dots + p_6 = 1\}$. Aleshores, la regió d'acceptació és:

$$A_0 = \{x \mid \lambda(x) \geq c_\alpha\} = \left\{x \mid \frac{n^n}{6^n n_1^{n_1} \cdots n_6^{n_6}}\right\},$$

de la qual no resulta fàcil trobar la constant c_α tal que $P_{\theta_0}(A_0^c) = \alpha$.

$$P_{\theta_0}(A_0^c) = P_{\theta_0}(\lambda(x) < c_\alpha) = P_{\theta_0}\left(\frac{n^n}{6^n \cdot n_1^{n_1} \cdots n_6^{n_6}} < c_\alpha\right) = P_{\theta_0}(-2\ln\lambda(x) > k_\alpha) \approx P(\chi^2_{(5)} > k_\alpha)$$

Pel que $P(\chi^2_{(5)} > k_\alpha) = \alpha$ i $P(\chi^2_{(5)} \leq k_\alpha) = 1 - \alpha$.

Exemple 4.2.9 (del control de qualitat)**.** En el model estadístic associat a l'exemple del control de qualitat fixem un valor del paràmetre $\theta_0 \in (0,1)$ i contrastem les hipòtesis $H_0 : \theta \leq \theta_0$ contra $H_1 : \theta > \theta_0$. Recordem la funció de versemblança $\mathscr{L}(x,\theta) = \theta^{n\bar{x}_n}(1-\theta)^{n(1-\bar{x}_n)}$ i l'estimador de màxima versemblança $\hat{\theta}(x) = \bar{x}_n$. Aleshores:

$$\sup_{\theta\in[0,1]} \mathscr{L}(x,\theta) = \mathscr{L}(x,\hat{\theta}(x)) = \bar{x}_n^{n\bar{x}_n}(1-\bar{x}_n)^{n(1-\bar{x}_n)}.$$

D'altra banda, com la funció $\mathscr{L}$ és creixent fins a $\bar{x}_n$ i després decreix, el suprem de $\mathscr{L}$ quan $\theta \in [0,\theta_0]$ depèn de si $\bar{x}_n$ és major o menor que θ_0, és a dir:

$$\hat{\theta}_0(x) = \begin{cases} \theta_0, & \text{si } \bar{x}_n \geq \theta_0; \\ \bar{x}_n, & \text{si } \bar{x}_n < \theta_0. \end{cases} \implies \sup_{\theta\leq\theta_0} \mathscr{L}(x,\theta) = \begin{cases} \theta_0^{n\bar{x}_n}(1-\theta_0)^{n(1-\bar{x}_n)}, & \text{si } \bar{x}_n \geq \theta_0; \\ \bar{x}_n^{n\bar{x}_n}(1-\bar{x}_n)^{n(1-\bar{x}_n)}, & \text{si } \bar{x}_n < \theta_0. \end{cases}$$

Per tant, la raó de versemblança val:

$$\lambda(x) = \begin{cases} \left(\frac{\theta_0}{\bar{x}_n}\right)^{n\bar{x}_n} \left(\frac{1-\theta_0}{1-\bar{x}_n}\right)^{n(1-\bar{x}_n)}, & \text{si } \bar{x}_n \geq \theta_0; \\ 1, & \text{si } \bar{x}_n < \theta_0. \end{cases}$$

Com aquesta funció $\lambda(x)$ respecte de $\bar{x}_n$ és constant igual a 1 fins a arribar a θ_0 i després decreix, tenim:

$$A_0 = \{x \mid \lambda(x) \geq c_\alpha\} = \{x \mid n\bar{x}_n \leq k_\alpha\} = \{x \mid S_n(x) \leq k_\alpha\},$$

on $S_n(x) = n\bar{x}_n = \sum_{i=1}^n x_i \sim B(n,\theta)$. Finalment s'obté $\sup_{\theta\leq\theta_0} P_\theta(A_0^c) = P_{\theta_0}(A_0^c)$.

Exemple 4.2.10 (del mesurament amb error). En aquest model, fixada $\mu_0 \in \mathbb{R}$, volem contrastar les hipòtesis:

$$H_0 : \mu = \mu_0 \quad \text{contra} \quad H_1 : \mu \neq \mu_0,$$

amb $\sigma > 0$ desconeguda. En aquest exemple tenim que

$$\Theta = \left\{\theta = \left(\mu, \sigma^2\right) \in \mathbb{R} \times (0, \infty)\right\},$$
$$\Theta_0 = \left\{\theta = \left(\mu_0, \sigma^2\right); \sigma^2 \in (0, \infty)\right\},$$

i

$$\mathscr{L}(x, \theta) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{ns_n^2(\mu)}{2\sigma^2}\right\},$$

on $s_n^2(\mu) = \frac{1}{n}\sum_{i=1}^n (x_i - \mu)^2$. L'estimador de màxima versemblança de $\theta = (\mu, \sigma^2)$ és $\hat{\theta}(x) = (\bar{x}_n, s_n^2)$ i

$$\sup_{\theta \in \Theta} \mathscr{L}(x, \theta) = \mathscr{L}(x, \hat{\theta}) = \frac{1}{(2\pi s_n^2)^{\frac{n}{2}}} e^{-\frac{n}{2}}.$$

El màxim de $\mathscr{L}$ en Θ_0 s'assoleix en $\widehat{\sigma^2} = s_n^2(\mu_0)$ i, per tant,

$$\sup_{\theta \in \Theta_0} \mathscr{L}(x, \theta) = L\left(x, \left(\mu_0, \widehat{\sigma^2}\right)\right) = \frac{1}{(2\pi s_n^2(\mu_0))^{\frac{n}{2}}} e^{-\frac{n}{2}}.$$

Com que $s_n^2(\mu_0) = s_n^2 + (\bar{x}_n - \mu_0)^2$, tenim que

$$\lambda(x) = \left(\frac{s_n^2(\mu_0)}{s_n^2}\right)^{-n/2} = \left(1 + \frac{T_n^2}{n-1}\right)^{-n/2},$$

amb $T_n = \sqrt{n-1}\frac{\bar{x}_n - \mu_0}{s_n}$, que sota la hipòtesi nul·la segueix una distribució $t_{(n-1)}$. Aleshores,

$$A_0 = \{x \mid \lambda(x) \geq c_\alpha\} = \{x \mid T_n(x)^2 \leq k_\alpha\} = \{x \mid |T_n(x)| \leq k'_\alpha\}$$

i, per tant, hem de trobar una constant k'_α tal que:

$$P_{\mu_0}(A_0^c) = P(|T_n(x)| > k'_\alpha) = \alpha,$$

amb $T_n(x) \sim t_{(n-1)}$.

4.3 Testos d'ajustament

Com podem comprovar si les nostres dades provenen d'una distribució o família de distribucions concreta? Contrastarem una distribució o família de distribucions fixada contra totes les altres. Per exemple, si suposem que les nostres observacions provenen d'una distribució contínua i volem saber si es tracta d'una distribució normal, plantejarem la prova d'hipòtesis següent:

$$H_0 : P \in \mathscr{P}_0 \text{ contra } H_1 : P \in \mathscr{P}_1 = \mathscr{P} \setminus \mathscr{P}_0,$$

on $\mathscr{P} = \{\text{distribucions contínues}\}$ i $\mathscr{P}_0 = \{N(\mu, \sigma^2) \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$.

4.3.1 Test de la χ^2

El test de la χ^2 d'ajustament es construeix sota la hipòtesi que les variables aleatòries que generen les dades tenen una llei **discreta**.

4.3.1.1 Llei multinomial

Considerem n realitzacions independents d'un experiment aleatori que té m resultats possibles excloents, això ens dona una partició de l'espai Ω amb m conjunts disjunts A_i, $\Omega = A_1 \sqcup \cdots \sqcup A_m$. La mostra $x_1, \ldots, x_n$ es pot resumir en el vector de les freqüències absolutes $f = (n_1, \ldots, n_m)$, on n_i és la freqüència d'A_i. Observem que $n_1 + \cdots + n_m = n$. Volem contrastar si la probabilitat de cadascun dels A_i és p_i^0 ($\sum_{i=1}^m p_i^0 = 1$), és a dir, plantegem les hipòtesis:

$$H_0 : p_i = p_i^0,\ \forall i = 1, \ldots, m \text{ contra } H_1 : \exists i \bigm| p_i \neq p_i^0.$$

De fet, aquest problema és paramètric ja que es tracta de decidir si els paràmetres de la distribució multinomial són $p_1 = p_1^0, \ldots, p_m = p_m^0$ (hipòtesi simple) o no (hipòtesi composta).

Definició 4.3.1 (Estadístic χ^2 de Pearson)**.** Per a construir la regió d'acceptació s'usa l'anomenat estadístic khi-quadrat (χ^2, χ-quadrat) de Pearson:

$$D_n \equiv D_n(x) = \sum_{j=1}^m \frac{(n_j - np_j^0)^2}{np_j^0} = \sum_{j=1}^m \frac{n}{p_j^0}\left(\frac{n_j}{n} - p_j^0\right)^2.$$

Obviarem a D_n la dependència de la nostra per a simplificar la notació. Acceptarem la hipòtesi nul·la si D_n és petit amb regió d'acceptació $A_0 = \{D_n \leq d_\alpha\}$. *Com es determina d_α?*

Teorema 4.3.2. *Sigui $(x_n)_n$ una successió de vectors aleatoris independents i idènticament distribuïts amb llei de Bernoulli m-dimensional de paràmetre $\theta = (\theta_1, \ldots, \theta_m)$ amb $\theta_j \neq 0$, per a tot j. Sigui $f = (n_1, \ldots, n_m) = \sum_{i=1}^n x_i$ el vector de les freqüències absolutes. Aleshores:*

$$D_n = \sum_{j=1}^m \frac{n}{p_j^0}\left(\frac{n_j}{n} - p_j^0\right)^2 \xrightarrow[n\to\infty]{\mathscr{L}} \chi^2_{(m-1)}.$$

Per tant:

$$P(D_n(x) \leq d_\alpha) \xrightarrow[n\to\infty]{\mathcal{L}} P(\chi^2_{(m-1)} \leq d_\alpha)$$

i així, sota H_0:

$$P_{H_0}(A_0^c) = P(\chi^2_{(m-1)} > d_\alpha) = \alpha.$$

Observació 4.3.3. En general, es considera que l'aproximació és vàlida quan $np_j^0 \geq 5$, i $n_j \geq 5$, per a tot $j = 1, \ldots, m$.

Considerem el model estadístic associat a multinomial $(\Omega, \mathscr{F}, \mathscr{P} = \{P_\theta \mid \theta \in \Theta\})$ amb $\Theta = \{\theta = (p_1, \ldots, p_m) \in [0,1]^m \mid \sum_{j=1}^m p_j = 1\}$. Si $(n_1, \ldots, n_m)$ és el vector de freqüències absolutes, aleshores la funció de versemblança es pot escriure com $\mathscr{L}(x, \theta) = \prod_{j=1}^m p_j^{n_j}$.

Calculant el màxim de versemblança sota la condició $\sum_{j=1}^m p_i = 1$ (no ho farem), obtenim que $\hat{\theta} = (\frac{n_1}{n}, \ldots, \frac{n_m}{n})$. Considerem $\theta_0 = (p_0^0, \ldots, p_m^0) \in \Theta$ i plantegem $H_0 : \theta = \theta_0$ contra $H_1 : \theta \neq \theta_0$. Aleshores, la raó de versemblança és:

$$\lambda_n = \frac{\mathscr{L}(x, \theta_0)}{\sup_{\theta \in \Theta} \mathscr{L}(x, \theta)} = \frac{\mathscr{L}(x, \theta_0)}{\mathscr{L}(x, \hat{\theta})} = \frac{\prod_{j=1}^m (p_j^0)^{n_j}}{\prod_{j=1}^m (\frac{n_j}{n})^{n_j}}.$$

Utilitzant el comportament asimptòtic de λ_n, la regió d'acceptació del test és donada per $A_0 = \{-2 \ln \lambda_n \leq c_\alpha\}$, on c_α és el quantil $1 - \alpha$ de la $\chi^2_{(m-1)}$. Bàsicament ens hem de quedar amb què aquesta regió **és similar a la del test de la** χ^2, ja que $-2 \ln \lambda_n$ és un valor pròxim a D_n.

Observació 4.3.4. Pel teorema del límit central,

$$\frac{n_j - np_j^0}{\sqrt{np_j^0(1 - p_j^0)}} \xrightarrow[n \to \infty]{\mathscr{L}} N(0, 1).$$

4.3.1.2 Llei multinomial parametritzada

Plantegem ara el mateix però suposant que les probabilitats de la llei multinomial depenen d'un paràmetre $\theta \in \Theta \subset \mathbb{R}^d$, és a dir, $\mathscr{P}_0$ està formada per les lleis multinomials $M(m, n, p(\theta))$ amb $\{p(\theta) = (p_1(\theta), \ldots, p_m(\theta)), \forall \theta \in \Theta\}$. Ara, la prova d'hipòtesi queda:

$$H_0 : (p_1, \ldots, p_m) \in \{(p_1(\theta), \ldots, p_m(\theta) \mid \theta \in \Theta\}$$

contra

$$H_1 : (p_1, \ldots, p_m) \notin \{(p_1(\theta), \ldots, p_m(\theta) \mid \theta \in \Theta\}.$$

Igual que abans, $f = (n_1, \ldots, n_m)$ és el vector de les freqüències absolutes. Amb aquesta notació la funció de versemblança es pot escriure com

$$L(x, \theta) = \prod_{j=1}^m p_j(\theta)^{n_j}$$

Definició 4.3.5 (Estadístic χ^2 de Pearson, multinomial)**.** Sigui $\hat{\theta}$ el màxim de versemblança del paràmetre θ. Definim l'estadístic de la khi-quadrat de Pearson per a aquest model com

$$D_n = \sum_{j=1}^m \frac{\left(n_j - np_j(\hat{\theta})\right)^2}{np_j(\hat{\theta})} \xrightarrow[n \to \infty]{\mathcal{L}} \chi^2_{(m-d-1)} \tag{4.1}$$

Com que valors petits de l'estadístic D_n indiquen proximitat a la hipòtesi nul·la, aquest resultat asimptòtic permet construir per a aquest test una regió d'acceptació a nivell α de la forma $A_0 = \{D_n \leq d_\alpha\}$.

Exemple 4.3.6 (χ^2 en els tests d'ajustament).

1. Suposem que volem contrastar si la nostra mostra prové d'una distribució discreta amb un nombre finit de valors possibles $\{a_j \mid j = 1, \ldots, m\}$. Posant $p_j = P(a_j)$, estem en el model multinomial i utilitzem la regió d'acceptació $A_0 = \{D_n \leq d_\alpha\}$ on d_α és el quantil $1 - \alpha$ d'una llei $\chi^2_{(m-1)}$.
2. Si volem utilitzar el test de la χ^2 quan la mostra prové d'una distribució contínua o d'una distribució discreta amb un nombre numerable de valors possibles, haurem d'agrupar les dades en intervals disjunts $I_1, \ldots, I_m$, $\bigsqcup I_j = \mathbb{R}$. Aleshores, considerant $p_j = P(I_j)$, podem situar-nos en el model multinomial i obtenir la mateixa regió d'acceptació que en el cas anterior.
3. Un dels casos més comuns és contrastar si la mostra prové d'una família de distribucions paramètrica $\{P_\theta \mid \theta \in \Theta\}$. Per això, utilitzarem els resultats obtinguts per la llei multinomial parametritzada. Es tracta d'agrupar les dades en intervals disjunts que recobreixin tots els valors possibles de la distribució, $I_1, \ldots, I_m$, $\bigsqcup I_j = \mathbb{R}$. **Els intervals s'han de triar independentment de la mostra**. Sota la hipòtesi nul·la, les probabilitats de cadascun d'aquests intervals dependran del paràmetre θ de la família:

$$p_1(\theta) = P_\theta(I_1), \ldots, p_m(\theta) = P_\theta(I_m).$$

   Per usar (4.1), cal *estimar* per al màxim de versemblança el paràmetre θ; però aquesta estimació en general no és senzilla.

Com que en general és molt més fàcil, o com a mínim coneguda, l'estimació màxim versemblant dels paràmetres utilitzant la mostra completa, val la pena tenir en compte el resultat següent:

Teorema 4.3.7. *Siguin $\hat{\theta}_1, \ldots, \hat{\theta}_d$ els estimadors de màxima versemblança utilitzant la mostra completa. Aleshores,* (4.1) *segueix asimptòticament una llei amb funció de distribució F tal que:*

$$F_{\chi^2_{(m-1)}}(x) \leq F(x) \leq F_{\chi^2_{(m-d-1)}}(x).$$

A partir d'aquí, es pot establir la regla de decisió següent:

1. *Calculem els valors crítics c_α i C_α utilitzant les lleis $\chi^2_{(m-d-1)}$ i $\chi^2_{(m-1)}$, respectivament.*
2. *Si $D_n > C_\alpha$, rebutgem la hipòtesi nul·la.*
3. *Si $D_n < c_\alpha$, acceptem la hipòtesi nul·la.*
4. *Si $c_\alpha < D_n < C_\alpha$ no podem prendre cap decisió.*

4.3.2 Test de Kolmogorov

Aquest test d'ajustament es basa en la funció de distribució empírica i les seves propietats asimptòtiques. Es tracta de valorar si les diferències entre les funcions de distribució *empírica* i *teòrica* són significatives.

Considerem el model estadístic $(\Omega, \mathscr{A}, \mathscr{P})$, on $\mathscr{P}$ és la família de totes les lleis contínues. Sigui una mostra aleatòria simple $x_1, \ldots, x_n$ d'una llei $P \in \mathscr{P}$ amb funció de distribució F (que suposarem contínua). Al capítol següent veurem com es defineix una distribució empírica. Per ara, necessitem aquest concepte així que l'introduïm breument:

Definició (Funció de distribució empírica). Es defineix:

$$F_n(y) = \sum_{i=1}^{n} \mathbb{1}_{(-\infty,y]}(x_i), \ \forall y \in \mathbb{R}.$$

Suposem que volem contrastar si les nostres dades provenen d'una probabilitat $P_0 \in \mathscr{P}$ amb funció de distribució F_0. Plantegem el contrast d'hipòtesi següent:

$$H_0 : P = P_0 \text{ contra } H_1 : P \neq P_0 \iff H_0 : F = F_0 \text{ contra } H_1 : F \neq F_0$$

Definició 4.3.8 (Estadístic de Kolmogorov). L'estadístic de Kolmogorov mesura la diferència màxima entre les dues funcions de distribució, F_n i F_0:

$$\Delta_n = \sup_{y\in\mathbb{R}} |F_0(y) - F_n(y)|.$$

Sota la hipòtesi nul·la, el teorema de Glivenko-Cantelli afirma que $\Delta_n \xrightarrow[n\to\infty]{\text{QS}} 0$.

Utilitzant aquest resultat asimptòtic, és natural proposar com a regió d'acceptació:

$$A_0 = \{\Delta_n \leq c_{\alpha,n}\}, \ P_0(A_0^c) = \alpha,$$

però per a poder determinar $c_{\alpha,n}$ ens cal saber la distribució de Δ_n sota la hipòtesi nul·la.

Teorema 4.3.9. *La distribució de Δ_n sota la hipòtesi nul·la és la mateixa per a tota funció de distribució contínua F_0, és a dir, no depèn de P_0.*

Demostració. Sigui F la funció de distribució de les variables X_i, $i = 1, \ldots, n$. Considerem l'estadístic d'ordre $(X_{(1)}, \ldots, X_{(n)})$. Com que la funció de distribució F és una funció contínua i creixent, mentre que F_n és una funció esglaonada, les diferències més grans es troben en els salts o just abans i, per tant:

$$\begin{aligned}\Delta_n = \sup_{y\in\mathbb{R}} |F_0(y) - F_n(y)| &= \sup_{1\leq i\leq n} \{|F(X_{(i)}) - F_n(X_{(i)})|, |F(X_{(i)}^-) - F_n(X_{(i)}^-)|\} \\ &= \sup_{1\leq i\leq n} \left\{\left|F(X_{(i)}) - \frac{i}{n}\right|, \left|F(X_{(i)}^-) - \frac{i-1}{n}\right|\right\}.\end{aligned}$$

És clar, doncs, que la llei de Δ_n depèn només del vector $(F(X_{(1)}), \dots, F(X_{(n)}))$. Com les variables $X_1, \dots, X_n$ són VAIID amb funció de distribució contínues, les variables aleatòries definides per $Y_i = F(X_i)$ també ho són i $Y_i \sim U(0,1)$, per a tot i. És a dir, $(F(X_1), \dots, F(X_n)) \sim U(0,1)^n$ i $U(0,1)^n$ és independent de F.

$$(F(X_{(1)}), \dots, F(X_{(n)})) \sim \frac{1}{n!}\mathbb{1}_D(x),\ X = (X_1, \dots, X_n),\ D = \{(X_1, \dots, X_n) \in \mathbb{R}^n \mid X_1 \leq \dots \leq X_n\}.$$

On aquest últim càlcul correspon al càlcul explícit de dita llei. ■

Observació 4.3.10. La funció de distribució de Δ_n està tabulada per a valors petits d'n i per a valors grans existeix un comportament asimptòtic:

$$P(\Delta_n > c_{\alpha,n}) = \alpha \iff c_{\alpha,n} = \sqrt{\frac{-\ln(\frac{\alpha}{2})}{2n}}.$$

4.4 Tests d'homogeneïtat

Suposem que tenim dues mostres independents $(x_1, x_2, \dots, x_{n_1})$ i $(y_1, \dots, y_{n_2})$ que provenen de dues llei de probabilitat P_1, P_2, respectivament. Ens interessa contrastar si aquestes probabilitats són iguals, és a dir, si tenen la mateixa distribució.

4.4.1 Test de la χ^2

Per al test de la χ^2 suposarem o bé que les variables prenen un nombre finit de valors o bé que tenim les dades agrupades. Sigui $A_1, \dots, A_m$ una partició de Ω i sigui $(p_1^1, \dots, p_m^1)$ i $(p_1^2, \dots, p_m^2)$ les probabilitats dels A_i respecte de P_1, P_2. Si les dades són agrupades, els A_i corresponen als intervals; si són discretes, els A_i corresponen als possibles valors d'aquestes variables. Podem resumir-ho en la *taula de contingència*:

Classe	A_1	A_2	$\cdots$	A_m	Total
Mostra 1	n_{11}	n_{12}	$\cdots$	n_{1m}	n_1
Mostra 2	n_{21}	n_{22}	$\cdots$	n_{2m}	n_2
Total	$n_{\cdot,1}$	$n_{\cdot,2}$	$\cdots$	$n_{\cdot,m}$	$n = n_1 + n_2$

La hipòtesi nul·la assumeix que les dades provenen de la mateixa distribució. Sota aquesta hipòtesi, ajuntant les dues mostres, tenim una mostra de mida n d'una multinomial i, per tant, l'estimador de màxima versemblança de les probabilitats $p_j = P_1(A_j) = P_2(A_j)$ és:

$$\hat{p}_j = \frac{n_{1j} + n_{2j}}{n_1 + n_2} = \frac{n_{\cdot,j}}{n},\ \forall j = 1, \dots, m.$$

Podem usar ara l'equivalent a l'estadístic de la χ^2 de Pearson:

$$D_{n_1,n_2} = \sum_{j=1}^{m} \frac{(n_{1j} - n_1\hat{p}_j)^2}{n_1\hat{p}_j} + \sum_{j=1}^{m} \frac{(n_{2j} - n_2\hat{p}_j)^2}{n_2\hat{p}_j}.$$

Es pot demostrar asimptòticament que D_{n_1,n_2} segueix una llei $\chi^2_{(m-1)}$. Utilitzant aquest resultat asimptòtic és natural proposar la regió d'acceptació $A_0 = \{D_{n_1,n_2} \leq c_{\alpha,m}\}$ on $c_{\alpha,m}$ és el quantil $1-\alpha$ d'una llei $\chi^2_{(m-1)}$.

Observació 4.4.1. Als exercicis usarem la següent expressió:

$$D_n(x,y) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\left(n_{ij} - \frac{n_{i,\cdot}n_{\cdot,j}}{n}\right)^2}{\frac{n_{i,\cdot}n_{\cdot,j}}{n}}.$$

4.4.2 Test de Kolmogorov-Smirnov

En el cas que tinguem dues mostres de lleis amb distribucions contínues podem utilitzar les funcions de distribució empíriques per contrastar si tenen o no la mateixa distribució.

Siguin dues mostres independents $(x_1, x_2, \ldots, x_{n_1})$ i $(y_1, y_2, \ldots, y_{n_2})$, que provenen de dues lleis de probabilitat P_1 i P_2 amb funcions de distribució contínues F_1 i F_2, respectivament. Volem contrastar:

$$H_0 : P_1 = P_2 \,(\iff F_1 = F_2) \text{ contra } H_1 : P_1 \neq P_2 \,(\iff F_1 \neq F_2)\,.$$

Considerem les dues funcions de distribució empíriques

$$F_n^1(y) = \sum_{i=1}^{n_1} \mathbb{1}_{(-\infty,y]}(x_i)\,,$$
$$F_n^2(y) = \sum_{i=1}^{n_2} \mathbb{1}_{(-\infty,y]}(y_i)\,.$$

L'estadístic de Kolmogorov-Smirnov es defineix com la diferència màxima entre les dues funcions de distribució empíriques

$$\Delta_{n_1,n_2} = \sup_{y\in\mathbb{R}} \left|F_n^1(y) - F_n^2(y)\right|$$

La llei de Δ_{n_1,n_2} no depèn ni de P_1 ni de P_2, sinó només de n_1 i n_2[3]. La regió d'acceptació d'aquest test és de la forma

$$A_0 = \{\Delta_{n_1,n_2} \leq c_{\alpha,n_1,n_2}\}\,, \quad \text{amb}\, P\left(A_0^c\right) = \alpha$$

on P és la llei de Δ_{n_1,n_2}. Per a valors petits de n_1 i n_2 la llei de l'estadístic de Kolmogorov-Smirnov està tabulada i per a valors grans tenim resultats asimptòtics que ens permeten trobar la constant c_{α,n_1,n_2}.

[3] 4.3.9 ens diu que la distribució de Δ_n sota la hipòtesi nul·la és la mateixa per a tota funció de distribució contínua F_0, és a dir, no depèn de P_0.

V

Teorema fonamental de l'estadística

5.1 Distribució empírica

Considerem n variables aleatòries $X_1, \dots, X_n$ definides en $(\Omega, \mathscr{A}, P)$ i prenent valors en $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$.

Definició 5.1.1 (Probabilitat empírica). Donat $w \in \Omega$, definim la probabilitat empírica associada a les variables $X_1, \dots, X_n$ com la llei de probabilitat discreta en $\mathbb{R}$ que assigna una probabilitat $\frac{1}{n}$ a cadascun dels punts $X_i(w)$, $i \in \{1, \dots, n\}$, és a dir, la probaiblitat empírica sobre $B \in \mathscr{B}(\mathbb{R})$ val:

$$p_n^w(B) = \frac{1}{n} \#\{i \in \{1, \dots, n\} \mid X_i(w) \in B\} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_B(X_i(w)).$$

La dependència del valor w s'ometrà a partir d'ara per no sobrecarregar de notacions. Segons el context s'entendrà clarament si estem parlant d'una probabilitat empírica o no.

Definició 5.1.2 (n-èsima probabilitat empírica). Per a una successió $(X_k)_k$ de VAIID amb llei p definides en $(\Omega, \mathscr{A}, \mathcal{P})$, podem considerar la n-èsima probabilitat empírica p_n associada al vector $(X_1, \dots, X_n)$.

Definició 5.1.3 (n-èsima funció de distribució empírica). La n-èsima funció de distribució empírica associada a la successió $(X_k)_k$ és la funció de distribució de la n-èsima probabilitat empírica associada a $(X_k)_k$ i això vol dir, per a tot $t \in \mathbb{R}$:

$$F_n(t) = p_n((-\infty, t]) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(X_i) = \begin{cases} 0, & \text{si } t < X_{(1)}, \\ \frac{i}{n}, & \text{si } t \in [X_{(i)}, X_{(i+1)}),\ 1 \leq i \leq n; \\ 1, & \text{si } t \geq X_{(n)}. \end{cases}$$

Notació 5.1.4. En aquestes definicions, Ω seria l'equivalent a $\tilde{\Omega}$ i $\mathbb{R} = \Omega$; per tant, $X_i : \tilde{\Omega} \longrightarrow \Omega = \mathbb{R}$. D'aquesta forma, en la definició no s'ha pres Ω per referenciar l'espai d'arribada de la variable aleatòria com s'ha fet durant el curs.

5.2

Teorema de Glivenko-Cantelli

Fixat un borelià B, la variable aleatòria $np_n(B)$ compta el nombre d'observacions en B. En efecte:

$$np_n(B) = \#\{i \in \{1, \dots, n\} \mid X_i(w) \in B\} = \sum_{i=1}^{n} \mathbb{1}_B(X_i(w)).$$

Com que $\mathbb{1}_B(X_i(w))$ segueix una Bernoulli de paràmetre $p(B)$ i les variables aleatòries són VAIID tenim que $np_n(B) \sim B(n, p(B))$. Per tant, la llei forta dels grans nombres implica que:

$$p_n^w(B) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_B(X_i(w)) \xrightarrow[n\to\infty]{\text{QS}} \mathbb{E}(\mathbb{1}_B(X_i)) = p_{X_i}(B).$$

És a dir, $p_n^w(B)$ convergeix quasi segurament cap a $p(B)$.

$$\exists N_B \in \mathscr{C} \mid P(N_B) = 0,\ P_n^w(B) \xrightarrow{n\to\infty} P(B),\ \forall w \notin N_B.$$

Una qüestió interessant i a la vegada molt important és determinar algun tipus de classe $\mathscr{C}$ de borelians que compleixi la propietat següent:

$$\sup_{B\in\mathscr{C}} |p_n(B) - p(B)| \xrightarrow[\text{QS}]{n\to\infty} 0.$$

Aquesta pregunta no té una resposta immediata.

Teorema 5.2.1 (Teorema de Glivenko-Cantelli). *Sigui $X = (X_k)_k$ una successió de* VAIID *en $(\Omega, \mathscr{A}, P)$ amb funció de distribució F. Aleshores:*

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \xrightarrow[QS]{n\to\infty} 0,$$

on F_n és la n-èsima funció de distribució empírica de $(X_k)_k$.

Definició 5.2.2. És una funció de distribució que es defineix com:

$$\begin{aligned} G : \ [0,1] &\longrightarrow \bar{\mathbb{R}} \\ u &\longmapsto \inf\{x \mid F(x) \geq u\}, \end{aligned}$$

amb els convenis que $G(0) = -\infty$. Sabem que $F_Y(y) = P(F(x) \leq y) = P(X \leq F^{-1}(y))$. Pel que si $\{x \mid F(x) \geq 1\} = \emptyset$, aleshores $G(1) = +\infty$. Aquesta funció compleix que $F(G(u)^-) \leq u \leq F(G(u))$, per a tot $u \in (0,1]$. En efecte, d'una banda:

$$F(G(u) - \varepsilon) = F(\inf\{x \mid F(x) \geq u\} - \varepsilon) < u \implies F(G(u)^-) \leq u.$$

D'altra banda, la segona és conseqüència immediata de la continuïtat per la dreta de la funció de distribució, que fa que l'ínfim pertanyi al conjunt.

Demostració, no entra. Fixat $t \in \mathbb{R}$, la llei forta dels grans nombres ens permet dir que existeix $W_t \in \mathscr{A}$ amb $P(\bigcup_{t\in\mathbb{R}} W_t) = 0$ tal que $\forall w \notin W_t$ tenim que:

$$F_n(t) \xrightarrow[n\to\infty]{\text{QS}} F(t),\ F_n(t^-) \xrightarrow[n\to\infty]{\text{QS}} F(t^-).$$

On $F_n(t^-)$ és la freqüència relativa a l'esdeveniment $\{X < t\}$. Com no podem utilitzar un argument per a tot $t \in \mathbb{R}$, perquè la unió no numerable de conjunt de probabilitat zero no ha de tenir necessàriament probabilitat zero, la demostració d'aquest teorema consistirà a obtenir una fita de $|F_n(t) - F(t)|$ sobre una col·lecció numerable d'intervals. Per a tot $m \geq 1$ i $k = 1, \ldots, m$, podem definir $y_{m,k} = G(\frac{k}{m})$ que, com veurem, tenen les propietats següents:

Propietat 5.2.3.

1. $F(y_{m,k}^-) - F(y_{m,k-1}) \leq \frac{1}{m}$, $k = 2, \ldots, m$.
2. $F(y_{m,1}^-) \leq \frac{1}{m}$.
3. $F(y_{m,m}) \geq 1 - \frac{1}{m}$.

Aplicant que $F(G(u)^-) \leq u \leq F(G(u))$, obtenim:

$$F(y_{m,k}^-) = F\left(G\left(\frac{k}{m}^-\right)^-\right) \leq \frac{k}{m},\ F(y_{m,k-1}) = F\left(G\left(\frac{k-1}{m}^-\right)\right) \leq \frac{k-1}{m},$$

que implica immediatament *1.*. Podem deduir *2.* i *3.* fent servir els mateixos arguments:

$$F(y_{m,1}^-) = F\left(G\left(\frac{1}{m}^-\right)\right) \leq \frac{1}{m},\ F(y_{m,m}) = F(G(1)) \geq 1 \geq 1 - \frac{1}{m}.$$

Aleshores, definint $A_t = \tilde{A}_t \cup \tilde{\tilde{A}}_t$, per a tot $t \in \mathbb{R}$ tenim que $F_n(t) \to F(t)$ per a tot $w \in \tilde{A}_t$ (i $P(\tilde{A}_t) = 0$). Anàlogament, $F_n(t^-) \to F(t^-)$ per a tot $w \in \tilde{\tilde{A}}_t$ (i $P(\tilde{\tilde{A}}_t) = 0$). Definint per a $n, m \geq 1$:

$$\Delta_{n,m}^w = \sup_{k\in\{1,\ldots,m\}} \left\{|F_n(y_{m,k}) - F(y_{m,k})|, |F_n(y_{m,k}^-) - F(y_{m,k}^-)|\right\}.$$

Per tant,

$$\lim_{n\to\infty} \Delta_{n,m}^w,\ \forall m \geq 1,\ w \in A_t.$$

Fixem ara un valor de $m \geq 1$. Observem que els conjunts $(-\infty, y_{m,1}), [y_{m,1}, y_{m,2}) \cup [y_{m,2}, y_{m,3}) \cup \cdots \cup [y_{m,m-1}, y_{m,m}), [y_{m,m}, +\infty)$ formen una partició de $\mathbb{R}$. Sigui $t \in \mathbb{R}$. Si $t \in [y_{m,k-1}, y_{m,k})$ per a algun $k = 2, \cdots, m$, utilitzant el creixement de les funcions F i F_n, la definició de Δ i la propietat 1), tenim que

$$\begin{aligned}
F_n(t) &\leq F_n\left(y_{m,k}^-\right) \leq F\left(y_{m,k}^-\right) + \Delta_{n,m}^w \leq F\left(y_{m,k-1}\right) + \frac{1}{m} + \Delta_{n,m}^w \\
&\leq F(t) + \frac{1}{m} + \Delta_{n,m}^w, \\
F_n(t) &\geq F_n\left(y_{m,k-1}\right) \geq F\left(y_{m,k-1}\right) - \Delta_{n,m}^w \geq F\left(y_{m,k}^-\right) - \frac{1}{m} - \Delta_{n,m}^w \\
&\geq F(t) - \frac{1}{m} - \Delta_{n,m}^w
\end{aligned}$$

d'on

$$|F_n(t) - F(t)| \leq \frac{1}{m} + \Xi^w_{n,m}$$

Utilitzant la segona i tercera propietat de manera anàloga podem tractar els casos $t \in (-\infty, y_{m,1})$ i $t \in (y_{m,m}, +\infty)$ respectivament, i obtindrem la mateixa fita. Conseqüentment,

$$\sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \leq \frac{1}{m} + \Delta^w_{n,m} \implies \lim_{n\to\infty} \sup_{t\in\mathbb{R}} |F_n(t) - F(t)| \leq \frac{1}{m}.$$

Com que el valor de m és arbitrari, si el fem tendir a infinit acabarem la prova d'aquest teorema.

■

VI

Model de regressió lineal simple

Una de les relacions deterministes més simples entre dues variables (x, y) és mitjançant una línia recta. Volem desenvolupar un model en què el lligam entre les variables s'expressarà amb una relació lineal més un terme aleatori. Aquest model s'anomena *model de regressió lineal simple.*

6.1 Model de regressió

Definició 6.1.1 (Model de regressió lineal). Dues variables (x, y) segueixen un model de regressió lineal simple si existeixen dues constants β_0 i β_1 i una variable aleatòria ε, centrada i amb variància finita σ^2, tal que per a cada valor fixat de x, el valor associat de y compleix

$$y = \beta_0 + \beta_1 x + \varepsilon.$$

Definició 6.1.2. La variable x s'anomena variable *independent*, predictora o explicativa, no és aleatòria i els seus valors són usualment fixats per l'experimentador. Donada la variable x, la variable y és aleatòria i s'anomena variable dependent o de resposta. A la variable ε se l'anomena error.

Una mostra de mida n estarà formada per n parells $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, on $x_1, \ldots, x_n$ són els diferents valors de la variable independent x-que es poden repetir- i $y_1, \ldots, y_n$ són els corresponents valors obtinguts de la variable dependent y. Abans d'ajustar la mostra a un model de regressió és convenient representar els parells $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ en una gràfica per observar si la tendència lineal pot ser adequada.

6.2 Estimació dels coeficients

Suposem que la nostra mostra prové d'un model lineal de regressió simple, és a dir, $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, $1 \leq i \leq n$, on $\varepsilon_1, \ldots, \varepsilon_n$ són VAIID centrades i amb variància finita σ^2. Els paràmetres del model β_0, β_1 se'ls anomena coeficients de regressió.

Utilitzant les propietats d'ε obtenim que els variables $y_1, \ldots, y_n$ compleixen, per a qualssevol $i, j \in \{1, \ldots, n\}$,

$$\begin{aligned} \mathbb{E}(y_i) &= \beta_0 + \beta_1 x_i + \mathbb{E}(\varepsilon_i) = \beta_0 + \beta_1 x_i, \\ \mathrm{Var}(y_i) &= \mathrm{Var}(\beta_0 + \beta_1 x_i + \varepsilon_i) = \mathrm{Var}(\varepsilon_i) = \sigma^2, \\ \mathrm{Cov}(y_i, y_j) &= \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \ i \neq j. \end{aligned}$$

Com no establim cap distribució per als errors, no podem usar el mètode de màxima versemblança per fer les estimacions dels paràmetres. El mètode que es fa servir és l'anomenat *mètode dels mínims quadrats.*

6.3 Mínims quadrats i els seus coeficients

Aquest mètode consisteix a trobar $\hat{\beta}_0$ i $\hat{\beta}_1$ tals que la recta $\hat{\beta}_0 + \hat{\beta}_1 x$ faci mínimes les distàncies verticals entre els punts observats i la recta estimada.

Definició 6.3.1. Per a cada punt observat (x_i, y_i) considerem el punt de la recta (x_i, y_i') tal que $y_i' = \hat{\beta}_0 + \hat{\beta}_1 x_i$, es tracta de minimitzar la distància entre aquests punts[1]:

$$\sum_{i=1}^{n} (y_i - y_i')^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \tag{6.1}$$

1. Mitjana empírica: $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$.
2. Covariància empírica: $S_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$.
3. Coeficient de correlació empíric: $S_{xx} = \sum_i (x_i - \bar{x})^2$, $S_{yy} = \sum_i (y_i - \bar{y})^2$.

Teorema 6.3.2. *Els valors dels coeficients de regressió que minimitzen l'expressió* (6.1) *són* $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ *i* $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

Sigui per a tot $i = 1, \ldots, n$, $\hat{y}_i$ el valor estimat de la variable predictora y_i per la recta de regressió, és a dir, $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Definim també per a tot $i = 1, \ldots, n$, l'estimació de l'error ε_i com $\hat{\varepsilon}_i = y_i - \hat{y}_i$. Tenim les propietats següents:

1. La suma de les estimacions dels errors és zero, és a dir

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} (y_i - \hat{y}_i) = 0$$

[1] De fet, definim $g(\beta_0, \beta_1) = \sum_i |y_i - \hat{y}_i|$ i «serviria», però com no és derivable ho fem amb el quadrat.

Demostració. En efecte, per 6.3.2 (en particular per les equacions normals) tenim $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$, i utilitzant que $\sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n (y_i - \bar{y}) = 0$, resulta

$$\begin{aligned}\sum_{i=1}^n (y_i - \hat{y}_i) &= \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \sum_{i=1}^n \left(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i\right) \\ &= \sum_{i=1}^n (y_i - \bar{y}) + \hat{\beta}_1 \sum_{i=1}^n (x_i - \bar{x}) = 0.\end{aligned}$$

I ha quedat demostrat. ■

2. La suma de les variables predictores i de les seves estimacions coincideixen

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{y}_i$$

Observem que això implica que tant les variables predictores com les seves estimacions tenen la mateixa mitjana.

3. La recta de regressió passa pel punt $(\bar{x}, \bar{y})$, ja que $\hat{\beta}_0 + \hat{\beta}_1\bar{x} = \bar{y} - \hat{\beta}_1\bar{x} + \hat{\beta}_1\bar{x} = \bar{y}$.

4. Es tenen les igualtats següents:

$$\sum_{i=1}^n x_i\hat{\varepsilon}_i = 0 \quad \text{i} \quad \sum_{i=1}^n \hat{y}_i\hat{\varepsilon}_i = 0$$

Demostració. Vegem la primera igualtat. Recordem que $\bar{x}\sum_i (y_i - \hat{y}_i) = 0$. Així,

$$\begin{aligned}\sum_{i=1}^n x_i\hat{\varepsilon}_i &= \sum_{i=1}^n x_i(y_i - \hat{y}_i) = \sum_i (x_i - \bar{x})(y_i - \hat{y}_i) = \sum_i (x_i - \bar{x})(y_i - \bar{y}) + \sum_i (x_i - \bar{x})(\bar{y} - \hat{y}_i) \\ &= \sum_i (x_i - \bar{x})(y_i - \bar{y}) - \hat{\beta}_1 \sum_i (x_i - \bar{x})^2 = S_{xy} - \frac{S_{xy}}{S_{xx}} S_{xx} = 0.\end{aligned}$$

Per a l'altra igualtat només cal usar la definició de $\hat{y}_i$, la primera propietat i el que acabem de demostrar,

$$\sum_i \hat{y}_i\hat{\varepsilon}_i = \hat{\beta}_0 \sum_{i=1}^n \hat{\varepsilon}_i + \beta_0 \sum_{i=1}^n x_i\hat{\varepsilon}_i = 0.$$ ■

5. El valor mínim de $\sum_i (y_i - \beta_0 - \beta_1 x_i)^2$ és $S_{xy}(1 - R_{xy}^2)$.

Demostració. Efectivament,

$$\begin{aligned}\sum_{i=1}^n (y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i)^2 &= \sum_{i=1}^n (y_i - \bar{y} - \hat{\beta}_1(x_i - \bar{x}))^2 = S_{yy} + \hat{\beta}_1^2 S_{xx} - 2\hat{\beta}_1 S_{xy} \\ &= S_{yy} + \frac{S_{xy}^2}{S_{xx}} - 2\frac{S_{xy}^2}{S_{xx}} = S_{yy}(1 - R_{xy}^2).\end{aligned}$$

Per tant, com més pròxim sigui R_{xy}^2 d'1, més petita serà la suma de les distàncies verticals entre els punts i la recta de regressió. ■

Definició 6.3.3. El valor R_{xy}^2 se l'anomena *coeficient de determinació* i s'interpreta com el quocient entre la variabilitat explicada per la recta de regressió i la variabilitat de les dades:

$$R_{xy}^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Tenint en compte que les variables $x_1, \ldots, x_n$ no són aleatòries, els estimadors $\hat{\beta}_0$ i $\hat{\beta}_1$ es poden expressar com a combinació lineal de les variables dependents $y_1, \ldots, y_n$. Fixem-nos-hi,

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})\, y_i - \frac{1}{S_{xx}} \bar{y} \sum_{i=1}^n (x_i - \bar{x}) = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i,$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \sum_{i=1}^n \frac{y_i}{n} - \bar{x} \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} y_i = \sum_{i=1}^n \left(\frac{S_{xx} - n\bar{x}\,(x_i - \bar{x})}{nS_{xx}} \right) y_i = \sum_{i=1}^n \left(\frac{\sum_{j=1}^n x_j^2 - n\bar{x}x_i}{nS_{xx}} \right) y_i.$$

A partir de les expressions anteriors i de les propietats de les variables $y_1, \ldots, y_n$ descrites en 6.3.1, deduïm les propietats següents:

1. Els estimadors $\hat{\beta}_0$ i $\hat{\beta}_1$ són no esbiaixats.

   Demostració. En efecte,

   $$\mathbb{E}\left[\hat{\beta}_1\right] = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} \mathbb{E}\left[y_i\right] = \sum_{i=1}^n \frac{x_i - \bar{x}}{S_{xx}} \left(\beta_0 + \beta_1 x_i\right) = \beta_1 \frac{\sum_{i=1}^n x_i^2 - n\hat{x}^2}{S_{xx}} = \beta_1,$$
   $$\mathbb{E}\left[\hat{\beta}_0\right] = \mathbf{E}[y] - \bar{x}\mathbb{E}\left[\hat{\beta}_1\right] = \beta_0 + \beta_1\bar{x} - \bar{x}\beta_1 = \beta_0.$$

   Com que $\mathbb{E}\left[\hat{\beta}_1\right] = \beta_1$ i $\mathbb{E}\left[\hat{\beta}_0\right] = \beta_0$, ja hem acabat. ∎

2. Les variàncies dels estimadors $\hat{\beta}_0$ i $\hat{\beta}_1$ són:

   $$\operatorname{Var}\left(\hat{\beta}_1\right) = \sigma^2 \sum_{i=1}^n \left(\frac{x_i - \bar{x}}{S_{xx}}\right)^2 = \frac{\sigma^2}{S_{xx}^2} S_{xx} = \frac{\sigma^2}{S_{xx}}$$
   $$\operatorname{Var}\left(\hat{\beta}_0\right) = \frac{\sigma^2}{n^2 S_{xx}^2} \left[n \left(\sum_{j=1}^n x_j^2\right)^2 - n^2\bar{x}^2 \left(\sum_{j=1}^n x_j^2\right) \right] = \frac{\sigma^2 \sum_{j=1}^n x_j^2}{nS_{xx}^2} \left[\sum_{j=1}^n x_j^2 - n\bar{x}^2\right]$$
   $$= \frac{\sigma^2 \sum_{j=1}^n x_j^2}{nS_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_x}\right)$$

3. La covariància entre $\hat{\beta}_0$ i $\hat{\beta}_1$ val:

   $$\operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma^2}{nS_{xx}^2} \sum_{i=1}^n (x_i - \bar{x}) \left(\sum_{j=1}^n x_j^2 - n\bar{x}x_i\right) = -\frac{\sigma^2}{nS_{xx}^2} \sum_{i=1}^n n\bar{x}(x_i - \bar{x})^2 = -\frac{\sigma^2\bar{x}}{S_{xx}},$$

   on hem usat que $\sum_i (x_i - \bar{x}) = 0$.

6.4

Gauss-Markov i estimador de la variància

El següent teorema no entra a examen.

Teorema 6.4.1 (de Gauss-Markov). *Els estimadors $\hat\beta_0$ i $\hat\beta_1$ són els estimadors lineals en y i no esbiaixats —de β_0 i β_1, respectivament— amb variància mínima. És a dir, per a tot T, S estimadors lineals en y i no esbiaixats –de β_0 i β_1, respectivament—, es compleix que:*

$$\operatorname{Var}(T) \geq \operatorname{Var}(\hat\beta_0), \ \operatorname{Var}(S) \geq \operatorname{Var}(\hat\beta_1).$$

A més, tenim la igualtat si, i només si, coincideixen; és a dir,

$$\operatorname{Var}(T) = \operatorname{Var}(\hat\beta_0) \iff T = \beta_0, \ \operatorname{Var}(S) = \operatorname{Var}(\hat\beta_1) \iff S = \beta_1\,.$$

Proposició 6.4.2 (Estimador de la variància σ^2). *L'estimador proposat és:*

$$\hat\sigma^2 = \frac{1}{n-2}\sum_i \hat\varepsilon_i = \frac{1}{n-2}\sum_i (y_i - \hat y_i)^2 = \frac{1}{n-2} S_{yy}(1 - R^2_{xy}),$$

el qual no té biaix respecte σ^2;

Demostració. En efecte,

$$\sum_i (y_i - \hat y_i)^2 \overset{2}{=} \sum_i (y_i - \hat y_i) y_i = \sum_i y_i^2 - \sum_i y_i \hat y_i = \sum_i y_i^2 - \sum_i \hat y_i^2.$$

Anàlogament:

$$\begin{aligned}
\mathbb{E}\left(\sum_i y_i^2\right) &= \sum_i \left(\operatorname{Var}(y_i) + (\mathbb{E}(y_i))^2\right) = n\sigma^2 + \sum_i \mathbb{E}(y_i)^2,\\
\mathbb{E}\left(\sum_i \hat y_i^2\right) &= \sum_i \left(\operatorname{Var}(\hat y_i) + (\mathbb{E}(\hat y_i))^2\right) = \sum_i \operatorname{Var}(\hat y_i) + \sum_i \mathbb{E}(y_i)^2 = \sum_i \operatorname{Var}(\hat y_i) + \sum_i \mathbb{E}(\hat y_i),
\end{aligned}$$

pel que

$$\mathbb{E}\left(\sum_i (y_i - \hat y_i)^2\right) = n\sigma^2 - \sum_i \operatorname{Var}(\hat y_i).$$

I, finalment:

$$\sum_i \operatorname{Var}(\hat y_i) = n \operatorname{Var}(\hat\beta_0) + \sum_i x_i^2 \operatorname{Var}(\hat\beta_1) + 2\sum_i \operatorname{Cov}(\hat\beta_0, \hat\beta_1) = \cdots = 2\sigma^2,$$

ja que $\hat y_i = \hat\beta_0 + \hat\beta_1 x_i$. ■

[2] Usarem iterativament que $\sum_i x_i \hat\varepsilon_i = 0$ i $\sum_i y_i \hat\varepsilon_i = 0$.

A

Algunes variables aleatòries

A.1 Distribucions discretes

Descriurem aquestes variables tal com les hem anomenat al llarg del llibre, i en donarem la funció de probabilitat, l'esperança, la variància i en quasi tots els casos, la funció característica.

Distribució	**PMF/PDF/CDF**	**Esperança**	**Variança**	φ
Bernoulli(p) $p \in (0,1)$	$P(X=1)=p$ $P(X=0)=q=1-p$	p	pq	$pe^{it}+q$
B(n,p) $n \in \mathbb{N}$, $p \in (0,1)$	$P(X=k)=\binom{n}{k}p^k q^{n-k}$	np	npq	$(pe^{it}+q)^n$
Geom(p) $p \in (0,1)$	$P(X=k)=p\cdot q^{k-1}$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-q\cdot e^{it}}$
Poiss(λ) $\lambda \in (0,+\infty)$	$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ	$\exp\{\lambda(e^{it}-1)\}$

Figura A.1: Distribucions discretes.

Propietat A.1.1.

1. *La Bernoulli és un cas particular de binomial quan $n=1$, és a dir, $B(p) \sim B(1,p)$.*
2. *Siguin $X_1,\dots,X_n$ VAIID amb llei $B(p)$ (Bernoulli), aleshores $X_1+\cdots+X_n \sim B(n,p)$ (Binomial).*
3. *Siguin $X \sim B(n_1,p)$ i $Y \sim B(n_2,p)$ dues variables aleatòries independents; aleshores, $X+Y \sim B(n_1+n_2,p)$. A més, la llei de X condicionada per $\{X+Y=k\}$, $0<k<n_1+n_2$, és una hipergeomètrica $H(n_1+n_2,n_1,k)$. Siguin $X \sim \text{Pois}(\lambda_1)$ i $Y \sim \text{Pois}(\lambda_2)$, dues variables aleatòries independents. Aleshores, $X+Y \sim \text{Poiss}(\lambda_1+\lambda_2)$. En aquest mateix cas, la llei de la variable X condicionada per $\{X+Y=k\}$, $k \geq 0$, és una binomial $B(k,\frac{\lambda_1}{\lambda_1+\lambda_2})$.*
4. *Sigui $(X_n)_n$ una successió de variables aleatòries amb lleis $B(n,p_n)$ on els paràmetres satisfan $\lim_n np_n=\lambda$; aleshores $X_n \xrightarrow[n\to\infty]{\mathcal{L}} \text{Pois}(\lambda)$.*
5. *La geomètrica no té memòria, ja que compleix que $P(X=n+k|X\geq n)=P(X=k)$, per a tot $n,k\geq 1$.*

A.2

Distribucions absolutament contínues

Distribució	PMF/PDF/CDF	Esperança	Variança	φ
$\mathrm{U}(a,b)$ $a<b$	$f(x)=\frac{1}{b-a}\cdot\mathbb{1}_{(a,b)}(x)$ $F(x)=\frac{x-a}{b-a}\cdot\mathbb{1}_{(a,b)}(x)$	$m_1=\frac{a+b}{2}$ $m_n=\frac{\sum_{i=0}^n a^i b^{n-i}}{n+1}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb}-e^{ita}}{it(b-a)}$
$\mathrm{Exp}(\lambda)$ $\lambda\in(0,+\infty)$	$f(x)=\lambda e^{-\lambda x}\cdot\mathbb{1}_{(0,+\infty)}(x)$ $F(x)=(1-e^{-\lambda x})\cdot\mathbb{1}_{(0,+\infty)}(x)$	$m_1=\frac{1}{\lambda}$ $m_n=\frac{n!}{\lambda^n}$	$\frac{1}{\lambda^2}$	$\left(1-\frac{it}{\alpha}\right)^{-1}$
$\mathrm{Gamma}(\alpha,\beta)$ $\alpha,\beta\in(0,+\infty)$	$f(x)=\frac{\alpha^\beta}{\Gamma(\beta)}\cdot x^{\beta-1}\cdot e^{-\alpha x}\cdot\mathbb{1}_{(0,\infty)}(x)$	$\frac{\beta}{\alpha}$	$\frac{\beta}{\alpha^2}$	$\left(1-\frac{it}{\alpha}\right)^{-\beta}$

Figura A.2: Distribucions absolutament contínues.

Propietat A.2.1.

1. *Sigui X una variable aleatòria amb funció de distribució contínua. Aleshores, $Y=F(X)\sim U(0,1)$.*
2. *Si $X\sim U(0,1)$, aleshores $Z=-\ln X\sim \mathrm{Exp}(1)$.*
3. *L'exponencial és un cas particular de Gamma quan $\beta=1$; és a dir, $\mathrm{Exp}(\alpha)\sim\mathrm{Gamma}(\alpha,1)$.*
4. *Es diu que l'exponencial no té memòria perquè $P(X\geq x+y|X\geq x)=P(X\geq y)$, per a tot $x,y\geq 0$.*
5. *Siguin $X\sim\mathrm{Exp}(\alpha)$ i $Y\sim\mathrm{Exp}(\alpha)$ dues variables aleatòries independents. Aleshores, $X+Y\sim\mathrm{Gamma}(\alpha,2)$. Siguin ara $X_1,\dots,X_n$ VAIID tals que $X_i\sim\mathrm{Exp}(\alpha)$, aleshores $\sum_{i=1}^n X_i=\mathrm{Gamma}(\alpha,n)$.*
6. *Siguin $X\sim\mathrm{Exp}(\alpha_1)$ i $Y\sim\mathrm{Exp}(\alpha_2)$ dues variables aleatòries independents, aleshores: $\frac{X}{X+Y}\sim B(\alpha_1,\alpha_2)$.*
7. *Siguin $X\sim\mathrm{Gamma}(\alpha,\beta_1)$ i $Y\sim\mathrm{Gamma}(\alpha,\beta_2)$ dues variables aleatòries independents. Aleshores, $X+Y\sim\mathrm{Gamma}(\alpha,\beta_1+\beta_2)$.*

Ara ens centrarem en la distribució normal i les derivades de la normal.

Definició A.2.2 (Distribució normal). Una variable es diu que té una distribució (o llei) normal $X\sim N(0,1)$ si la seva densitat és de la forma $f(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, $-\infty<x<\infty$. En aquest cas, $\mathbb{E}(X)=0, \mathrm{Var}(X)=1$ i $\varphi(t)=e^{-\frac{t^2}{2}}$. De manera més general, podem definir una variable amb distribució (o llei) normal $X\sim N(\mu,\sigma^2)$, $\mu\in\mathbb{R}$ i $\sigma\in(0,+\infty)$, si la seva densitat és:

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

Per aquesta generalització es té que:

1. $\mathbb{E}(X) = \mu$,
2. $\mathrm{Var}(X) = \sigma^2$ i
3. $\varphi(t) = e^{it\mu}e^{-\frac{\sigma^2t^2}{2}}$.

La variable aleatòria constant es pot considerar com un cas particular de distribució normal quan $\sigma^2 = 0$.

Propietat A.2.3 (Propietats de la distribució normal)**.**

1. *Si $X \sim N(0,1)$, aleshores per a $n \geq 1$:*

$$\mathbb{E}(X^{2n}) = \frac{(2n)!}{2^n \cdot n!} \text{ i } \mathbb{E}(X^{2n-1}) = 0.$$

2. *Evidentment, $N(0,1)$ és un cas particular de $N(\mu, \sigma^2)$ prenent els valor $\mu = 0$ i $\sigma^2 = 1$.*
3. *Si $X \sim N(0,1)$, aleshores $Y = \mu + \sigma X \sim N(\mu, \sigma^2)$. Això ens dona un mètode per pivotar, ja que si $Y \sim N(\mu, \sigma^2)$, aleshores $X = \frac{Y-\mu}{\sigma} \sim N(0,1)$.*
4. *Siguin $X \sim N(\mu_1, \sigma_1^2)$ i $Y \sim N(\mu_2, \sigma_2^2)$ dues variables aleatòries independents, aleshores, per a tot $a, b, c \in \mathbb{R}$:*

$$aX + bY + c \sim N(a\mu_1 + b\mu_2 + c, a^2\sigma_1^2 + b^2\sigma_2^2),$$

sempre i quan els coeficients a i b no siguin zero a la vegada.

5. *Si la suma de n variables aleatòries independents segueix una distribució normal, aleshores els sumands són necessàriament variables aleatòries normals.*

Ara parlarem de distribucions derivades de la normal, la Student i la χ^2 de Pearson.

Definició A.2.4 (t de Student)**.** Direm que una variable aleatòria X té llei t de Student amb n graus de llibertat, i ho denotarem per $X \sim t_{(n)}$, si té per densitat:

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

La rellevància la té la propietat 2..

Propietat A.2.5.

1. *Siguin $X \sim N(0,1)$ i $Y \sim \chi^2_{(n)}$ dues variables aleatòries independents, aleshores $\frac{\sqrt{n}X}{\sqrt{Y}} \sim t_{(n)}$.*
2. *Siguin $Y, X_1, \ldots, X_n$, $n+1$ VAIID amb llei $N(0, \sigma^2)$, aleshores $\frac{\sqrt{n}Y}{\sqrt{X_1^2 + \cdots + X_n^2}} \sim t_{(n)}$.*
3. *Tenim que $\mathbb{E}(X) = 0$ per a $n > 1$ i $\mathrm{Var}(X) = \frac{n}{n-2}$, $n > 2$.*
4. *Asimptòticament es comporta com una $N(0,1)$.*

Definició A.2.6 (χ^2 de Pearson)**.** Direm que una variable aleatòria té llei χ^2 de Pearson amb n graus de llibertat, i la denotarem per $\chi^2_{(n)}$ si té per densitat:

$$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \mathbb{1}_{(0,+\infty)}(x).$$

Propietat A.2.7.

1. *En particular, es té que* $\chi^2_{(n)}$ $\text{Gamma}(\frac{1}{2}, \frac{n}{2})$.
2. *És fàcil comprovar que si* $X_1, \ldots, X_n$ *són* VAIID *amb llei* $N(0,1)$*, aleshores* $X_i^2 \sim \chi^2_{(1)} \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$*. Conseqüentment,* $X_1^2 + \cdots + X_n^2 \sim G(\frac{1}{2}, \frac{n}{2}) \sim \chi^2_{(n)}$.
3. *Siguin* $X_1, \ldots, X_n$*,* n VAIID *amb llei* $N(0, \sigma^2)$*, aleshores* $\frac{X_1^2 + \cdots + X_n^2}{\sigma^2} \sim \chi^2_{(n)}$.
4. *Siguin* $X \sim \chi^2_{(n_1)}$ *i* $Y \sim \chi^2_{(n_2)}$ *dues variables aleatòries independents. Aleshores,* $X + Y \sim \chi^2_{(n_1+n_2)}$.
5. *Si* $X \sim \chi^2_{(n)}$*, com a conseqüència de les propietats de la Gamma tenim que* $\mathbb{E}(X) = n$*,* $\text{Var}(X) = 2n$ *i* $\varphi(t) = (1-2it)^{-\frac{n}{2}}$.
6. *Si* $X \sim \chi^2_{(n)}$*, aleshores la variable aleatòria* $Y = \sqrt{2X} - \sqrt{2n-1}$ *es comporta asimptòticament com una normal* $N(0,1)$*. Aquesta aproximació es considera acceptable per a* $n > 30$.

Distribució	PMF/PDF/CDF	Esperança	Variança	φ
$\mathcal{N}(0,1)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$	0	1	$e^{-\frac{t^2}{2}}$
$\mathcal{N}(\mu, \sigma^2)$ $\mu \in \mathbb{R}$, $\sigma \in (0, +\infty)$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{(2\sigma^2)}}$	$m_1 = \mu$ $m_2 = \mu^2 + \sigma^2$	σ^2	$e^{it\mu} e^{-\frac{\sigma^2 t^2}{2}}$
$\mathcal{LN}(\mu, \sigma^2)$ $\mu \in \mathbb{R}$, $\sigma \in (0, +\infty)$	$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\frac{\log x - \mu)^2}{2\sigma^2}}$	$\theta = e^{\mu + \frac{\sigma^2}{2}}$	$\theta^2(e^{\sigma^2} - 1)$	
χ^2_n $n \in \mathbb{N}$	$f(x) = \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} \cdot \mathbb{1}_{(0,\infty)}(x)$	n	$2n$	$(1-2it)^{-\frac{n}{2}}$
t_n $n \in \mathbb{N}$	$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} (1 + \frac{x^2}{n})^{-\frac{n+1}{2}}$	0 if $n > 1$	$\frac{n}{n-2}$ if $n > 2$	

Figura A.3: Distribucions normals i derivades de la normal.

B

Resultats rellevants de Probabilitats

B.1

Variables i vectors aleatoris

Definició B.1.1 (Àlgebra de Borel). La σ-àlgebra de Borel és la σ-àlgebra generada pels conjunts oberts de $\mathbb{R}$; és a dir, la més petita de les σ-àlgebres de parts de $\mathbb{R}$ que contenen tots els conjunts oberts de $\mathbb{R}$ respecte a la topologia euclidiana. La denotarem per $\mathcal{B}$.

Definició B.1.2 (Variable aleatòria). Una variable aleatòria és una aplicació $X : \Omega \longrightarrow \mathbb{R}$ que és *mesurable*; és a dir, $\forall B \in \mathcal{B}$ es dona que $X^{-1}(B) = \{\omega \in \Omega \bigm| X(\omega) \in B\} \in \mathcal{F}$. Una variable aleatòria proporciona, doncs, una assignació numèrica als elements de l'espai mostral. Equivalentment, per a tot $t \in \mathbb{R}$:

$$\begin{aligned} X^{-1}\left((-\infty, t]\right) &\in \mathcal{F} \\ X^{-1}\left((-\infty, t)\right) &\in \mathcal{F} \\ X^{-1}\left((t, \infty)\right) &\in \mathcal{F} \\ X^{-1}\left([t, \infty)\right) &\in \mathcal{F} \end{aligned}$$

Definició B.1.3 (Llei d'una variable aleatòria). Sigui $B \in \mathcal{B}$. La llei d'una variable aleatòria $X : (\Omega, \mathcal{F}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ és la probabilitat associada:

$$P_X(B) = P(X^{-1}(B)) = P \circ X^{-1}(B) = P(X \in B) = P(\{\omega \in \Omega \bigm| X(\omega) \in B\}).$$

És clar que $P_X(\mathbb{R}) = 1$, de manera que és una probabilitat. Finalment, sigui $\{B_n\}_{n \geq 1}$ una col·lecció de conjunts disjunts de B. Aleshores:

$$X^{-1}\left(\bigcup_{n \geq 1} B_n\right) = \bigcup_{n \geq 1} X^{-1}(B_n).$$

Proposició B.1.4. *Siguin X, Y dues variables aleatòries definida en l'espai de probabilitat $(\Omega, \mathcal{F}, P)$ i $a, b \in \mathbb{R}$. Aleshores:*

1. *$aX + bY$ i XY són també variables aleatòries.*
2. *$\max(X, Y)$ i $\min(X, Y)$ són variables aleatòries.*
3. *Si $Y : \Omega \longrightarrow (0, +\infty)$, el quocient és una variable aleatòria.*

Ara introduirem una sèrie d'exemples importants, canònics si es vol, de variables aleatòries, des de B.1.6 fins a B.1.9, són totes variables aleatòries *discretes*.

Definició B.1.5 (Variable aleatòria discreta)**.** Una variable aleatòria X direm que és discreta si la seva llei està concentrada en un conjunt numerable $B_0 \in \mathcal{B}$.

Exemple B.1.6 (Variable aleatòria de Bernoulli)**.** Considerem una experiència aleatòria amb dos resultats possibles a_1, a_2 amb probabilitats respectives p i $q = 1 - p$, $p \in (0,1)$. Associem a aquesta experiència la variable aleatòria definida per $X(a_1) = 1$ i $X(a_2) = 0$. La llei d'aquesta variable aleatòria és una probabilitat concentrada en els punts 1,0 amb intensitats p, q, respectivament. S'anomena *llei de Bernoulli* de paràmetre p i es denota per $B(1,p)$. Formalment:

$$\begin{array}{rccc} X: & \{a_1, a_2\} & \longrightarrow & \{0,1\} \\ & \{a_1\} & \longmapsto & 1 \\ & \{a_0\} & \longmapsto & 0 \end{array} \qquad F(X) = \begin{cases} 0, & \text{si } x < 0; \\ 1 - p, & \text{si } 0 \leq x < 1; \\ 1, & \text{si } x \geq 1. \end{cases}$$

Aleshores, $P_X(\{1\}) = P(X = 1) = p$ i $P_X(\{0\}) = P(X = 0) = q$. L'esperança és $E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = p$.

Exemple B.1.7 (Variable aleatòria binomial)**.** Considerem aquí la repetició, n vegades *de manera independent*, d'una experiència aleatòria com la descrita en l'exemple B.1.6. Ens interessa estudiar el nombre de vegades que hem obtingut el resultat a_1. L'espai mostral és

$$\Omega = \{\omega = (\omega_1, \cdots, \omega_n) \mid \omega_i = a_1 \text{ o } \omega_i = a_2, i = 1, 2, \ldots, n\}.$$

Fixem $\omega \in \Omega$ i sigui $n_1(\omega)$ el nombre de components a_1 en ω. Aleshores, els arguments explicitats en l'exemple anterior ens porten a assignar la probabilitat següent als elements de Ω,

$$P(\{\omega\}) = p^{n_1(\omega)}(1-p)^{n-n_1(\omega)}.$$

Definim $X : \Omega \to \{0, 1, \ldots, n\}, X(\omega) = n_1(\omega)$. La llei d'aquesta variable aleatòria és una probabilitat Q concentrada en $0, 1, \ldots, n$ tal que

$$Q(\{k\}) = P\{\omega : n_1(\omega) = k\} = \binom{n}{k} p^k (1-p)^{n-k}.$$

En efecte, el conjunt $\{\omega : n_1(\omega) = k\}$ té cardinal $\binom{n}{k}$ i cada element del conjunt té una probabilitat $p^k(1-p)^{n-k}$. És evident, a partir del desenvolupament del binomi de Newton, que $\sum_{k=0}^{n} Q(\{k\}) = 1$. La llei de la variable aleatòria X s'anomena llei binomial amb paràmetres (n,p) i es denota per $B(n,p)$.

Pel que fa a l'esperança, tenim:

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} kP(X=k) = \sum_{k=0}^{n} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \sum_{k=1}^{n} \frac{n!}{(k-1)! \cdot (n-k)!} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \frac{(n-1)!}{k-1)! \cdot (n-k)!} p^{k-1} (1-p)^{n-k} \overset{\ell=k-1}{=} \sum_{\ell=0}^{n-1} \frac{(n-1)!}{\ell! \cdot (n-1-\ell)!} p^{\ell} (1-p)^{n-1-\ell} \\ &= np(p+1-p)^{n-1} = np. \end{aligned}$$

La llei binomial és utilitzada, per exemple, en models probabilístics per al *control de qualitat.* L'observació de la qualitat d'un producte porta a dos resultats: bo o defectuós. L'estimació del nombre de productes defectuosos serveix en el control del procés de producció.

Exemple B.1.8 (Variable aleatòria geomètrica)**.** Considerem una successió de repeticions independents d'una experiència aleatòria com la de l'exemple B.1.6. Ens aturem la primera vegada que obtenim el resultat a_1 (*cara* en el llançament de la moneda). Designem per X la variable aleatòria que descriu el nombre de jugades fins que ha aparegut a_1 per primera vegada. L'espai mostral associat a aquesta experiència és:

$$\Omega = \{\omega = (\omega_1, \ldots, \omega_n) \bigm| \omega_i = a_1 \text{ o } \omega_i = a_2, i = 1, 2, \ldots\}.$$

La llei de X és una probabilitat Q concentrada en $\{1, 2, \ldots\}$ tal que $P(X = k) = Q(\{k\}) = p(1-p)^{k-1}$. S'anomena *llei geomètrica de paràmetre* p. Observem que $\sum_{k=1}^{\infty} Q(\{k\}) = 1$ car, pels resultats de la suma de la sèrie geomètrica:

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

Pel que fa l'esperança (cf. secció B.2):

$$E(X) = \sum_{k=1}^{\infty} k \cdot P(X = k) = p \sum_{k=1}^{\infty} k(1-p)^{k-1} = p \frac{1}{(1-(1-p))^2} = \frac{1}{p}.$$

En la penúltima igualtat hem usat que si tenim la sèrie $f(x) = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$, $|x| < 1$ i f derivable, tenim que $f'(x) = \sum_{k=1}^{\infty} kx^{k-1} = \frac{1}{(1-x)^2}$.

Exemple B.1.9 (Variable aleatòria de Poisson)**.** És la probabilitat de tenir èxit amb un sol experiment. Tenim $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, $k = 0, 1, 2, \ldots$ i $\lambda > 0$. En efecte, la suma de totes les probabilitats ens dona 1:

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \overset{\text{Taylor}}{=} e^{-\lambda} e^{\lambda} = 1.$$

A més, l'esperança és:

$$E(X) = \sum_{k=1}^{\infty} k \cdot P(X = k) = e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^k}{(k-1)!} = e^{-\lambda} \lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \overset{\text{Taylor}}{=} e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \lambda.$$

Matemàticament, la distribució de Poisson definida pot obtenir-se mitjançant un pas al límit en l'esquema que ens ha portat a definir la llei binomial. En efecte, considerem una successió de lleis binomials $B(n, p_n)$ i suposem que els valors del paràmetre p_n decreixen de manera que

$\lim_n np_n = \lambda > 0$. Observem que això implica que $\lim_n p_n = 0$. Aleshores, per a tot nombre natural $k \geq 0$ es té que:

$$\begin{aligned}\lim_n B(n, p_n)(\{k\}) &= \lim_n \binom{n}{k} p_n^k (1-p_n)^{n-k} = \lim_n \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \lim_n \frac{(np_n)^k}{k!} \cdot \frac{n!}{n^k (n-k)!} \left((1-p_n)^{\frac{1}{p_n}}\right)^{np_n} (1-p_n)^{-k} = \frac{\lambda^k}{k!} e^{-\lambda},\end{aligned}$$

ja que $\lim_n (1-p_n)^{-k} = 1$ i $\lim_n \frac{(n-k+1)\cdots n}{n^k} = 1$. La llei de Poisson de paràmetre λ serà denotada per $\mathrm{Poiss}(\lambda)$.

Definició B.1.10 (Funcions de distribució). La funció de distribució associada a una variable aleatòria X és la funció $F : \mathbb{R} \longrightarrow [0,1]$ definida per:

$$\begin{aligned} F_X : \ \mathbb{R} &\longrightarrow [0,1] \\ x &\longmapsto F_X := P(X \leq x) = P \circ X^{-1}((-\infty, x]). \end{aligned}$$

És a dir, considerem la família de les semirrectes $(-\infty, x], x \in \mathbb{R}$ i el valor que pren $P \circ X^{-1}$ sobre aquesta família.

Proposició B.1.11. *Sigui F la funció de distribució d'una variable aleatòria X. Es compleix que:*

1. *F és creixent.*
2. *F és contínua per la dreta, és a dir,* $\lim_{x_n \to x^+} F(x_n) = F(x)$.
3. $\lim_{x \to +\infty} F(x) = 1$ *i* $\lim_{x \to -\infty} F(x) = 0$.

La implicació contrària també és certa (però no ho demostrarem).

Propietat B.1.12.

1. $P(X < a) = F(a^-)$,
2. $P(a \leq X \leq b) = F(b) - F(a^-)$,
3. $P(a < X \leq b) = F(b) - F(a)$,
4. $P(a \leq X < b) = F(b^-) - F(a^-)$,
5. $P(a < X < b) = F(b^-) - F(a)$.
6. $P(X = a) = F(a) - F(a^-) \geq 0$.

Definició B.1.13 (Densitat). Una funció $f : \mathbb{R} \longrightarrow \mathbb{R}$ s'anomena *densitat* si compleix les condicions següents:

1. $f \geq 0$,
2. f és integrable (en el sentit de Riemann) en $\mathbb{R}$,
3. Es té que $\int_{-\infty}^{+\infty} f(x)\, dx = 1$.

Definició B.1.14 (Variable aleatòria absolutament contínua)**.** Es diu que una variable aleatòria X és absolutament contínua amb densitat f si la seva funció de distribució F es pot escriure com:

$$F(x) = \int_{-\infty}^{x} f(y)\, dy,$$

per a tot $x \in \mathbb{R}$, on la funció f satisfà les condicions de la definició B.1.13.

Exemple B.1.15 (Llei uniforme)**.** Siguin $a, b \in \mathbb{R}$ de manera que X té llei $U(a,b)$ (això és, $X \sim U(a,b)$). Definim:

$$f(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \geq 0. \tag{B.1}$$

És evident que (B.1) defineix una funció de densitat. Notem que:

$$P(X = x) = \int_{-\infty}^{x} f(y)\, dy = \int_{-\infty}^{x} \frac{1}{b-a} \mathbb{1}_{[a,b]}(y)\, dy = \int_{a}^{x} \frac{1}{b-a}\, dy = \frac{x-a}{b-a}.$$

La funció de distribució associada és, doncs:

$$F(x) = \int_{-\infty}^{x} f(y)\, dy = \begin{cases} 0, & \text{si } x \leq a, \\ \frac{x-a}{b-a}, & \text{si } x \in [a,b], \\ 1, & \text{si } x \geq b. \end{cases}$$

També tenim:

$$P(X \leqslant x) = \int_{-\infty}^{x} f(y)\, dy = \int_{-\infty}^{x} \frac{1}{b-a} \mathbb{1}_{[a,b]}\, dy = \int_{a}^{x} \frac{1}{b-a}\, dy$$

L'esperança és

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx = \int_{a}^{b} x \frac{1}{b-a}\, dx = \frac{b+a}{2}.$$

Aquest exemple correspon a una distribució de probabilitat amb densitat *constant*, és a dir, uniforme en un interval de la recta real. S'anomena la distribució *uniforme* en $[a,b]$ i es designa per $\mathcal{U}[a,b]$.

Exemple B.1.16 (Llei exponencial)**.** Fixem un $\lambda > 0$, $\lambda \in \mathbb{R}$. La llei *exponencial* amb paràmetre λ, $\exp(\lambda) \equiv e^{\lambda}$, correspon a la funció de densitat següent:

$$f(x) = \lambda \exp(-\lambda x) \cdot \mathbb{1}_{[0,+\infty)}(x).$$

La funció de distribució associada a aquesta densitat és:

$$F(x) = \begin{cases} 0, & \text{si } x \leq 0, \\ 1 - \exp(-\lambda x), & \text{si } x > 0. \end{cases}$$

Un model per a la llei exponencial és el d'una variable aleatòria X positiva que descrigui el temps de vida d'un cert sistema amb les hipòtesis següents:

1. La funció de distribució $F(x)$, $x \geq 0$, és contínua i no nul·la.
2. No hi ha envelliment. Més precisament, suposem que:

$$P\{X \geq x+y | X \geq x\} = P\{X \geq y\}, \forall x, y \geq 0.$$

Dit altrament, el temps de vida residual a l'instant x segueix la mateixa llei de probabilitat que el temps de vida des de l'instant inicial. Denotem per $\varphi(x) = 1 - F(x)$. Aleshores, $\varphi(x+y) = \varphi(x)\varphi(y)$. Tenint en compte que $\varphi(0) = 1$, resulta que $\varphi(x) = \exp(-\lambda x)$, $\lambda > 0$ car $\lim_{n\to\infty} \varphi(x) = 0$.

Exemple B.1.17 (Llei Gamma)**.** Fixem $p \in (0, +\infty)$. Es defineix la funció gamma mitjançant la integral.

$$\Gamma(p) = \int_0^\infty x^{p-1} e^{-x}\, dx, \tag{B.2}$$

en particular, $\Gamma(1) = 1$ i $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Fixem ara $p \in (0, +\infty)$ i $\lambda \in (0, +\infty)$. Definim, al seu torn:

$$f(x) = \begin{cases} \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}, & \text{si } x \geq 0; \\ 0, & \text{en cas contrari.} \end{cases}$$

Fent un canvi de variables i utilitzant (B.2) és fàcil veure que $f(x)$ defineix una funció de densitat. La llei corresponent s'anomena la *llei gamma* amb paràmetres (p, λ). La llei exponencial amb paràmetre λ es pot també descriure com una llei $\Gamma(1, \lambda)$. Pel que fa a l'esperança (cf. secció B.2):

$$J = \int_0^\infty x \frac{\lambda^p}{\Gamma(p)} x^{p-1} e^{-\lambda x}\, dx = \frac{\Gamma(p+1)}{\lambda \Gamma(p)}$$

Exemple B.1.18 (Llei normal estàndard)**.** Considerem la funció $g(x) = e^{-\frac{x^2}{2}}$, amb $x \in \mathbb{R}$. Se sap que $\int_{-\infty}^{+\infty} g(x) = \sqrt{2\pi}$. Considerem la funció de densitat $f(x) = \frac{1}{\sqrt{2\pi}} g(x)$, és a dir,

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{B.3}$$

Una variable aleatòria amb distribució normal estàndard és la que té per funció de densitat la donada en (B.3). Denotarem la llei normal estàndard per $\mathcal{N}(0, 1)$. L'esperança (cf. secció B.2) d'X és finita i $E(X) = 0$, ja que:

$$\int_{-\infty}^{+\infty} |x| \cdot e^{-\frac{x^2}{2}}\, dx \overset{1}{=} 2 \int_0^{+\infty} x e^{-\frac{x^2}{2}}\, dx = 2.$$

A partir d'aquí, és clar que $\int_{-\infty}^{+\infty} x \cdot e^{-\frac{x^2}{2}}\, dx = 0$ per la paritat de l'integrand.

[1] Hem usat que $|x| \cdot e^{-\frac{x^2}{2}}$ és parella.

Proposició B.1.19. *Y té una densitat donada per:*

$$f_Y(y) = f_X(g^{-1}(y)) \cdot |(g^{-1})'(y)| = f_X(g^{-1}(y)) \frac{1}{|g'(g^{-1}(y))|}, y \in J.$$

Definició B.1.20 (Vector aleatori). Un vector aleatori m-dimensional és una aplicació $X : \Omega \longrightarrow \mathbb{R}^m$, $X = (X_1, \ldots, X_m)$ tal que cadascun dels components $X_i : \Omega \longrightarrow \mathbb{R}$ és una variable aleatòria.

1. Per a definir la llei d'un vector aleatori cal introduir una σ-àlgebra associada a $\mathbb{R}^n$. La llei d'un vector X és una probabilitat sobre $(\mathbb{R}^n, B \in \mathbb{R}^n)$ definida per a tot $B \in \mathcal{B}(\mathbb{R}^n)$ de la manera següent: $P_X(B) = P(X^{-1}(B)) = P(\omega \in \Omega \mid X(\omega) = B)$.
2. La funció de distribució d'un vector X és:

$$\begin{array}{rccc} F: & \mathbb{R}^n & \longrightarrow & [0,1] \\ & x & \longmapsto & P(X = x) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) = P(\{X_1 \leq x_1\} \cap \cdots \cap \{X_n \leq x_n\}). \end{array}$$

Definició B.1.21 (Densitat, d'un vector). Una funció $f : \mathbb{R}^m \to \mathbb{R}$ és una densitat en $\mathbb{R}^m$ si es compleixen les condicions següents:

1. $f \geq 0$,
2. existeix la integral

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x)\, dx$$

en el sentit de Riemann, *i*

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x)\, dx = 1$$

Definició B.1.22 (Vector absolutament continu). Un vector aleatori X m-dimensional és absolutament continu (o té llei absolutament contínua) amb densitat f, si la seva funció de distribució F es pot escriure com

$$F(x) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_m} f(x)\, dx$$

on f és una funció de densitat en $\mathbb{R}^m$, és a dir, compleix les condicions de la definició de densitat d'un vector, B.1.21, i $x = (x_1, \ldots, x_m)$. En els punts $x \in \mathbb{R}^m$ on la funció de densitat f és contínua es té:

$$f(x) = \frac{\partial^m F(x)}{\partial x_1, \cdots, \partial x_m}.$$

Proposició B.1.23. *Sigui X un vector aleatori amb llei absolutament contínua. Aleshores, cadascuna de les variables aleatòries components X_i, $i = 1, \ldots, m$, són també absolutament contínues i les seves densitats respectives, que denotarem per f_i (densitats marginals), s'expressen com:*

$$f_i(y) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_m)\, dx_1 \cdots dx_{i-1}\, dx_{i+1} \cdots dx_m,$$

per a $i=1, x_1=y$.

Exemple B.1.24 (Llei normal bidimensional)**.**

$$f_{(X,Y)}(x,y)=\frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)},\ x,y\in\mathbb{R}.$$

En efecte, podem calcular $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}f(x,y)\,dx\,dy=1$. Les densitats marginals corresponen a lleis $\mathcal{N}(0,1)$.

Suposem que X té llei absolutament contínua i considerem el cas $m=d$: U,V dos oberts de $\mathbb{R}^m$ i $g:U\longrightarrow V$ és un difeomorfisme. Aleshores, el vector aleatori Y té llei absolutament contínua i la seva densitat és:

$$f_Y(y)=f_X(g^{-1}(y))\cdot|J_{g^{-1}}(z)|.$$

$J_{g^{-1}}$ és el jacobià de la transformació g^{-1}.

Definició B.1.25 (Variables aleatòries independents)**.** Siguin $X_1,\dots,X_m$ variables aleatòries definides en un espai de probabilitat $(\Omega,\mathcal{F},P)$. Direm que són *independents* si, per a conjunts qualssevol $B_1,\dots,B_m\in\mathcal{B}$ es compleix que:

$$P_1(\{X_1\in B_1,\dots,X_m\in B_m\}=\prod_{i=1}^{m}P(\{X_i\in B_i\});$$

en altres paraules, si els esdeveniments $\{X_1\in B_1\},\dots,\{X_m\in B_m\}$ són independents. Un conjunt $\{X_n \mid n\geq 1\}$ infinit té esdeveniments independents si qualsevol subcol·lecció finita ho és.

Teorema B.1.26. *Considerem un vector aleatori* $X=(X_1,\dots,X_m)$*. Denotem per* F *la seva funció de distribució i per* F_i*,* $i=1,\dots,m$ *les funcions de distribució respectives de les variables aleatòries components. Suposem que les variables aleatòries* $X_1,\dots,X_m$ *són independents. Aleshores, es compleix que* $F(x_1,\dots,x_m)=F_1(x_1)\cdots F_m(x_m)$*, per a tot* $x=(x_1,\dots,x_m)\in\mathbb{R}^m$*. El recíproc també és cert.*

Proposició B.1.27. *Siguin* $X_1,\dots,X_m$ *variables aleatòries. Suposem que el vector aleatori* $X=(X_1,\dots,X_m)$ *té llei absolutament contínua amb densitat* f *i densitats marginals* $f_1,\dots,f_m$*. Suposem que es compleix:*

$$f(x_1,\dots,x_m)=f_1(x_1)\cdots f_m(x_m).$$

Aleshores, les variables $X_1,\dots,X_m$ *són independents.*

B.2

ESPERANÇA

Proposició B.2.1. *Siguin X, Y variables aleatòries* simples *i $a, b \in \mathbb{R}$. Aleshores:*

1. $E(aX + bY) = aE(X) + bE(Y)$,
2. *Si* $X, Y \geq 0$ *i* $X \leq Y$, *es té* $E(X) \leq E(Y)$.

Definició B.2.2. Donada una variable aleatòria X qualsevol, direm que es té esperança matemàtica finita si, i només si, $E(|X|) < \infty$ i, en aquest cas, definim $E(X) = E(X^+) - E(X^-)$.

Proposició B.2.3. *Siguin* $X, Y \in L^1(\Omega)$, $a, b \in \mathbb{R}$.

1. *La variable aleatòria* $aX + bY \in L^1(\Omega)$ *i* $E(aX + bY) = aE(X) + bE(Y)$.
2. *Si* $X \leq Y$, *aleshores* $E(X) \leq E(Y)$.
3. *Propietat de convexitat. Es compleix que* $|E(X)| \leq E(|X|)$.

Proposició B.2.4 (Esperança d'una variable absolutament contínua). *Una variable aleatòria X amb densitat f pertany a $L^1(\Omega)$ si, i només si,* $\int_{-\infty}^{+\infty} |x| \cdot f(x)\, dx < +\infty$. *En aquest cas,* $E(X) = \int_{-\infty}^{+\infty} x \cdot f(x)\, dx$.

Procés B.2.5 (Càlcul de moments). *Si existeix $\mu_1 = E(X)$, els altres moments es calculen tenint en compte la definició d'esperança d'una funció d'una variable aleatòria; així, existeix el moment no centrat d'ordre k, donat per:*

$$m_k = \begin{cases} \sum_{x \in E_X} x^k P(X = x), & \textit{si } X \textit{ és discreta.} \\ \int_{\mathbb{R}} x^k f_X(x)\, dx, & \textit{si } X \textit{ és contínua.} \end{cases}$$

Proposició B.2.6.

1. *Desigualtat de Minkowskii. Si X, Y són dues variables aleatòries amb moment d'ordre k finit, la variable aleatòria $X + Y$ té també moment d'ordre k finit.*
2. *Desigualtat de Hölder. Sigui X una variable aleatòria amb moment d'ordre p finit, sigui Y una variable aleatòria amb moment d'ordre q finit, tal que* $\frac{1}{p} + \frac{1}{q} = 1$. *Aleshores:*

$$E(|XY|) \leq (E|X|^p)^{\frac{1}{p}} + (E|Y|^q)^{\frac{1}{q}}$$

Proposició B.2.7 (Desigualtat de Jensen). *Sigui g una funció real convexa i X una variable aleatòria de $L^1(\Omega)$ tal que $g(X) \in L^1(\Omega)$. Aleshores,* $g(E(X)) \leq E(g(X))$.

Proposició B.2.8. *Siguin X, Y dues variables aleatòries independents amb esperança finita. Aleshores, la variable aleatòria producte XY té també esperança finita i* $E(XY) = E(X)E(Y)$.

Proposició B.2.9. *Siguin X, Y dues variables aleatòries independents amb moments de segon ordre finits. Aleshores,* $\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$.

Definició B.2.10 (Covariància)**.** Per mesurar el grau de dependència entre dues variables aleatòries X i Y s'introdueix la noció de covariància. Suposem que les variables tenen moments de segon ordre finits. La covariància entre X i Y es defineix per

$$\operatorname{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$$

B.3

Successions de variables aleatòries

Definició B.3.1 (Convergència quasi segura)**.** Direm que una successió de variables aleatòries $\{X_n\}_n$ convergeix *quasi segurament* cap a una variable aleatòria X si existeix un conjunt $N \in \mathcal{A}$ de probabilitat zero tal que $\lim_n X_n(\omega) = X(\omega)$, per a tot $\omega \notin N$.

Si X_n convergeix quasi segurament cap a X quan $n \to \infty$, escriurem $X_n \to X$ QS o bé $\lim_n X_n = X$, QS.

1. Es tracta, doncs, d'una convergència puntual de la successió numèrica $X_n(\omega)$, en tots els punts $\omega \in \Omega \setminus N$.
2. La variable límit és única llevat el conjunts de probabilitat zero. Naturalment, que la successió $\{X_n\}_n$ convergeix QS si, i només si, és de Cauchy amb probabilitat 1.

Definició B.3.2 (Convergeix en probabilitat)**.** Direm que una successió de variables aleatòries $(X_n)_n$ convergeix en probabilitat *cap a una variable aleatòria* X si per a tot $\varepsilon > 0$ tenim $\lim_n P(\{|X_n - X| > \varepsilon\}) = 0$. Escriurem $P - \lim_n X_n = X$. En augmentar n és cada vegada menys probable que X_n i X difereixin en més de ε, per a tot $\varepsilon > 0$, fixat.

Definició B.3.3 (Convergeix en mitjana d'ordre p)**.** Sigui $(X_n)_n$ una successió de variables aleatòries de $L^p(\Omega, \mathcal{A}, P)$. Direm que aquesta successió convergeix en mitjana d'ordre p cap a una variable aleatòria X, que també té moment d'ordre p finit:

$$\lim_{n\to\infty} E(|X_n - X|^p) = 0.$$

Escriurem $L^p - \lim_n X_n = X$.

Definició B.3.4 (Convergència en llei)**.** Sigui $(X_n)_n$ una successió de variables aleatòries i $(F_n)_n$ la successió de les seves funcions de distribució respectives. Direm que la successió $(X_n)_n$ convergeix en llei cap a una variable aleatòria X que té funció de distribució F si es compleix que $\lim_n F_n(x) = F(x)$, per a tot $x \in \mathbb{R}$, on la funció F sigui contínua. Denotarem aquest tipus de convergència mitjançant $\mathscr{L} - \lim_n X_n = X$.

Teorema B.3.5. *Sigui $(X_n)_n$ una successió de variables aleatòries i X una altra variable aleatòria.*

1. *Si $X_n \to x$ QS, quan $n \to \infty$, aleshores $P - \lim_n X_n = X$.*
2. *Si $P - \lim_n X_n = x$, existeix una subsuccessió $(X_{n_k})_k$ que convergeix QS cap a la variable X quan $k \to \infty$.*

Corol·lari B.3.6. *Sigui $(X_n)_n$ una successió de variables aleatòries i X una altra variable aleatòria. Si la successió de variables és monòtona (creixent o decreixent), aleshores la convergència en probabilitat implica la quasi segura.*

Proposició B.3.7. *Sigui $(X_n)_n$ una successió de variables aleatòries que convergeix en probabilitat cap a una variable aleatòria X. Aleshores, la successió convergeix també en llei cap a X. L'afirmació recíproca també és certa si la variable aleatòria X és constant quasi segurament, és a dir, $X(\omega) = k$ per a tot $\omega \notin N$, on N és un conjunt d'$\mathcal{A}$ de probabilitat zero i $k \in \mathbb{R}$.*

Teorema B.3.8. *Sigui $(Y_n)_n$ una successió de variables aleatòries que convergeixen en llei cap a una variable aleatòria Y. Existeixen una successió de variables aleatòries $(X_n)_n$ i una variable aleatòria X que compleixen les següents propietats:*

1. *La llei de Y_n és la mateixa que la de X_n, per a tot $n \geq 1$ i la llei de Y la mateixa que la de X.*
2. *La successió $(X_n)_n$ convergeix QS cap a la variable aleatòria X.*

Teorema B.3.9 (Llei forta dels grans nombres per a variables incorrelacionades)**.** *Sigui $(X_n)_n$ una successió de variables aleatòries incorrelacionades amb moments de segon ordre fitats uniformement per una constant C. Aleshores, en la convergència QS tenim:*

$$\lim_{n\to\infty} \frac{S_n - E(S_n)}{n} = 0.$$

Teorema B.3.10. *Sigui $(X_n)_n$ una successió de variables aleatòries i.i.d.:*

1. *Suposem que $E(|X_1|) < \infty$. Aleshores, $\lim_n \frac{S_n}{n} = E(X_1)$, QS.*
2. *Recíprocament, si $E(|X_1|) = \infty$. Aleshores, $\limsup_n \frac{S_n}{n} = \infty$, QS.*

Teorema B.3.11 (del límit central de Lévy-Lindeberg)**.** *Sigui (Z_n) una successió de variables aleatòries independents, idènticament distribuïdes i de quadrat integrable. Designem per m i σ^2 els valors de la mitjana i la variància de la distribució comuna. Suposem $\sigma^2 > 0$. Si $S_n = Z_1 + \cdots + Z_n$, es compleix que:*

$$\mathcal{L} - \lim_{n\to\infty} \frac{S_n - nm}{\sigma\sqrt{n}} = Y,$$

on Y és una variable aleatòria amb distribució $\mathcal{N}(0,1)$.

Petit disclaimer

Qui hagi fet l'assignatura i llegit aquestes notes, s'adonarà que falten alguns continguts (pocs, però certament alguns). L'autor n'és conscient i, per manca de temps, quedaran així com a mínim durant unes setmanes. I d'aquí l'etiqueta d'*Incompletes*. Val a dir que el criteri d'omissió no és pas aleatori. Tot allò que el lector no trobi és perquè no s'ha fet en les classes magistrals de teoria d'aquest semestre. A continuació es llisten les seccions que no hi són:

1. Tests d'independència.
2. Altres testos no paramètrics.
3. Model normal de regressió simple.

Bibliografia

[San99] Marta Sanz-Solé. *Probabilitats*. cat. UB; 28. Barcelona: Edicions Universitat de Barcelona, 1999. isbn: 8483380919.

[JM10] Olga Julià de Ferran i David Márquez-Carreras. *Un Primer curs d'estadística*. cat. Universitat ; 48. Barcelona: Publicacions i Edicions Universitat de Barcelona, 2010. isbn: 9788447534838.

[CV21] Eloi Castaño Camps i Víctor Valldoseria i Socías. *Estadística*. Català. Barcelona, maig de 2021.

Apunts de l'assignatura d'Estadística, impartida per David Márquez durant la primavera de 2021.

[JMR21] Olga Julià de Ferran, David Márquez-Carreras i Carles Rovira Escofet. *Estadística: problemes i més problemes*. cat. Textos docents; 431. Barcelona: Edicions de la Universitat de Barcelona, 2021. isbn: 9788491686453.