

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

**Evolució dels algorismes MCMC i
la seva implementació en
programació probabilística**

Autor: Cristina Rosich Solé

Director: Dr. Jordi Vitrià
Realitzat a: Departament de
Matemàtiques i Informàtica

Barcelona, 27 de juny de 2018

Abstract

The first goal of this paper is to explain how the Markov Chain Monte Carlo algorithms were born and follow their evolution throughout the 20th century to the present day. Firstly, we are going to discuss different types of MCMC algorithms and explain how they were created, and secondly, we are going to introduce the large amount of different applications they have. The other main goal is analyze their implementation in the field of programming, which has resulted in the creation of several probabilistic programming languages, and use them in a real case in which we are going to examine the client flow of a certain company.

Resum

El primer objectiu d'aquest document és explicar com es van originar els algorismes de *Markov Chain Monte Carlo* i com han anat evolucionant al llarg del segle XX fins a l'actualitat. Comentarem, d'una banda, els tipus d'algorismes que hi ha i com es van crear, i d'altra banda, introduïrem el gran ventall d'aplicacions que tenen. L'altre objectiu principal és veure com s'han implementat aquests algorismes en programació, donant lloc als llenguatges de programació probabilístics, i aplicar-los en un cas real, estudiant l'afluència de clients d'una certa empresa.

Agraïments

Agraeixo al Dr. Jordi Vitrià per acceptar dirigir aquest treball així com per guiar-me i inspirar-me durant tot el seu desenvolupament.

També agraeixo als propietaris del restaurant El Vinyet, Germà Solé i Montse Rañé, per accedir a cedir-me les dades que han fet possible la realització de la part pràctica del treball.

Índex

1	Introducció	1
2	Teoria de MCMC	3
2.1	Cadenes de Markov	3
2.2	Teoria de camps aleatòris de Markov	5
2.3	Inferència bayesiana	6
2.3.1	Inferència bayesiana aplicada a l'anàlisi de dades	7
2.4	Què són els algorismes MCMC	8
2.4.1	Tipus d'algorismes MCMC	8
2.4.2	Aplicació dels algorismes MCMC en inferència bayesiana	9
3	Origen i evolució dels algorismes MCMC	10
3.1	Context històric	10
3.2	Origen del mètode de Monte Carlo Ordinari (MCO)	10
3.3	L'algorisme de Metropolis	12
3.4	L'algorisme de Metropolis-Hastings	14
3.4.1	Mostreig preferencial	16
3.5	Mostreig de Gibbs	16
3.5.1	Article dels germans Geman	17
3.6	Implementació de la dinàmica hamiltoniana en MCMC	19
3.6.1	Dinàmica hamiltoniana	19
3.6.2	Algorisme de Hamilton Monte Carlo	20
3.7	Revolució dels anys 90	21
3.7.1	Article de Gelfand i Smith	21
3.7.2	Conseqüències de l'article de Gelfand i Smith	22
3.7.3	El mostreig per rebuig adaptatiu de Gilks i Wild	23
3.7.4	El salt reversible de Green	24
4	Aplicacions dels algorismes MCMC	25
5	Llenguatges de programació probabilístics	27
5.1	Estudi sobre l'afluència de clients	28
5.1.1	Estudi sobre els <i>esmorzars</i>	29
5.1.2	Estudi sobre els <i>dinars</i>	31

5.1.3	Estudi sobre els <i>sopars</i>	32
5.1.4	Estudi sobre l'afluència <i>total</i> de clients	33
5.2	Evolució de l'afluència de clients	36
5.2.1	Evolució dels <i>esmorzars</i>	38
5.2.2	Evolució dels <i>dinars</i>	40
5.2.3	Evolució dels <i>sopars</i>	42
5.2.4	Evolució de l'afluència <i>total</i> de clients	44
5.3	Optimització del mostreig per a l'estudi de l'afluència de clients . .	47
6	Conclusions	49

1 Introducció

La implementació de mètodes Markov Chain Monte Carlo (MCMC) per a la simulació de funcions de densitat de probabilitat ha suposat, des de principis dels anys 90, tota una revolució en una gran varietat de branques científiques, tals com en física computacional, bioestadística o inferència bayesiana. Algorismes que en els seus inicis podien tardar unes quantes hores en executar-se, avui en dia s'executen en qüestió de minuts o segons gràcies als llenguatges de programació probabilístics, permetent-nos resoldre en poc temps complexos problemes computacionals que anteriorment havien resultat intractables.

En termes generals, donada una certa distribució de probabilitat, un algorisme MCMC és la simulació d'una cadena de Markov a partir de la qual es recullen mostres que ens permeten fer estimacions sobre els paràmetres d'aquesta distribució, inicialment desconeguts. Aquest mètode és especialment útil en estadística bayesiana on, donat un conjunt de dades al qual se li vol associar un model probabilístic, s'obtenen fàcilment els valors més probables dels paràmetres que el determinen, problema que havia estat impossible de resoldre abans d'aparèixer els algorismes MCMC.

El seu origen es remunta a l'època de la Segona Guerra Mundial, on els investigadors de Los Álamos Stanislaw Ulam i John von Neumann creen l'algorisme Ordinari de Monte Carlo. Pocs anys més tard, apareix el primer algorisme MCMC en un article de Nicholas Metropolis, un dels companys de Ulam i von Neumann, a on s'introdueixen les cadenes de Markov. Donat que aquest algorisme estava originalment enfocat en l'estudi de substàncies amb les seves fases líquides i gasoses en equilibri, les primeres aplicacions que se'n van donar estaven especialment centrades en problemes físics i químics.

En els anys posteriors es van anar elaborant nous algorismes, com Metropolis-Hastings (1970) o el mostreig de Gibbs (1984), però no és fins als anys 90, amb la publicació d'un article de Gelfand i Smith i just després que l'accés als ordinadors es generalitzés amb l'arribada al mercat dels ordinadors personals, quan els algorismes MCMC marquen un abans i un després en la resolució de problemes computacionals, especialment en el camp de l'estadística. Des de llavors s'han anat publicant diferents variants d'aquests algorismes, com Hamilton Monte Carlo o Metropolis-Hastings-Green, i s'han publicat de manera massiva articles sobre diferents aplicacions en una gran varietat de camps [13].

El primer objectiu d'aquest document és explicar com han anat evolucionant els algorismes MCMC des de l'època de l'algorisme Ordinari de Monte Carlo, utilitzant com a font primària d'informació cadascun dels articles originals on es publiquen els diferents algorismes, així com analitzar les diferències i avantatges de cada un d'ells respecte la resta d'algorismes. El segon objectiu principal és comprovar-ne la seva gran eficiència en un cas pràctic utilitzant un d'aquests llenguatges de programació probabilístics que s'han comentat. Aquest treball es dividirà en tres parts:

Abans d'entrar en l'estudi de la seva evolució, primer s'introduiran alguns conceptes bàsics de processos estocàstics que ens permetin entendre la terminologia

utilitzada durant la definició dels algorismes. En aquesta primera part també es donarà una breu introducció a la inferència bayesiana i, per acabar, s'explicarà en què consisteixen, de manera general, els mètodes MCMC i com es poden aplicar en aquest tipus d'inferència.

La segona part del treball consistirà en explicar la història que envolta els mètodes MCMC. Com es van originar, quins són els algorismes MCMC més rellevants parant atenció a la definició de cada un d'ells, i com han evolucionat fins a dia d'avui. A més a més, s'exposaran alguns exemples de la gran quantitat d'aplicacions que tenen en diferents camps de la ciència.

En la tercera i última part, es realitzarà un estudi exhaustiu sobre l'afluència de clients d'una certa empresa, implementant al llarg d'aquest gran part dels coneixements adquirits en les parts anteriors. A partir d'un llenguatge de programació probabilístic, en aquest cas la llibreria PyMC3 per a Python, s'elaboraran diferents models probabilístics els paràmetres dels quals es determinaran mitjançant la implementació de Metropolis-Hastings o de Hamilton Monte Carlo, que permetran analitzar la presència de possibles canvis en l'afluència de clients. A més a més, es realitzarà un petit exercici d'optimització del temps d'execució pel qual es compararan els avantatges de Hamilton Monte Carlo respecte Metropolis-Hastings.

2 Teoria de MCMC

El primer que farem és recordar alguns conceptes bàsics que ens ajudaran a entendre tota la teoria que hi ha darrera els algorismes MCMC.

2.1 Cadenes de Markov

El primer que necessitem és definir què és un procés estocàstic i quina classe de procés és una cadena de Markov [47]:

Definició 2.1. *Sigui \mathcal{T} un subconjunt de $[0, \infty)$. Un **procés estocàstic** és una família de variables aleatòries $\mathbb{X} = \{X_t\}_{t \in \mathcal{T}}$. En el cas que $\mathcal{T} = \mathbb{N}_0$, direm que es tracta d'un procés estocàstic discret, mentre que per $\mathcal{T} = [0, \infty)$ direm que és un procés estocàstic continu.*

De manera senzilla, podem definir una cadena de Markov com un procés estocàstic amb la propietat de Markov, és a dir, que els estats del futur només depenen de l'estat present, sense tenir en compte els estats passats. A continuació, donarem una definició més precisa [47], tenint en compte que el procés $\{X_n\}_{n \in \mathbb{N}_0}$ pren valors d'un cert conjunt E , generalment numerable, que es coneix com l'espai d'estats.

Definició 2.2. *Direm que un procés estocàstic $\mathbb{X} = (X_n)_{n \in \mathbb{N}_0}$ és una **cadena de Markov** amb*

- *Distribució inicial $\lambda := (\lambda_i, i \in E)$, on $\lambda_i = \mathbb{P}(X_0 = i)$*
- *Matriu de transició $P := (p_{i,j}, i, j \in E)$, on $p_{i,j} = \mathbb{P}(X_{n+1} = j | X_n = i)$*

si es compleix que

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n, \dots, X_0 = i_0) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n) = p_{i_n, i_{n+1}} \quad (2.1)$$

Per a tot $n \in \mathbb{N}_0$ i tot $\{i_0, \dots, i_{n+1}\} \subset E$, sempre que $\mathbb{P}(X_n = i_n, \dots, X_0 = i_0) > 0$.

La condició de l'equació (2.1) rep el nom de propietat de Markov, ja que és la que defineix aquest tipus de procés.

Definició 2.3. *Una cadena de Markov \mathbb{X} es diu que es **irreductible** si des d'un estat qualsevol $i \in E$ es pot accedir a qualsevol altre estat $j \in E$.*

Definició 2.4. *Sigui \mathbb{X} un procés estocàstic qualsevol, direm que \mathbb{X} és **estacionari** si per tot subconjunt de temps $\{t_1, \dots, t_n\} \subset \mathcal{T}$ i tot subconjunt d'estats $\{i_1, \dots, i_n\} \subset E$, es compleix que*

$$\mathbb{P}(X_{t_1+h} = i_1, \dots, X_{t_n+h} = i_n) = \mathbb{P}(X_{t_1} = i_1, \dots, X_{t_n} = i_n) \quad (2.2)$$

per a qualsevol $h \in \mathcal{T}$ [13].

Per tant, en el cas d'una cadena de Markov, direm que aquesta és estacionària si, i només si, la distribució marginal de X_n , $\mathbb{P}(X_n = i)$, no depèn de n .

Definició 2.5. Donada una certa cadena de Markov \mathbb{X} , direm que les seves **probabilitats de transició són estacionàries** si la distribució condicional

$$\mathbb{P}(X_{n+1} = j | X_n = i) \quad (2.3)$$

no depèn de n [13].

Es veu ràpidament que si una cadena de Markov és estacionària, les seves probabilitats de transició també ho seran ja que

$$\begin{aligned} \mathbb{P}(X_{n+1} = j | X_n = i) &= \frac{\mathbb{P}(X_{n+1} = j, X_n = i)}{\mathbb{P}(X_n = i)} \\ &= \frac{\mathbb{P}(X_{k+1} = j, X_k = i)}{\mathbb{P}(X_k = i)} = \mathbb{P}(X_{k+1} = j | X_k = i) \end{aligned} \quad (2.4)$$

per a qualsevol $k = n + h$, però la implicació no és certa en l'altre sentit [47].

Definició 2.6. Sigui \mathbb{X} una cadena de Markov, direm que la seva distribució inicial λ és **invariant** si $\lambda P = \lambda$. En aquest cas es diu que la cadena de Markov té una distribució en equilibri.

En general, totes les cadenes de Markov que s'utilitzen en algorismes MCMC tenen distribucions inicials en equilibri. A continuació introduïrem els camins aleatòris [23]:

Definició 2.7. Sigui \mathbb{X} una cadena de Markov amb un espai d'estats E tal que es pugui definir la posició relativa entre dos estats qualssevol de E . Siguin $i, j \in E$ i k una translació. Direm que la cadena és **homogènia en l'espai** si la probabilitat de transició depèn únicament de la posició relativa entre els estats i, j . És a dir:

$$p_{i,j} = p_{i+k,j+k} \quad (2.5)$$

Definició 2.8. Un **camí aleatòri** és una cadena de Markov estacionària i homogènia en l'espai, amb probabilitats de transició $p_{i,j} = p(j - i)$.

En el cas concret que les probabilitats de transició només siguin diferents de zero pels estats veïns, es diu que el camí aleatòri és simple. Per exemple, per un espai d'estats unidimensional, tenim

$$p(1) = p \quad p(-1) = q \quad (2.6)$$

de manera que $p + q = 1$ i $p(k) = 0$ per $k \neq \pm 1$. Pel cas en que $p = q = \frac{1}{2}$, a més, es diu que el camí és simètric.

També recordarem el concepte de reversibilitat i en provarem la seva condició tal com es veu a [16].

Definició 2.9. Donada una certa cadena de Markov \mathbb{X} , direm que la seva distribució de probabilitats de transició és reversible respecte a la distribució inicial si per tot $n \in \mathbb{N}_0$ i per qualsevol parella d'estats $i, j \in E$ es té

$$\mathbb{P}(X_n = i, X_{n+1} = j) = \mathbb{P}(X_n = j, X_{n+1} = i) \quad (2.7)$$

Aleshores direm que la cadena de Markov és **reversible**.

En particular, tenim que tot camí aleatòri simètric és reversible.

Teorema 2.10. Donada una certa cadena de Markov amb distribució inicial invariànt, la cadena serà reversible si, i només si,

$$\lambda_i p_{i,j} = \lambda_j p_{j,i} \quad (2.8)$$

per a qualsevol parella d'estats $i, j \in E$. Aquest teorema es coneix com la condició de reversibilitat.

Demostració.

$$\begin{aligned} \mathbb{P}(X_n = i, X_{n+1} = j) &= \mathbb{P}(X_n = i) \mathbb{P}(X_{n+1} = j | X_n = i) = \lambda_i p_{i,j} = \\ &= \lambda_j p_{j,i} = \mathbb{P}(X_n = j) \mathbb{P}(X_{n+1} = i | X_n = j) = \mathbb{P}(X_n = j, X_{n+1} = i) \end{aligned} \quad (2.9)$$

□

Un cop es coneix la reversibilitat d'una cadena de Markov, la invariància de la seva distribució inicial queda garantida.

Definició 2.11. Sigui $\mathbb{X} = \{X_t\}_{t \in \mathcal{T}}$ un procés estocàstic qualsevol i f una funció real definida en el seu espai d'estats. Aleshores, el conjunt de variables $\{f(X_t)\}_{t \in \mathcal{T}}$ és un procés estocàstic i rep el nom de **funcional** de \mathbb{X} .

Cal dir que, en general, el funcional d'una cadena de Markov no és una cadena de Markov, però sí que té una sèrie de propietats interessants que habitualment no tenen altres processos estocàstics [13].

2.2 Teoria de camps aleatoris de Markov

Anem a introduir alguns conceptes necessaris per entendre el mostreig de Gibbs, tots ells explicats a [10]:

Definició 2.12. Sigui \mathcal{S} un conjunt de posicions, direm que un conjunt $\mathcal{G} = \{\mathcal{G}_s : s \in \mathcal{S}\}$ és un **sistema de veïnats** de \mathcal{S} si es compleix:

(i) $s \notin \mathcal{G}_s$.

(ii) $s \in \mathcal{G}_r \Leftrightarrow r \in \mathcal{G}_s$.

Notem que el conjunt $\{\mathcal{S}, \mathcal{G}\}$ representa un graf.

Definició 2.13. Sigui $\mathbb{X} = \{X_s : s \in \mathcal{S}\}$ una família de variables aleatòries indexada pel conjunt de posicions \mathcal{S} , de dimensió N , i sigui E el seu espai d'estats. Sigui $\Omega = \{\omega = (x_{s_1}, \dots, x_{s_N}) : x_{s_i} \in E\}$ l'espai de configuracions de \mathbb{X} . Direm que \mathbb{X} és un **camp aleatòri de Markov** respecte $\{\mathcal{S}, \mathcal{G}\}$ si:

- (i) $\mathbb{P}(\mathbb{X} = \omega) > 0$ per a tot $\omega \in \Omega$.
- (ii) $\mathbb{P}(X_s = x_s | X_r = x_r, r \neq s) = \mathbb{P}(X_s = x_s | X_r = x_r, r \in \mathcal{G}_s)$.

Notem que aquesta propietat és una generalització de la propietat de Markov en dimensió N on la probabilitat condicionada d'un cert estat només depèn dels estats veïns. És a dir, les cadenes de Markov són un cas particular de camp aleatòri de Markov en dimensió 1.

Definició 2.14. Una **distribució de Gibbs canònica** respecte a $\{\mathcal{S}, \mathcal{G}\}$ és una mesura de probabilitat $p : \Omega \rightarrow [0, 1]$ definida per

$$p(\omega) = \frac{1}{Z} e^{-U(\omega)/T} \quad (2.10)$$

on Z és la constant de normalització, U és la funció energia i T és la temperatura.

En particular, qualsevol distribució contínua, com per exemple una distribució normal, es pot reescriure com una distribució de Gibbs [39]. A continuació anunciarem un teorema que ens permet establir una connexió entre els camps aleatòris de Markov i les distribucions de Gibbs:

Teorema 2.15. Una família de variables \mathbb{X} és un camp aleatòri de Markov respecte $\{\mathcal{S}, \mathcal{G}\}$ si, i només si, $p(\omega) = \mathbb{P}(\mathbb{X} = \omega)$ és una distribució de Gibbs respecte $\{\mathcal{S}, \mathcal{G}\}$.

2.3 Inferència bayesiana

A diferència de la inferència freqüentista, que és la més tradicional i amb la que estem més habituats a treballar, la probabilitat de que succeeixi un cert esdeveniment en inferència bayesiana no ve donada per la freqüència de vegades que ha passat aquest a llarg termini, sinó que la probabilitat s'interpreta com una mesura de la nostra creença en que aquest esdeveniment succeeixi [6]. No obstant, aquesta creença és d'origen individual i, per tant, la probabilitat que se li associa a un cert esdeveniment variarà d'una persona a una altra.

Utilitzant notació freqüentista, la nostra creença de que succeeixi l'esdeveniment \mathcal{A} l'escriuim com $\mathbb{P}(\mathcal{A})$ i rep el nom de probabilitat *a priori*. D'altra banda, un cop coneixem una nova informació o evidència \mathcal{X} , la nova creença de que succeeixi \mathcal{A} coneixent \mathcal{X} l'escriuim com $\mathbb{P}(\mathcal{A}|\mathcal{X})$ i es coneix com probabilitat *a posteriori*. Així doncs, a mesura que anem afegint informació, anem actualitzant les nostres creences sobre l'esdeveniment en qüestió. Anem a il·lustrar-ho amb un exemple:

Exemple 2.16. Considerem dos individus, A i B , i una baralla de cartes francesa, on tenim 52 cartes repartides en 4 pals. Suposem que l'individu A agafa una carta

i, donant-ne a conèixer el pal, l'individu B l'ha d'endevinar. Inicialment, els dos individus atribuiran la mateixa probabilitat a qualsevol carta \mathcal{C} , $\mathbb{P}(\mathcal{C}) = \frac{1}{52}$. Suposant que l'individu A agafa, a l'atzar, el 4 de piques, i li diu a l'individu B el pal de la carta, piques, els dos individus passaran a tenir informació diferent, \mathcal{X}_A i \mathcal{X}_B . D'una banda, l'individu A li atribuirà al 4 de piques probabilitat 1, $\mathbb{P}(4P|\mathcal{X}_A) = 1$, i a la resta de cartes probabilitat 0. D'altra banda, per a l'individu B, que té menys informació, el 4 de piques tindrà la mateixa probabilitat que la resta de cartes de piques, $\mathbb{P}(4P|\mathcal{X}_B) = \frac{1}{13}$, i tota la resta de cartes probabilitat 0.

Per tal de connectar les nostres probabilitats *a posteriori*, $\mathbb{P}(\mathcal{A}|\mathcal{X})$, amb les probabilitats *a priori*, $\mathbb{P}(\mathcal{A})$, s'utilitza el teorema de Bayes, molt utilitzat també en inferència freqüentista:

$$\mathbb{P}(\mathcal{A}|\mathcal{X}) = \frac{\mathbb{P}(\mathcal{X}|\mathcal{A})\mathbb{P}(\mathcal{A})}{\mathbb{P}(\mathcal{X})} \quad (2.11)$$

A mesura que anem acumulant grans quantitats d'evidències, les nostres creences cada vegada seran més objectives i s'acabaran alineant amb els resultats obtinguts en inferència freqüentista. No obstant, quan el conjunt d'evidències és relativament petit, el més aconsellable és utilitzar inferència bayesiana, ja que el fet de retornar distribucions de probabilitat en lloc d'escalars, com és el cas dels freqüentistes, es conserva la incertesa que s'associa a un conjunt d'evidències petit.

2.3.1 Inferència bayesiana aplicada a l'anàlisi de dades

La inferència bayesiana té una gran quantitat d'aplicacions en diversos camps, ja sigui en medicina, intel·ligència artificial o en dret, però és particularment coneguda en el camp de l'anàlisi de dades. Donat un conjunt de dades, li associem un model probabilístic que dependrà d'una sèrie de paràmetres [6]. Per exemple, en el cas que les dades siguin de tipus discret, com és el cas de l'afluència de clients que veurem en la part pràctica, podem associar-li una distribució de Poisson, que depèn d'un únic paràmetre λ :

$$\mathbb{P}(D = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \mathbb{E}(D|\lambda) = \lambda \quad (2.12)$$

per a valors $k = 0, 1, 2$, etc., o en el cas de tractar-se d'una distribució de dades contínua, podem associar-li una distribució normal, $\mathcal{N}(\mu, \sigma)$ o una distribució exponencial,

$$f_D(\alpha) = \alpha e^{-\alpha d} \quad \mathbb{E}(D|\alpha) = \frac{1}{\alpha} \quad d \geq 0 \quad (2.13)$$

No obstant, no hi ha cap manera directa de determinar els valors d'aquests paràmetres a partir de les dades. El que es fa en inferència bayesiana és atribuir a aquests paràmetres una distribució *a priori* del que creiem que poden valdre. A mesura que anem extraient informació de les dades, aquesta distribució s'anirà actualitzant de manera que la nostra distribució *a posteriori* ens indicarà quins són els valors més probables de cada paràmetre.

Pel que fa a la manera com s'obté la distribució *a posteriori*, aquí és on entren els mètodes MCMC.

2.4 Què són els algorismes MCMC

Els mètodes MCMC són tècniques que, mitjançant la construcció d'una cadena de Markov \mathbb{X} -en la majoria de casos un camí aleatòri- dins d'un cert espai d'estats, ens permeten simular sistemes amb distribucions canòniques de Gibbs:

$$p(x) = \frac{1}{Z} e^{-U(x)/T} \quad (2.14)$$

Com ja s'ha dit anteriorment, qualsevol tipus de distribució de probabilitat es pot reescriure d'aquesta manera i, per tant, qualsevol distribució associada als paràmetres d'un cert model estadístic es pot interpretar com una distribució de Gibbs. Aquesta simulació s'utilitza per estimar l'esperança d'un cert funcional d'interès $f(\mathbb{X})$, $\mathbb{E}(f(\mathbb{X}))$, que en el cas d'inferència bayesiana sol ser l'esperança dels valors dels paràmetres del model [39].

Partint d'una certa distribució blanc, durant la simulació es construeix una cadena de Markov que té aquesta distribució blanc com a distribució invariant. Simultàniament amb la construcció del nostre camí aleatòri, es va calculant la mitjana dels valors $f(x_i)$ de N mostres x_i del camí aleatòri \mathbb{X} i s'estima el valor de l'esperança:

$$\mathbb{E}(f(\mathbb{X})) \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (2.15)$$

A més, el ritme al qual aquesta mitjana de valors tendeix cap a $\mathbb{E}(f(\mathbb{X}))$ coincideix amb el ritme al qual la distribució invariant del camí convergeix cap a la distribució blanc.

Per tant, un cop acabada la simulació tindrem, d'una banda, una distribució que s'aproxima a la nostra distribució blanc, i d'altra, un estimador de l'esperança d'un cert funcional d'interès. Com veurem més endavant, els mètodes MCMC són una gran eina per fer simulacions de distribucions de probabilitat posteriors en inferència probabilística [21].

2.4.1 Tipus d'algorismes MCMC

De manera general podem classificar els algorismes MCMC en dues categories, una formada per tots els algorismes basats en Metropolis-Hastings, i una altra on s'engloben tots els algorismes basats en el mostreig de Gibbs.

La característica principal de l'algorisme de Metropolis-Hastings és que la construcció del camí aleatòri ve determinada per un pas d'acceptació. És a dir, primer s'escull d'una certa distribució proposta un possible candidat per a ser el següent estat del camí i, a continuació, aquest s'accepta amb una certa probabilitat, avançant cap al següent estat del camí. Si pel contrari, el candidat és rebutjat, s'escull un altre candidat, repetint el procés.

D'aquesta manera sempre s'afavoreixen els candidats que tenen una probabilitat més elevada en la distribució blanc, mentre que es perjudica els estats que tenen menor probabilitat que l'estat actual. No obstant, requereix una certa preparació ja que s'ha d'escollir una bona distribució proposta que tingui un alt ràtio d'acceptació i de manera que sigui fàcil implementar el mostreig [12].

Algunes de les variants més destacables de Metropolis-Hastings són:

- Metropolis-Hastings-Green, també conegut com a *Reversible Jump*, que és una generalització de l'anterior a on es poden variar les dimensions.
- Hamilton-Monte Carlo, també anomenat Híbrid Monte Carlo i generalment abreujat HMC, on s'introdueix un vector moment i s'incorpora la dinàmica Hamiltoniana. No obstant, aquest últim no segueix un camí aleatòri, com veurem més endavant.

En el cas del mostreig de Gibbs, que tot i tenir la mateixa justificació en teoria de Markov que l'algorisme de Metropolis-Hastings, té un origen totalment diferent, el seu tret característic és que utilitza les distribucions condicionals completes durant la simulació amb l'objectiu de trobar la distribució conjunta. Aquest algorisme aprofita al màxim les propietats de la distribució conjunta i és més fàcil d'implementar per requerir menys preparació que Metropolis-Hastings [37].

2.4.2 Aplicació dels algorismes MCMC en inferència bayesiana

Tornant al cas que s'ha exposat en l'apartat 2.3.1, un cop definida la distribució *a priori*, apliquem un algorisme MCMC de manera que la distribució blanc d'aquest correspongui a la distribució *a posteriori* dels valors més probables del paràmetre en qüestió. És a dir, l'algorisme MCMC genera un camí aleatòri que convergeix cap a les zones on la probabilitat *a posteriori* és màxima.

És molt probable que a l'inici de l'algorisme, durant els primers passos del camí aleatòri, es passi per estats de molt baixa probabilitat en la distribució *a posteriori*, de manera que, habitualment, aquests primers passos del camí s'acostumen a *cremar*, i es comencen a agafar mostres més endavant, quan el camí entra en una zona amb probabilitats relativament altes. Aquestes mostres que ens retorna l'algorisme, són les que s'utilitzen per calcular els estimadors dels paràmetres, com en l'equació (2.15).

Per últim, cal dir que els mètodes MCMC no són els únics utilitzats per trobar les distribucions *a posteriori* d'aquest tipus de problemes. Junt amb MCMC, els mètodes més utilitzats són l'aproximació de Laplace i el variacional de Bayes, cadascun amb els seus punts forts i febles, i que s'escolliran segons el tipus de problema i les preferències de l'autor de l'estudi [6].

3 Origen i evolució dels algorismes MCMC

3.1 Context històric

Arran de l'esclat de la Segona Guerra Mundial (1939-1945) i del recent descobriment de la fissió nuclear en mans dels químics Otto Hahn i Fritz Strassmann al 1938, s'inicia als Estats Units el Projecte Manhattan, actiu entre el 1942 i el 1946, que té com a objectiu el desenvolupament d'armament nuclear. Tot i que hi contribuïen laboratoris d'arreu del país, tant la investigació principal com la construcció de les bombes es duu a terme al Laboratori Nacional de Los Álamos, situat a Nou Mèxic. Allí s'hi desenvolupen les bombes *Little Boy*, a base d'urani-235, i *Fat Man*, a base de plutoni, llençades a Hiroshima i Nagasaki respectivament a l'agost de 1945 i que desencadenen en el final de la guerra. No obstant, aquestes bombes són peces de laboratori de gran complexitat i poca fiabilitat. Així doncs, tot i haver posat fi al Projecte Manhattan, l'inici de la Guerra Freda pressiona als Estats Units per seguir investigant i desenvolupant bombes nuclears més simples, segures i eficaces, per tal de contrarestar les amenaces de la Unió Soviètica [15].

D'altra banda, és en aquesta mateixa època quan es comença a desenvolupar el primer ordinador a la Universitat de Pennsylvania amb l'objectiu d'agilitzar els farragosos càlculs amb la calculadora i elaborar sofisticades taules sobre el comportament de projectils per al Laboratori de Reserca Balística. Aquest primer ordinador, conegut com ENIAC, es posa en marxa el 15 de febrer de 1946 i, com veurem més endavant, té un paper fonamental en el desenvolupament del mètode Monte Carlo [37].

3.2 Origen del mètode de Monte Carlo Ordinari (MCO)

El descobriment d'aquest mètode recau en les mans dels matemàtics Stanislaw Ulam i John von Neumann, els quals treballaven a Los Álamos en el Projecte Manhattan. Una senzilla pregunta que va assaltar Ulam mentre jugava a cartes durant una convalescència va ser-ne el desencadenant: "Quina probabilitat hi ha de guanyar al solitari Canfield?". Després de fer una infinitat de càlculs combinatoris sense donar amb la resposta, es va preguntar si no hi hauria una altra manera més ràpida i senzilla d'estimar aquesta probabilitat, fent ús del recent estranat ENIAC. Immediatament ho va relacionar amb els problemes de difusió de neutrons en les bombes nuclears que intentava resoldre al laboratori. De seguida va comentar aquesta idea amb el seu company de recerca von Neumann i junts van idear un mètode basat en processos aleatòris i en el mostreig estadístic [8].

Al març de 1947, von Neumann envia una carta al cap de la Divisió Teòrica, Robert Richtmyer, explicant detalladament aquest mètode i la seva aplicació en el problema de la difusió de neutrons. Aquest mètode va ser batejat més tard amb el nom Monte Carlo, a proposta del seu company Nicholas Metropolis, fent referència a una anècdota sobre l'oncle d'Ulam, que sempre estava demanant diners als seus familiars per anar-se'ls a jugar al Casino Monte Carlo. En aquesta carta fa les

següents suposicions:

- Es té un nucli esfèric de material fissionable envoltat d'una capa de material resistent.
- Els neutrons es generen isotròpicament, amb un espectre de velocitats conegut, i totes les interaccions d'absorció, dispersió i fissió es poden implementar com una funció de la velocitat.
- S'ignoren la radiació i els efectes hidrodinàmics.

En cada etapa de la vida del neutró s'hauran de prendre una sèrie de decisions. Per començar, a $t = 0$ s'hauran d'assignar valors aleatoris a la posició i la velocitat del neutró en qüestió. Després s'hauran de decidir la posició i la naturalesa de les primeres col·lisions. En el cas de tractar-se d'una fissió, s'hauran de donar el nombre de neutrons resultants i fer el mateix procediment que amb els originals. En el cas d'una dispersió, haurem de calcular-ne els nous moments. D'aquesta manera s'elabora un historial geneològic de cada neutró. A base de repeticions s'acaba obtenint una imatge del seu comportament. Anàlegament, cada repetició del problema de difusió de neutrons es pot identificar amb una partida de Canfield, i els valors aleatòris associats a cada decisió corresponen als valors aleatòris de les cartes [8].

A més, von Neumann proposa els díigits de quadrat mig com a font de nombres pseudo-aleatòris uniformement distribuïts [44]. Aquest mètode consisteix en agafar un enter de n díigits i elevar-lo al quadrat. Si el nombre resultant té menys de $2n$ díigits, s'afegeixen zeros a l'esquerra fins a arribar a $2n$. A continuació s'agafen els n díigits centrals del nombre resultant i es repeteix el procés. D'aquesta manera s'obté una cadena d'enters de n díigits que s'acaba repetint en algun punt. A aquest conjunt d'enters se li associa una distribució no uniforme g que indica les probabilitats de les diferents possibilitats de cada decisió.

Al 1947 von Neumann exposa en una nova carta dues tècniques per obtenir distribucions no uniformes [8][32]. La primera d'elles, anomenada tècnica d'inversió, consisteix en aplicar la funció inversa $g^{-1} = f$ a una distribució uniforme en l'interval $(0, 1)$. Així, per exemple, per obtenir una exponencial decreixent $g(x)$, aplicariem la seva inversa $f(x) = -\ln x$ sobre la distribució uniforme. En la segona tècnica, anomenada d'acceptació-rebuig, es consideren dues distribucions uniformes i independents (x^i) i (y^i) i s'accepten els valors de x^i pels quals es compleix $y^i \leq f(x^i)$, on f és la densitat de la distribució desitjada g , $g(x) = \int f(x)d(x)$. Així doncs, amb aquesta tècnica s'acceptaran només aquells punts que quedin per sota de la gràfica de la densitat.

A finals de 1947, l'ENIAC es trasllada al Laboratori de Recerca Balística i es fa un test del mètode Monte Carlo sobre el problema dels neutrons [32]. S'executen nou problemes diferents corresponents a diferents distribucions inicials i configuracions de materials, els resultats dels quals evidencien la gran eficàcia del mètode i de seguida s'estén originant un gran nombre d'articles. Al 1949 s'organitzen dos simposis sobre el mètode Monte Carlo a Los Angeles i a Gainesville i, al 1952,

amb l'aparició del nou ordinador MANIAC a Los Álamos, es comencen a executar diferents aplicacions del mètode en diversos camps de la física i de la matemàtica experimental.

3.3 L'algorisme de Metropolis

Metropolis, junt amb la col·laboració d'Arianna i Marshall Rosenbluth i Augusta i Edward Teller, publica al 1953 l'article *Equation of State Calculations by Fast Computing Machines* [31], on es proposa el primer algorisme MCMC ideat per ser utilitzat amb el recent estrenat MANIAC.

En aquest article es descriu una modificació del ja conegut mètode MCO incorporant la simulació de camins aleatòris construïts a partir de passos d'acceptació basats en el canvi d'energies entre estats. Aquesta modificació permet tractar amb problemes de dimensions elevades, amb els quals el mètode MCO no acostuma a tenir bons resultats, com pot ser el cas de l'estudi de les propietats de qualsevol conjunt de molècules interactuant [37]. A continuació anem a comentar, doncs, l'algorisme descrit per aquest article [31].

Es considera un conjunt θ de N partícules dins d'una substància periòdica distribuïdes en un quadrat d' \mathbb{R}^2 i sigui \bar{d}_{ij} la distància mínima entre les partícules i i j en aquest conjunt θ . El primer punt de l'article es sobre la computació dels valors d'equilibri d'una funció qualsevol F :

$$\mathbb{E}(F(\theta)) = \frac{\int F(\theta) \exp(E(\theta)/KT) d\theta}{\int \exp(E(\theta)/KT) d\theta} \quad (3.1)$$

en \mathbb{R}^{2N} , on $\exp(E(\theta)/KT)$ és una distribució de Gibbs -sense normalitzar- d'energia $E(\theta)$ definida a partir del potencial V com

$$E(\theta) = \frac{1}{2} \sum_{i=1}^N \sum_{j=1, j \neq i}^N V(\bar{d}_{ij}) \quad (3.2)$$

Per evitar càlculs numèrics pesats, seria lògic implementar el mètode MCO, que assignaria a cada una de les N partícules una posició aleatòria de les $2N$ possibles posicions del quadrat. Com la probabilitat d'escollir una configuració amb un $\exp(E(\theta)/KT)$ petit és molt gran, Metropolis proposa que en lloc d'escollir les configuracions aleatòriament i ponderar-les amb $\exp(E(\theta)/KT)$, s'escullin configuracions amb probabilitats $\exp(E(\theta)/KT)$ grans i ponderar-les uniformement. Per implementar-ho, es col·loquen les N partícules en una configuració qualsevol i a continuació es mou una de les partícules segons el camí aleatòri següent:

$$\begin{aligned} X &\rightarrow X + \alpha \xi_1 \\ Y &\rightarrow Y + \alpha \xi_2 \end{aligned} \quad (3.3)$$

On α és el desplaçament màxim permès i $-1 \leq \xi_i \leq 1$ és aleatòri. Aquest moviment produirà un canvi en l'energia ΔE . Si aquest és negatiu, el moviment estarà

permès amb probabilitat 1, mentre que si és positiu, estarà permès amb probabilitat $\exp(\Delta E/KT)$. Aleshores, $\mathbb{E}(F(\theta))$ es pot aproximar com

$$\mathbb{E}(F(\theta)) = \frac{1}{M} \sum_{j=1}^M F_j \quad (3.4)$$

on F_j correspon al valor de F en el j -èssim moviment i M és el nombre de moviments. Per M prou grans, aquest mètode permetrà passar per qualsevol punt de l'espai de configuracions i, per tant, el mètode és ergòdic.

Per comprovar que aquest mètode escull configuracions amb probabilitat $\exp(E(\theta)/KT)$, considerem un conjunt molt gran de sistemes i definim v_r com el nombre de sistemes en una certa configuració r . Haurem de demostrar doncs que a mesura que augmentem els moviments, es té

$$v_r \propto \exp(-E(r)/KT) \quad (3.5)$$

Sigui \mathbb{P}_{rs} la probabilitat *a priori* de passar de la configuració r a una altra s , amb $\mathbb{P}_{rs} = \mathbb{P}_{sr}$. Suposem $E(r) > E(s)$. Llavors el nombre de moviments $r \rightarrow s$ serà $v_r \mathbb{P}_{rs}$, mentre que per a $s \rightarrow r$ seran $v_s \mathbb{P}_{sr} \exp\{-(E(r) - E(s))/KT\}$. El nombre total de moviments $r \rightarrow s$ vindrà donat per

$$\mathbb{P}_{rs}(v_r - v_s \exp\{-(E(r) - E(s))/KT\}) \quad (3.6)$$

i si

$$\frac{v_r}{v_s} > \frac{\exp(-E(r)/KT)}{\exp(-E(s)/KT)}, \quad (3.7)$$

tindrem més moviments $r \rightarrow s$ que no pas $s \rightarrow r$, com era d'esperar. Aquesta condició junt amb l'ergodicitat del mètode ens garanteix que el nostre conjunt tendirà cap a la distribució canònica.

A continuació, partint del Teorema del Virial $\langle V \rangle = -2\langle E_{kin} \rangle$, Metropolis obté l'equació d'estat d'un sistema d'esferes rígides en dues dimensions en funció de la densitat mitjana de partícules a la superfície d'una d'elles, $\bar{n} = \mathbb{E}(n)$,

$$PA = E_{kin} \left(1 + \frac{\pi d_0^2 \bar{n}}{2} \right) \equiv NKT \left(1 + \frac{\pi d_0^2 \bar{n}}{2} \right) \quad (3.8)$$

On P és la pressió, A l'àrea i $E_{kin} = Nm\bar{v}^2/2$ és l'energia cinètica. Així doncs, només calculant \bar{n} segons (3.4) podem determinar completament l'equació d'estat per esferes rígides.

Finalment, es presenten els resultats obtinguts pel cas d'un sistema de $N = 224$ partícules disposades en una xarxa trigonal 14×16 d'àrea A unitat i distàncies entre veïns $d = 1/14$. Definint A_0 com l'àrea que ocupen les partícules en empaquetament compacte i el paràmetre auxiliar $0 \leq \nu \leq 7$, escrivim

$$(A/A_0) = \frac{1}{\sqrt{3}d^2(1 - 2^{\nu-8})^2 N/2} \quad (3.9)$$

ν	(A/A_0)	X_1	X_2	X_3	(PA_0/NkT)
2	1.04269	49.17	47.35	9.77	48.11
4	1.14957	13.95	13.85	7.55	13.01
5	1.31966	6.43	6.72	5.35	5.63
5.5	1.4909	4.41	4.53	4.02	3.63
6	1.7962	2.929	2.939	2.680	2.187
6.25	2.04616	2.186	2.323	2.065	1.557
6.5	2.41751	1.486	1.802	1.514	1.028
7	4.04145	0.6766	0.990	0.667	0.4149

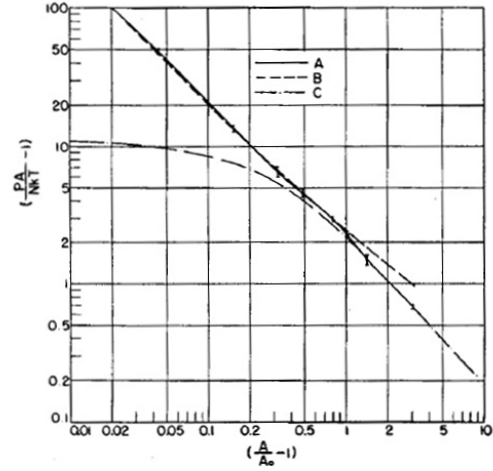


Figura 1: Resultats obtinguts per l'experiment realitzat a Metropolis (1953) per a un conjunt de $N=224$ partícules. La curva A representa els resultats obtinguts, la curva B correspon als resultats que s'obtidrien utilitzant teoria de volum lliure i la curva C els resultats pels quatre primers coeficients virials.

Menyspreant els setze primers cicles -cada cicle equival a moure cada una de les N partícules una vegada- de l'algorisme descrit per Metropolis, s'agafen entre 48 i 64 cicles per als valors

$$\nu = 0, 1, 2, 3, 4, 5, 5.5, 6, 6.25, 6.5, 7 \quad (3.10)$$

Cada cicle tarda en executar-se aproximadament tres minuts al MANIAC, i el conjunt dels cicles per cada un dels valors tarda entre quatre i cinc hores a completar-se. Amb totes les dades obtingudes, Metropolis elabora una gràfica logarítmica de $(PA/NKT) - 1$ en funció de $(A/A_0) - 1$ (Figura 1). Observem que els resultats són pràcticament idèntics que els de la teoria de volum lliure de Wood fins a $\nu = 6$, que correspon a $A/A_0 \simeq 1.8$. A partir de $A/A_0 = 2.5$, els resultats s'ajusten més a l'expansió de coeficients virial

$$\left(\frac{PA}{NKT} \right) - 1 = C_1(A_0/A) + C_2(A_0/A)^2 + C_3(A_0/A)^3 + C_4(A_0/A)^4 + 0(A_0/A)^5 \quad (3.11)$$

On els coeficients C_i s'obtenen per integració Monte Carlo.

Per tant, tot i que el nombre de passos de l'algorisme de Metropolis està limitat i té un temps d'execució molt gran -en total se simulen un màxim de 80 cicles per a $N=224$ que tarden quasi 5 hores- s'obtenen molt bons resultats en problemes on es tracten dimensions considerablement grans, que resulten impossibles de treballar amb MCO.

3.4 L'algorisme de Metropolis-Hastings

L'any 1970 Wilfred K. Hastings, amb la col·laboració del seu alumne Peter Peskun, publica l'article *Monte Carlo sampling methods using Markov chains and*

their application [18], on es presenta una generalització de l'algorisme descrit per Metropolis a on es treballa amb cadenes de Markov finites i reversibles i presenta probabilitats d'acceptació més genèriques. Aquest rep el nom d'algorisme de Metropolis-Hastings, MH, i és un dels mètodes MCMC més utilitzats avui en dia [37].

Donada una certa densitat de probabilitat *blanc* λ , computable fins a una certa constant de proporcionalitat, aquest algorisme construeix una cadena de Markov amb matriu de transició P tal que λ és la seva distribució invariant, és a dir, que $\lambda = \lambda P$. Aquesta contrucció progressiva cap a la distribució blanc ens permet agilitzar certs problemes on els mostreigs genèrics requereixen massa esforç. Prenent únicament cadenes de Markov reversibles, garantim que λ sigui invariant:

$$\lambda_i p_{ij} = \lambda_j p_{ji} \Rightarrow \sum_i \lambda_i p_{ij} = \sum_i \lambda_j p_{ji} = \lambda_j \sum_i p_{ji} = \lambda_j \quad (3.12)$$

S'assumeix que els p_{ij} són de la forma

$$p_{ij} = q_{ij} \alpha_{ij} \quad (i \neq j) \quad p_{ii} = 1 - \sum_{j \neq i} p_{ij} \quad (3.13)$$

on $Q = \{q_{ij}\}$ és una matriu de transició d'una cadena de Markov arbitrària, que anomenem *proposta*, i α_{ij} ve donada per

$$\alpha_{ij} = \frac{s_{ij}}{1 + \frac{\lambda_i q_{ij}}{\lambda_j q_{ji}}} \quad (3.14)$$

on $s_{ij} = s_{ji}$. En l'algorisme descrit per Hastings se segueixen, per a cada temps t , els passos següents:

- (i) **Pas de proposta:** Prenem $X(t) = i$ i escollim un cert estat j utilitzant la i -èsima fila de Q .
- (ii) **Pas d'acceptació:** El següent estat serà $X(t+1) = j$ amb probabilitat α_{ij} o $X(t+1) = i$ amb probabilitat $1 - \alpha_{ij}$, és a dir, α_{ij} representa la probabilitat d'acceptació del moviment $i \rightarrow j$. Durant la simulació només es considera el terme $(\lambda_j q_{ji} / \lambda_i q_{ij})$.

En general, prenem

$$s_{ij} = g \left[\min \left\{ \frac{\lambda_i q_{ij}}{\lambda_j q_{ji}}, \frac{\lambda_j q_{ji}}{\lambda_i q_{ij}} \right\} \right] \quad (3.15)$$

on $g(x)$ és tal que $0 \leq g(x) \leq 1 + x$ per $0 \leq x \leq 1$. Pel cas concret de l'algorisme de Metropolis tenim

$$s_{ij}^{(M)} = \begin{cases} 1 + \frac{\lambda_i q_{ij}}{\lambda_j q_{ji}} & \text{si } \frac{\lambda_j}{\lambda_i} \geq 1 \\ 1 + \frac{\lambda_j q_{ji}}{\lambda_i q_{ij}} & \text{si } \frac{\lambda_j}{\lambda_i} < 1 \end{cases} \xrightarrow{q_{ij}=q_{ji}} \alpha_{ij}^{(M)} = \begin{cases} 1 & \text{si } \frac{\lambda_j}{\lambda_i} \geq 1 \\ \frac{\lambda_j}{\lambda_i} & \text{si } \frac{\lambda_j}{\lambda_i} < 1 \end{cases} \quad (3.16)$$

Tant a l'hora d'escollir Q com després de cada simulació, el ritme de rebuigs, que està determinat per la proporció de temps t pels quals $X(t+1) = X(t)$, s'ha de mantenir en valors baixos, ja que valors alts indicarien una mala elecció de Q o un error en el programa.

Per tant, l'algorisme MH no és més que una generalització de l'algorisme de Metropolis, on també es tracten problemes de grans dimensions a diferència de MCO, i que permet modificar la probabilitat d'acceptació a la nostra conveniència, englobant diversos algorismes com el de Metropolis, o com veurem més endavant, el mostreig de Gibbs.

3.4.1 Mostreig preferencial

En aquest mateix article [18] també s'explica la implementació del mostreig preferencial (en anglès conegut com *importance sampling*) amb els mètodes MCMC. La idea general d'aquest és estimar les propietats d'una certa distribució d'interès a partir de mostres extretes d'una altra distribució d'on resulta més fàcil generar-les. Suposem que per a una certa funció f definida en tots els estats d'una certa cadena \mathbb{X} , el mostreig preferencial busca un estimador de

$$\mathbb{E}_\lambda(f) = \frac{\sum_{i=0}^S f(i)\lambda_i}{\sum_{i=0}^S \lambda_i} \quad (3.17)$$

que es pot reescriure en funció d'una nova distribució λ'

$$\frac{\sum_{i=0}^S \{f(i)\lambda_i/\lambda'_i\}\lambda'_i / \sum_{i=0}^S \lambda'_i}{\sum_{i=0}^S (\lambda_i/\lambda'_i)\lambda'_i / \sum_{i=0}^S \lambda'_i} = \frac{E_{\lambda'}\{f(i)\lambda_i/\lambda'_i\}}{E_{\lambda'}(\lambda_i/\lambda'_i)} \quad (3.18)$$

sempre i quan el mostreig des de λ' sigui més senzill que des de λ . L'estimador que busquem ve donat llavors per

$$\hat{I} = \frac{\sum_{t=1}^N [f\{X(t)\}\lambda_{\mathbb{X}}/\lambda'_{\mathbb{X}}]/N}{\sum_{t=1}^N \{\lambda_{\mathbb{X}}/\lambda'_{\mathbb{X}}\}/N} \quad (3.19)$$

Si al mostreig preferencial hi afegim els mètodes MCMC, ens podem estalviar el pas de la construcció de λ' . Podem construir una cadena de Markov P tal que tingui λ com a distribució estacionària i prendre $q_{ij} = \lambda'_j$. D'aquesta manera podem estimar directament $\mathbb{E}_\lambda(f)$ sense haver de recórrer als pesos $\lambda_{X(t)}/\lambda'_{X(t)}$.

3.5 Mostreig de Gibbs

Un altre exemple d'algorisme MCMC, que es pot considerar un cas particular de MH però que té un origen totalment diferent, és el mostreig de Gibbs, publicat a mitjans dels anys 80 pels germans Geman.

Donat un cert conjunt de variables aleatòries $\mathbb{X} = \{X_s : s \in \mathcal{S}\}$, indexat per un conjunt de posicions \mathcal{S} de dimensió N , la idea general del mostreig de Gibbs consisteix en generar mostres de la distribució posterior, que és la distribució conjunta,

a partir de les distribucions condicionals completes de cada una de les variables [45]. És a dir, inicialitzant les variables a partir d'una certa distribució *a priori*, $\mathbf{X}(0) = \{X_s(0) = x_s : s \in \mathcal{S}\}$, a cada temps $t \in \mathcal{T}$ agafem una mostra d'una d'aquestes variables a partir de la seva distribució condicional:

$$x_{s_i}^{(t)} \sim \mathbb{P}(X_{s_i} = x_{s_i} | X_{s_1} = x_{s_1}^{(t)}, \dots, X_{s_{i-1}} = x_{s_{i-1}}^{(t)}, X_{s_{i+1}} = x_{s_{i+1}}^{(t-1)}, \dots, X_{s_N} = x_{s_N}^{(t-1)}) \quad (3.20)$$

Després d'un gran nombre d'iteracions, la distribució estacionària generada per aquestes mostres s'aproxima a la distribució blanc, que és la distribució conjunta. En aquest tipus de mostreig es busca calcular l'estimador *Maximum a Posteriori* (MAP), és a dir, busca la configuració $\omega = (x_{s_1}, \dots, x_{s_N})$ per la qual es maximitza la probabilitat conjunta:

$$\mathbb{P}(\mathbf{X} = \omega) = \mathbb{P}(X_{s_1} = x_{s_1}, \dots, X_{s_N} = x_{s_N}) \quad (3.21)$$

El mostreig de Gibbs és, doncs, un cas particular de MH on el pas d'acceptació és 1 -sempre s'accepten els nous candidats- i els nous estats s'obtenen a partir de les distribucions condicionals de l'equació (3.20). El mostreig de Gibbs requereix menys preparació que l'algorisme MH i, sempre i quan es coneixin totes les distribucions condicionals completes, acostuma a tenir temps d'execució inferiors, sent així més pràctic. No obstant, no sempre es coneixen les distribucions condicionals i, per tant, si només es té la distribució conjunta del model que es vol estudiar, és molt més pràctic aplicar directament MH que no pas derivar totes les condicionals analíticament.

3.5.1 Article dels germans Geman

En l'article de 1984 *Stochastic relaxation, Gibbs distributions and the bayesian restoration of images* dels germans Stuart i Donald Geman, es presenta per primera vegada el mostreig de Gibbs amb l'objectiu d'aplicar-lo a un problema de creixent interès en aquella època: la restauració d'imatges malmeses [37]. Anem a veure el plantejament d'aquest problema en termes de distribucions i com es resol amb el mostreig de Gibbs [10].

Aquest mètode consisteix en, donada una certa imatge degradada, aplicar una relaxació estocàstica junt amb un paràmetre de temperatura que disminueix gradualment, la qual genera una sèrie d'imatges que convergeix cap a l'estimador MAP de la seva imatge original. Aquest procés es descriu a partir dels següents passos:

- (i) Es fa un canvi local aleatòri en la imatge degradada basat en els valors dels pixels veïns, generat pel mostreig d'una distribució de probabilitat condicional local.
- (ii) S'introdueix un nou paràmetre de control global anomenat *temperatura*, T . Per a valors de T baixos, les distribucions condicionals locals es concentren en estats que augmenten la distribució posterior i generen una imatge regular, mentre que per a valors alts la distribució és uniforme i generen imatges caòtiques.

- (iii) *Annealing*. Es prenen inicialment temperatures altes on alguns dels canvis poden disminuir la distribució posterior. A mesura que avança la relaxació, es va disminuint gradualment la temperatura i el procés adopta el comportament d'una millora iterativa, és a dir, utilitzant les solucions aproximades d'un cert pas de la iteració, per calcular les del següent.

En primer lloc, es considera el conjunt de posicions $\mathcal{S} = \mathcal{Z}_m \cup \mathcal{D}_m$, on $\mathcal{Z}_m = \{(i, j) : 0 \leq i, j \leq m\}$ representa una xarxa d'enters $m \times m$ que corresponen a la xarxa de pixels, i \mathcal{D}_m és la seva matriu dual, és a dir, la xarxa formada pels llocs situats a la meitat de cada parella de pixels de \mathcal{Z}_m . Sigui $\mathcal{G} = \{\mathcal{G}_s : s \in \mathcal{S}\}$ un sistema de veïnats de \mathcal{S} (definició 2.12).

Es representa la imatge original, la que volem recuperar, com una parella de variables $\mathbb{X} = (\mathbb{F}, \mathbb{L})$, on \mathbb{F} és la matriu de intensitats de gris observables i \mathbb{L} és la matriu dual dels elements de vora no observables, coneguts com elements de línia. Aleshores, la imatge degradada serà de la forma $\mathbb{G} = \phi(H(\mathbb{F})) \odot \mathbb{N}$, on H és la matriu borrosa, ϕ una possible transformació no lineal, \mathbb{N} la matriu de soroll provinent del sensor i \odot una operació invertible.

Donada doncs, una certa imatge degradada \mathbb{G} , es busca la distribució posterior $\mathbb{P}(\mathbb{F} = f, \mathbb{L} = l | \mathbb{G} = g)$ de la imatge original. Suposem que \mathbb{X} és un camp aleatòri de Markov respecte $\{\mathcal{S}, \mathcal{G}\}$. Pel teorema 2.15 tenim que

$$\mathbb{P}(\mathbb{F} = f, \mathbb{L} = l) = \frac{1}{Z} e^{-U(f,l)/T}. \quad (3.22)$$

Utilitzant el mostreig de Gibbs, es va creant un procés estocàstic que va fent petits canvis locals en funció dels valors dels pixels veïns fins que convergeix cap a l'estimador MAP de la imatge original, el qual es donarà en configuracions per les quals es minimitzi l'energia de l'equació 3.22, $U(f, l)$. A mesura que anem disminuint T , com s'explica en el pas (iii), els valors pels quals l'energia és mínima cada vegada seran més fàcils de trobar i, per tant, s'obtindrà una imatge més regular.

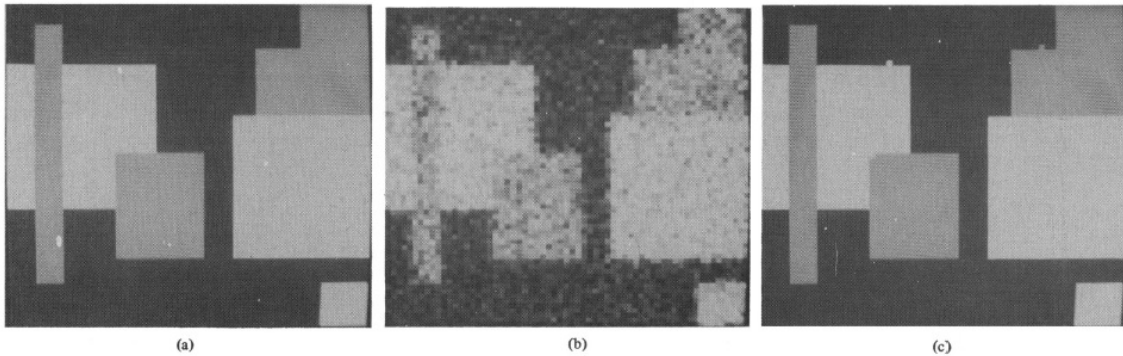


Figura 2: Resultats obtinguts en un dels experiments realitzats a l'article Geman i Geman (1984)[10]. La imatge (a) correspon a la imatge original, a (b) es té la imatge degradada, borrosa, on hi actua una transformació no lineal i soroll multiplicatiu. Per últim a (c) s'observa la imatge restaurada després de 1000 iteracions del mostreig de Gibbs .

3.6 Implementació de la dinàmica hamiltoniana en MCMC

Al 1987 Duane, Kennedy, Pendleton i Roweth publiquen l'article *Hybrid Monte Carlo* on s'uneix, per primera vegada, els algorismes MCMC amb dinàmica hamiltoniana, amb l'objectiu d'aplicar-ho en teoria de camps reticulats i, més concretament, en cromodinàmica quàntica, camp centrat en l'estudi de l'estructura de neutrons i protons dins dels nuclis dels àtoms [35]. Tot i que inicialment els autors van anomenar aquest nou algorisme Hybrid Monte Carlo, a Mackay (2003) se'l rebateja com Hamilton Monte Carlo (HMC) i des d'aleshores se'l coneix com a tal.

A continuació es presenta una breu introducció a la dinàmica de Hamilton i de quina manera s'implementa per donar lloc aquest nou algorisme MCMC [35].

3.6.1 Dinàmica hamiltoniana

Considerem un cert objecte que es mou per una superfície de dimensió d i d'alçada variable, amb una matriu de massa M diagonal d'elements m_1, \dots, m_d . La posició d'aquest objecte vindrà donada per un cert vector q de dimensió d , i tindrà un moment associat $p = mv = m\dot{q}$. L'energia de l'objecte és la suma de l'energia potencial, $U(q)$, que depèn de l'alçada, i l'energia cinètica, $K(p) = \sum_{i=1}^d \frac{p_i^2}{2m_i}$, que depèn de la velocitat.

$$H(q, p) = U(q) + K(p) \quad (3.23)$$

Suposant que no hi ha fregament ni interacció de cap tipus, l'objecte conservarà l'energia, de manera que dins d'una mateixa alçada l'objecte sempre tindrà la mateixa velocitat, i si augmentem l'alçada l'energia cinètica s'anirà convertint en energia potencial fins que l'objecte s'aturi i torni a baixar superfície avall.

Les equacions de moviment del sistema, és a dir, la manera com varien q i p al llarg del temps, estan determinades per les equacions de Hamilton de cada coordenada:

$$\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i} \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} \quad (3.24)$$

Donat que la seqüència d'estats que es construeix en qualsevol algorisme MCMC té associat un temps discret, haurem de discretitzar les equacions 3.24 utilitzant una certa mida de pas ϵ , que serà un dels paràmetres de HMC, de manera que canviarem d'estat a temps $\epsilon, 2\epsilon, 3\epsilon$, etc. Una de les maneres més habituals per fer-ho és amb el mètode de salt de granota, que reescriu les equacions de Hamilton com:

$$\begin{aligned} p_i(t + \frac{\epsilon}{2}) &= p_i(t) - \frac{\epsilon}{2} \frac{\partial U}{\partial q_i}(q(t)) \\ q_i(t + \epsilon) &= q_i(t) + \epsilon \frac{p_i(t + \frac{\epsilon}{2})}{m_i} \\ p_i(t + \epsilon) &= p_i(t + \frac{\epsilon}{2}) - \frac{\epsilon}{2} \frac{\partial U}{\partial q_i}(q(t + \epsilon)) \end{aligned} \quad (3.25)$$

i consisteix en la repetició dels següents passos:

- (i) S'avança mig pas amb les variables de moment utilitzant la primera eq. (3.25).
- (ii) S'avança un pas per les variables de posició segons la segona equació, utilitzant les variables de moment calculades en el pas anterior.
- (iii) S'avança mig pas més les variables de moment utilitzant les variables de posició del pas anterior.

No obstant, a excepció de la primera i la última iteració del mètode, el tercer i el primer pas de la següent iteració es poden ajuntar, de manera que una vegada s'ha fet el pas per les variables de posició, podem fer un pas complet per les variables de moment utilitzant la tercera equació de 3.25.

En el cas de tractar-se d'un sistema físic, la interpretació és immediata ja que se simulen les posicions i els moments de l'objecte en estudi, i se li associen una energia potencial i una energia cinètica. No obstant, en casos no relacionats amb la física, com pot ser l'estudi d'un cert model probabilístic, les nostres variables d'interès correspondran a les variables de posició, mentre que les variables de moment són només una eina introduïda per poder utilitzar la dinàmica hamiltoniana, les quals no tenen cap relació directa amb cap variable del model estudiat. A continuació s'explica com s'implementa tota aquesta dinàmica hamiltoniana amb els algorismes MCMC.

3.6.2 Algorisme de Hamilton Monte Carlo

De la mateixa manera que es fa en els altres algorismes MCMC, l'objectiu de HMC és simular una distribució de Gibbs canònica que s'ajusti a la distribució que es busca [35]. En aquest cas, però, l'energia vindrà donada pel hamiltonià escrit en l'equació (3.23):

$$\mathbb{P}(q, p) = \frac{1}{Z} e^{-H(q,p)/T} = \frac{1}{Z} e^{-U(q)/T} e^{-K(p)/T} \quad (3.26)$$

Notem que, com l'energia cinètica és quadràtica, la distribució de p es correspondrà amb un producte de gaussianes de mitjana zero i variància m_i . Tot seguit s'exposa l'algorisme HMC, que consta de dos passos que es repeteixen fins a obtenir una mostra d'estats suficientment gran:

- (i) S'assigna un valor aleatòri a les variables de moment d'acord amb la seves distribucions normals, independentment dels valors de les variables de posició.
- (ii) Partint de l'estat actual (q, p) , se simulen L passos de mida ϵ del mètode de salt de granota, arribant a un nou estat candidat (q^*, p^*) , que s'acceptarà amb probabilitat:

$$\min[1, e^{H(q,p)-H(q^*,p^*)}] = \min[1, e^{U(q)-U(q^*)} e^{K(p)-K(p^*)}] \quad (3.27)$$

Si, pel contrari, el candidat és rebutjat, aleshores s'agafa com a nou estat l'estat actual (q, p) i es computa dues vegades en l'estimador de l'esperança del funcional d'interès, que apareix en l'equació (2.15).

El fet que a cada iteració de l'algorisme s'assigni un nou valor aleatòri a les variables de moment evita que la simulació quedi estancada en algun mínim local i, per tant, no pugui retornar una mostra representativa de l'espai d'estats. No obstant, per tal d'obtenir bons resultats, s'hauran d'afinar els valors dels dos paràmetres L i ϵ de què depèn l'algorisme HMC. Cal tenir en compte que aquest algorisme només s'utilitza per a distribucions contínues en \mathbb{R}^d i, donat que implementa el gradient d'energia, s'ha de complir l'existència d'aquest i la possibilitat de calcular-lo en un temps raonable.

A diferència de tots els algorismes MCMC que s'han vist fins ara, HMC no construeix camins aleatoris i és per aquesta raó que la distància entre estats successius sol ser molt més gran que en la resta d'algorismes, permetent així obtenir mostres representatives amb moltes menys iteracions que amb qualsevol altre mètode MCMC [39]. El segon avantatge que trobem en HMC és que la probabilitat d'acceptació és elevada, permetent obtenir més mostres diferents que en un altre algorisme en el mateix nombre d'iteracions. Per tant, en la majoria de casos HMC és més eficient i ràpid que la resta, sempre i quan es compleixin les condicions necessàries per poder-lo aplicar.

3.7 Revolució dels anys 90

Durant la dècada dels anys 80 van arribar al mercat els ordinadors personals, els quals van tenir gran popularitat i, a diferència de l'època en què va néixer l'algorisme de Metropolis, pràcticament qualsevol persona podia tenir accés a un ordinador. Aquesta revolució en el camp de la informàtica, junt amb tots els articles que s'havien anat publicant sobre els mètodes MCMC, van provocar al seu torn una revolució dins de l'estadística a l'inici dels anys 90 [37]. El principal desencadenant va ser, però, l'article de Gelfand i Smith (1990) [9] on s'introduïa per primera vegada els mètodes MCMC dins del camp de l'estadística, demostrant especialment la seva gran utilitat en estadística bayesiana.

3.7.1 Article de Gelfand i Smith

El juny de 1990 Alan E. Gelfand i Adrian F.M. Smith publiquen a *Journal of the American Statistical Association* l'article *Sampling-Based Approaches to Calculating Marginal Densities* [9]. En aquest article s'utilitzen fins a tres algorismes MCMC diferents amb l'objectiu de calcular estimadors numèrics de distribucions de probabilitat marginals. Més concretament, els algorismes utilitzats són el mostreig de Gibbs, ja comentat en la secció anterior, l'algorisme de substitució, de Tanner i Wong (1987) [42], i el mostreig preferencial de Rubin (1988) [40], que és una modificació de l'algorisme introduït per Hastings (1970) [18].

En qualsevol dels tres casos, partint d'una família de variables aleatòries X_1, \dots, X_k , suposem que tenim una de les següents informacions:

- (a) Es coneixen totes les distribucions condicionals de la forma $\mathbb{P}(X_i|X_j, i \neq j)$.

- (b) Es coneix la forma funcional de la distribució conjunta $\mathbb{P}(X_1, \dots, X_k)$ i com a mínim una distribució condicional del tipus $\mathbb{P}(X_i|X_j, i \neq j)$.

L'objectiu és, doncs, a partir de (a), en els casos de mostreig de Gibbs i de substitució, o de (b), en el cas del mostreig preferencial, calcular estimadors de les marginals $\mathbb{P}(X_i)$. Donat que el mostreig de Gibbs és l'algorisme amb més repercussió i ja se n'ha explicat el funcionament, en aquest document només reproduïrem l'estimació per a aquest algorisme.

Com ja s'ha vist en l'apartat 3.5, el mostreig de Gibbs genera un esquema markovià de la següent manera: donat un conjunt de variables aleatòries inicials $X_1^{(0)}, \dots, X_k^{(0)}$, es van agafant valors de $X_i^{(t)}$ tal com indica l'equació 3.20, variant una única variable en cada iteració t . D'aquesta manera després d'un nombre τ d'iteracions s'obté un nou conjunt $X_1^\tau, \dots, X_k^\tau$. Generant m rèpliques d'aquestes τ iteracions s'obtenen m k -plets de la forma $(X_{1l}^\tau, \dots, X_{kl}^\tau)$ per $l = 1, \dots, m$. Aleshores, escrivint l'estimador d'una certa variable aleatòria a la iteració τ com la integració de Monte Carlo, s'obté que la distribució marginal d'aquesta serà:

$$\mathbb{P}(\hat{X}_i)_\tau = \mathbb{E}(\mathbb{P}(X_i)_\tau) = \frac{1}{m} \sum_{l=1}^m \mathbb{P}(X_i|X_j = X_{jl}^\tau, j \neq i) \quad (3.28)$$

L'avantatge d'utilitzar el mostreig de Gibbs en lloc de l'algorisme de substitució és que es requereix menys informació, ja que només es necessiten els k condicionals complets, mentre que per substitució es requereixen $k(k-1)$ condicionals, entre ells condicionals parcials que s'hauran de calcular a partir dels complets. No obstant, si es disposa de condicionals parcials ja d'entrada, que no sempre és possible, l'algorisme de substitució, que només necessita τ iteracions, convergirà més ràpid que no pas Gibbs, que n'ha de fer $\tau(k-1)$.

3.7.2 Conseqüències de l'article de Gelfand i Smith

Arran de la introducció dels mètodes MCMC en el món de l'estadística, que en certa manera ja s'havia començat a introduir a Geman i Geman (1984)[10] i a Tanner i Wong (1987) [42] però sense tanta repercussió com a Gelfand i Smith (1990) [9], es va desencadenar la publicació massiva d'articles i treballs al llarg dels anys 90 en diferents camps.

En primer lloc, durant aquest període es publiquen un gran nombre d'articles que contribueixen a aprofundir en la teoria de MCMC [37]. Una petita mostra d'aquestes publicacions serien Tierney (1994) [43], on s'estudien detalladament les cadenes de Markov i les seves propietats de convergència; Liu, Wong i Kong (1995) [27], on s'estudia la covariància del mostreig de Gibbs; Mengersen i Tweedie (1996) [30], on s'estudia la velocitat de convergència dels algorismes MCMC; o Hobert i Casella (1996) [20], on es determinen les condicions necessàries per a que el mostreig de Gibbs convergeixi, entre molts d'altres. A més a més, a principis dels 90 s'organitzen un gran nombre de conferències i congressos sobre aquesta temàtica.

En segon lloc, sorgeixen els primers llenguatges de programació probabilístics, llenguatges orientats a treballar amb models estadístics i que habitualment ja incorporaven algorismes MCMC per treballar més còmodament. Tot i que la majoria derivaven de llenguatges de programació ja existents, en alguns casos, com en BUGS, es creaven llenguatges completament nous destinats únicament a aquesta finalitat.

En tercer lloc, gràcies la gran popularitat dels algorismes MCMC en aquella època i a la recent creació de llenguatges de programació probabilístics, la comunitat científica es va adonar del gran potencial que tenien el mostreig de Gibbs i MH, que podien resoldre pràcticament qualsevol problema computacional que anteriorment havia suposat un malson pels programadors, i de manera senzilla i ràpida. Això va provocar una allau d'articles sobre la gran multitud de possibilitats que tenien aquests algorismes, i no només en el camp de l'estadística, sinò que es va estendre a altres camps on encara es desconeixien. Alguns exemples d'aquests articles són Churchill (1995) [4], que utilitza algorismes MCMC per a la seqüenciació d'ADN humà, o George i McCulloch (1993) [11], on s'utilitza el mostreig de Gibbs per les variables amb més probabilitat dins d'un model de regressió múltiple. També destaca l'article de Neal (1996) [34], on s'utilitza per primera vegada HMC dins del camp de l'estadística, més concretament, en l'estudi de models de xarxes neuronals.

Per últim, se segueixen publicant generalitzacions i modificacions d'algorismes ja coneguts, com el cas de el mostreig per rebuig adaptatiu aplicat al mostreig de Gibbs descrit per Gilks i Wild (1992) [14], però que es pot aplicar també a MH, o de Metropolis-Hastings-Green (MHG), generalització de MH descrita a Green (1995) [17]. Moltes de les aplicacions anteriors van seguir evolucionant a arrel del mostreig per rebuig adaptatiu.

A continuació es comenten breument aquests dos articles.

3.7.3 El mostreig per rebuig adaptatiu de Gilks i Wild

L'any 1992, W.R. Gilks i P.Wild publiquen l'article *Adaptive Rejection Sampling for Gibbs Sampling* [14] a on introdueixen un nou mètode de mostreig per rebuig. En general, un mostreig per rebuig consisteix en, donada una certa densitat de probabilitat $f(x)$, escollir una funció envolvent adequada $g_s(x)$, que actua com a límit superior, i agafar mostres sota la restricció $f(x) < M g_s(x)$ per una certa $M > 1$.

En el cas del mostreig per rebuig adaptatiu, a més a més s'afegeix una certa funció d'ajustament $g_i(x)$, que actua com a límit inferior, i després de cada rebuig s'actualitzen les funcions $g_s(x)$ i $g_i(x)$ de manera que aquestes convergeixen cap a la densitat $f(x)$ a mesura que avança el mostreig.

Aquest tipus de mostreig es pot utilitzar per a qualsevol densitat logarímicament còncaua d'una sola variable i és especialment útil en els casos de models no conjugats, que acostumen a donar problemes en el mostreig de Gibbs. Recordem que un model és conjugat si les distribucions *a priori* i *a posteriori* pertanyen a la mateixa família de distribucions de probabilitats, és a dir, si la *a priori* és una distribució normal, la *a posteriori* també ho serà. De forma general, totes les densitats que s'acostumen

a utilitzar en mostreig de Gibbs són logàrítmicament còncaves respecte algun dels seus paràmetres o alguna transformació d'aquests.

3.7.4 El salt reversible de Green

Peter J. Green publica l'any 1995 l'article *Reversible Jump Markov chain Monte Carlo computation and Bayesian model determination* [17] a on es proposa per primera vegada la construcció de cadenes de Markov reversibles que salten entre subespais de paràmetres de diferents dimensions. Aquest nou mètode es pot interpretar com una extensió de l'algorisme MH per a espais de paràmetres de dimensió variable i de fet, també rep el nom de Metropolis-Hastings-Green. A continuació s'explica breument el seu funcionament.

Sigui $\{\mathcal{X}_i\}_{i \in I}$ un conjunt d'espais d'estats disjunts i de diferents dimensions, es defineixen l'espai d'estats de la cadena de Markov i l'espai d'estats argumentat com:

$$\mathcal{X} = \bigcup_{i \in I} \mathcal{X}_i \quad \mathcal{Z} = \bigcup_{i \in I} \mathcal{Z}_i \quad (3.29)$$

on cada \mathcal{Z}_i ve donat pel producte $\mathcal{X}_i \times \mathcal{Y}_i$, per algun cert subespai d'estats \mathcal{Y}_i de manera que \mathcal{Z}_i sigui de la mateixa dimensió per a tot $i \in I$. Aleshores, donada una certa transformació $g : \mathcal{Z} \rightarrow \mathcal{Z}$, que té g com a inversa, per a cada $i \in I$ i cada $x \in \mathcal{X}_i$ es genera una certa variable aleatòria $y \in \mathcal{Y}_i$ i s'aplica la transformació, de manera que per cada parella $(x, y) \in \mathcal{Z}$ es genera una altra parella $(x^*, y^*) = g(x, y)$ [12]. Amb aquestes dues parelles es calcula el ràtio de Green, que depèn del gradient de g , i que determinarà amb quina probabilitat s'accepta o es rebutja el nou estat x^* .

4 Aplicacions dels algorismes MCMC

Per últim, abans d'entrar en la part pràctica del treball, s'exposen alguns exemples de la gran diversitat d'aplicacions dels algorismes MCMC que s'han anat publicant en una gran varietat de camps des del seu auge durant els anys 90. Cal tenir en compte que pràcticament en qualsevol branca de la ciència es pot aplicar, d'una manera o d'una altra, els mètodes MCMC en una gran diversitat de problemes de diferent naturalesa. El que ve a continuació només en representa una petita mostra.

Estadística

Un dels camps on més rellevància tenen els algorismes MCMC és, sense cap dubte, en estadística, i en especial, en estadística bayesiana. La conjunció d'aquests algorismes amb la inferència bayesiana és el centre de molts estudis que es donen en altres rams de la ciència, on s'utilitzen els primers per obtenir els resultats desitjats i el segon per interpretar-los.

Com ja s'ha dit anteriorment, l'elaboració d'un model probabilístic associat a un conjunt de dades i la posterior simulació amb un algorisme MCMC, permet extreure'n conclusions molt interessants, i són aplicables a una gran diversitat d'estudis, ja sigui a nivell científic, empresarial o, fins i tot, judicial.

Una altra de les seves principals aplicacions és, per exemple, en filtres de partícules, que són mètodes de simulació aplicats a conjunts de dades que es van recopilant progressivament en el temps [37]. Aquests mètodes utilitzen mostres amb pesos, els quals es van actualitzant a mesura que s'obtenen noves dades, i donen lloc a aproximacions Monte Carlo de la distribució *a posteriori*. Aquest procediment també es coneix amb el nom de Monte Carlo Seqüencial, descrit per primera vegada a Liu i Wong (1994) i posat en pràctica a Doucet, de Freitas i Gordon (2001).

Aplicacions en altres camps de les matemàtiques

- **Equacions en derivades parcials.** L'estudi del comportament de solucions d'equacions diferencials a llarg termini està estretament relacionat amb la teoria de cadenes de Markov i, a més, s'utilitzen algorismes MCMC per a provar certes desigualtats. De fet, a Zegarliniski (1992) [46], per exemple, s'utilitza el mostreig de Gibbs per demostrar la desigualtat logarítmica de Sobolev.
- **Geometria algebraica.** Una altra aplicació dins del camp de les matemàtiques és la resolució de certs problemes computacionals de geometria algebraica [7]. Per exemple, considerem el cas en que es vol generar, de manera aleatòria, un conjunt de matrius de coeficients enters no negatius tals que totes comparteixen les mateixes sumes per files i per columnes. Aleshores, donada una matriu inicial, s'aplica un algorisme MCMC de manera que es van construint noves matrius, que s'accepten o rebutjen segons si compleixen els requisits o no. Per exemple, en el cas des matrius 2×2 , donada una certa matriu inicial,

se li suma una de les següents matrius:

$$\begin{pmatrix} +1 & -1 \\ -1 & +1 \end{pmatrix} \text{ o } \begin{pmatrix} -1 & +1 \\ +1 & -1 \end{pmatrix} \quad (4.1)$$

En cas que tots els coeficients siguin sent enters no negatius, la nova matriu s'accepta i es repeteix el procediment. Si, per contra, no ho compleix, es retorna a la matriu actual. Aquest procés és especialment útil per matrius de major dimensió.

- **Teoria de grups computacional.** Tot i que en menor mesura, s'acostumen a utilitzar certs algorismes, entre ells MCMC, per tal de generar diferents tipus de grups [3]: finits, policíclics, resolubles, de permutació, etc.
- **Criptografia.** Els mètodes MCMC també resulten útils a l'hora de desxifrar missatges encriptats. Com es pot veure a [7], Marc Coram i Phil Beineke, en aquella època estudiants de la Universitat de Stanford, desxifren amb èxit un missatge codificant suposant que cada caràcter del codi representa una lletra, un signe de puntuació o un espai en blanc. A aquesta relació se li associa una certa funció de prova de manera que, durant la simulació per MCMC, es va modificant fins a arribar a la funció blanc que relaciona cada caràcter del codi amb cada caràcter del text desxifrat, permetent així desxifrar-lo amb només 2000 passos.

Física

Des dels seus inicis, moltes de les grans transformacions que han sofert alguns dels algorismes MCMC han estat orientats en l'estudi de problemes d'aquesta temàtica. Per començar, l'algorisme ordinari de Monte Carlo disenyat per Ulam i von Neumann tenia per objectiu simular la difusió de neutrons, mentre que Metropolis desenvolupa el seu algorisme amb l'objectiu d'estudiar el comportament d'una substància formada per molècules que interactuen entre sí. Un altre clar exemple és l'algorisme HMC, una part dels fonaments del qual està basat en conceptes físics i el qual va ser ideat per estudiar del comportament dels gluons i quarks que constitueixen els protons i neutrons.

En els últims anys han anat sorgint altres aplicacions com per exemple en el camp de l'astronomia [41], on destaquen alguns articles com Christensen i Meyer (1998) [5], on apliquen MCMC en l'estudi de radiació gravitacional, o Lewis i Bridle (2002) [25], on se centren en l'estudi de radiació còsmica de fons. A Heckman, Bernstein i Vigoda (2017) [19] també s'utilitza l'algorisme de Metropolis-Hastings en teoria de cordes i branques.

Biologia

Els algorismes MCMC també tenen una gran presència en bioestadística. Alguns articles rellevants on s'implementen algorismes MCMC són: Newton, Mau i Larget (1999), [29] i [36], en teoria de l'evolució, on s'estudien els arbres filogenètics de diferents taxons, és a dir, els arbres evolutius de diferents grups d'organismes; Liang i Wong (2001) [26] a nivell biomolecular, on s'estudia del plegament de proteïnes; o Zinc (2013) en medicina, on s'estudien els efectes adversos de medicaments.

5 Llenguatges de programació probabilístics

Degut a la gran popularitat que van tenir els mètodes MCMC, especialment durant els anys 90, una de les principals preocupacions en aquells moments era buscar una manera senzilla, ràpida i eficaç d'implementar-los, la qual cosa va desencadenar la creació de llenguatges de programació probabilístics.

Aquests llenguatges tenen com a objectiu crear models probabilístics els paràmetres dels quals corresponen a variables de programació [6]. Des de la seva creació, han despertat l'interès de sectors tan diversos com en estadística, intel·ligència artificial, aprenentatge automàtic o neurociència, entre molts d'altres.

El projecte BUGS va ser pioner en aquest camp, ja que el 1989 va iniciar el desenvolupament d'un nou programa que permetés treballar amb models bayesians utilitzant el mostreig de Gibbs (d'aquí el seu nom, Bayesian inference Using Gibbs Samplig), que es va dur a terme paral·lelament i independentment de l'article de Gelfand i Smith, que van ser els que van introduir els mètodes MCMC dins de l'estadística [28]. El primer prototip es va presentar al quart Congrés internacional d'Estadística Bayesiana de València, l'Abril de 1991 [37], i va seguir evolucionant fins a la versió per a Windows, WinBUGS, presentada l'any 1997 i la qual ja permetia treballar també amb l'algorisme MH.

Des d'aleshores s'han anat presentant gran multitud de llenguatges de programació probabilístics, la majoria basats en llenguatges ja existents, com ara Probabilistic C per a C o Edward i PyMC per a Python, els quals es poden incloure en forma de llibreries, i d'altres completament nous, que no tenen relació amb cap llenguatge ja existent, com és el cas de BUGS.

Per il·lustrar els mètodes MCMC de manera pràctica, s'utilitzarà PyMC3, la versió més recent de PyMC, que permet implementar algorismes MCMC junt amb models bayesians. Per a la part pràctica es realitza un estudi sobre l'afluència de clients del restaurant El Vinyet, el qual serveix esmorzars i dinars els caps de setmana i els festius i serveix sopars els dissabtes a la nit. A més a més s'haurà de tenir en compte que des de la primera setmana de febrer fins a Setmana Santa, coincidint amb la temporada alta de calçotades, també s'obre de dimecres a divendres per a servir dinars, la qual cosa quedarà reflectida en les gràfiques. Les dades que estudiarem s'agrupen setmanalment i es divideixen en quatre categories: *esmorzars*, *dinars*, *sopars* i el *total* de clients. Per últim, afegir que la capacitat màxima del restaurant és de 50 comensals, la qual limita el nombre màxim de *dinars* i *sopars*, mentre que durant els *esmorzars* es realitzen diferents torns i el seu nombre és més variable. Aquesta part pràctica consta de tres parts:

1. Estudi sobre possibles canvis rellevants en el flux de clients al llarg de l'any mitjançant l'estimació dels paràmetres associats al model.
2. Comparativa entre el flux de clients de diferents anys mitjançant inferència bayesiana.
3. Optimització del temps d'execució del mostreig de la primera part comparant MH amb HMC.

Durant l'estudi, és d'esperar que en les setmanes que tinguin algun dia festiu apareguin pics més pronunciats que, ens segons quins casos, es poden considerar casos puntuals. Descomptant les vacances, les dades es reparteixen en 48 setmanes per any. Per últim, notar que en les representacions gràfiques dels resultats les setmanes venen numerades del 0 al 47, que és l'enumeració utilitzada en el programa, però que interpretem del 1 al 48 a l'hora d'analitzar els resultats.

5.1 Estudi sobre l'afluència de clients

El primer objectiu és analitzar si algun dels quatre grups de dades que tenim presenta algun canvi rellevant al llarg de l'any 2017 i determinar en quin moment succeeix. Tenint en compte la percepció i l'opinió dels propietaris i treballadors, s'espera que l'afluència de clients sigui més elevada les primeres 14-15 setmanes, que corresponen a la temporada de calçotades, que és quan hi ha més volum de feina i es treballa amb més intensitat.

Per analitzar si realment és així, se seguirà el procediment següent per cada un dels grups. Sigui C_i el nombre de clients de la setmana i (de qualsevol de les quatre categories), s'assumeix que les dades segueixen una distribució de Poisson, que s'ha descrit a l'equació (2.12) i que és la més habitual per a aquest tipus de dades:

$$C_i \sim Pois(\lambda) \quad (5.1)$$

Es fa el supòsit que el paràmetre λ pot canviar de valor després d'un cert temps τ , de manera que

$$\lambda = \begin{cases} \lambda_1 & t < \tau \\ \lambda_2 & t \geq \tau \end{cases} \quad (5.2)$$

Donat que no es coneixen els valors de τ i de λ_j per $j = 1, 2$, utilitzant estadística bayesiana s'associa a cada un dels paràmetres una distribució *a priori* del que es creu que poden valdre [6]. En el cas de les λ_j , com ha de ser un valor real positiu, se li associa una distribució exponencial que dependrà, al seu torn, d'un hiperparàmetre α que es pren com la inversa de la mitjana de clients de cada grup al llarg de l'any. D'altra banda, a τ se li associa una distribució *a priori* uniforme discreta distribuïda al llarg de les 48 setmanes.

$$\lambda_j \sim Exp(\alpha) \quad \alpha = \left(\frac{1}{48} \sum_{i=1}^{48} C_i \right)^{-1} \quad \tau_i \sim UDiscreta(1, 48) \quad (5.3)$$

Com a observació notar que si no hi hagués cap canvi, les distribucions *a posteriori* de les λ_j serien pràcticament iguals.

Així doncs, s'elabora un model on a cada paràmetre se li associa una variable estocàstica de PyMC3 segons el procés que acabem de descriure. Ajudant-se de la funció *switch()* es defineix λ com una funció a trossos que dependrà de tots tres paràmetres. Com es pot veure al codi de la Figura 2, es defineix, a més, una variable

observacio que combina les dades reals del restaurant amb les dades generades pel model que s'ha proposat.

A continuació, fent ús de l'algorisme de MH, les distribucions de les λ_j i de τ es van modificant durant el mostreig fins a convergir cap a la distribució blanc, la distribució *a posteriori*. Això permetrà saber quins són els valors més probables de cada un dels paràmetres i per tant, concloure si hi ha hagut algun canvi rellevant o no i quan s'ha produït aquest. A continuació es presenten els resultats obtinguts amb les gràfiques corresponents per cada una de les categories utilitzant el codi següent:

```
import numpy as np
import pymc3 as pm

with pm.Model() as model_d:

    alpha = 1/d

    lambda_d1 = pm.Exponential('lambda_d1', alpha)
    lambda_d2 = pm.Exponential('lambda_d2', alpha)

    tau_d = pm.DiscreteUniform('tau_d', 0, n)

    index=np.arange(n)

    lambda_d=pm.math.switch(tau_d >= index, lambda_d1, lambda_d2)

    observacio_d=pm.Poisson('Obs_d', lambda_d, observed = din)

with model_d:
    trace_d = pm.sample(5000, tune = 2500, step = pm.Metropolis())

lambda_d1_samples = trace_d['lambda_d1']
lambda_d2_samples = trace_d['lambda_d2']
tau_d_samples = trace_d['tau_d']

Multiprocess sampling (4 chains in 4 jobs)
CompoundStep
>Metropolis: [tau_d]
>Metropolis: [lambda_d2_log_]
>Metropolis: [lambda_d1_log_]
The number of effective samples is smaller than 25% for some parameters.
```

Figura 3: Exemple del codi programat en Python on s'elabora un model a on a cada paràmetre se li associa una distribució de probabilitat i a continuació s'efectua la simulació mitjançant Metropolis-Hastings.

5.1.1 Estudi sobre els *esmorzars*

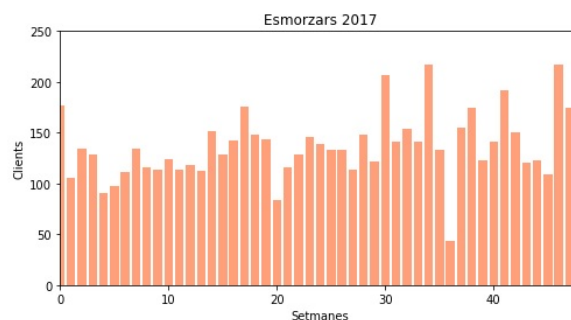


Figura 4: Dades de la categoria *esmorzars*, representades setmanalment.

Segons la Figura 4, el nombre d'*esmorzars* manté una certa regularitat la primera meitat de l'any, mentre que la segona meitat és més irregular, presentant una sèrie de pics que podrien indicar la presència d'un canvi, fent que la mitjana d'*esmorzars* per setmana pugés en algun moment d'aquesta segona meitat.

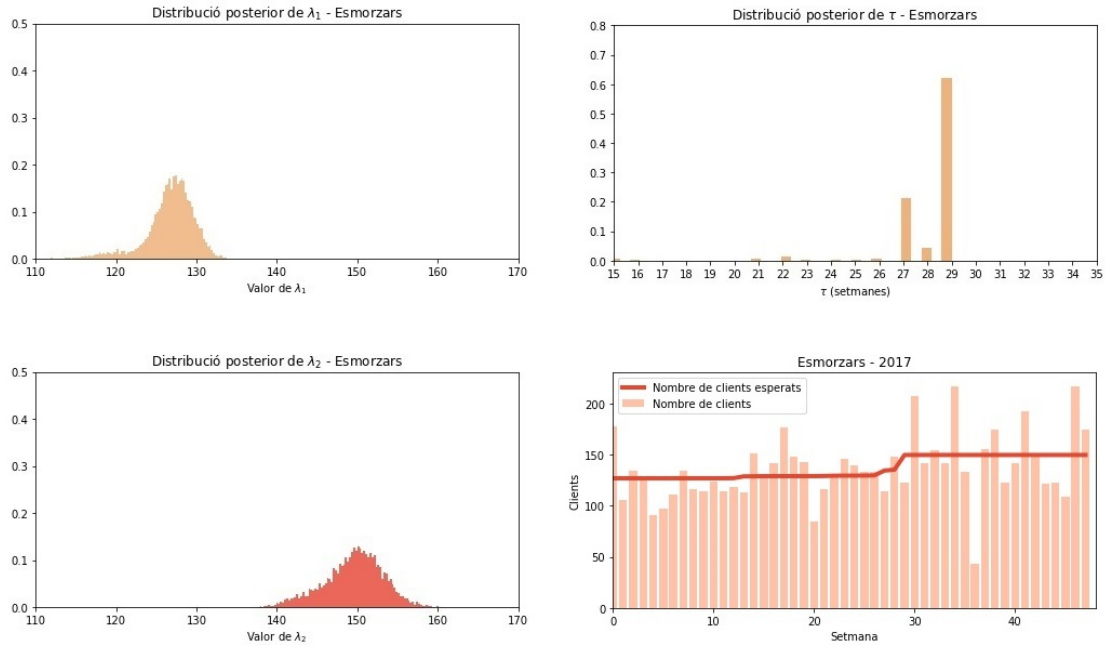


Figura 5: Resultats del mostreig per a la categoria *esmorzars*. A l'esquerra s'observen les distribucions *a posteriori* de les λ_j i a la dreta la distribució *a posteriori* de la τ i la taula dels *esmorzars* esperats respecte als *esmorzars* servits.

Efectivament, un cop feta la simulació, s'observa que els valors més probables de les λ_j difereixen bastant. El valor més probable de λ_1 és $\lambda_1 = 127.49$, mentre que el valor més probable de λ_2 és $\lambda_2 = 150.22$, suposant així un increment d'un 17.8% en la mitjana setmanal d'*esmorzars*. El valor més probable pel que es produeix aquest canvi és a la setmana 30, tot i que també existeix certa possibilitat de que sigui durant la setmana 28. A continuació es representen les distribucions associades al nombre d'*esmorzars* abans i després del canvi.

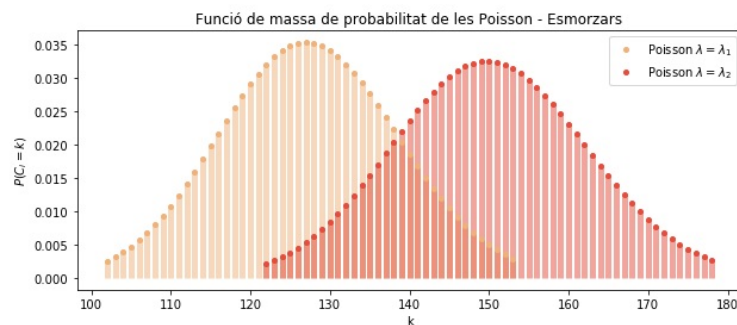


Figura 6: Funcions de massa de probabilitat per als valors més probables de les λ_j , on k representa el nombre d'*esmorzars* i l'eix y representa la probabilitat que es doni k .

5.1.2 Estudi sobre els *dinars*

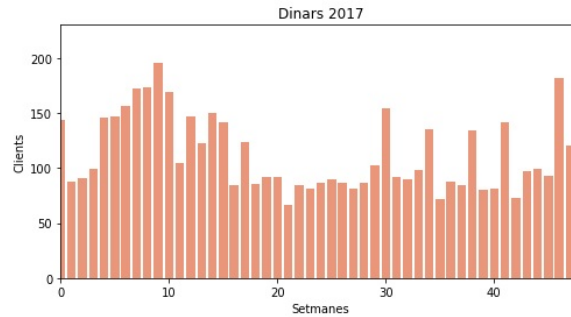


Figura 7: Dades de la categoria *dinars*, representades setmanalment.

Observant la gràfica dels *dinars* es veu que clarament es produeix un canvi entre les setmanes 15 i 20, on la mitjana de *dinars* disminueix i manté una certa regularitat però amb la presència d'alguns pics esporàdics. Aquest canvi és d'esperar tenint en compte que entre les setmanes 5 i 14 el restaurant també obra de dimecres a divendres (*dinars* exclusivament) degut a la temporada de calçotades, en la qual s'especialitza el restaurant.

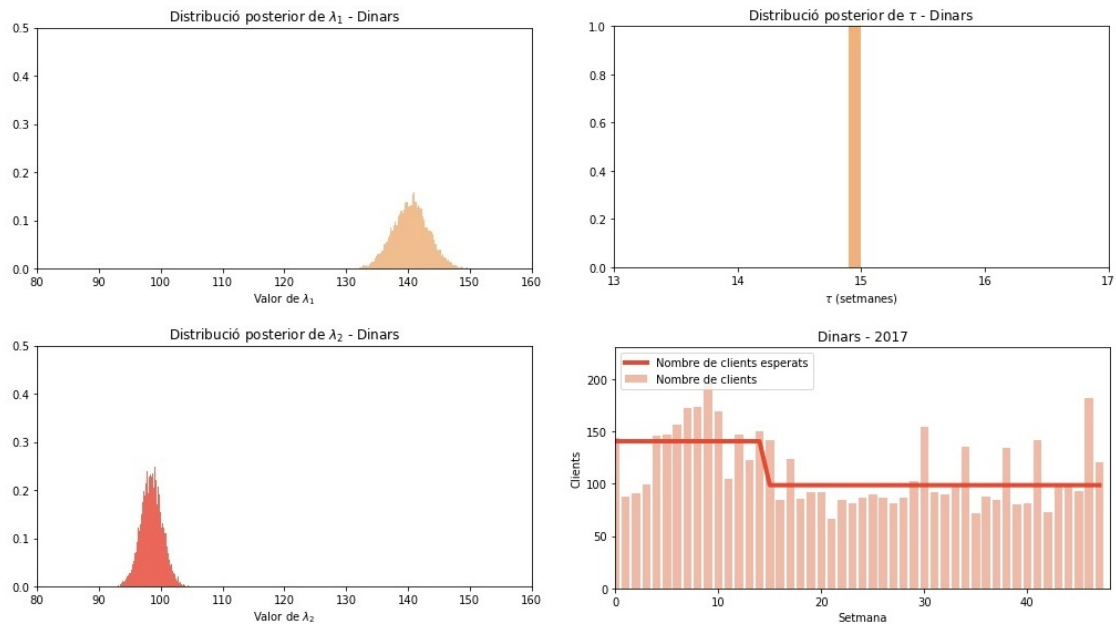


Figura 8: Resultats del mostreig per a la categoria *dinars*. S'observen les distribucions *a posteriori* de les λ_j i de τ , junt amb la taula dels *dinars* esperats respecte als *dinars* servits.

Després de realitzar la simulació s'obtenen els resultats de la Figura 8, on el valor més probable de λ_1 és $\lambda_1 = 140.35$, mentre que per λ_2 es té $\lambda_2 = 98.48$. Per tant, es pot afirmar que durant la temporada de calçotades hi ha un augment relatiu del 42.5% en la mitjana de *dinars* respecte als valors habituals de la resta de l'any. Notem que el canvi es produeix durant la setmana 16 amb probabilitat 1, que es

just la setmana que es deixa d'obrir entre setmana. A continuació es mostren les distribucions de probabilitat del nombre de *dinars* abans i després del canvi de la setmana 15, on clarament s'aprecia una gran diferència ja que la zona d'intersecció és molt petita:

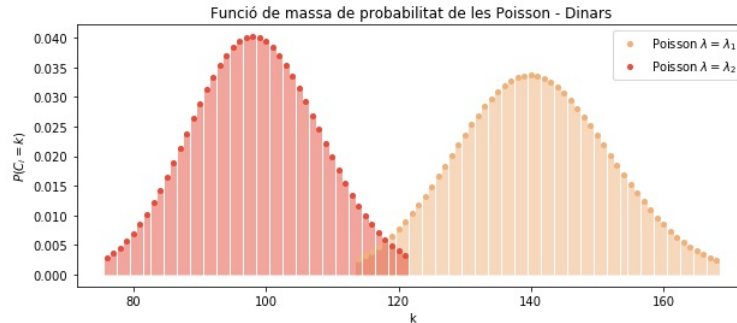


Figura 9: Funcions de massa de probabilitat per als valors més probables de les λ_j , on k representa el nombre d'*dinars* i l'eix y representa la probabilitat que es doni k .

5.1.3 Estudi sobre els *sopars*

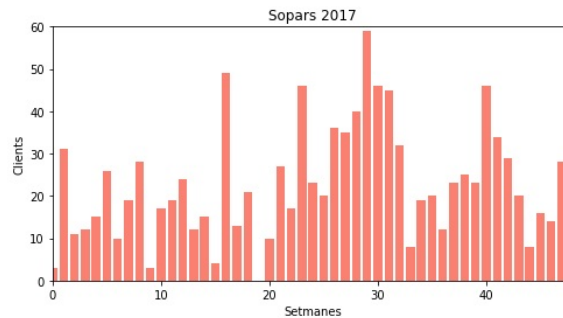


Figura 10: Dades de la categoria *sopars*, representades setmanalment.

Cal tenir en compte que en aquest cas es treballen amb valors molt més baixos respecte a les categories anteriors ja que només se serveixen sopars el dissabte a la nit. En la gràfica de dades es pot apreciar una activitat més aviat baixa fins al voltant de la setmana 20 seguida d'un període de màxima aflluència que correspon a l'estiu. A mitjans de setembre, cap a les setmanes 33-35, es torna a apreciar una petita baixada que va seguida d'una sèrie de pics irregulars que eleven la mitjana.

Com es pot veure a la Figura 11, es tenen dos valors ben diferenciats de cada λ , $\lambda_1 = 16.50$ i $\lambda_2 = 27.98$, el qual representa un increment relatiu del 69.6% de la mitjana de *sopars* que es produeix, o bé a la setmana 21, o bé a la 23, amb probabilitats semblants, que correspon a principis d'estiu, com era d'esperar.

En aquest cas, al tractar amb valors molt més baixos, notem que les distribucions associades al nombre de *sopars* de la Figura 12 estan molt més concentrades en intervals més petits i, a diferència dels dos casos anteriors, que el màxim de

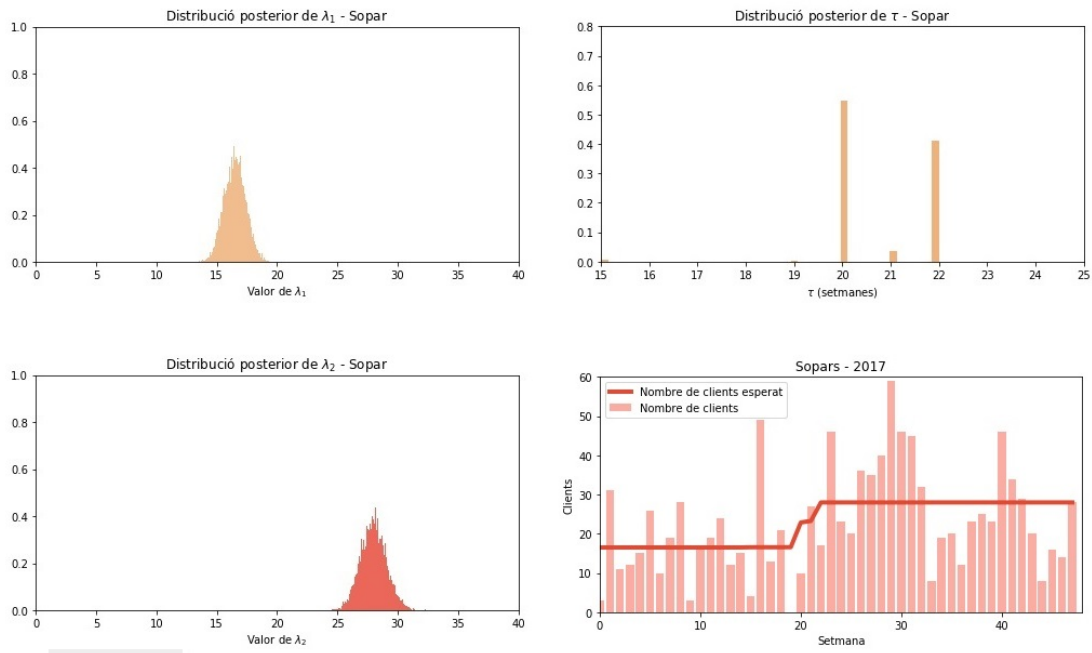


Figura 11: Resultats del mostreig per a la categoria *sopars*. S'observen les distribucions *a posteriori* de les λ_j i de τ , junt amb la taula dels *sopars* esperats respecte als *sopars* servits.

probabilitat es troba per sota de 0.04, en aquest cas s'observen probabilitats més altes, amb màxims al voltant de 0.1.

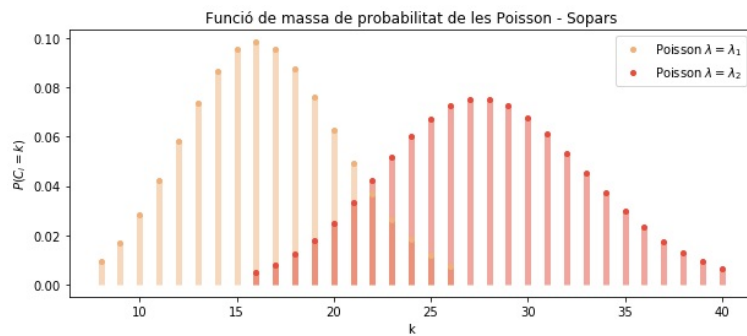


Figura 12: Funcions de massa de probabilitat per als valors més probables de les λ_j , on k representa el nombre d'*dinars* i l'eix y representa la probabilitat que es doni k .

5.1.4 Estudi sobre l'afluència *total* de clients

Per últim, s'analitzaran possibles canvis en el flux *total* de clients que passa pel restaurant setmanalment.

En aquest cas, segons la Figura 13, s'observa una certa semblança amb el cas dels *esmorzars*. Primer es dona un període amb una certa regularitat que es manté durant les primeres 27-28 setmanes i a la segona meitat de l'any comencen a aparèixer una sèrie de pics que despunten sobre la resta de dades, tot i que en aquest cas

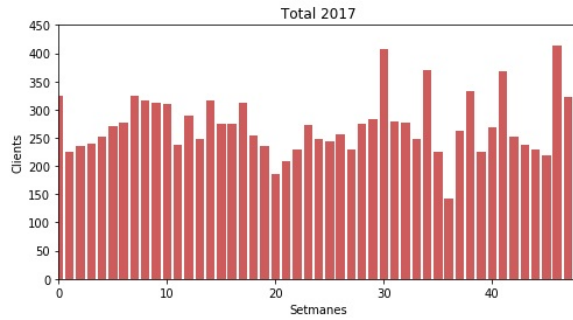


Figura 13: Dades de la categoria *total*, representades setmanalment.

també s'observen més valors per sota la mitjana i, per tant, a simple vista no es pot afirmar que es produeixi cap canvi rellevant.

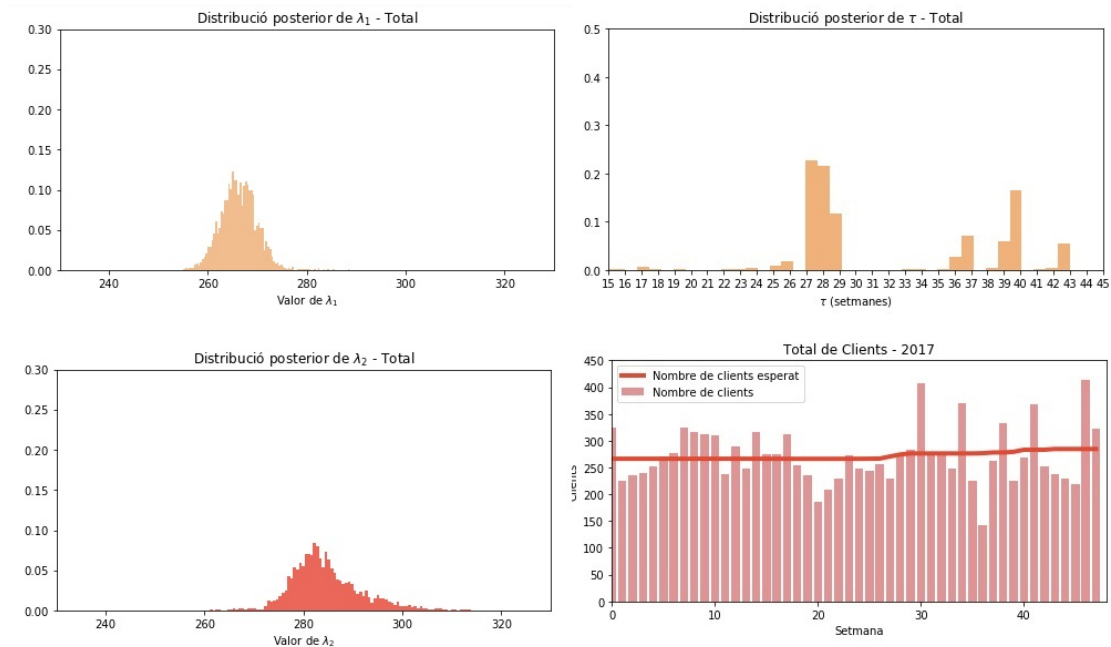


Figura 14: Resultats del mostreig per a la categoria *total*. S'observen les distribucions *a posteriori* de les λ_j i de τ , junt amb la taula del *total* de clients esperats respecte al *total* de clients que van anar al restaurant.

En aquest cas, per tal d'evitar l'efecte de vora degut a la presència de diversos casos puntuals relativament alts, s'han obviat les 5 primeres setmanes i les 5 últimes durant la simulació de τ . Sense aquesta consideració, els resultats que s'obtenien atribuïen els canvis més probables a la setmana 1 i a la setmana 46, els quals no corresponen als canvis globals que s'estan buscant al llarg de l'any, sinó a casos puntuals degut a la presència de diversos festius.

Així doncs, els resultats obtinguts després de la simulació indiquen que el valor més probable per a λ_1 correspon a $\lambda_1 = 266.59$, mentre que per a λ_2 el més probable és $\lambda_2 = 285.60$, però amb una dispersió considerable, suposant un increment relatiu de només 6.86%, el qual pot passar fàcilment desapercbut en els valors que s'estan

treballant. Aquest canvi es produeix molt probablement entre les setmanes 28 i 30, coincidint amb l'increment d'*esmorzars*, tot i que també hi ha certa probabilitat que s'hagi produït a les setmanes 37, 38, 40, 41 o 44. És precisament la gran inestabilitat que s'observa durant la segona meitat de l'any la que provoca que sigui difícil de determinar quan es produeix aquest canvi exactament així com el causant de la gran dispersió de λ_2 , de 7.4, en comparació a la dispersió de 3.5 de λ_1 .

En aquest cas s'observa que les distribucions associades al *total* dels clients tenen una zona d'intersecció molt gran, degut a que el canvi en els valors de les λ_j és molt subtil.

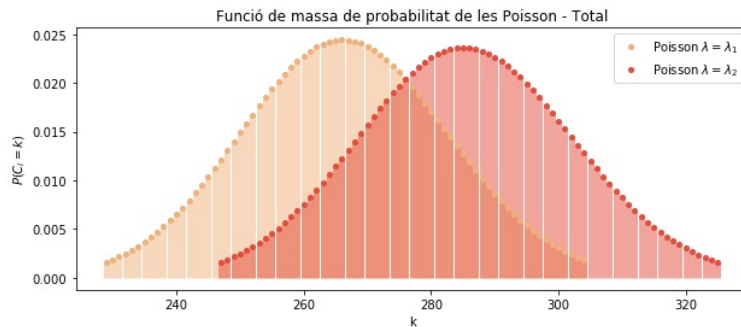


Figura 15: Funcions de massa de probabilitat per als valors més probables de les λ_j , on k representa el nombre de clients *totals* i l'eix y representa la probabilitat que es doni k .

Finalment, observant els resultats de les quatre categories se'n poden extreure algunes conclusions interessants.

- (i) Pel que fa referència als *esmorzars*, es pot observar a la Figura 4 que tot hi haver festius distribuïts al llarg de tot l'any, només destaquen respecte a la resta de setmanes aquells que cauen a partir del mes d'agost (setmanes 29-30). Això indica que no només hi ha un major flux d'*esmorzars* els caps de setmana de forma regular, sinó que a més a més, també se serveixen més *esmorzars* els dies festius.
- (ii) Com era d'esperar, durant les primeres 15 setmanes de l'any, durant la temporada de calçotades, la mitjana de *dinars* esperats es considerablement més alta que a la resta de l'any. D'altra banda, els *sopars* augmenten considerablement a partir del mes de maig, i tot i que decauen una mica a finals d'estiu, segueixen tenint més pes que a principis d'any.
- (iii) Tot i que en les primeres tres categories s'observen canvis rellevants en diferents èpoques de l'any, la mitjana *total* de clients pràcticament no canvia al llarg d'aquest, ja que només experimenta un canvi molt subtil pràcticament imperceptible. És a dir, la davallada que s'observa als *dinars* a la setmana 16 es compensa amb l'increment de *sopars* de la setmana 21 i la lleu baixada en el nombre de *sopars* a finals d'estiu es compensa amb l'augment d'*esmorzars* a la setmana 30, fent així que el nombre *total* de clients es mantingui pràcticament constant durant tot l'any.

Es pot concloure, doncs, que els diferents canvis que es produeixen en cada un dels serveis (*esmorzars*, *dinars* i *sopars*), mantenen un cert equilibri entre ells fent que la mitjana *total* de clients es mantingui constant. Així doncs, en contraposició a l'opinió inicial, notem que tot i que la percepció és que la intensitat de feina disminueix fora de temporada de calçotades, l'afluència de clients no disminueix en cap moment, si no que queda més repartida entre els tres serveis en lloc de centrar-se exclusivament en els *dinars*.

5.2 Evolució de l'afluència de clients

A continuació s'analitzarà l'evolució del flux de clients en els últims 10 anys. En aquest cas, s'utilitzen les dades del 2007, 2010, 2013, 2016 i 2017 i les es comparen consecutivament. El més habitual per comparar dos grups de dades seria utilitzar un test estadístic, com pot ser el test T, però aquests obliguen a fer una sèrie de suposicions i habitualment aporta una informació incompleta que s'ha de saber interpretar correctament [24]. Així que en aquest cas s'utilitza una estimació bayesiana que, en lloc de determinar si dos grups de dades són diferents o no, proporciona informació sobre com de diferents són aquests dos grups.

El primer pas és representar les dades de cada categoria per cada un dels cinc anys en forma d'histograma i associar-los-hi un model probabilístic. Habitualment, s'acostumen a utilitzar distribucions normals per a aquesta finalitat, de manera que la diferència entre les mitjanes ens indica la magnitud de la diferència entre les tendències centrals dels dos grups comparats. No obstant, com s'explica a Kruschke (2013) [24], el més adient per a grups de dades on hi ha presents diversos casos atípics que s'escapen fora d'aquesta distribució normal, com és el cas de setmanes on es treballa algun dia extra o ocasions en les quals es tanca per força major, és utilitzar una distribució *t*. Aquesta distribució depèn de tres paràmetres, la mitjana μ i la desviació σ , que ja coneixiem de la distribució normal, i el paràmetre de normalitat ν que pot prendre valors continus entre 1 i l'infinit. Aquest últim ens permet determinar l'alçada del pic i, per tant, l'amplada de la base, de manera que per a valors petits tenim distribucions semblants a una normal però amb bases més amples que ens permeten incloure els casos atípics dins la distribució de probabilitat. D'altra banda, a mesura que ν s'acosta a infinit, recuperem la distribució normal.

S'assumeix, doncs, que a cada histograma li correspon una distribució *t* amb mitjanes i desviacions característiques, però amb un mateix paràmetre de normalitat, ja que el nombre de casos atípics és semblant en els dos casos i relativament baix. D'aquesta forma tenim que per cada parella de grups a comparar es tindrà

$$C_i^{(1)} \sim T(\nu, \mu_1, \sigma_1) \quad C_i^{(2)} \sim T(\nu, \mu_2, \sigma_2) \quad (5.4)$$

Donat que en aquest cas, igual que en la part anterior, no se sap quins poden ser els possibles valors dels paràmetres, associem a cada un d'ells unes distribucions de probabilitat que s'acostin al que es pensa que poden valdre, que corresponen a les distribucions *a priori*.

En primer lloc, donat que les mitjanes han de tenir valors reals, se'ls hi associen

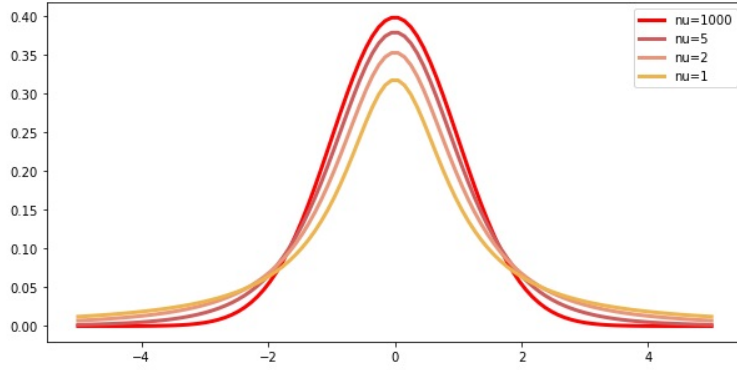


Figura 16: Representació gràfica de diferents distribucions t en funció del paràmetre de normalitat ν . Per a $\nu = \infty$ es recupera la distribució normal.

distribucions normals els hiperparaàmetres de les quals seran la mitjana dels valors de les dades i dues vegades la desviació estàndard [22]. De manera que queda

$$\mu_j \sim N(\bar{C}, 2s) \quad (5.5)$$

En segon lloc, a les desviacions se'ls hi associaran distribucions uniformes dins del rang de valors que es cregui adequat per a cada categoria, ja que per a *sopars* podem agafar un interval més petit que no pas per a *esmorzars* i *dinars*, mentre que per a la categoria *total* haurà de ser major.

$$\sigma_j \sim U[0, l_j] \quad (5.6)$$

Per últim, donat que ν pren valors continus entre 1 i infinit, se li associa una distribució exponencial amb una mitjana de 30, de manera que la major part de probabilitat es concentra en valors relativament petits de ν que originen distribucions t amb bases amples. Com la distribució exponencial pren valors entre 0 i infinit, se li suma 1 per tal que correspongui amb els valors que pot adoptar ν .

$$\nu \sim 1 + \exp\left(\frac{1}{30}\right) \quad (5.7)$$

Una vegada s'ha creat aquest model on a cada paràmetre se li ha associat una distribució de probabilitat, es crea una variable a la qual se li atribueix una distribució t amb cada un d'aquests paràmetres definits i a la qual s'inclouen les dades. Com es pot veure a la Figura 12, també es defineix dins del model algunes variables que ens determinin la diferència entre les mitjanes i les desviacions de cada parell de grups, així com una variable que ens dóna la magnitud de l'efecte, on hi intervé la diferència de mitjanes i l'arrel quadrada de la suma de les desviacions al quadrat. Totes aquestes variables donaran informació addicional que permetran analitzar les diferències i semblances entre els dos grups. A continuació s'aplica l'algorisme HMC i obtenim les distribucions *a posteriori* de cada un dels paràmetres, d'on se'n podran extreure les corresponents conclusions.

```

d1 = np.array(din1)
d2 = np.array(din2)
d3 = np.array(din3)
d4 = np.array(din4)
d5 = np.array(din5)

d = pd.DataFrame(dict(value = np.r_[d1, d2, d3, d4, d5], group =
np.r_[['Dinars 2007']*len(din1), ['Dinars 2010']*len(din2),
      ['Dinars 2013']*len(din3), ['Dinars 2016']*len(din4),
      ['Dinars 2017']*len(din5)]))

mu_m_d = d.value.mean()
mu_s_d = d.value.std()*2

sig_d_d = 1
sig_u_d = 50

with pm.Model() as model_d:
    mu_1 = pm.Normal('mu_1', mu_m_d, sd = mu_s_d)
    mu_2 = pm.Normal('mu_2', mu_m_d, sd = mu_s_d)
    mu_3 = pm.Normal('mu_3', mu_m_d, sd = mu_s_d)
    mu_4 = pm.Normal('mu_4', mu_m_d, sd = mu_s_d)
    mu_5 = pm.Normal('mu_5', mu_m_d, sd = mu_s_d)

    sigma_1 = pm.Uniform('sigma_1', sig_d_d, sig_u_d)
    sigma_2 = pm.Uniform('sigma_2', sig_d_d, sig_u_d)
    sigma_3 = pm.Uniform('sigma_3', sig_d_d, sig_u_d)
    sigma_4 = pm.Uniform('sigma_4', sig_d_d, sig_u_d)
    sigma_5 = pm.Uniform('sigma_5', sig_d_d, sig_u_d)

    lambda_1 = sigma_1**-2
    lambda_2 = sigma_2**-2
    lambda_3 = sigma_3**-2
    lambda_4 = sigma_4**-2
    lambda_5 = sigma_5**-2

    nu_d = pm.Exponential('nu_menys_1', 1/30) + 1

    din_1 = pm.StudentT('Dinars 2007', nu_d, mu_1, lambda_1, observed=d1)
    din_2 = pm.StudentT('Dinars 2010', nu_d, mu_2, lambda_2, observed=d2)
    din_3 = pm.StudentT('Dinars 2013', nu_d, mu_3, lambda_3, observed=d3)
    din_4 = pm.StudentT('Dinars 2016', nu_d, mu_4, lambda_4, observed=d4)
    din_5 = pm.StudentT('Dinars 2017', nu_d, mu_5, lambda_5, observed=d5)

    dif_mu1 = pm.Deterministic('Diferència mitjanes - Dinars 07-10', mu_2 - mu_1)
    dif_mu2 = pm.Deterministic('Diferència mitjanes - Dinars 10-13', mu_3 - mu_2)
    dif_mu3 = pm.Deterministic('Diferència mitjanes - Dinars 13-16', mu_4 - mu_3)
    dif_mu4 = pm.Deterministic('Diferència mitjanes - Dinars 16-17', mu_5 - mu_4)
    dif_mu5 = pm.Deterministic('Diferència mitjanes - Dinars 07-17', mu_5 - mu_1)

    dif_sigma1 = pm.Deterministic('Diferència desviacions - Dinars 07-10', sigma_2 - sigma_1)
    dif_sigma2 = pm.Deterministic('Diferència desviacions - Dinars 10-13', sigma_3 - sigma_2)
    dif_sigma3 = pm.Deterministic('Diferència desviacions - Dinars 13-16', sigma_4 - sigma_3)
    dif_sigma4 = pm.Deterministic('Diferència desviacions - Dinars 16-17', sigma_5 - sigma_4)
    dif_sigma5 = pm.Deterministic('Diferència desviacions - Dinars 07-17', sigma_5 - sigma_1)

    efecte1 = pm.Deterministic('Efecte - Dinars 07-10', dif_mu1 / np.sqrt((sigma_1**2 + sigma_2**2) / 2))
    efecte2 = pm.Deterministic('Efecte - Dinars 10-13', dif_mu2 / np.sqrt((sigma_2**2 + sigma_3**2) / 2))
    efecte3 = pm.Deterministic('Efecte - Dinars 13-16', dif_mu3 / np.sqrt((sigma_3**2 + sigma_4**2) / 2))
    efecte4 = pm.Deterministic('Efecte - Dinars 16-17', dif_mu4 / np.sqrt((sigma_4**2 + sigma_5**2) / 2))
    efecte5 = pm.Deterministic('Efecte - Dinars 07-17', dif_mu5 / np.sqrt((sigma_1**2 + sigma_5**2) / 2))

with model_d:
    trace_d = pm.sample(1000, tune = 500, cores=5)

Auto-assigning NUTS sampler...
Initializing NUTS using jitter+adapt_diag...
Multiprocess sampling (5 chains in 5 jobs)
NUTS: [nu_menys_1_log__, sigma_5_interval__, sigma_4_interval__, sigma_3_interval__, sigma_2_interval__, mu_5, mu_4, mu_3, mu_2, mu_1]

```

Per acabar, a més de realitzar un anàlisi global de l'evolució durant aquests deu anys i com afecta l'evolució de cada una de les tres primeres categories sobre l'evolució del *total* de clients que ha rebut el restaurant en aquest període, afegir que també s'avaluarà quin efecte pot haver tingut la crisi econòmica que es va iniciar el 2008 a Espanya sobre l'afluència de clients al restaurant.

5.2.1 Evolució dels *esmorzars*

A primer cop d'ull, s'observa a la Figura 17 que la suposició sobre la distribució t és correcta, ja que apareixen diversos casos atípics que s'allunyen d'una distribució normal típica. Tant al 2007 com al 2010 la major part de setmanes s'acostumen a servir al voltant de 100-110 *esmorzars*, mentre que al 2013 sembla que recoli lleugerament fins a 80-90. També s'observa una remuntada al 2016, on la major part de pes està al voltant de 120 i, al 2017, puja lleugerament fins a 125.

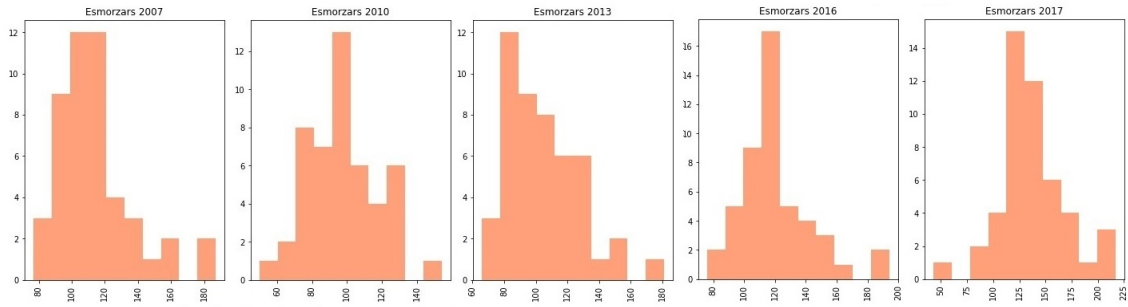


Figura 17: Histogrames dels *esmorzars* durant els anys 2007, 2010, 2013, 2016 i 2017.

Una vegada aplicat l'algorisme HMC s'obtenen els següents resultats:

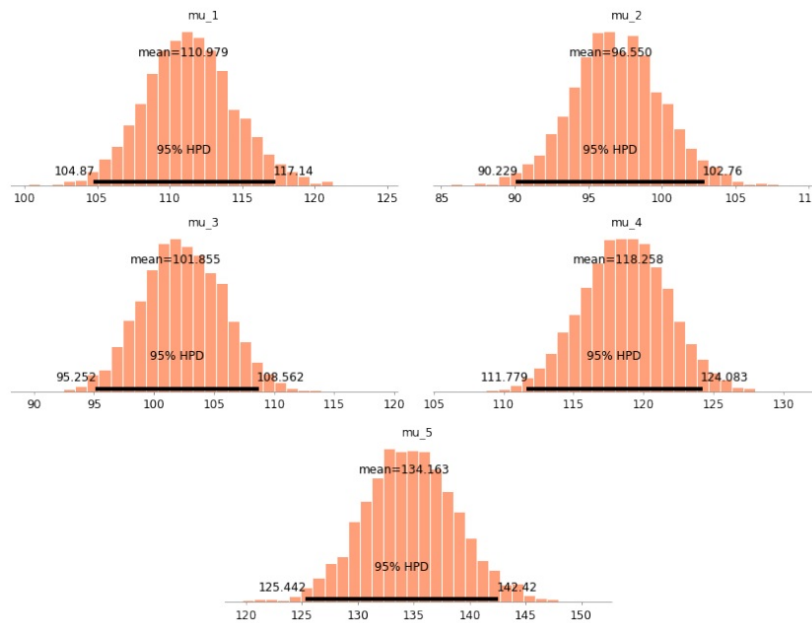


Figura 18: Distribució de probabilitats *a posteriori* de les mitjanes de les distribucions t que descriuen els *esmorzars*.

Tenint en compte que μ_1 , μ_2 , μ_3 , μ_4 i μ_5 , corresponen respectivament als *esmorzars* de 2007, 2010, 2013, 2016 i 2017, s'observa que entre els anys 2007 i 2010 hi ha una baixada important, a diferència de la interpretació inicial, ja que es passa d'una mitjana de $\mu_1 = 111.0$ *esmorzars* a només $\mu_2 = 96.6$. A continuació es dona una petita remuntada al 2013, pujant fins a $\mu_3 = 101.9$, mentre que al 2016 i 2017 es veu clarament un increment important, amb mitjanes de $\mu_4 = 118.3$ i $\mu_5 = 134.2$. A continuació es presenten les diferències de mitjanes representades junt amb els corresponents intervals de confiança.

El primer canvi que s'observa en la Figura 19 es dona entre els anys 2007 i 2010, on la diferència de mitjanes és de -14.4 i donat que aquesta mitjana cau fora de l'interval de confiança, es pot assegurar que efectivament hi ha hagut una davallada important en aquest període, que es correspon amb l'inici de la crisi econòmica. No obstant, el petit increment que s'havia observat entre les mitjanes de 2010 i 2013,

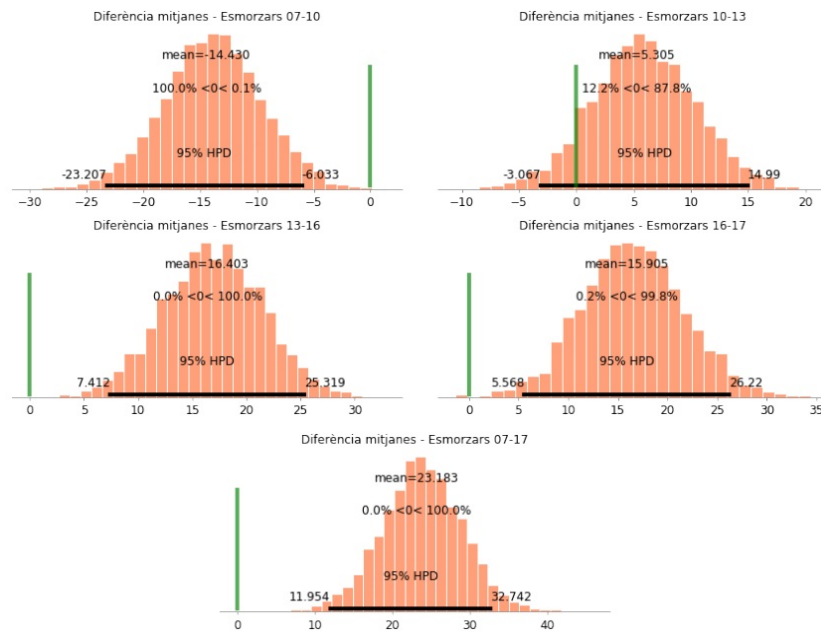


Figura 19: Representació de la diferència de mitjanes respecte als intervals de confiança.

que és de 5.3, no és suficientment gran com per dir que hi ha hagut un canvi real. Com veiem, aquest cau dins l'interval de confiança i, per tant, aquests dos grups de dades es poden considerar pràcticament iguals. Això indica que, pel que fa als *esmorzars*, al 2013 encara es mantenia la davallada causada per la crisi.

No obstant, entre 2013 i 2016 la mitjana augmenta 16.4, quedant fora de l'interval de confiança. Per tant es pot assegurar que hi ha hagut un increment, que no pot ser degut a la casuística del nostre model estadístic ni del mostreig, recuperant el nivell del 2007 i fins i tot, passant lleugerament al davant. A més a més, entre el 2016 i el 2017 torna a haver-hi una diferència de 15.9, igualant pràcticament l'augment que hi ha entre 2013 i 2016, en un sol any. Per últim, si es calcula la diferència entre 2007 i 2017, aquesta és de 23.2, que resulta en un increment del 20.9% respecte a la mitjana d'*esmorzars* que es servien al 2007, un augment molt important tenint en compte la davallada inicial deguda a la crisi.

5.2.2 Evolució dels *dinars*

Pel que es pot veure als histogrames de les dades que corresponen als *dinars*, s'observa que en tots ells la major part del pes es reparteix entre 80-100, mentre que els *dinars* que estan més amunt de 100 varien de forma irregular. A priori no es podria afirmar que hi hagi hagut cap canvi rellevant.

En aquest cas, segons la Figura 21, es té que no hi ha gaire variació al llarg dels anys. Podria haver-hi una lleu disminució entre els anys 2007 i 2013, seguit d'un augment molt progressiu fins que el 2017 es recuperen els valors de 2007.

Observant la figura 22 es pot concloure que tot hi haver-hi petites variacions entre les mitjanes, no hi ha cap que caigui fora de l'interval de confiança i, per tant,

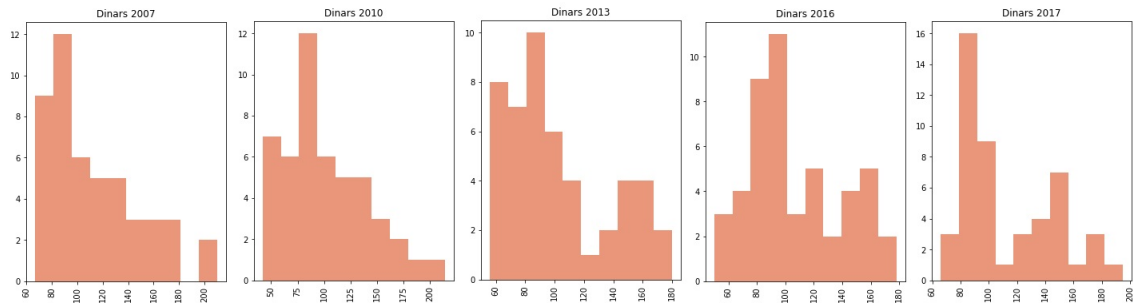


Figura 20: Histogrames dels *dinars* durant els anys 2007, 2010, 2013, 2016 i 2017.

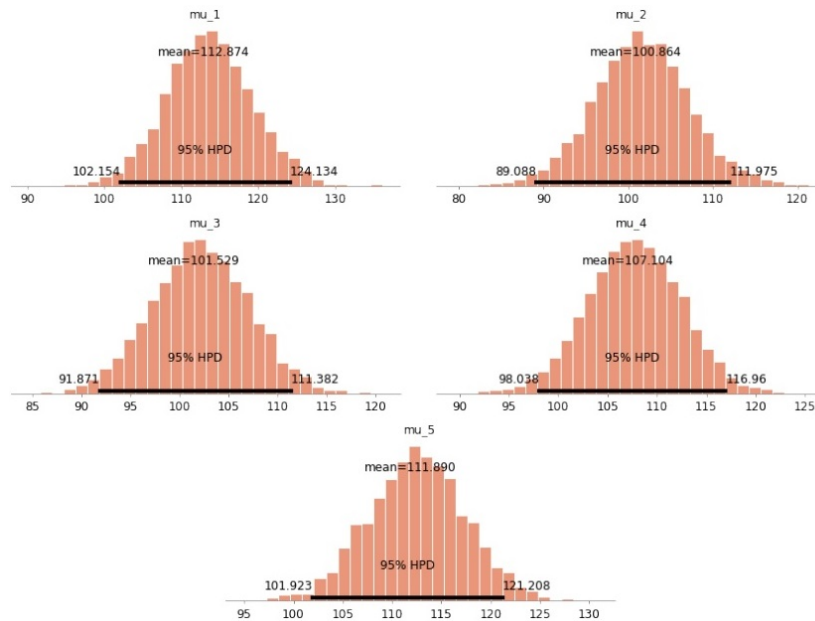


Figura 21: Distribució de probabilitats *a posteriori* de les mitjanes de les distribucions t que descriuen els *dinars*.

podem dir que tots els grups són, a trets generals, iguals. Així doncs, en el cas dels *dinars*, no hi ha hagut cap canvi rellevant pel que fa a la mitjana de *dinars* en els últims 10 anys tot i la presència de la crisi.

El que sí es pot afirmar si es torna a mirar la figura 20 és que si bé al principi els *dinars* es distribuïen de forma més o menys regular al voltant del pic central, a mesura que passen els anys s'observa com la major part del pes dels histogrames se centra en un pic principal molt pronunciat al voltant de 90-100 i en un pic secundari al voltant de 150. Això és especialment evident en els anys 2016 i 2017.

Per explicar aquest fet utilitzaré part de la meua experiència com a treballadora del restaurant durant gran part d'aquest període. A diferència dels *esmorzars* en els quals es fan diferents torns i, per tant, el seu nombre pot variar molt segons el dia, durant els *dinars* es fa un únic torn, i el nombre màxim de persones ve limitat per la capacitat del restaurant, que varia entre 45 i 55 segons la disposició que es requereixi. No obstant, aquesta xifra pot variar lleugerament si excepcionalment es

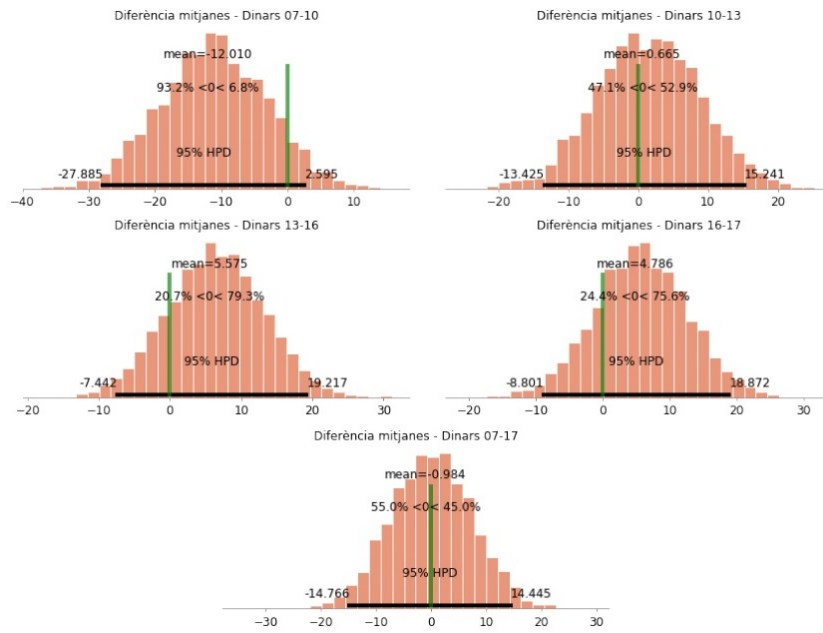


Figura 22: Representació de la diferència de mitjanes respecte als intervals de confiança.

dóna alguna taula per segona vegada. Així doncs, mentre que a principis d'aquest període que estem estudiant el nombre de *dinars* varia molt segons la setmana, durant els últims anys el restaurant fa ple quasi a diàri durant la major part de l'any, fent que majoritàriament se serveixin entre 90 i 100 *dinars* els caps de setmana i al voltant de 150 les setmanes amb algun dia festiu.

Per tant, tot i que no es pugui apreciar cap canvi en la mitjana de *dinars* al llarg d'aquesta dècada, si que es pot observar un canvi en el comportament en l'afluència de clients, passant de ser un flux irregular en el qual podies passar de fer només 20 *dinars* un dissabte i fer-ne 60 el diumenge següent, a fer-ne pràcticament 45-55 diàris. No obstant, no sembla que la crisi hagi tingut cap mena d'impacte directe en aquest canvi, ja que els histogrames de 2007 i 2010 són bastant similars.

5.2.3 Evolució dels *sopars*

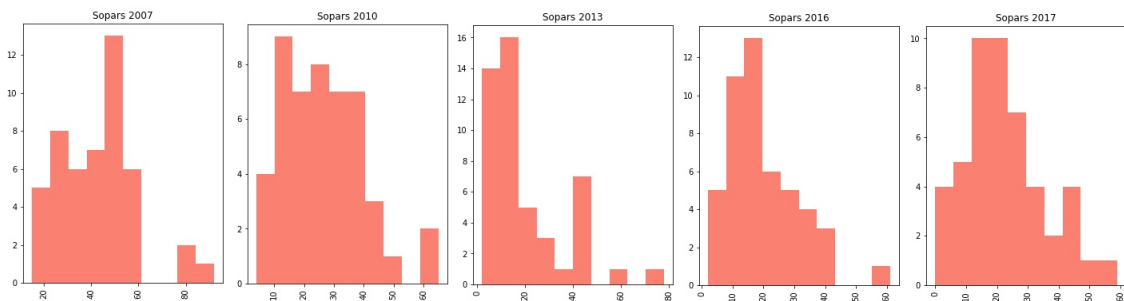


Figura 23: Histogrames dels *sopars* durant els anys 2007, 2010, 2013, 2016 i 2017.

En aquest cas es veu clarament que els histogrames són bastant diferents. D'una

banda, al 2007 la major part de setmanes es feien entre 25 i 60 *sopars*, predominant especialment les setmanes que se'n feien 50. Al 2010 es feien, de manera més o menys uniforme, entre 10 i 35, mentre que al 2013 predominen les setmanes amb menys de 15 *sopars*. D'altra banda, el 2016 remunta lleugerament fent majoritàriament entre 10 i 30 *sopars* per setmana, mentre que finalment el 2017 el pic es troba al voltant de 20 *sopars*. Per tant, es d'esperar que en aquest cas es trobi algun canvi important.

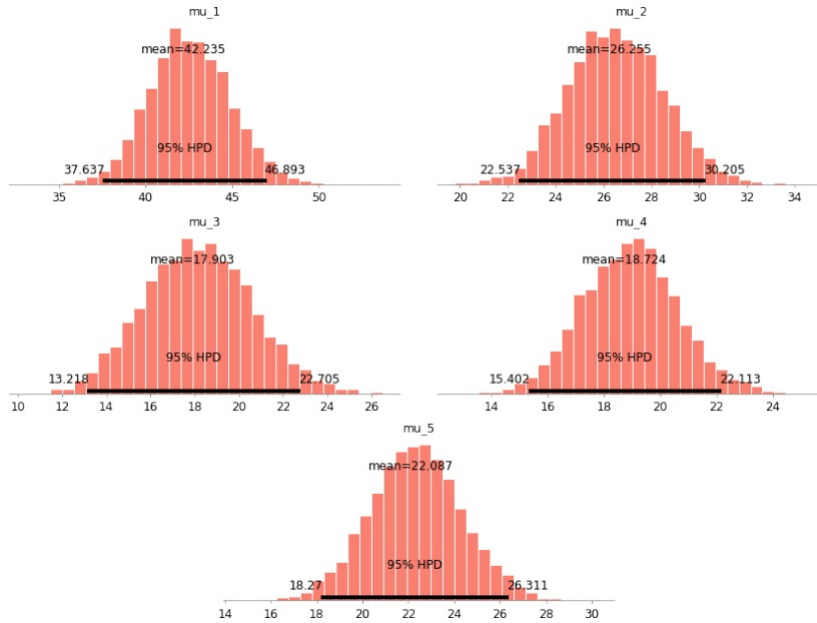


Figura 24: Distribució de probabilitats *a posteriori* de les mitjanes de les distribucions t que descriuen els *sopars*.

Un cop s'ha executat HMC s'observa que, efectivament, els valors més probables de les mitjanes de cada un dels grups difereixen considerablement. En primer lloc, es passa de $\mu_1 = 42.2$ a $\mu_2 = 26.3$, entre 2007 i 2010 i baixa fins a $\mu_3 = 17.9$ el 2013. A continuació les mitjanes augmenten lleugerament fins a arribar a $\mu_5 = 22.1$ el 2017. Anem a veure quins d'aquests canvis són reals i quins poden ser degut a les fluctuacions del nostre model estadístic.

En la figura 25, d'una banda, s'observa que tant la diferència de mitjanes entre el 2007 i el 2010, que és de -16.0, com la diferència entre 2010 i 2013, que és de -8.4, queden fora de l'interval de confiança i, per tant, es pot assegurar que hi ha una caiguda important en la mitjana de *sopars* entre 2007 i 2013. D'altra banda, les diferències entre 2013 i 2016 i entre 2016 i 2017, que són de 0.8 i 3.4 respectivament, indiquen que en aquest període no hi ha cap canvi i els tres grups de dades són pràcticament iguals. Per últim, l'evolució entre 2007 i 2017, que tenen una diferència de mitjanes de -20.1, és clarament decreixent, suposant una caiguda del 47.6% respecte els valors de 2007. És el cas més clar de pèrdua de clients durant el període 2007-2013, que ha sigut impossible recuperar i que coincideix amb l'inici de la crisi econòmica, la qual es podria considerar causa directa en aquest cas.

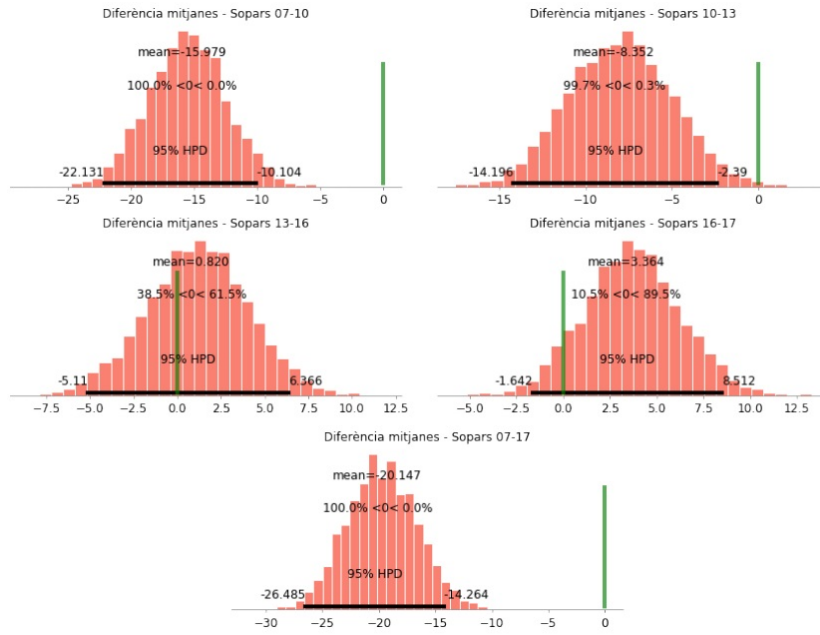


Figura 25: Representació de la diferència de mitjanes respecte als intervals de confiança.

5.2.4 Evolució de l'aflluència *total* de clients

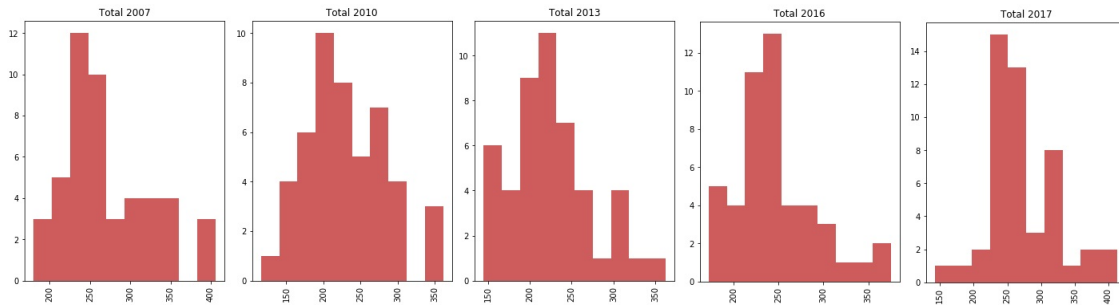


Figura 26: Histogrames dels clients *totals* durant els anys 2007, 2010, 2013, 2016 i 2017.

En el cas del nombre *total* de clients, l'any 2007 el valor més habitual estava al voltant de 250 i estava distribuït de forma regular entre 200 i 350, mentre que al 2010 i al 2013 es veu que els valors més habituals voltaven entre 150 i 275, amb el seu màxim pròxim a 200. No obstant, el 2016 el valor més probable remunta fins a 250 altra vegada i es manté durant el 2017, on es veuen clarament dos pics ben definits, un de principal a 250 i un de secundari a 325, probablement per la mateixa raó que s'ha comentat a la secció 5.2.2.

Com s'observa a la figura 27, en el primer període la mitjana decau de $\mu_1 = 268.2$ fins a $\mu_2 = 227.3$, mentre que entre 2010 i 2013 pràcticament no varia. Al 2016 es recupera fins a $\mu_4 = 245.2$ i el 2017 torna a pujar fins a $\mu_5 = 269.8$, valor semblant al que teníem inicialment el 2007.

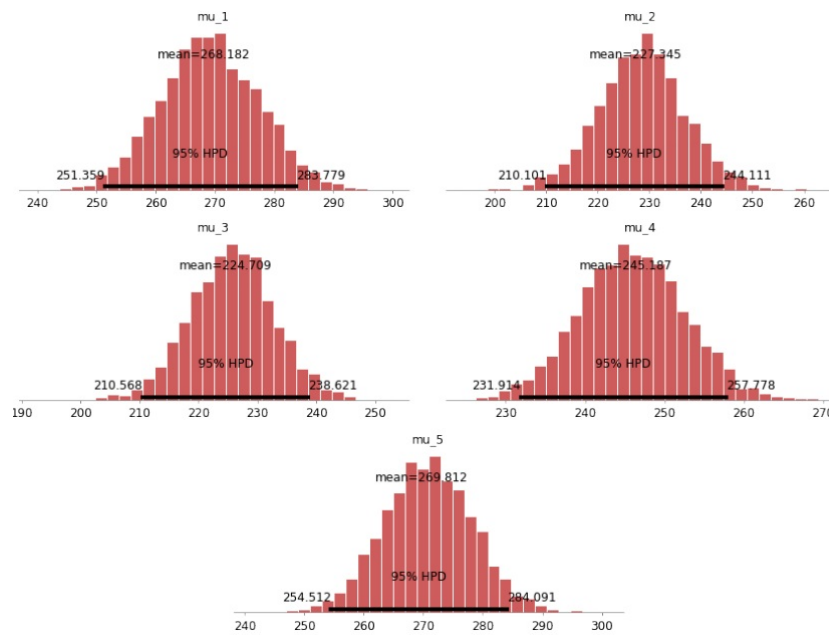


Figura 27: Distribució de probabilitats *a posteriori* de les mitjanes de les distribucions *t* que descriuen els *totals*.

A continuació analitzem la figura 28. Donat que la diferència de mitjanes entre el 2007 i el 2010, que és de -40.8 , cau fora de l'interval de confiança, es pot afirmar amb total certesa que hi ha una davallada important en aquest període, mentre que la mínim diferència entre 2010 i 2013, de -2.6 , que cau just al centre de l'interval, ens indica que les dades d'aquests dos anys són pràcticament idèntiques. Per un marge mínim, es podria dir que entre el 2013 i el 2016 hi va haver un augment en l'afluència *total* de clients, que augmenta en 20.5 , mentre que es pot assegurar més convençudament la recuperació que es dona entre 2016 i 2017, de 24.6 , superant així la remuntada dels tres anys anteriors, en un de sol.

Per últim, notar que si es comparen els anys 2007 i 2017, la diferència és ínfima, indicant que les mitjanes del nombre total de clients setmanals entre aquests dos anys és pràcticament la mateixa. En termes generals es té, doncs, que la sobtada caiguda que apareix a l'inici de la crisi, entre 2007 i 2010, i que representa un 15.2% de la mitjana de clients que passa pel restaurant, es recupera íntegrament durant el període 2013-2017. En especial, cal destacar que un 9.2% d'aquesta caiguda inicial es recupera sorprenentment durant l'últim any.

Per acabar amb aquest apartat, s'analitzarà la influència de les evolucions dels *esmorzars*, *dinars* i *sopars* sobre l'evolució del *total* de clients. Donat que en la categoria *dinars* no s'ha detectat cap canvi rellevant a partir de la diferència de les mitjanes, els canvis obtinguts en el *total* només dependran dels canvis que s'han donat en les categories *esmorzars* i *sopars*.

En ambdós casos es produeix una caiguda inicial en el període 2007-2010, que sumant-se donen lloc a una important davallada en el nombre *total* de clients en aquest període. La segona caiguda en el nombre de *sopars* durant el 2010-2013

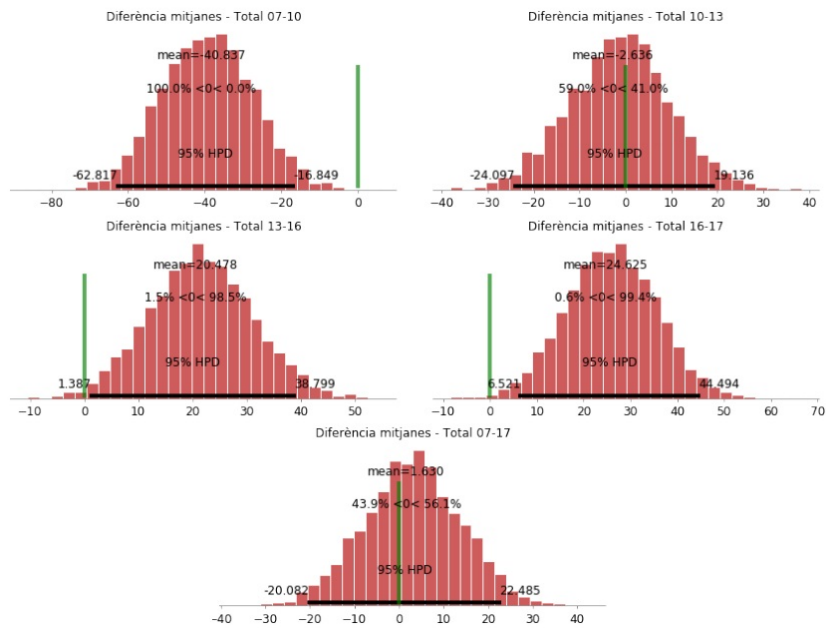


Figura 28: Representació de la diferència de mitjanes respecte als intervals de confiança.

queda compensada amb la lleugera diferència entre mitjanes que experimenta la categoria *esmorzars*, que no es prou gran per considerar-la un canvi important de per sí, però és suficient per contrarestar-la. Això dona lloc a una certa estabilitat els anys 2010-2013.

Entre els anys 2013 i 2017, els *esmorzars* tornen a pujar mentre que els *dinars* i *sopars* experimenten petites fluctuacions positives que no arriben a ser cap canvi important, però que se sumen a l'augment dels *esmorzars*. La suma d'aquests diversos elements fa que el nombre *total* recuperi els seus valors inicials. Per tant, es pot dir que la disminució que encara no s'ha recuperat al 2017 en el nombre de *sopars*, queda compensada amb l'augment d'*esmorzars*, que ha superat els seus valors inicials, mantenint així un equilibri en el nombre *total* de clients que han passat pel restaurant.

Per tant, les conclusions que en podem extreure es resumeixen en:

- (i) La crisi econòmica ha afectat directament el negoci, com tants d'altres, provocant una caiguda continuada entre els anys 2007 i 2013, en especial pel que fa al nombre de *sopars*, que no s'han recuperat des d'aleshores.
- (ii) L'única categoria que no s'ha vist afectada ni alterada en cap moment a nivell de mitjana de clients setmanals és la de *dinars*, tot i que si s'observen les dades, es pot percebre un canvi a nivell diàri, passant d'un comportament errant a un de completament regular, on pràcticament s'arriba a la capacitat màxima a diàri.
- (iii) Tot i haver-hi hagut la crisi econòmica, el nombre mig d'*esmorzars* ha evolucionat favorablement fins a superar els valors anteriors a la crisi, permetent

així recuperar el nombre *total* de clients mitjà.

- (iv) La recuperació que experimenta l'afluència *total* de clients durant l'últim any supera, encara que per poc, la recuperació que es dona en el període anterior, que abarca un interval de tres anys. Així doncs, cal destacar que el major augment en el nombre de clients en els últims deu anys s'ha produït entre 2016 i 2017.

5.3 Optimització del mostreig per a l'estudi de l'afluència de clients

En aquesta última part es compararan els resultats obtinguts en l'apartat 5.1 utilitzant, d'una banda, l'algorisme de MH, i d'altra banda, l'algorisme HMC. Com a observació comentar que en l'apartat 5.2 s'ha utilitzat directament HMC degut a que es tracten paràmetres amb distribucions contínues, però la convergència es tan ràpida que és pràcticament indiferent escollir HMC o MH.

En primer lloc es presenten els temps d'execució obtinguts per l'algorisme de Metropolis-Hastings en cada un dels models treballats, junt amb els paràmetres que s'han utilitzat en la seva preparació. Aquests paràmetres s'han escollit basant-se en els millors resultats obtinguts en el mínim temps possible.

Model	Nombre de passos	Nombre de passos cremats	Temps d'execució
Esmorzars	5000	2500	2m 56s
Dinars	5000	2500	2m 53s
Sopars	5000	2500	3m 47s
Total	5000	2500	2m 50s

Figura 29: Taula de paràmetres i temps d'execució corresponents per a Metropolis-Hastings.

Substituint l'algorisme de MH per HMC, es té que, pel mateix nombre de passos, el temps d'execució és més gran ja que els passos de HMC són més complexos i requereixen de més temps que els de MH. No obstant, si es va disminuint el valor dels paràmetres de preparació de HMC s'observa que la qualitat dels resultats obtinguts no disminueix. Per aquest motiu, utilitzant només 1000 passos -més 500 passos cremats-, s'obtenen els mateixos resultats que s'obtenen amb 5000/2500 a MH, en un temps molt menor.

Recordem però que HMC només es pot utilitzar amb distribucions de probabilitat contínues i que, per tant, no l'ha podem utilitzar en el cas de la τ . Com es pot observar en el codi següent, de manera automàtica el programa utilitza MH per calcular la distribució d'aquesta τ i genera els mateixos resultats que si s'apliqués directament Metropolis-Hastings:

A continuació es representa la taula comparativa amb els temps d'execució per cada un dels diferents models per ambdós algorismes:

```

with model_d2:
    trace_d2 = pm.sample(1000, tune = 500, step = pm.NUTS())

lambda2_d1_samples = trace_d2['lambda2_d1']
lambda2_d2_samples = trace_d2['lambda2_d2']
tau2_d_samples = trace_d2['tau2_d']

Multiprocess sampling (4 chains in 4 jobs)
CompoundStep
>NUTS: [lambda2_d2_log__, lambda2_d1_log__]
>Metropolis: [tau2_d]

```

Model	Algorisme	Nombre de passos	Nombre de passos cremats	Temps d'execució
Esmorzars	Metropolis	5000	2500	2m 56s
	NUTS	1000	500	0m 47s
Dinars	Metropolis	5000	2500	2m 53s
	NUTS	1000	500	0m 51s
Sopars	Metropolis	5000	2500	3m 47s
	NUTS	1000	500	0m 56s
Total	Metropolis	5000	2500	2m 50s
	NUTS	1000	500	0m 48s

Figura 30: Taula de paràmetres i temps d'execució corresponents per HMC.

Per tant, tot i que aplicant MH no es recomenable -en aquest cas- baixar per sota de 5000/2500, amb HMC es pot baixar tranquil·lament fins a 1000/500 i segueix donant els mateixos resultats, disminuint considerablement el nombre de mostres generades per la simulació i el temps d'execució. Es pot apreciar, doncs, que efectivament l'algorisme HMC és més eficient que MH, com ja s'havia comentat anteriorment.

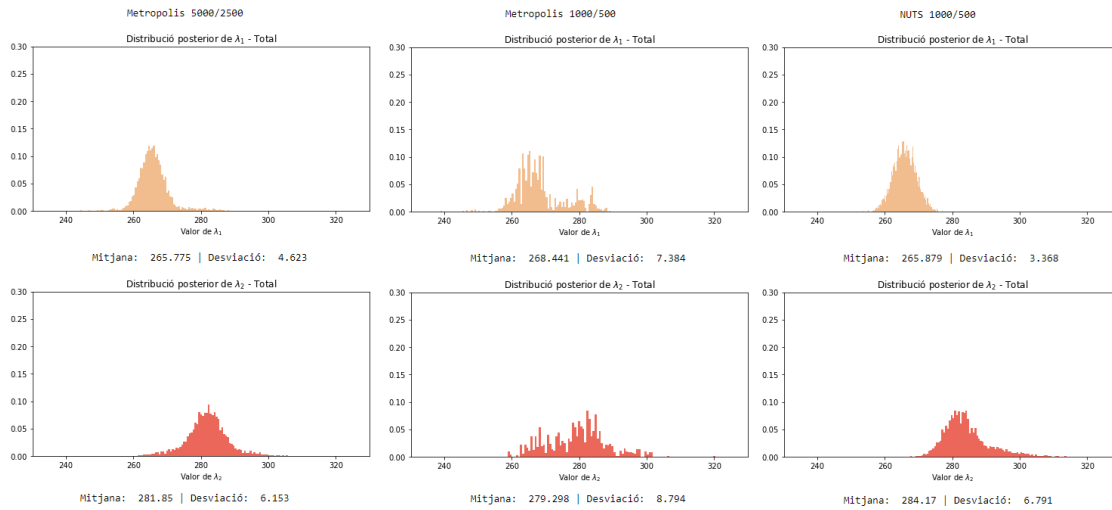


Figura 31: Distribucions de probabilitat de les λ_j a posteriori. D'esquerra a dreta, els resultats obtinguts aplicant MH 5000/2500, MH 1000/500 i HMC 1000/500 en el model on s'estudia el *total* de clients.

En l'exemple de Figura 31 s'observa que la simulació realitzada mitjançant HMC per 1000/500 és molt similar a la que es dona per MH 5000/2500, i fins i tot disminueix considerablement la desviació de λ_1 . No obstant, utilitzant 1000/500 passos per MH, s'observa clarament que la distribució encara no acaba de convergir, resultant en desviacions molt grans i distribucions irregulars.

6 Conclusions

L'aparició dels primers algorismes, com MCO i Metropolis, arriba immediatament després de la creació dels primers ordinadors, i a mesura que aquests milloren, també ho fan tots els algorismes MCMC que van sorgint durant la segona meitat del segle XX. D'altra banda, a mesura que s'aprofundeix en la teoria que envolta aquests algorismes i se'n van descobrint aplicacions, la presència dels algorismes MCMC també impulsa, en certa manera, la construcció d'ordinadors cada vegada més potents així com el desenvolupament de noves eines, com els llenguatges de programació probabilístics, que permetin utilitzar aquests algorismes de la forma més ràpida i eficient possible, i representar-ne els resultats d'una forma clara i entenedora, que pugui ser interpretada fàcilment.

El desenvolupament de tot aquest entramat d'algorismes MCMC ha permès, tant dins de qualsevol branca de la comunitat científica com fora d'ella, enfrontar problemes que anteriorment s'havien donat per impossibles. Problemes tan enrevessats com pot ser l'assignació d'un model probabilístic de paràmetres variables, com s'ha vist en l'estudi de l'afluència de clients del restaurant, avui en dia es resolen en qüestió de segons gràcies a una senzilla simulació programada amb unes poques línies de codi. La combinació d'aquests algorismes junt amb la inferència bayesiana ha suposat un abans i un després en la comunitat estadística, i arran de la gran popularitat que van tenir als anys 90, a dia d'avui encara es publiquen centenars d'articles trobant només aplicacions en pràcticament qualsevol àmbit.

L'impacte que han representat els algorismes MCMC no només es limita en ciències ja existents, sinó que a més a més han col·laborat en el desenvolupament de nous camps més recents, com és el cas de la ciència de dades, on l'estadística bayesiana hi té un paper rellevant. Les possibilitats de MCMC són infinites i el nombre d'aplicacions que se li donen segueix augmentant dia rere dia.

No obstant, els algorismes MCMC no són infal·libles i en algunes ocasions poden presentar certs problemes associats. Un dels més comuns és trobar-se amb casos on el ritme de convergència és tan lent que els temps d'execució són extremadament grans. Aquest problema queda reflectit, per exemple, en l'article de Mossel i Vigoda (2007)[33] on, retornant a l'aplicació dels mètodes MCMC en l'estudi d'arbres filogenètics -estudi iniciat a [29] i [36]-, s'observa clarament que quan el conjunt de dades s'agafa aleatòriament de la barreja de dos arbres, els temps d'execució creixen exponencialment amb la mida del conjunt de dades, assolint així casos en els quals la simulació és extremadament lenta, resultant inviàbles. Esperem que, amb aquesta evolució contínua, problemes que encara no tenen solució avui en dia, com el que s'acaba d'explicar, siguin capaços de resoldre's amb unes poques línies de codi en el futur, tal com es fa ara amb problemes que en el passat eren impensables de resoldre.

Referències

- [1] Betancourt, M.; Byrne, S.; Livingstone, S.; Girolami, M.: The Geometric Foundations of Hamilton Monte Carlo. [arXiv:1410.5110v1 \[stat.ME\]](#), Octubre de 2014.
- [2] Betancourt, M.: A Conceptual Introduction to Hamilton Monte Carlo. [arXiv:1701.02434 \[stat.ME\]](#), Gener de 2017.
- [3] Bollobás, B.; Fulton, W.; Katok, A.; Kirkwan, F.; Sarnak, P.: *152 Permutation Group Algorithms*. 1a edició, Cambridge University Press, 2002.
- [4] Churchill, G.: Accurate restoration of DNA sequences. A: Gatsonis, C.; Hodges, J.S. (ed.): *Case Studies in Bayesian Statistics*, 2a edició, Springer, Nova York: 90-148, 1995.
- [5] Christensen, N.; Meyer, R: Markov chain Monte Carlo methods for Bayesian gravitational radiation data analysis. *Phys. Rev.*, Vol. 58, 1998.
- [6] Davidson-Pilon, C.: *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference*. Boston, Massachusetts: Addison-Wesley Professional, 2015. Recuperat 6 juny 2018, des de <http://camdavidsonpilon.github.io/Probabilistic-Programming-and-Bayesian-Methods-for-Hackers/>.
- [7] Diaconis, P.: The Markov Chain Monte Carlo Revolution. *Bull. Amer. Math. Soc.*, No. 46: 179-205, 2009.
- [8] Eckhardt, R.: Stan Ulam, John von Neumann and the Monte Carlo Method. *Los Alamos Science*, No. 15: 131-137, 1987.
- [9] Gelfand, A.E.; Smith, A.F.: Sampling-Based Approaches to Calculating Marginal Densities. *J. Am. Stat. Assoc.*, Vol. 85, No. 410: 398-409, 1990.
- [10] Geman, S.; Geman, D.: Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. PAMI-6, No. 6: 721-741, 1984.
- [11] George, E.; McCulloch, R.: Variable selection via Gibbs sampling. *J. Amer. Statist. Assoc.*, Vol. 88: 881-889, 1993.
- [12] Geyer, C.J.: The Metropolis-Hastings-Green Algorithm, [lecture notes], Desembre de 2003.
- [13] Geyer, C.J.: Introduction to Markov Chain Monte Carlo. A: Brooks, S.; Gelman, A.; Jones, G.; Meng, X.L. (ed.): *Handbook of Markov Chain Monte Carlo.*, Chapman & Hall / CRC Press, Cap. 1, 2011.
- [14] Gilks, W.R; Wild, P.: Adaptive Rejection Sampling for Gibbs Sampling. *Appl. Statist.*, Vol. 41, No. 2: 337-348, 1992.
- [15] Gosling, F.G.: *The Manhattan Project: Making the Atomic Bomb*. DOE/MA-0002 Revised. Washington, D.C.: Department of Energy, 2010.
- [16] Gravner, J.: Lecture notes for Introductory Probability, [lecture notes], Desembre 2017.

- [17] Green, P.J.: Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*, Vol. 82, No. 4: 711-732, 1995.
- [18] Hastings, W.K.: Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, Vol. 57, No. 1: 97-109, 1970.
- [19] Heckman, J.J.; Bernstein, J.G.; Vigoda, B.: MCMC with Strings and Branes: The Suburban Algorithm. [arXiv:1605.05334v2\[physics.comp-ph\]](https://arxiv.org/abs/1605.05334v2), Setembre de 2017.
- [20] Hobert, J.P.; Casella, G.: The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *J. Amer. Statist.*, Vol. 91: 1.461-1.473, 1996.
- [21] Hogg, D.W.; Foreman-Mackey, D.: Data Analysis Recipes: Using Markov Chain Monte Carlo. [arXiv:1710.06068v1\[astro-ph.IM\]](https://arxiv.org/abs/1710.06068v1), Octubre de 2017.
- [22] Fonnesbeck, C.: *Bayesian Estimation Supersedes the T-Test*, 2015. Recuperat 2 Juny 2018 des de <http://docs.pymc.io/notebooks/BEST.html>.
- [23] Konstantopoulos, T.: Markov Chains and Random Walks, [lecture notes], Tardor 2009.
- [24] Kruschke, J.K.: Bayesian Estimation Supersedes the t Test. *J. Exp. Psychol. Gen.*, Vol. 142, No. 2: 573-603, 2013. DOI:10.1037/a0029146.
- [25] Lewis, A.; Bridle, S.: Cosmological parameters from CMB and other data: a Monte-Carlo approach. [arXiv:astro-ph/0205436v3](https://arxiv.org/abs/astro-ph/0205436v3), Octubre de 2002.
- [26] Liang, F.; Wong, W.H.: Evolutionary Monte Carlo for protein folding simulations. *J. Chem. Phys.*, Vol. 115, No. 7: 3.374-3.380, 2001. DOI:10.1063/1.1387478.
- [27] Liu, J.S.; Wong, W.H.; Kong, A.: Covariance structure and convergence rate of the Gibbs sampler with various scans. *J. Roy. Statist. Soc. Ser. B*, Vol. 57: 157-169, 1995.
- [28] Lunn, D.; Spiegelhalter, D.; Thomas, A.; Best, N.: The BUGS project: Evolution, critique and future directions *Statist. Med.*, DOI:10.1002/sim.3680, Octubre de 2009.
- [29] Mau, B.; Newton, M.A; Larget, B.: Bayesian phylogenetic inference via Markov chain Monte Carlo methods. *Biometrics*, Vol. 55, No.1:1-12, 1999. DOI:10.1111/j.0006-341X.1999.00001.x.
- [30] Mengersen, K.L.; Tweedie, R.L.: Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.*, Vol. 24: 101-121, 1996.
- [31] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E.: Equations of State Calculations by Fast Computing Machines. *J. Chem. Phys.*, Vol. 21, No. 6: 1.087-1.092, 1953.
- [32] Metropolis, N.: The Beginning of the Monte Carlo Method. *Los Alamos Science*, No. 15: 125-130, 1987.

- [33] Mossel, E.;Vigoda, E.: Limitations of Markov Chain Monte Carlo Algorithms for Bayesian Inference of Phylogeny. [arXiv:q-bio/0505002v5](https://arxiv.org/abs/q-bio/0505002v5) [q-bio.PE], Febrer de 2007.
- [34] Neal, R.M.: Sampling from multimodal distributions using tempered transitions. *Statist. Comput.*, Vol. 6: 353-356, 1996.
- [35] Neal, R.M.: MCMC using Hamiltonian dynamics. A: Brooks, S.; Gelman, A.; Jones, G.; Meng, X.L. (ed.): *Handbook of Markov Chain Monte Carlo.*, Chapman & Hall / CRC Press, Cap. 5, 2011.
- [36] Newton, M.A.; Mau, B.; Larget, B.: Markov chain Monte Carlo for the Bayesian analysis of evolutionary trees from aligned molecular sequences. *F. Seillier-Moseiwitch (Ed.), Statistics in Molecular Biology and Genetics. IMS Lecture Notes-Monograph Series*, Vol. 33: 143-162, 1999. DOI:10.1214/lnms/1215455551.
- [37] Robert, C.; Casella, G.: A Short History of Markov Chain Monte Carlo: Subjective Recollections from Incomplete Data. *Stat. Sci.*, Vol. 26, No. 1: 102-115, 2011. [arXiv:0808.2902v7](https://arxiv.org/abs/0808.2902v7) [stat.CO].
- [38] Robert, C.: The Metropolis-Hastings algorithm. [arXiv:1504.01896v3](https://arxiv.org/abs/1504.01896v3) [stat.CO], Gener de 2016.
- [39] Rogozhnikov, A.: *Hamiltonian Monte Carlo explained*, 2016. Recuperat 10 Juny 2018 des de http://arogozhnikov.github.io/2016/12/19/markov_chain_monte_carlo.html.
- [40] Rubin, D.B.: Using the SIR algorithm to Simulate Posterior Distributions. A: Bernardo, J.M; DeGroot, M.H.; Lindley, D.V.; Smith, A.F.M. (ed.): *Bayesian Statistics 3*, Oxford, U.K., Oxford University Press: 395-402, 1988.
- [41] Sharma, S.: Markov Chain Monte Carlo Methods for Bayesian Data Analysis in Astronomy. [arXiv:1706.01629v1](https://arxiv.org/abs/1706.01629v1) [astro-ph.IM], Juny de 2017.
- [42] Tanner, M.A.; Wong, W.H.: The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.*, Vol. 82: 528-550, 1987.
- [43] Tierney, L.: Markov chains for exploring posterior distributions. *Ann. Statist.*, Vol. 22: 1.701-1.786, 1994.
- [44] Warnock, T.: Random-Number Generators. *Los Alamos Science*, No. 15: 137-141, 1987.
- [45] Yildirim, I.: Bayesian Inference: Gibbs Sampling, [lecture notes], Agost 2012.
- [46] Zegarlinski, B.: Dobrushin Uniqueness Theorem and Logarithmic Sobolev Inequalities. *J. Funct. Anal.*, Vol. 105: 77-111, 1992.
- [47] Žitković, G.: Introduction to Stochastic Processes, [lecture notes], Desembre 2010.