

Quants rècords veurem al llarg de la nostra vida?*

Xavier Bardina

1 Introducció i exemple

Les situacions extremes sempre han estat observades amb gran interès. En manuscrits antics ja es troben textos sobre estius extraordinàriament freds, sobre inundacions, sobre temperatures molt elevades, etc...

Naturalment hom s'interessa també en els rècords d'altres disciplines. Per exemple, *El llibre Guinness dels Rècords* (veure [5]) és molt popular a tot el món i sovint se'n treuen noves versions.

El punt de vista de la teoria probabilística dels rècords permet demostrar alguns resultats curiosos. Per exemple, en aquest treball veurem que si suposem que la variable *cota màxima de neu que cau a l'hivern* s'ajusta al nostre model, de mitjana, un nen d'onze anys haurà vist en la seva vida 3 rècords mentre que un avi de 83 anys n'haurà viscut només 5.

D'altra banda, aquest estudi no inclou els rècords de la majoria de disciplines esportives on els individus s'esforcen sistemàticament per obtenir cada cop millors resultats, és a dir, van a la recerca del rècord. Hi ha models per a estudiar aquest tipus de situacions però no els tractarem en aquesta nota.



Jou. Fotografia: Elisenda Vila

*Aquest treball està basat en la conferència inaugural del curs 2006–07 de la Llicenciatura de Matemàtiques de la Universitat Autònoma de Barcelona impartida per l'autor, precisament el dia 18 d'octubre.

La teoria dels rècords s'ha utilitzat com a model d'esdeveniments extrems en matemàtica financera (veure [4]). El llibre d'Arnold, Balakrishnan i Nagaraja (veure [2]) és el llibre de referència de la teoria probabilística dels rècords, però en [1] trobem una introducció d'aquesta teoria adreçada a una públic menys especialitzat. Si teniu interès en aprofundir una mica més en el tema també podeu consultar [3] que és una extensió d'aquest treball dirigida a lectors familiaritzats en el llenguatge de les demostracions matemàtiques.

Quan tenim una sèrie de nombres reals el primer valor el considerarem un rècord i a partir d'aquí, un valor és un rècord si és més gran que tots els valors anteriors. Per exemple, una temperatura és un rècord màxim si és la temperatura més elevada des de que tenim dades.

La **Taula 1**, conté les temperatures màximes que s'han assolit a Barcelona el dia 17 d'octubre des de l'any 1900 fins al 2006. Les dades s'han obtingut consultant l'hemeroteca del diari **La Vanguardia**. Hi ha alguns valors perduts corresponents a anys entre el 1935 i el 1947. El motiu és que el diari no va publicar la informació meteorològica.

Temperatures màximes a Barcelona el dia 17 d'octubre

	0	1	2	3	4	5	6	7	8	9
190_	24,3	17,2	21,8	23,7	21,3	20,7	19,6	17,0	21,7	23,4
191_	22,8	22,4	20,3	23,1	17,2	20,5	24,9	21,5	17,5	19,0
192_	23,2	24,4	21,5	23,8	20,7	23,1	25,2	18,7	20,2	22,8
193_	24,0	22,8	23,6	25,0	18,4	*	21,0	*	*	*
194_	*	20,6	*	19,4	*	*	*	*	24,8	22,4
195_	22,8	17,7	22,4	21,2	22,6	19,8	22,9	24,5	17,6	19,9
196_	16,6	22,5	19,0	22,0	18,6	19,6	22,0	25,2	20,2	20,2
197_	21,1	21,0	19,5	22,0	17,4	23,7	20,6	21,0	22,5	21,6
198_	16,0	25,4	23,4	18,6	19,0	20,0	20,6	20,2	25,4	21,8
199_	23,3	22,3	16,3	21,3	21,2	21,5	20,1	23,5	23,5	22,9
200_	21,2	23,5	21,5	18,8	20,2	22,1	21,9			

Taula 1: Temperatures màximes a Barcelona el dia 17 d'octubre en graus centígrads¹.

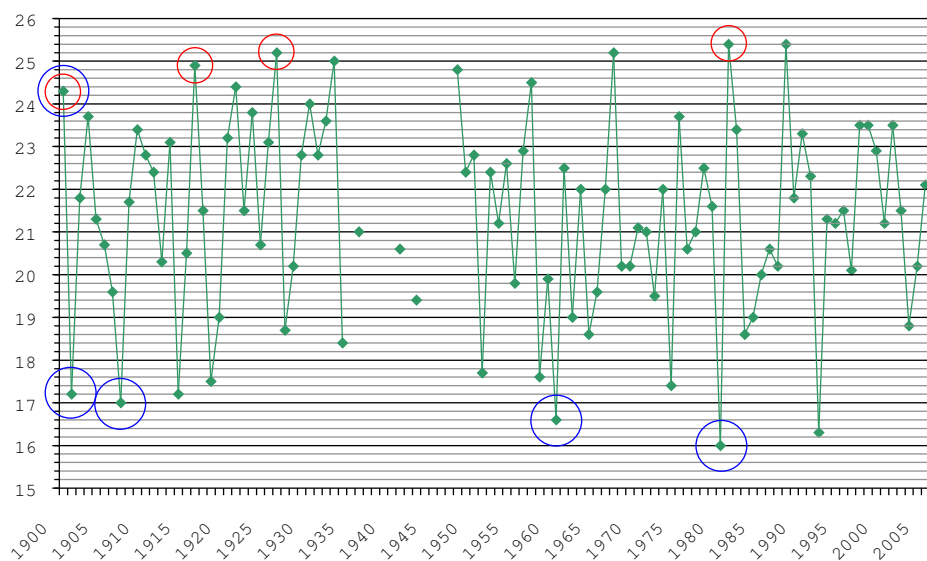
¹**Font:** Hemeroteca de **La Vanguardia**. Les temperatures són les publicades al diari el dia 18 d'octubre. Els dies en què el 18 d'octubre va caure en dilluns i no es va publicar el diari s'han agafat les temperatures del dia següent.

Si busquem a la taula, veurem que hi ha 4 rècords màxims (són els valors encerclats de color vermell) que corresponen als anys 1900, 1916, 1926 i 1981 i que prenen els valors 24,3; 24,9; 25,2 i 25,4.

Una observació important és que el fet que un valor sigui o no un rècord depèn de l'instant en el que comencem a comptar. Per exemple, en la **Taula 1** si enlloc de tenir dades des del 1900 tinguéssim dades des del 1901 tindríem 6 rècords enlloc de 4 que correspondrien als anys 1901, 1902, 1903, 1916, 1926 i 1981 i els valors serien 17,2; 21,8; 23,7; 24,9; 25,2 i 25,4.

De la mateixa forma també té sentit considerar els rècords mínims de la sèrie. És a dir, en el nostre exemple, els dies 17 d'octubre en què la temperatura màxima a Barcelona ha estat més petita que tots els anys anteriors dels que disposem de dades. En aquest cas són els anys 1900, 1901, 1907, 1960 i 1980 i els valors són 24,3; 17,2; 17,0; 16,6 i 16,0 (marcats en blau en la taula excepte el primer).

En el gràfic següent s'observen aquests rècords marcats amb cercles vermells (els màxims) i blaus (els mínims).



Temperatures màximes a Barcelona el dia 17 d'octubre en graus centígrads.

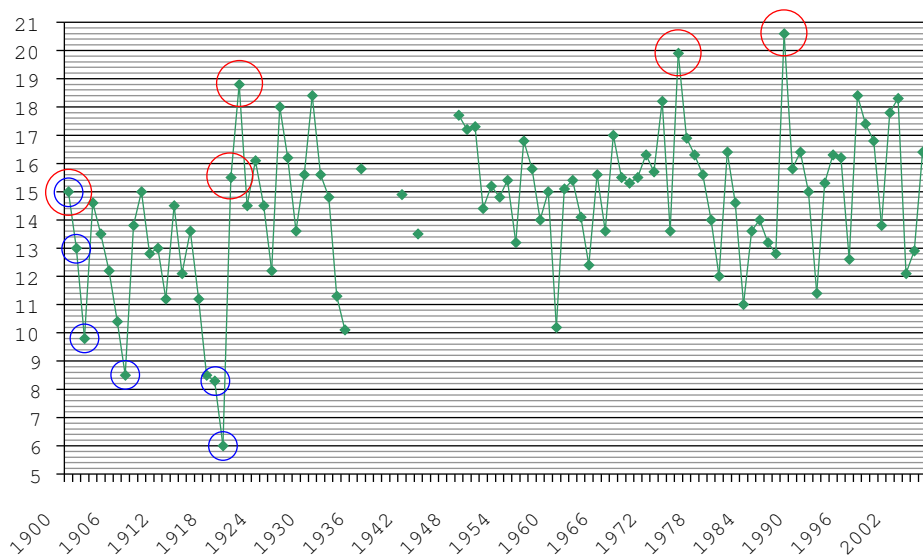
Considerem ara la sèrie de temperatures mínimes, vegeu la **Taula 2**, dels mateixos dies.

Temperatures mínimes a Barcelona el dia 17 d'octubre

	0	1	2	3	4	5	6	7	8	9
190_	15,0	13,0	9,8	14,6	13,5	12,2	10,4	8,5	13,8	15,0
191_	12,8	13,0	11,2	14,5	12,1	13,6	11,2	8,5	8,3	6,0
192_	15,5	18,8	14,5	16,1	14,5	12,2	18,0	16,2	13,6	15,6
193_	18,4	15,6	14,8	11,3	10,1	*	15,8	*	*	*
194_	*	14,9	*	13,5	*	*	*	*	17,7	17,2
195_	17,3	14,4	15,2	14,8	15,4	13,2	16,8	15,8	14,0	15,0
196_	10,2	15,1	15,4	14,1	12,4	15,6	13,6	17,0	15,5	15,3
197_	15,5	16,3	15,7	18,2	13,6	19,9	16,9	16,3	15,6	14,0
198_	12,0	16,4	14,6	11,0	13,6	14,0	13,2	12,8	20,6	15,8
199_	16,4	15,0	11,4	15,3	16,3	16,2	12,6	18,4	17,4	16,8
200_	13,8	17,8	18,3	12,1	12,9	16,4	17,5			

Taula 2: Temperatures mínimes a Barcelona el dia 17 d'octubre en graus centígrads.

Observant la Taula 2, veiem que hi ha 6 rècords mínims (marcats en blau) corresponents als anys 1900, 1901, 1902, 1907, 1918 i 1919 (aquest darrer any la temperatura mínima va ser de només 6,0 graus). També observem que hi ha 5 rècords màxims (marcats en vermell excepte el primer) que corresponen als anys 1900, 1920, 1921, 1975 i 1988 (aquest darrer any la temperatura mínima va ser de 20,6 graus). Com abans, aquesta informació queda reflectida molt clarament en el gràfic següent:



Temperatures mínimes a Barcelona el dia 17 d'octubre en graus centígrads.

En aquesta nota volem calcular quants rècords s'obtenen, de mitjana, en una sèrie com aquesta. Abans però cal fixar la notació i les hipòtesis amb les que treballarem. En l'última secció trobareu un altre exemple amb dades reals de la pluviometria a Palma de Mallorca.

2 Hipòtesis de treball

Considerem una sèrie de nombres reals que denotarem X_1, X_2, \dots, X_n . El primer valor X_1 el considerarem un rècord. A partir d'aquí, X_i serà considerat un rècord si es el valor més gran obtingut fins a l'instant i . És a dir, X_i serà un rècord si

$$X_i > \max(X_1, X_2, \dots, X_{i-1}) \quad \text{per a } i \geq 2.$$

Abans de continuar fixarem algunes hipòtesis per poder tractar aquests problemes:

- (i) Suposarem que X_1, X_2, \dots, X_n són variables aleatòries independents i idènticament distribuïdes (v.a.i.i.d.)
- (ii) Suposarem que la distribució d'aquestes variables és contínua.

Vegem que signifiquen aquestes hipòtesis:

- Què X_1, X_2, \dots, X_n siguin variables aleatòries vol dir que, a priori, no coneixem quins valors prenen. Nosaltres volem calcular quants rècords hi ha, de mitjana, en una sèrie de mida n . Per això necessitem que els valors siguin aleatoris, que hi hagi incertesa en el valor que poden prendre, perquè per un exemple concret ja sabem comptar el nombre de rècords.
- Que aquestes variables siguin idènticament distribuïdes vol dir que, a priori, totes tenen les mateixes probabilitats de prendre determinats valors. És a dir, en el nostre exemple estem suposant que, a priori, la probabilitat que la temperatura màxima del 17 d'octubre estigui, per exemple, entre 20 i 23 graus és la mateixa per a tots els anys.
- Què aquestes variables siguin independents vol dir que el valor concret que pugui prendre una d'elles no em diu res sobre el valor que pugui prendre una altra variable. És a dir, estem suposant que el fet de saber

que el 17 d'octubre de l'any 2006 la temperatura màxima a Barcelona ha estat exactament de 21,9 graus no em modifica les probabilitats sobre les temperatures de l'any vinent. És a dir, estem suposant que és igual de probable que la temperatura màxima el 17 d'octubre de l'any 2007 estigui entre 20 i 23 graus tant si se quin ha estat el valor d'aquesta temperatura l'any 2006, o l'any 1957, com si no tinc aquesta informació.

- Finalment estem suposant que aquestes variables són contínues. És a dir, en el nostre exemple tenim mesurades les temperatures amb una sola xifra decimal, però suposem que, teòricament, si tinguéssim instruments de mesura més bons podríem afinar prou en la mesura de les temperatures, amb prou xifres decimals, per tal que no es produïssin empats.

És precisament la hipòtesis (i) la que exclou d'aquest estudi els rècords de la majoria de disciplines esportives.

La hipòtesi (ii) és simplement tècnica i és per a excloure la possibilitat que un rècord pugui ser igualat; en ser les variables aleatòries X_1, X_2, \dots, X_n contínues sabem que la probabilitat que siguin diferents és igual a 1. De fet, aquest és el motiu que fa que el cas discret sigui molt més complicat de tractar.

Nosaltres ens interessarem per la variable aleatòria:

$$R_n = \text{nombre de rècords de la sèrie } X_1, X_2, \dots, X_n.$$

Una observació important és que el valor d'aquesta variable que acabem d'introduir només depèn de com estan ordenades les variables X_1, X_2, \dots, X_n en la sèrie.

En aquesta nota estudiarem els rècords màxims, però l'estudi també serveix per als rècords mínims ja que els rècords mínims de la sèrie X_1, X_2, \dots, X_n són rècords màxims si considerem la sèrie que resulta de canviar tots els signes $-X_1, -X_2, \dots, -X_n$.

Un exemple d'aquest cas és el següent: podem considerar que les velocitats (a la que circularien si no hi hagués cap obstacle) dels vehicles que van pel carril addicional d'una autopista segueixen les hipòtesis amb les que estem treballant. És a dir, aquestes velocitats X_1, X_2, \dots, X_n són v.a.i.i.d. i la seva distribució és contínua.

Passat un cert nombre de quilòmetres, quan la velocitat d'un vehicle és un rècord mínim crea una caravana i, al no poder avançar, aquesta caravana estarà formada per tots els cotxes fins que aparegui un altre rècord mínim de velocitat que crearà novament una altra caravana, etc...

Segons l'estudi que veurem, si el flux de vehicles que entren al carril adicional és important, cada cop les caravanes de vehicles tindran una mida més gran.

3 El concepte d'esperança matemàtica

El concepte de mitjana teòrica es coneix amb el nom de nombre esperat o esperança matemàtica.

Si llancem un dau normal esperem que, de mitjana, en un de cada sis llançaments obtindrem un quatre. Això no vol dir que si llancem 60 vegades el dau obtindrem exactament 10 cops un quatre i 50 vegades un número diferent de quatre. Sinó que admetem que si repetíssim molts cops l'experiment de llançar 60 vegades un dau i comptar el nombre de quatres, de vegades n'obtidríem exactament 10, de vegades més de 10 i de vegades menys de 10. Però si féssim la mitjana del nombre de quatres obtinguts esperem que doni un nombre molt proper a 10.

En matemàtiques diem que si considerem la variable aleatòria

$Y =$ nombre de quatres obtinguts en el llançament de 60 cops un dau

l'esperança matemàtica d'aquesta variable és 10 i ho escriurem

$$\mathbb{E}(Y) = 10.$$

Quan tenim una variable aleatòria discreta X que pot prendre n valors diferents amb diferents probabilitats:

$$X = \begin{cases} x_1 & \text{amb probabilitat } p_1 \\ x_2 & \text{amb probabilitat } p_2 \\ \vdots & \vdots \\ x_n & \text{amb probabilitat } p_n, \end{cases}$$

l'esperança matemàtica es calcula de la forma següent:

$$\mathbb{E}(X) = x_1 \times p_1 + x_2 \times p_2 + \cdots + x_n \times p_n.$$

Una propietat important és que l'esperança matemàtica de la suma de variables aleatòries és la suma d'esperances. Això ens permet calcular de forma fàcil l'esperança en l'exemple anterior.

Si llancem un cop un dau i considerem la variable:

$$Y_1 = \text{nombre de quatres obtinguts en llançar un cop un dau.}$$

Aquesta variable només pot prendre dos valors diferents:

$$Y_1 = \begin{cases} 1 & \text{amb probabilitat } p_1 = \frac{1}{6} \\ 0 & \text{amb probabilitat } p_0 = \frac{5}{6}. \end{cases}$$

En efecte, quan llancem un dau pot ser que obtinguem un quatre (cosa que passa amb probabilitat $\frac{1}{6}$, una de cada sis vegades) o que no obtinguem cap quatre (és a dir, que surti un nombre diferent de 4, cosa que passa cinc de cada sis cops).

Aquest tipus de variables aleatòries que només poden prendre dos valors diferents s'anomenen variables de tipus Bernoulli.

Segons la definició que hem donat, l'esperança matemàtica d'aquesta variable serà:

$$E(Y_1) = 1 \times p_1 + 0 \times p_0 = 1 \times \frac{1}{6} + 0 \times \frac{5}{6} = \frac{1}{6}.$$

Que expressa que de mitjana s'obtenen $\frac{1}{6}$ quatres en llançar un dau. És a dir, que de mitjana s'obté un quatre en un de cada sis llançaments.

La variable aleatòria Y la podem posar com la suma de 60 variables com l'anterior:

$$Y = Y_1 + Y_2 + \dots + Y_{60}.$$

El nombre de quatres en llançar 60 cops un dau és la suma del nombre de quatres obtinguts en cada un dels 60 llançaments. O dit d'una altra manera, si sumem 1 quan surt un quatre i 0 quan no, el total serà el nombre de quatres dels 60 llançaments.



Jakob Bernoulli (1654-1705)

Utilitzant la propietat que hem comentat abans, que l'esperança de la suma és la suma d'esperances, tenim que:

$$\mathbb{E}(Y) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \cdots + \mathbb{E}(Y_{60}) = \frac{1}{6} + \frac{1}{6} \cdots + \frac{1}{6} = 60 \times \frac{1}{6} = 10.$$

Per calcular el nombre esperat de rècords d'una sèrie farem un càlcul molt semblant.

4 Nombre esperat de rècords

Sigui X_1, X_2, \dots, X_n una sèrie de v.a.i.i.d. amb distribució contínua. Començarem calculant la probabilitat que el valor X_n sigui un rècord.

Les variables aleatòries X_1, X_2, \dots, X_n són diferents amb probabilitat 1, i per tant hi ha $n!$ formes diferents d'ordenar-les. En efecte, això seria equivalent a comptar de quantes formes diferents es poden ordenar n llibres en una lleixa: escollim un llibre i el col.loquem, això ho podem fer de n formes diferents. Ara hem d'escollir un segon llibre i ens queden $n - 1$ llibres, per tant tenim $n - 1$ opcions. I així successivament fins que ens queda un sol llibre. Per tant, podem ordenar els n llibres en una lleixa de

$$n \times (n - 1) \times (n - 2) \times \cdots \times 1 = n!$$

formes diferents.

El fet que aquestes variables siguin independents i idènticament distribuïdes implica que la seva llei conjunta és simètrica. Això significa que totes les ordenacions, recordem que n'hi havia $n!$, són equiprobables.

D'altra banda, que X_n sigui un rècord vol dir que és el valor més gran, és a dir, que X_n sigui un rècord és equivalent a que la variable més gran aparegui en darrer lloc. Volem ara comptar en quantes de les diferents ordenacions passa això. Hem de fixar la variable més gran per a que aparegui la última i comptar de quantes formes es poden ordenar les altres variables. Ens queden per col.locar $n - 1$ variables i per tant es podran ordenar de $(n - 1)!$ formes diferents.

Així, si denotem per p_n la probabilitat que X_n sigui un rècord, podem calcular-la de la següent forma,

$$p_n = \frac{\text{nombre d'ordenacions amb les quals } X_n \text{ és un rècord}}{\text{nombre d'ordenacions possibles}} = \frac{(n - 1)!}{n!} = \frac{1}{n}.$$

Introduïm ara les següents variables aleatòries,

$$Y_i := \begin{cases} 1 & \text{si } X_i \text{ es un rècord} \\ 0 & \text{altrament} \end{cases}$$

per $i = 1, \dots, n$.

Observem que $\mathbb{E}(Y_i) = 1 \cdot p_i + 0 \cdot (1 - p_i) = p_i = \frac{1}{i}$.

De manera que el nombre total de rècords de la sèrie X_1, X_2, \dots, X_n , que denotarem per R_n , serà igual a

$$R_n = Y_1 + Y_2 + \dots + Y_n$$

i per tant,

$$\mathbb{E}(R_n) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \dots + \mathbb{E}(Y_n) = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{n} = \sum_{i=1}^n \frac{1}{i}.$$

Veiem que la suma anterior $\sum_{i=1}^n \frac{1}{i}$ es va fent arbitràriament gran a mesura que n creix. Per això observem el següent:

$$\begin{aligned} \sum_{i=1}^{\infty} \frac{1}{i} &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots + \frac{1}{m} + \dots \\ &= 1 + \frac{1}{2} + \overbrace{\frac{1}{3} + \frac{1}{4}}^{2 \text{ termes}} + \overbrace{\frac{1}{5} + \dots + \frac{1}{8}}^{4 \text{ termes}} + \overbrace{\frac{1}{9} + \dots + \frac{1}{16}}^{8 \text{ termes}} + \dots \\ &> 1 + \frac{1}{2} + \overbrace{\frac{1}{4} + \frac{1}{4}}^{2 \text{ termes}} + \overbrace{\frac{1}{8} + \dots + \frac{1}{8}}^{4 \text{ termes}} + \overbrace{\frac{1}{16} + \dots + \frac{1}{16}}^{8 \text{ termes}} + \dots \\ &= 1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots = \infty. \end{aligned}$$

Per tant direm que la suma de la sèrie harmònica $1 + \frac{1}{2} + \frac{1}{3} + \dots$ és infinit. Però la convergència d'aquesta sèrie és molt lenta: si un ordinador sumés 10^6 termes de la sèrie per segon, després de $3,17 \times 10^{85}$ anys hauria sumat fins

al terme $\frac{1}{10^{99}}$ i el resultat que hauria obtingut seria²

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{10^{99}} = 228,5331\dots$$

És a dir, en una sèrie formada per 10^{99} termes, el nombre de rècords esperat és de només 228,5. Així doncs, encara que matemàticament parlant el nombre de rècords que s'obté creix indefinidament, la convergència és molt i molt lenta.

En la següent taula veiem, per a determinats valors de N , el valor de n per al qual la suma de la sèrie és per primer cop més gran o igual que N . És a dir, per a cada N el valor de n per al qual el nombre esperat de rècords, $E(R_n)$, supera N per primera vegada:

N	2	3	4	5	6	7	8	9	10	100
n	4	11	31	83	227	616	1674	4550	12367	$1,5 \times 10^{43}$

Taula 3: Per a cada N la taula mostra el valor de n per al qual el nombre esperat de rècords supera N per primera vegada.

Suposem que X_i representa, per exemple, la cota màxima de neu caiguda durant l'hivern de l'any i -èssim de vida d'una persona, de mitjana, un nen d'11 anys hauria vist en la seva vida 3 rècords, mentre que als 31 anys hauria viscut 4 rècords i la mitjana de rècords que hauria vist un avi de 83 anys seria només de 5. La pregunta natural és si això no tindrà alguna cosa a veure amb el fet que algunes persones pensin que abans nevava més, feia més calor, etc...

En l'exemple de les temperatures, que hem vist en la Secció 1, teníem $n = 97$ i el nombre esperat de rècords és

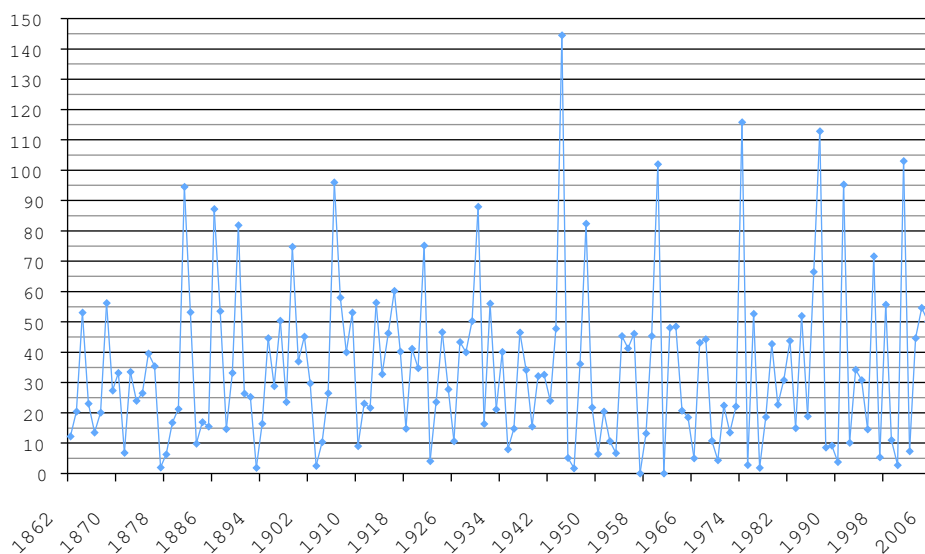
$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{97} = 5,1571.$$

És a dir, en els quatre exemples concrets que hem vist (temperatures màximes i mínimes i rècords màxims i mínims en cada cas) hem obtingut 4, 5, 6 i 5 rècords. Però sota les hipòtesis que hem imposat, de mitjana, en una sèrie com aquesta formada per 97 observacions, apareixen 5,1571 rècords.

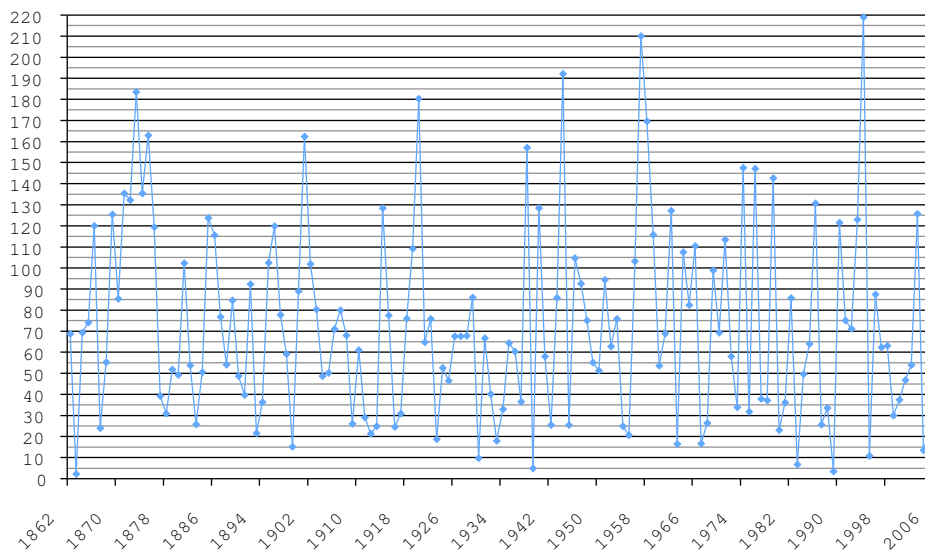
²Aquesta suma ha estat calculada usant els manipuladors algebraics Maple i Mathematica obtenint en els dos casos el mateix resultat. Vegeu la pagina web de Wolfram MathWorld (<http://mathworld.wolfram.com/HarmonicSeries.html>) per a més detalls sobre els mètodes que s'usen per a calcular-la.

5 Un altre exemple: la pluja caiguda a Palma de Mallorca

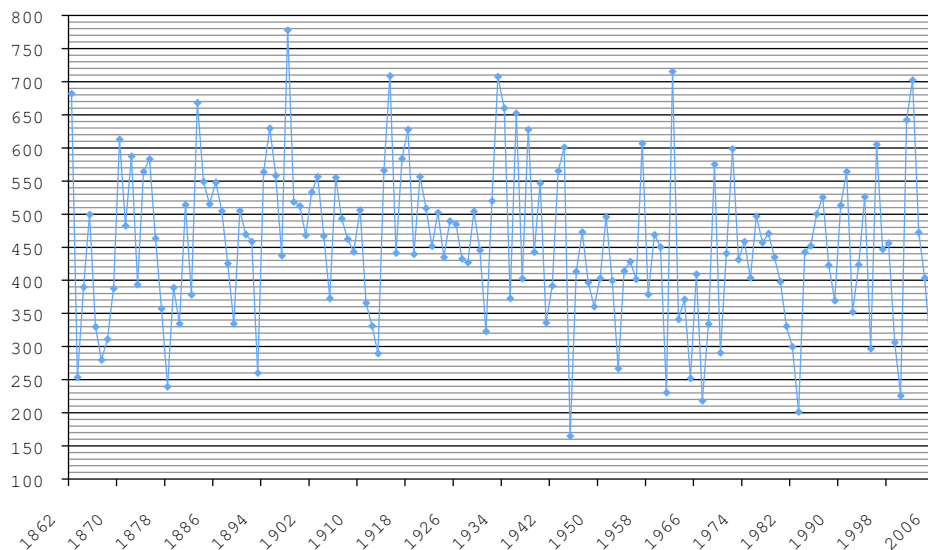
En els gràfics següents es mostra l'evolució de la pluviositat (litres per metre quadrat) a Palma de Mallorca des de l'any 1862 fins al 2006 dels mesos de febrer i octubre així com l'acumulada de cada any.



Pluviositat dels mesos de febrer.



Pluviositat dels mesos d'octubre.



Pluviositat anual.

Observeu que en les tres sèries de 145 dades hi ha respectivament 7, 10 i 2 rècords màxims i 6, 2 i 4 rècords mínims. Això dóna una mitjana d'una mica més de 5 rècords entre les 6 sèries. Aquest resultat s'ajusta força al nombre de rècords esperats per a sèries d'aquesta longitud que podeu consultar en la **Taula 3**.

Agraïments: El meu agraïment per a l'Elisenda Vila per la fotografia del nostre poble nevada. També per a en Joan Guasp Alemany per les seves gestions amb l'“Instituto Nacional de Meteorología”.

Nota dels editors: La informació meteorològica de la Secció 5 ha estat facilitada per l'“Instituto Nacional de Meteorología. Ministerio de Medio Ambiente”. No està permesa la seva reproducció total o parcial.

Referències

- [1] Jirí Andel, *Mathematics of Chance*. Wiley Series in Probability and Statistics, John Wiley & Sons, Inc. (2001)

- [2] Barry C. Arnold, N. Balakrishnan i H. N. Nagaraja, *Records*. Wiley series in probability and statistics (1998)
- [3] Xavier Bardina, *Rècords: Quina és la probabilitat d'obtenir-ne? Quan apareixen? Quins valors prenen?*, Butlletí de la Societat Catalana de Matemàtiques (per a aparèixer)
- [4] Paul Embrechts, Claudia Klüppelberg i Thomas Mikosch, *Modelling extremal events for insurance and finance*. Springer (1997)
- [5] Diversos autors, *Llibre Guinness dels Rècords*. Versió electrònica; <http://www.guinnessworldrecords.com>



Departament de Matemàtiques
Universitat Autònoma de Barcelona
Xavier.Bardina@uab.cat

Publicat el 7 de març de 2007