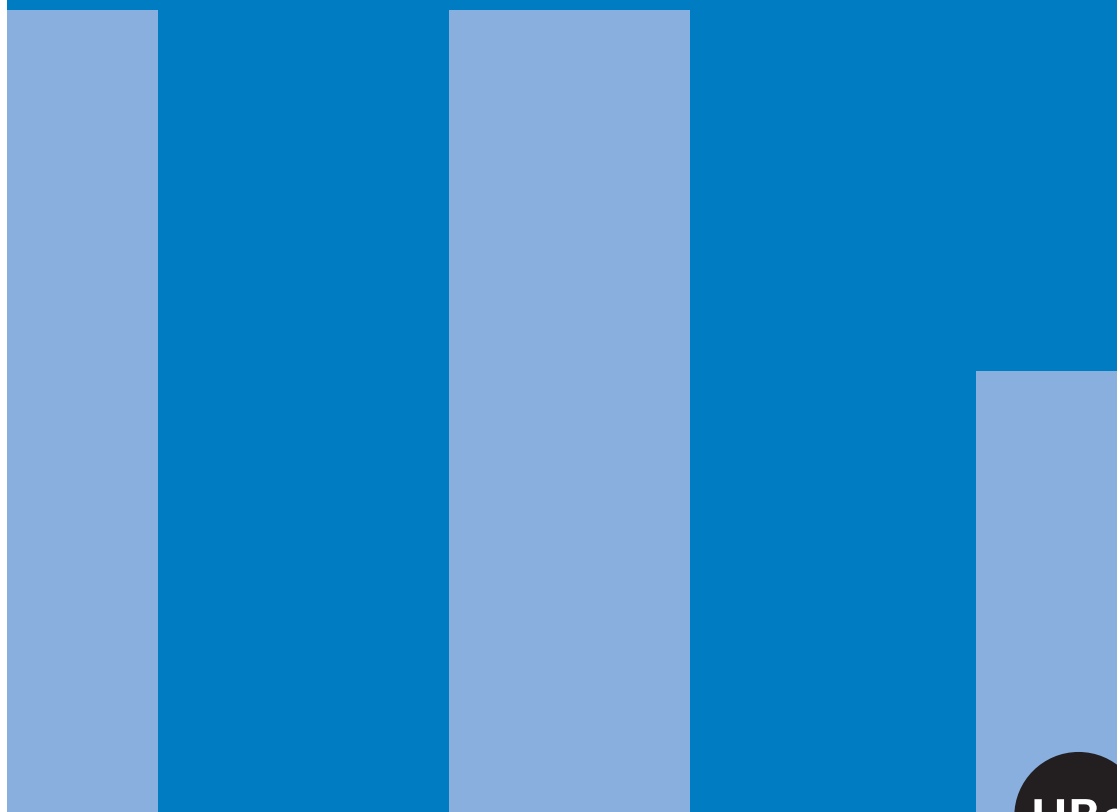


Un primer curs d'estadística

OLGA JULIÀ I DAVID MÁRQUEZ-CARRERAS



UN PRIMER CURS D'ESTADÍSTICA

UN PRIMER CURS D'ESTADÍSTICA

Olga Julià
David Márquez-Carreras

Publicacions i Edicions



UNIVERSITAT DE BARCELONA



ÍNDEX

<i>Introducció</i>	13
1. L'estadística	15
1.1. La seva definició	15
1.2. Una mica d'història	16
1.2.1. El càlcul de probabilitats	16
1.2.2. L'estadística	18
1.2.3. L'estadística avui dia	20
1.3. El present i el futur	21
1.4. Referències complementàries	22
2. Resultats teòrics de probabilitats	23
2.1. Espai de probabilitat	23
2.2. Probabilitat condicionada i independència	24
2.3. Variables aleatòries	25
2.3.1. Variables aleatòries discretes	26
2.3.2. Variables aleatòries absolutament contínues	27
2.3.3. Definicions i propietats relacionades amb l'esperança matemàtica ..	29
2.4. Vectors aleatoris	30
2.4.1. Independència	31
2.4.2. Vectors aleatoris discrets	32
2.4.3. Vectors aleatoris absolutament continus	33
2.4.4. Definicions i propietats relacionades amb l'esperança matemàtica ..	35
2.5. Successions de variables aleatòries	36
2.5.1. Convergència quasi segura	37
2.5.2. Convergència en probabilitat	38
2.5.3. Convergència en L^p	39
2.5.4. Convergència en llei	39
2.5.5. Relacions entre les convergències	40

2.6. Llei forta dels grans nombres	41
2.7. Funcions característiques	41
2.8. Teorema del límit central	45
2.9. Referències complementàries	45
3. Estadística descriptiva	47
3.1. Conceptes generals bàsics	47
3.2. Tipus de mostreig	48
3.3. Dades univariants	49
3.3.1. Característiques numèriques	49
3.3.2. Dades univariants en taules de freqüències	52
3.3.3. Dades agrupades per intervals	54
3.3.4. Representació de dades	56
3.4. Dades bivariants	59
3.4.1. Taules de freqüències	59
3.4.2. Característiques numèriques	60
3.5. Referències complementàries	61
4. El model estadístic	63
4.1. El model estadístic	63
4.1.1. Procediments de l'estadística	64
4.2. Exemples de models estadístics	65
4.2.1. Exemple de control de qualitat	65
4.2.2. Exemple de succés poc freqüent	66
4.2.3. Exemple de tria de nombres reals	67
4.2.4. Exemple de temps de vida	67
4.2.5. Exemple de mesurament amb error	68
4.3. Funció de versemblança	68
4.4. Models exponencials	70
4.4.1. Exemples de funció de versemblança i models exponencials	70
4.5. Referències complementàries	72
5. Suficiència i informació	75
5.1. Suficiència	75
5.1.1. Propietats	80
5.2. Informació	81
5.2.1. Informació de Fisher	81
5.2.2. Relació amb els models exponencials i la suficiència	87
5.2.3. Informació de Kullback	90
5.3. Referències complementàries	92
6. Teoria de la decisió	93
6.1. La formulació	93

6.2. Principis estadístics	95
6.3. Referències complementàries	97
7. Estimació puntual	99
7.1. Estimadors sense biaix	100
7.2. Estimadors uniformement de mínima variància	104
7.3. Cota de Cramer-Rao i eficiència	108
7.4. Propietats asimptòtiques dels estimadors	114
7.4.1. Consistència	114
7.4.2. Altres propietats asimptòtiques	116
7.5. Referències complementàries	116
8. Mètodes d'estimació puntual	119
8.1. Mètode dels moments	119
8.2. Mètode de la màxima versemblança	122
8.2.1. Propietats dels estimadors de màxima versemblança	126
8.3. Referències complementàries	132
9. Regions de confiança	133
9.1. Definició de regió de confiança	133
9.2. Mètode del pivot	134
9.3. Teorema de Fisher	137
9.4. Regions de confiança en poblacions normals	139
9.4.1. Intervals de confiança per a una mostra	139
9.4.2. Interval de confiança per a dues mostres independents	141
9.4.3. Exemples	144
9.5. Mètode de Neyman	145
9.6. Intervals de confiança asimptòtics	147
9.6.1. Interval de confiança per a la mitjana d'una població	148
9.6.2. Interval de confiança mitjançant l'EMV	149
9.6.3. Interval de confiança per una proporció	150
9.6.4. Interval de confiança per la diferència de proporcions	151
9.7. Interval de confiança utilitzant Txeibixev	152
9.8. Referències complementàries	152
10. Tests d'hipòtesis	153
10.1. Definicions i notacions bàsiques	153
10.2. Contrast entre hipòtesis simples	160
10.3. Extensió a contrastos d'hipòtesis unilaterals	166
10.4. Test de la raó de versemblança	168
10.5. Test de la raó de versemblança i normalitat	178
10.6. Referències complementàries	181

11. Teorema fonamental de l'estadística	183
11.1. La distribució empírica	183
11.2. El teorema de Glivenko-Cantelli	184
11.3. Referències complementàries	188
12. Ajustament, homogeneïtat i independència	189
12.1. Test d'ajustament	190
12.1.1. Test de la χ^2	190
12.1.2. Test de Kolmogorov	200
12.2. Tests d'homogeneïtat	203
12.2.1. Test de la χ^2 d'homogeneïtat	204
12.2.2. Test de Kolmogorov-Smirnov	206
12.3. Tests d'independència	207
12.4. Referències complementàries	210
13. Altres tests no paramètrics	211
13.1. Tests d'aleatorietat	211
13.1.1. Test de ratxes per a variables dicotòmiques	212
13.1.2. Test de ratxes per a variables quantitatives	213
13.1.3. Test de Von Neumann	213
13.2. Tests de localització o posició	213
13.2.1. Test de signes	214
13.2.2. Test dels rangs signats de Wilcoxon	215
13.2.3. Test de la mediana	217
13.2.4. Test de Wald-Wolfowitz	217
13.2.5. Test de Mann-Whitney	218
13.2.6. Test de Kruskal-Wallis	219
13.3. Tests d'associació	220
13.3.1. Test de correlació de rangs de Spearman	220
13.3.2. Test de Kendall	221
13.4. Referències complementàries	222
14. Model de regressió lineal simple	223
14.1. Estimació dels coeficients del model lineal	225
14.1.1. Mètode dels mínims quadrats	225
14.1.2. Propietats de l'ajust per mínims quadrats	227
14.1.3. Propietats dels estimadors dels coeficients de regressió	230
14.1.4. Estimador de la variància σ^2	234
14.2. Model normal de regressió simple	235
14.2.1. Intervals de confiança i proves d'hipòtesi	236
14.2.2. Predicció	237
14.2.3. Residus	238
14.3. Referències complementàries	240

15. Lleis més usals	241
15.1. Variables aleatòries discretes	241
15.2. Variables aleatòries absolutament contínues no relacionades amb la normal	244
15.3. La distribució normal	246
15.4. Distribucions derivades de la normal	247
15.4.1. χ^2 de Pearson	247
15.4.2. t de Student	248
15.4.3. F de Fisher-Snedecor	249
15.5. Vectors aleatoris més usals	249
15.5.1. Multinomial	249
15.5.2. Uniforme	250
15.6. La normal multidimensional	250
15.7. Referències complementàries	252
<i>Referències bibliogràfiques</i>	253

INTRODUCCIÓ

Aquest és un llibre dedicat bàsicament a la teoria de l'estadística, una de les branques de la matemàtica que més s'ha desenvolupat des de la meitat del segle XX. Aquest auge té dos motius fonamentals: d'una banda, l'aparició de l'ordinador digital, i, de l'altra, un canvi en la metodologia estadística.

Fins fa poques dècades l'estadística era vista amb escepticisme i no hi havia professionals que s'hi dediquessin ni estudis universitaris que possibilitessin obtenir-ne títols acadèmics. Actualment, la situació ha canviat radicalment. A més, l'estadística s'aplica a una infinitat de dominis: la sociologia, la indústria, la meteorologia, l'experimentació agrícola, l'epidemiologia, la farmacologia, la investigació de mercats, la codificació de missatges, el disseny i la planificació d'experiments científics, l'arqueologia i un llarg etcètera.

És fonamentalment un llibre de teoria sobre la inferència estadística però que també conté alguns exemples i alguns problemes per il·lustrar aquests aspectes teòrics. El llibre es pot seguir perfectament amb només certs coneixements d'anàlisi matemàtica. També seria apropiat tenir coneixements de probabilitats, però cal tenir en compte que el llibre presenta un recull de resultats de probabilitats resumits en un capítol.

El llibre està destinat a estudiants d'enginyeries o graus en matemàtiques o estadística. Això motiva que en certs resultats no es pugui aprofundir suficientment per manca dels prerequisits necessaris; per pal·liar aquesta mancança cada capítol conté una última secció en què s'ofereixen tot un seguit de referències complementàries per a l'estudiant que vulgui endinsar-se més en aquesta matèria.

El primer capítol presenta una visió global de l'estadística, en repassa la història i en prediu el futur. En el segon hi ha un recull de tots els resultats de probabilitats necessaris per poder seguir aquest llibre. I en el tercer s'hi troba un resum dels aspectes més importants de l'estadística descriptiva.

En el quart presentem el que entenem per model estadístic i donem exemples molt detallats per comprendre'l, així com altres conceptes relacionats amb els models estadístics. En el cinquè introduïm què vol dir tenir informació suficient d'un model. I en el sisè formulem a grans trets la teoria de la decisió.

Els capítols set i vuit estan dedicats a l'estimació puntual i als seus mètodes i el capítol nou tracta de les regions de confiança.

Tota la teoria dels tests d'hipòtesis està introduïda en el capítol deu. En l'onze parlem de la distribució empírica i demostrem l'anomenat teorema fonamental de l'estadística, el teorema de Glivenko-Cantelli.

Els dos capítols següents tracten de l'estadística no paramètrica. En el capítol dotze s'estudien els tests d'ajustament, homogeneïtat i independència més usuals; i en el tretze es parla d'altres tests coneguts no paramètrics.

Finalment, en l'últim capítol oferim un índex amb les variables i els vectors aleatoris més usuals.

1. L'ESTADÍSTICA

Aquest primer capítol és una introducció a l'estadística. Mitjançant una breu presentació històrica s'observa com l'estadística actual és fruit de la unió de dues disciplines que inicialment van evolucionar de manera independent fins a confluïr en el segle XIX: el càlcul de probabilitats i l'estadística o ciència d'estat.

Aquesta part del llibre només pretén situar el lector dins el món de l'estadística i donar a conèixer els noms dels científics i particularment dels matemàtics que més han fet perquè aquesta especialitat avanci; òbviament, si es desitja trobar informació més profunda i rigorosa, al final d'aquest capítol s'ofereixen tot un seguit de referències que tracten de la història de l'estadística.

Finalment, cal indicar que si algun lector només està interessat en la part purament matemàtica del llibre pot saltar-se tranquil·lament aquest capítol.

1.1. La seva definició

Els diferents autors no es posen d'acord sobre l'origen de la paraula *estadística*; n'hi ha que diuen que prové del llatí *status* 'estat o situació actual d'una cosa'; d'altres diuen que ve de l'alemany *staat* 'estat'; i uns pocs parlen d'una altra paraula llatina, *statera* 'balança', perquè mesura, pesa, els efectes, les causes i les forces socials. No obstant això, sembla que la paraula fou introduïda per Godofredo Achenwall cap el 1748.

Tenim també un bon grapat de definicions; de totes les que hem trobat, que són moltes, en donem dues, una per la seva procedència i l'altra per la seva bellesa:

- L'*Enciclopedia universal ilustrada* la definia com «La ciència que té per objecte aplicar les lleis de la quantitat als fets socials per mesurar-ne la intensitat, deduir-ne les lleis que els regeixen i fer-ne una predicció».

- L'any 1982 V. Barnett proposava la següent: «És la ciència que ens indica com ha de ser utilitzada la informació, amb l'objectiu de reflectir i donar una guia d'acció en situacions pràctiques que envolten incertesa».

La podem considerar com la ciència, el mètode o la tècnica que ens permet estudiar numèricament, amb la màxima precisió, els fenòmens col·lectius que no es coneixen d'una manera completa. El més raonable per estudiar un d'aquests fenòmens parcialment conegut és partir de la realitat i efectuar observacions directament o de forma planificada sobre una part dels individus –una mostra–, per poder establir així una teoria. Ordenant, escollint, agrupant convenientment i estudiant les dades observades, s'obindrà informació sobre el conjunt de tota la població. Aquest pas, de la mostra a tota la població, constitueix la *inferència estadística*, basada en el *càlcul de probabilitats*. Podríem dir, formalment, que actuen a l'inrevés, és a dir, el *càlcul de probabilitats* estudia els efectes quan les causes són conegudes i la *inferència estadística* intenta determinar certes causes quan els efectes han estat mesurats.

1.2. Una mica d'història

El coneixement de la història d'una disciplina és important, tant per comprendre'n l'estat actual de desenvolupament, com per preveure'n el futur, i en el seu aspecte més lúdic per adquirir una certa cultura general sobre el tema en qüestió.

1.2.1. El càlcul de probabilitats

El primer objecte d'estudi de la probabilitat van ser els jocs d'atzar. Tractar d'esbrinar quins van ser els primers és certament molt difícil, si bé podem aventurar que el joc de la taba (dau rudimentari obtingut a partir del taló de certs animals com l'ós, el cérvol o l'ovella) va ser un dels primers que va utilitzar l'ésser humà per divertir-se. Les excavacions arqueològiques indiquen una antiguitat d'uns 40.000 anys. També s'ha documentat l'ús de jocs d'atzar en les civilitzacions egípcia, grega, romana, tibetana, índia i jueva. En l'antiguitat l'atzar s'explicava sovint mitjançant la voluntat divina.

El Renaixement suposa un nou enfocament del món, ja que s'abandonen les explicacions teològiques. Els matemàtics italians de principis del segle XVI

comencen a interpretar els resultats d'experiments aleatoris simples. G. Cardano (1501-1576) estableix l'equiprobabilitat de l'aparició de cares d'un dau a llarg termini. N. Tartaglia (1499-1557) aborda diversos problemes de combinatòria. Galileo Galilei (1564-1642), mitjançant l'observació i l'estudi de la naturalesa amb el telescopi, va aconseguir progressos sense la necessitat de recórrer als jocs d'atzar.

El desenvolupament de l'anàlisi matemàtica dels jocs d'atzar es va produir molt lentament entre finals del segle XVI i principis del XVII. Molts historiadors consideren com a origen del càlcul de probabilitats la resolució l'any 1654 de certs problemes sobre alguns jocs d'atzar en la correspondència entre B. Pascal (1623-1662) i P. Fermat (1601-1665). C. Huygens (1629-1695) va publicar el 1657 el primer llibre de teoria de probabilitat, *De Rationiis in Ludo Aleae*, i va escriure sobre els problemes de Pascal i Fermat a un amic aquell mateix any: «M'agradaria creure que considerant de prop aquestes matèries, el lector observarà que estem tractant no només jocs d'atzar sinó més aviat amb els fonaments d'una nova teoria, a la vegada profunda i interessant...».

El càlcul de probabilitats es va consolidar com a disciplina independent entre mitjan segle XVII i començament del XVIII però s'aplicava encara bàsicament als jocs d'atzar.

Durant el segle XVIII les probabilitats es van estendre a problemes físics i marítims. Un factor que va incidir en aquesta direcció van ser els problemes que van sorgir a causa de la contrastació empírica de la teoria d'I. Newton (1642-1727), com per exemple el tractament dels errors de medicació.

És en aquesta època quan J. Bernoulli (1654-1705), A. de Moivre (1667-1754) i T. Bayes (1702-1761) van desenvolupar eines més potents i van estudiar nous problemes. El 1713 J. Bernoulli va publicar un tractat titulat *Ars Conjectandi*: el primer volum tracta de la teoria de probabilitats; en el segon i tercer treballa amb la teoria general de permutacions i combinacions; i en el quart hi ha el famós teorema que porta el seu nom, és a dir, la llei feble dels grans nombres. A. De Moivre va investigar les lleis d'atzar. Se li atribueixen, entre moltes altres aportacions, el teorema de la probabilitat composta i l'aproximació del límit d'una distribució binomial a una distribució normal. T. Bayes va introduir la fórmula que permet invertir l'ordre dels condicionats. Per acabar aquest segle cal destacar també G.W. Leibniz (1646-1716), que va mantenir correspondència amb tots ells i va participar en les seves nombroses troballes.

A la segona meitat del segle XVIII, els astrònoms i físics comencen a estimar magnituds a partir de diverses observacions. Necessitaven noves eines i amb aquesta finalitat els matemàtics van cercar una base teòrica. El primer impuls fo-

amental d'aquest període prové de P. Simon, marquès de Laplace (1749-1827). Va treure de l'oblit l'obra de T. Bayes i va estudiar el mètode de la probabilitat condicionada invertida, va fer la primera definició explícita de probabilitat i va introduir el teorema del límit central. La segona gran contribució és deguda a C.F. Gauss (1777-1855), el qual va treballar, entre altres temes, amb la teoria d'errors en les observacions. Una altra contribució important la trobem en A.M. Legendre (1752-1833), que va estudiar el mètode dels mínims quadrats.

Més endavant destaquen els treballs de P.L. Txebixev (1821-1894) i el seu deixeble A.A. Markov (1856-1922). Situant-nos ja en el segle passat, R.E. Von Mises (1883-1953) va intentar formular la definició de probabilitat, S.N. Bernstein (1880-1968) va iniciar el camí de l'axiomatització, i, finalment, A. Kolmogorov (1903-1987) va resoldre el problema establint els axiomes que avui en dia considerem la base del càlcul de probabilitats.

1.2.2. L'estadística

El primer objecte d'estudi de l'estadística van ser els censos. Les primeres estadístiques que ens consten són, tal vegada, osques en arbres fetes pels homes primitius per tenir un recompte del bestiar i altres possessions. Fa uns 4.000 anys es van portar a terme censos a la Xina, en fa 3.400 es van fer a Egipte, i una mica més tard a Roma i a l'Índia. El seu objectiu era o bé tributari (recollir impostos) o bé militar (calcular homes disponibles). Després hi va haver un buit en la recerca i l'avanç que va durar força segles amb molt poques activitats referents a l'estadística.

El 1662, J. Graunt (1620-1674) es va plantejar el problema de fer una estimació de la població anglesa de la seva època i, a partir d'una mostra, va fer per primera vegada una estimació de les taxes de mortalitat per edats i en va deduir la freqüència de naixements. Cal destacar també els estudis fets per W. Petty (1623-1687) i E. Halley (1656-1742).

Durant el segle XVIII els treballs estadístics es van fer sobre demografia, indústria, sanitat i economia i se situen principalment a Anglaterra i Alemanya. Com ja s'ha dit anteriorment, l'estadística evolucionava com a ciència separatament del càlcul de probabilitats i la teoria d'errors.

Una contribució cap a la interrelació entre aquestes disciplines la va generar el belga A. Quételet (1796-1874), el qual va demostrar la importància del càlcul de probabilitats per a l'estudi de dades humanes. Va emprar, per exemple, la llei normal per descriure la distribució de les alçades dels homes.

C. Babbage (1792-1871) va contribuir a fundar la Royal Statistical Society de Londres el 1834. Cinc anys més tard, el 1839, es va crear l'American Statistical Association. Mentrestant, tal com s'ha dit en la secció anterior, entre el 1800 i el 1820, P.S. Laplace, A.M. Legendre i C.F. Gauss van desenvolupar dos conceptes fonamentals per a la teoria estadística: la teoria dels errors en les observacions i la teoria dels mínims quadrats.

La unió entre el càlcul de probabilitats i l'estadística iniciada per A. Quételet es va produir a començament del segle XIX, i hi van tenir molt a veure els nous problemes teòrics i metodològics plantejats per a la contrastació empírica de la teoria de Darwin. El primer a destacar la necessitat de recórrer als mètodes estadístics per a aquestes contrastacions fou F. Galton (1822-1911), el qual va estudiar amb rigorositat la distribució normal i va introduir el concepte de línia de regressió.

Si F. Galton va ser el principal precursor de l'estadística moderna com a ciència independent, els actuals fonaments els devem a K. Pearson (1857-1936) i R.A. Fisher (1890-1962). K. Pearson va publicar taules estadístiques molt importants en la seva època per facilitar la difusió dels nous mètodes; va desenvolupar conceptes molt rellevants en la correlació i la regressió, així com moltes de les tècniques de l'estadística clàssica; i va fundar la famosa revista *Biometrika* el 1902. En el seu laboratori hi van treballar grans estadístics, com W.S. Gosset (1876-1937), descobridor de la distribució t de Student. A R.A. Fisher se li deuen molts mètodes d'inferència, contrastos d'hipòtesi, anàlisi de la variància i disseny d'experiments. Va desenvolupar el mètode de la màxima versemblança. Els seus descobriments van culminar amb la publicació de *Statistical Methods for Research Workers*, on apareixen el problema d'elegir un model a partir de dades empíriques, la seva deducció matemàtica, l'estimació dels paràmetres condicionats a la bondat del model i la seva validació final mitjançant contrastos d'hipòtesis.

Entre el 1920 i el final de la Segona Guerra Mundial s'estén l'aplicació dels mètodes estadístics en àrees tan diverses com l'enginyeria –control de qualitat per W.A. Shewhart (1891-1967), mètodes de predicció i control de processos i codificació de senyals per N. Wiener (1894-1964) i A. Kolmogorov–, la física –teoria cinètica de gasos–, l'antropologia, la medicina i la psicologia.

El desenvolupament posterior de l'estadística, cada vegada més relacionada amb la teoria de probabilitats, ha estat espectacular. Podem destacar, entre d'altres, E.S. Pearson (1895-1980) i J. Neyman (1894-1981) per la teoria general de contrastos d'hipòtesi, i A. Wald (1902-1950) per la renovació de molts conceptes clàssics.

La intenció d'aquestes dues seccions ha estat simplement fer un breu relat històric del càlcul de probabilitats i l'estadística tenint en compte el temari d'aquests apunts. És ben segur que molts lectors trobaran a faltar noms que també han contribuït de manera notable a l'avanç d'aquestes dues disciplines, com poden ser, per exemple, D. Bernoulli (1700-1782), S.D. Poisson (1781-1840), L.E. Boltzmann (1844-1906), J. Jensen (1859-1925), É. Borel (1871-1956), P. Lévy (1886-1971), H. Jeffreys (1891-1989), H. Cramer (1893-1985), A.Y. Khinchin (1894-1959), W. Feller (1906-1970), J.L. Doob (1910-2004), J.W. Tukey (1915-2000), L.J. Savage (1917-71), C.R. Rao (1920-) i molts d'altres. També ens agradaria destacar altres investigadors els treballs dels quals apareixen en aquest llibre, com C. Spearman (1863-1945), F.P. Cantelli (1875-1966), G.W. Snedecor (1881-1974), V.I. Smirnov (1887-1974), F. Wilcoxon (1892-1965), V. Glivenko (1897-1940), J. Von Neumann (1903-1957), M.B. Mann (1905-2000), M. Kendall (1907-1983), S. Kullback (1907-1994), J. Wolfowitz (1910-1981), W.A. Wallis (1912-1998), D.R. Whitney (1915-2007) i W.H. Kruskal (1919-2005).

1.2.3. L'estadística avui dia

A partir del 1950 podem considerar que comença l'època moderna de l'estadística. D'una banda, l'aparició de l'ordinador digital obre unes enormes possibilitats per a la construcció de models més complexos i permet tractar adequadament problemes dinàmics i multivariants; de l'altra, el canvi en la metodologia estadística. Sota la influència de J. Neyman, K. Pearson i A. Wald, als anys quaranta i cinquanta, la investigació teòrica es concentra en la recerca de procediments òptims d'estimació i contrast d'hipòtesi. Aquests procediments parteixen de dues premisses fonamentals: les dades de què disposem han estat generades per un model de distribució de probabilitat conegut excepte per un vector de paràmetres; i el model pertany a una família restringida de distribucions de probabilitat, matemàticament tractables. Aquests procediments no consideren, però, dos problemes centrals de l'anàlisi estadística: la identificació de l'estructura del model; i els contrastos –una vegada estimat el model– per tal de decidir si se'n pot rebutjar l'estructura bàsica mitjançant les dades empíriques. Actualment, la metodologia estadística es concep com un procés iteratiu d'aprenentatge, en lloc d'una aplicació única i directa d'un determinat procediment òptim.

Així, en els nostres dies, la recerca en estadística es desenvolupa utilitzant un cicle deductiu-inductiu que consta de les etapes següents:

- 1) Plantejament del problema. Partint del problema real es defineixen els objectius de l'estudi i es relacionen amb els valors numèrics de les variables observables.
- 2) Construcció del model estadístic.
- 3) Recollida de la informació mostral. Per mostreig o amb un disseny d'experiments.
- 4) Denejament de la mostra.
- 5) Estimació dels paràmetres o funcions d'interès.
- 6) Contrastos de simplificació (contrastos d'hipòtesi).
- 7) Crítica i diagnosi del problema.

Si acceptem el model com a correcte, l'utilitzarem per prendre decisions o fer previsions; altrament tornarem a construir un nou model estadístic i seguirem el mateix procés.

1.3. El present i el futur

No fa gaires dècades, el terme *estadística* no era ni gaire emprat ni ben comprès. Molt sovint, l'estadística era vista amb escepticisme. No hi havia professionals anomenats *estadístics*, excepte uns quants treballadors en departaments governamentals. Tampoc no hi havia cursos universitaris que possibilitessin obtenir títols acadèmics en estadística. Ara la situació ha canviat completament: hi comença a haver una certa demanda d'experts en estadística en molts camps del saber humà i aquesta demanda ben segur que no parerà de créixer.

En el domini de la sociologia, l'estadística s'utilitza per a la selecció psicotècnica dels treballadors, per obtenir enquestes de l'opinió pública, en sondejos sobre estudis polítics, comercials, socials i econòmics. En la indústria ha permès de resoldre econòmicament el problema del control de qualitat. La meteorologia, l'experimentació agrícola, l'epidemiologia, la farmacologia, la investigació de mercats, la codificació de missatges, el disseny i la planificació d'experiments científics, l'arqueologia, etc., també hi recorren. Malauradament també s'aplica en la recerca armamentista. Fins i tot en el camp de la poesia.

Per la quantitat de camps en què s'aplica alguns autors opinen que és una ciència amb una gran dependència de la resta de ciències; en canvi, hi ha un altre punt de vista que descrivia molt bé el gran matemàtic C.R. Rao: L'estadística és una disciplina independent, tal vegada la més independent de totes les disciplines».

Actualment l'estadística resol moltíssims i variats problemes com:

- la descripció de dades,
- l'anàlisi de mostres,
- el contrast d'hipòtesis,
- el mesurament de relacions,
- la predicció.

Per acabar aquest capítol, ens agradaria escriure com ubicava R.A. Fisher ja el 1952 l'estadística dins del segle passat: «La ciència estadística és l'aspecte peculiar del progrés humà que dona al segle XX el seu caràcter específic..., l'època actual gira al voltant de l'estadística per cercar el més essencial en totes les seves activitats més importants».

1.4. Referències complementàries

Per trobar més informació sobre la història de la probabilitat i l'estadística podeu consultar el llibre de Maistrov [49] o els dos volums d'Anders [32] i [33]. D'altra banda, com que antigament els grans noms de les matemàtiques treballaven en molts camps de la ciència i en particular en diverses branques de les matemàtiques, també se'n pot buscar informació en llibres sobre història de les matemàtiques en general, com per exemple, el llibre de Boyer [5] i els dos volums de Collette [14] i [15].

També voldríem citar el llibre de Lehmann i Casella [45], que, tot i ser una referència matemàtica molt útil per a la consulta d'aspectes estadístics relacionats amb el nostre llibre, al final de cada capítol també conté una breu cronologia històrica dels temes que ells han tractat.

Si, en canvi, l'interès del lector va dirigit a la lectura d'una dissertació de l'estadística, l'adreçem a la traducció feta per Cuadras i Oller del llibre de Rao [63].