Statistics and Mathematics

Nancy Reid, University of Toronto

McMaster University, November 21

`www.utstat.utoronto.reid/research`

0 What is the interface?

"Is statistics part of mathematics?": primary motivation from data, primary goal is inference from data, primary justification is empirical

"Statistics is a science in my opinion, and is no more a branch of mathematics than are physics, chemistry and economics, for if its methods fail the test of experience – not the test of logic – they are discarded" Tukey, 1953

Statistics has, and needs, theory:
-provides a framework for finding common features among problems with different specific details
-suggests approaches to new types of data

Mathematics is essential: for the development of the theory, in providing the tools and the language for progress

Mathematical tools and structures

Mathematical tools:
-calculus, matrix algebra, differential equations, combinatorics
-Taylor series, asymptotic expansions, numerical analysis
-real and complex analysis, functional analysis

Mathematical structures:
-group theory
-category theory
-differential geometry
-graph theory
-probability theory
-mathematical physics
-computational algebraic geometry

Level of complexity/sophistication:
-verify a technique or set of techniques
-elucidate some basic principles

Statistics $\leftrightarrow$ Mathematics

Illustrations

-differential geometry and statistical inference

Amari, Efron

-martingale theory and survival data analysis

ABGK

-category theory and statistical models

McCullagh

-functional analysis and theory of wavelets

Donoho & Johnstone

-asymptotic analysis and statistical inference

Barndorff-Nielsen & Cox

-graph theory and causality

Lauritzen

-quantum statistics

Barndorff-Nielsen, Gill & Jupp

-geometry of random fields; medical imaging

Worsley

-computational algebraic geometry and contingency tables

Diaconis & Sturmfels, Pistone & Wynn

Parametric models, Inference, Likelihood

model $f(y;\theta)$ $\theta \in R^d$, $\qquad y \in R$ or R^k

sample $\underline{y} = (y_1, \ldots, y_n)$ $\qquad f(\underline{y};\theta) = \prod f(y_i;\theta)$

goal is having observed y, to infer θ

Example: $Y \sim$ Poisson(θ); $\quad \theta = b + \mu$ $\mu > 0$?

$f(y;\theta) = \theta^y e^{-\theta}/y!, \quad y = 0, 1, \ldots; \theta > 0$

data: $y = 27, b = 6.7$

Example: $X \sim$ Poisson$(k\beta), Y \sim$ Poisson$(\beta+\mu)$

data: $y = 27, x/k = 6.7, k = 1.5$

Example: matched pairs $(Y_{1j}, Y_{2j}), j = 1, \ldots, n$

$$Y_{1j} = \begin{cases} 1 & \text{with prob. } p_{1j} \\ 0 & \text{with prob. } 1 - p_{1j} \end{cases}$$

$$Y_{2j} = \begin{cases} 1 & \text{with prob. } p_{2j} \\ 0 & \text{with prob. } 1 - p_{2j} \end{cases}$$

$$\begin{aligned} p_{1j} &= 1/\{1 + \exp(-\psi - \lambda_j)\} \\ p_{2j} &= 1/\{1 + \exp(-\lambda_j)\} \end{aligned}$$

$$\theta = (\psi, \lambda)$$

		Control day	
		on cellphone	not on cellphone
Day of	on cellphone	13	157
crash	not on cellphone	24	505

$Y_{1j} = 1$ if there was a cellphone call at the time of the collision

$Y_{2j} = 1$ if there was a cellphone call at a *control* time

λ_j measures probability that jth person uses the phone, ψ the increase on a collision day, $\exp(\psi)$ the odds of a call on the day of the collision versus the control time (day before the collision)

Redelmeier and Tibshirani, 1997, Cdn. J. Statist.

The New England
Journal of Medicine

© Copyright, 1997, by the Massachusetts Medical Society

VOLUME 336 FEBRUARY 13, 1997 NUMBER 7

ASSOCIATION BETWEEN CELLULAR-TELEPHONE CALLS AND MOTOR VEHICLE COLLISIONS

DONALD A. REDELMEIER, M.D., AND ROBERT J. TIBSHIRANI, PH.D.

ABSTRACT

Background Because of a belief that the use of cellular telephones while driving may cause collisions, several countries have restricted their use in motor vehicles, and others are considering such regulations. We used an epidemiologic method, the case–crossover design, to study whether using a cellular telephone while driving increases the risk of a motor vehicle collision.

Methods We studied 699 drivers who had cellular telephones and who were involved in motor vehicle collisions resulting in substantial property damage but no personal injury. Each person's cellular-telephone calls on the day of the collision and during the previous week were analyzed through the use of detailed billing records.

Results A total of 26,798 cellular-telephone calls were made during the 14-month study period. The risk of a collision when using a cellular telephone was four times higher than the risk when a cellular telephone was not being used (relative risk, 4.3; 95 percent confidence interval, 3.0 to 6.5). The relative risk was similar for drivers who differed in personal characteristics such as age and driving experience; calls close to the time of the collision were particularly hazardous (relative risk, 4.8 for calls placed within 5 minutes of the collision, as compared with 1.3 for calls placed more than 15 minutes before the collision; $P<0.001$); and units that allowed the hands to be free (relative risk, 5.9) offered no safety advantage over hand-held units (relative risk, 3.9; P not significant). Thirty-nine percent of the drivers called emergency services after the collision, suggesting that having a cellular telephone may have had advantages in the aftermath of an event.

Conclusions The use of cellular telephones in motor vehicles is associated with a quadrupling of the risk of a collision during the brief period of a call. Decisions about regulation of such telephones, however, need to take into account the benefits of the technology and the role of individual responsibility. (N Engl J Med 1997;336:453-8.)

©1997, Massachusetts Medical Society.

MOTOR vehicle collisions are a leading cause of death in North America; they are the single most frequent cause of death among children and young adults and account for one fatality every 10 minutes.[1-3] During an average year, about 1 person in 50 will be involved in a motor vehicle collision; 1 percent of them will die, 10 percent will be hospitalized, and 25 percent will be temporarily disabled.[4,5] Motor vehicle collisions often injure persons who are otherwise in good health. The causes of motor vehicle collisions are complicated, but error on the part of drivers contributes to over 90 percent of events.[6]

Cellular telephones can be used for placing and receiving telephone calls while in a motor vehicle. North American sales are enormous; for example, in 1995 the number of new subscribers in the United States exceeded the birth rate.[7,8] Many believe that telephones may contribute to collisions by distracting drivers,[9] and a few countries (such as Brazil, Israel, and Australia) have laws against using a cellular telephone while driving. Research with simulators suggests that use of the telephone can impair some aspects of driving performance.[10-14] However, industry-sponsored surveys have found no increased risk associated with car telephones.[15,16]

The most rigorous experimental method for testing the effects of cellular telephones on motor vehicle collisions is to assess outcomes for persons randomly assigned to use or not use the devices, but such a study would be very difficult to perform and possibly unethical. Instead, we used an epidemiolog-

From the Departments of Medicine (D.A.R.), Preventive Medicine and Biostatistics (R.J.T.), and Statistics (R.J.T.) and the Program in Clinical Epidemiology and Health Care Research (D.A.R.), University of Toronto, Toronto; and the Division of Clinical Epidemiology, Sunnybrook Health Science Centre, North York, Ont., Canada (D.A.R.). Address reprint requests to Dr. Redelmeier at Sunnybrook Health Science Centre, G-151, 2075 Bayview Ave., North York, ON M4N 3M5, Canada.

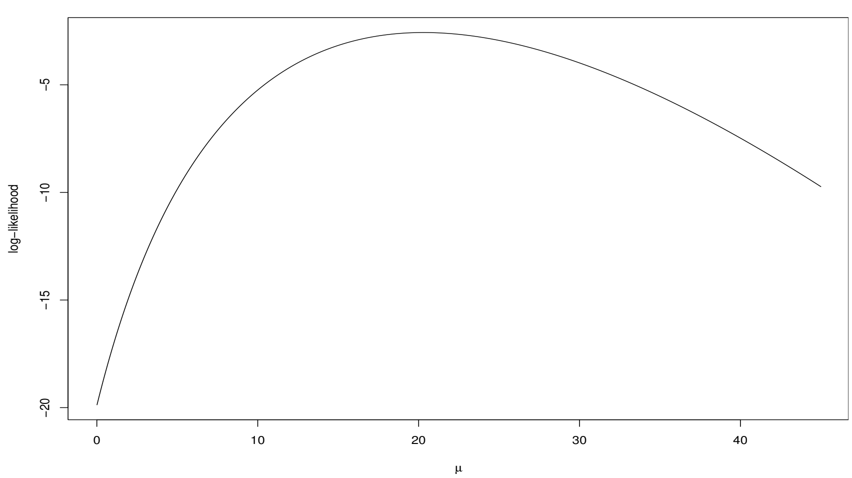
log likelihood function $\ell(\theta; \underline{y}) = \log f(\underline{y}; \theta) =$

maximum likelihood estimator $\widehat{\theta} = \widehat{\theta}(\underline{y})$

Fisher information $j(\widehat{\theta}\,) = -\ell''(\widehat{\theta}\,)$

Example: $Y \sim$ Poisson(θ); $\quad \theta = b + \mu \quad [\beta + \mu]$

data: $y = 27, b = 6.7$; $\widehat{\theta} = y$



pr($Y > 27; \mu = 0) \doteq 10^{-8}$ $\quad [2 \times 10^{-5}]$

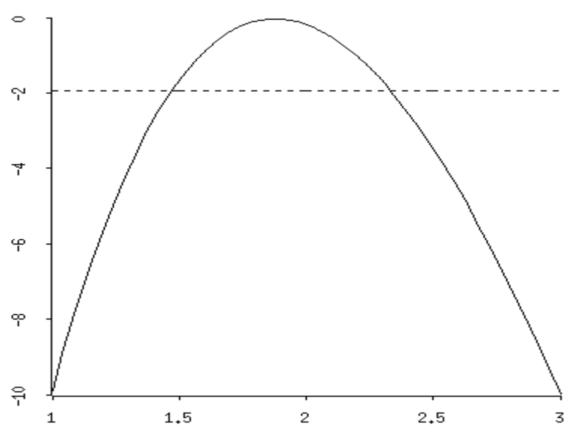
parameter of interest $\psi(\theta)$, dimension $<< p$

often $\theta = (\psi, \lambda)$

log profile likelihood function $\ell_P(\psi; \underline{y}) = \ell(\psi, \hat{\lambda}_\psi; \underline{y})$

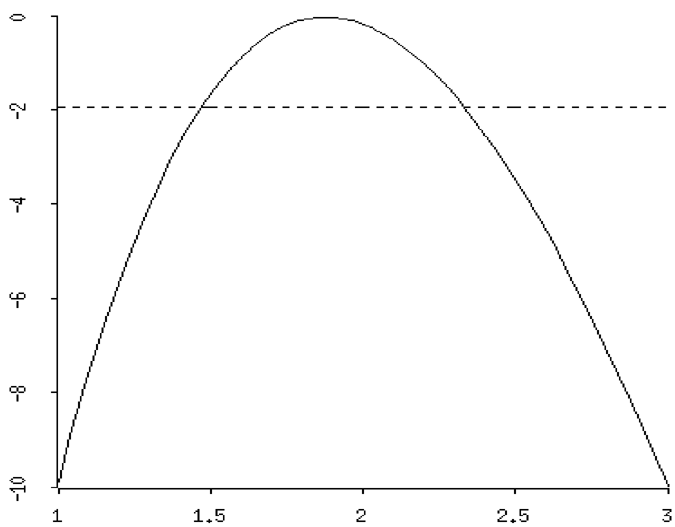
for matched pairs example use
$\ell_c(\psi; \underline{y}) = \sum \log \mathrm{pr}(y_{1j} = 1 \mid y_{1j} + y_{2j} = 1)$

$\hat{\psi}_c = 1.88$: relative risk $= 6.54$ (4.50, 9.99)



Inference based on the log (profile) likelihood

-point estimation $\widehat{\theta} = \widehat{\theta}(y)$, (mle); has precision $j(\widehat{\theta})^{-1}$ $\quad (\widehat{\psi};\ j_P(\widehat{\psi})^{-1})$



-values of θ compatible with the observed data $\ell(\widehat{\theta}) - \ell(\theta) > c$ $\quad (\ell_P(\widehat{\psi}) - \ell_P(\psi) > c)$

-these are often calibrated using the (possibly crude) approximation

$$\widehat{\theta} \,\dot\sim\, N\{\theta, j^{-1}(\widehat{\theta}\,)\}$$

justified by appeal to limiting distribution

assume $\ell(\theta; y) = O(n)$, $\theta - \widehat{\theta} = O(n^{-1/2})$

Differential geometry (Efron, Amari)

model $\mathcal{M} = \{f(\cdot;\theta);\theta \in \Theta\}$: differentiable manifold

with coordinates $\theta = (\theta^1, \ldots, \theta^p)$

tangent space spanned by $\{\partial\ell_r(\theta;\cdot), r = 1, \ldots, p\}$

Riemannian metric $< \partial\ell_r, \partial\ell_s >= \int \partial\ell_r \partial\ell_s f(y;\theta)dy$

Note: expected Fisher information $E\{-\partial_{rs}\ell(\theta;y)\} = Ej(\theta)$.
i.e. is related to the precision of the maximum likelihood estimate $\hat{\theta}$

affine connection (covariant derivative)

$$\Gamma^{(\alpha)}_{rst} = E(\partial_r\partial_s\ell\partial_t\ell) + \frac{1}{2}(1-\alpha)E(\partial_r\ell\partial_s\ell\partial_t\ell)$$

is an essentially unique family for statistical manifolds, in a certain sense
although is not compatible with the metric

Why might this be useful? –
Statistically important classes of models have nice geometric structure

-exponential family models

$$f(y;\theta) = \exp\{y^T\theta - c(\theta) - d(y)\}$$

are closed under sampling, and have p-dimensional sufficient statistics

$$f(\underline{y};\theta) = \exp\{s^T\theta - nc(\theta) - d(\underline{y})\} : \quad s = \sum y_i$$

-exponential family models are flat in the $\alpha = 1$ connection

-transformation family models

$$f(y;\theta) = f_0(gy; g^*\theta) \quad g \in G, \quad g^* \in \mathcal{G}$$

have a natural type of invariance, and admit a dimension reduction by conditioning

-transformation family models are (sort of) flat in the $\alpha = -1$ connection

A submanifold $\mathcal{M}_0$ of $\mathcal{M}$ can be obtained by fixing some components of θ, and we can then use the connections to describe both

- embedding curvature of $\mathcal{M}_0$

- intrinsic curvature of $\mathcal{M}_0$

These have natural statistical interpretations related to estimation by maximum likelihood, and conditioning on ancillary statistics.

For example, any *bias-corrected*, *first order efficient* estimator of θ has variance matrix

$$\begin{aligned} E\{(\tilde{\theta}^r - E\tilde{\theta}^r)(\tilde{\theta}^s - E\tilde{\theta}^s)\} &= i^{rs}n^{-1} \\ &+ (c_1 + c_2 + c_3)n^{-2} + O(n^- \end{aligned}$$

c_1 contracted from the $\alpha = -1$ connection coefficients in $\mathcal{M}_0$

c_2 contracted from the $\alpha = 1$ embedding curvature of $\mathcal{M}_0$ in $\mathcal{M}$

c_3 vanishes if $\tilde{\theta}$ is the maximum likelihood estimator

Many results of this type in Amari, Barndorff-Nielsen & Cox, and Murray & Rice

Geometrical notions of invariance and orthogonality have become embedded in current asymptotic theory, but impact of the more detailed development of connections and curvature components is as yet unclear

$$c_1 = ({}^{-1}\Gamma^2)^{rs} = {}^{-1}\Gamma^r_{tu}\ {}^{-1}\Gamma^s_{vw} i^{tv} i^{uw}$$

$$1 \le r, s, t, u, v, w, \le d,$$

$$c_2/2 = ({}^{1}H^2)^{rs} = {}^{1}H^r_{tu}\ {}^{1}H^s_{vw} i^{tv} i^{uw}$$

$$1 \le r, s, t, v \le d; d + 1 \le u, w, \le p$$

$$c_3 = ({}^{-1}H^2)^{rs} = {}^{-1}H^r_{tu}\ {}^{-1}H^s_{vw} i^{tv} i^{uw}$$

$$1 \le r, s \le d; d + 1 \le t, u, v, w, \le p$$

$$H^\alpha_{rst} = \Gamma^\alpha_{rst} \quad 1 \le r \le d, 1 \le s \le d, d + 1 \le t \le p$$

indices for H and Γ are raised via the Riemannian metric: ${}^{\alpha}\Gamma^r_{st} = \Gamma^\alpha_{stu} i^{ur}$

Category theory (McCullagh, 2000, 2002)

-model: $\{f(\cdot;\theta); y \in \mathcal{Y}, \theta \in \Theta\}$ is not enough

-more structure is needed, for 'sensible' models

-the structure should be a category, i.e. a collection of objects and of maps or morphisms between objects.

-the objects are themselves (simpler) categories, and the morphisms associate probability distributions with the objects

-the basic categories are
–cat$_{\mathcal{U}}$: each object $\mathcal{U}$ is a set of units: morphisms are injective maps
–cat$_{\Omega}$: each object Ω is a covariate space: morphisms are injective maps
–cat$_{\mathcal{V}}$: i each object is a response scale: morphisms are surjective maps

-the statistical categories are
– the sample space: $\mathcal{S} = \mathcal{V}^{\mathcal{U}}$
– the design: $\text{cat}_{\mathcal{D}}$; each object is a pair $(\mathcal{U}, \Omega)$
– the parameter $(\Theta, *)$: a contravariant functor $\text{cat}_{\Omega} \to \mathcal{K}$, associating with each Ω a parameter set Θ_{Ω}

Finally, "a statistical model is a functor on $\text{cat}_{\mathcal{D}}$ associating with each design object $\psi : \mathcal{U} \to \Omega$ a model object, which is a map $P_{\psi} : \Theta_{\Omega} \to \mathcal{P}(\mathcal{S})$ such that $P_{\psi}(\Theta)$ is a probability distribution on $\mathcal{S}$."

And associated with this is the notion of a "natural subparameter", which is a transformation with an associated map that is commutative

McCullagh argues that this ensures that statistical models have an appropriate notion of embedding, so that we may, for example, assume the model that is constructed for 5 equally spaced dose levels (the covariate) remains valid at doses between those used in the experiment, even though we will not have information about them.

Do we need all this?

Example: transformed linear regression

$$y_i^\lambda = x_i'\beta + \sigma e_i \quad i = 1, \ldots, n$$

-x_i is a known vector of covariates for the ith subject
-$\theta = (\beta, \sigma, \lambda)$, with interest usually in one or more components of β

Widely used in applied work:
-estimate λ from the data,
-compute $y^{\tilde{\lambda}}$,
-inference on β as if λ were known to be equal to $\tilde{\lambda}$.

Conventional statistical arguments suggest that uncertainty about λ should be incorporated into uncertainty statements about β, but this leads to absurd results.

From the category point of view, the parameter β above is not a natural parameter, although β/σ is, as is the pair (β, λ) and $\beta^{1/\lambda}$

Example: factorial models

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \text{error}$$

- α_i: parameters associated with I levels of factor A, etc.
- $(\alpha\beta_{ij})$: parameters associated with the interaction of factors A and B

widely used in industrial experimentation, often with many more than 3 factors

often of interest to assess the magnitude of, e.g., 2nd order interactions, or to test whether higher order interactions are needed in the model

category theory can be used to formalize the idea that all sensible models are hierarchical, i.e. if AB interaction is included in a model then A and B must be included as well

Category theory appears to offer some insights into modelling, extends work on group theory due to Fraser (1968) and others.

It might have most importance in analysis of spatial data, where it seems to be more difficult to construct realistic statistical models.

But,
"... one does not have to look far in probability, statistics or physics to see that today's concrete foundations are yesterday's abstractions. On the other hand, it must ruefully be admitted that most of yesterday's abstractions are buried elsewhere." (McCullagh, 2002; rejoinder)

Computational algebraic geometry

The statistical motivation: contingency tables

Example: cellphone study

		Control day	
		on cellphone	not on cellphone
Day of	on cellphone	13	157
crash	not on cellphone	24	505

Example: Queen Victoria's descendants

Month of Birth	Month of death: Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Jan	1	0	0	0	1	2	0	0	1	0
Feb	1	0	0	1	0	0	0	0	0	1
Mar	1	0	0	0	2	1	0	0	0	0
Apr	3	0	2	0	0	0	1	0	1	3
May	2	1	1	1	1	1	1	1	1	1
Jun	2	0	0	0	1	0	0	0	0	0
Jul	2	0	2	1	0	0	0	0	1	1
Aug	0	0	0	3	0	0	1	0	0	1
Sep	0	0	0	1	1	0	0	0	0	0
Oct	1	1	0	2	0	0	1	0	0	1
Nov	0	1	1	1	2	0	0	2	0	1
Dec	0	1	1	0	0	0	1	0	0	0

Month of Birth	Month of death											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Jan	1	0	0	0	1	2	0	0	1	0	1	0
Feb	1	0	0	1	0	0	0	0	0	1	0	2
Mar	1	0	0	0	2	1	0	0	0	0	0	1
Apr	3	0	2	0	0	0	1	0	1	3	1	1
May	2	1	1	1	1	1	1	1	1	1	1	0
Jun	2	0	0	0	1	0	0	0	0	0	0	0
Jul	2	0	2	1	0	0	0	0	1	1	1	2
Aug	0	0	0	3	0	0	1	0	0	1	0	2
Sep	0	0	0	1	1	0	0	0	0	0	1	0
Oct	1	1	0	2	0	0	1	0	0	1	1	0
Nov	0	1	1	1	2	0	0	2	0	1	1	0
Dec	0	1	1	0	0	0	1	0	0	0	0	0

-standard statistical arguments suggest that the relevant probability space for testing association in the table is the set of all tables of the same size with the same margins.

-Diaconis & Sturmfels (1998) devised a Markov chain based algorithm to find all such tables

-choose a pair of columns (at random) and a pair of rows (at random) and change the intersection elements using

$$\begin{array}{cc} + & - \\ - & + \end{array} \qquad \begin{array}{cc} - & + \\ + & - \end{array}$$

-this is an example of a Markov basis

Abstraction:

$\mathcal{X}$ is a finite set (the sample space)

$T = t(X)$ is a sufficient statistic (map on the sample space to N^d)

$k[\mathcal{X}]$ a ring of polynomials

$T : \mathcal{X} \to N^d$ represented by $\varphi_T : k[\mathcal{X}] \to k[t_1, \ldots t_d]$

$\mathcal{I}_T =$ kernel of φ_T is an ideal

the generators for this ideal correspond to a Markov basis

the Gröbner basis for $\mathcal{I}_T$ can be computed

in Queen Victoria example, $k[\mathcal{X}]$ is the ring of polynomial functions on $I \times J$ matrices, and $\mathcal{I}_T$ is generated by the 2×2 minors

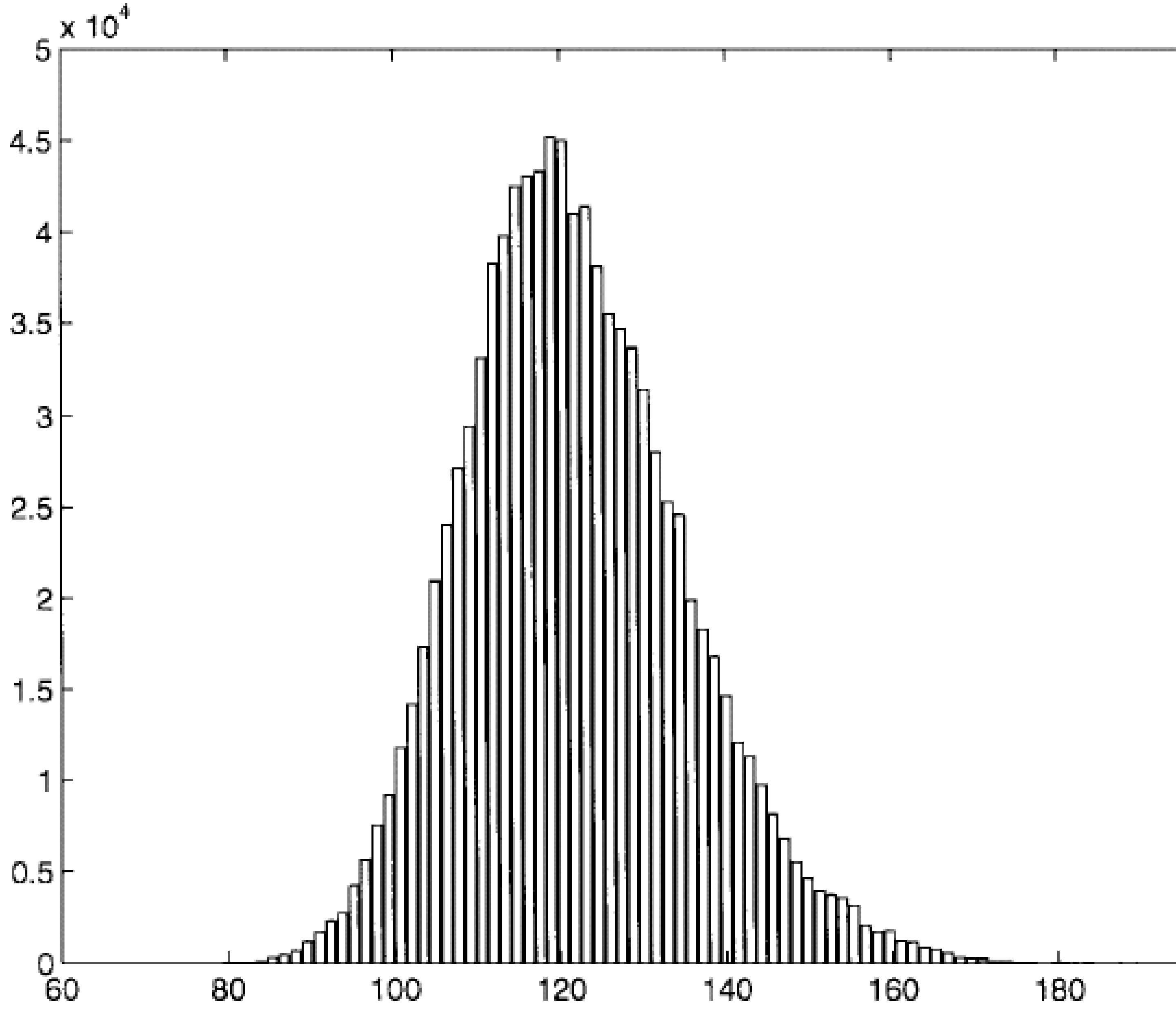


FIG. 2. *Histogram of the chi-square statistic for Table* 1.

Other examples in D&S include:
- higher way contingency tables
- graphical models
- Hardy-Weinberg equilibrium
- logistic (binary) regression
- spectral analysis of permutation data

Pistone & Wynn (1996)
-Gröbner bases for factorial designs
Pistone, Riccomagno & Wynn (2001)
-more on design, also probability models and statistical models

GroStat VI February 2003 Bayesian networks, causality, factorial design, contingency tables

Disclosure limitation in contingency tables (Serkan Hosten, Stephen Fienberg, ...)
–what entries are compatible with given marginals?

Example (Dobra, Fienberg, Trottini)

		Income level		
gender	race	$< 10,000$	$10,000 - 25,000$	$> 25,000(\$)$
	white	96	72	161
Male	black	10	7	6
	chinese	1	1	2
	white	186	127	51
female	black	11	7	3
	chinese	0	1	0

What are the bounds on the small entries, given the marginal totals from the two tables race $\times$ income and income $\times$ gender?

Conclusion

-differential geometry and statistical inference

-martingale theory and survival data analysis

-category theory and statistical models

-functional analysis and theory of wavelets

-asymptotic analysis and statistical inference

-graph theory and causality

-quantum statistics

-geometry of random fields; medical imaging

-computational algebraic geometry and contingency tables

–productive interface depends on a reasonably high level of mathematical sophistication on the part of statisticians
–productive interface depends on a reasonably broad approach to the subject on the part of mathematicians

if its methods fail the test of experience – not the test of logic – they are discarded

today's concrete foundations are yesterday's abstractions. On the other hand, it must ruefully be admitted that most of yesterday's abstractions are buried elsewhere."

Quantum statistics: new applications of differential geometry; relevant for quantum computing?

Asymptotic theory for $p > n$ (Portnoy; 1988)

False Discovery Rates: ecology, astronomy, microarrays

Motivation of students through research in other fields (?)

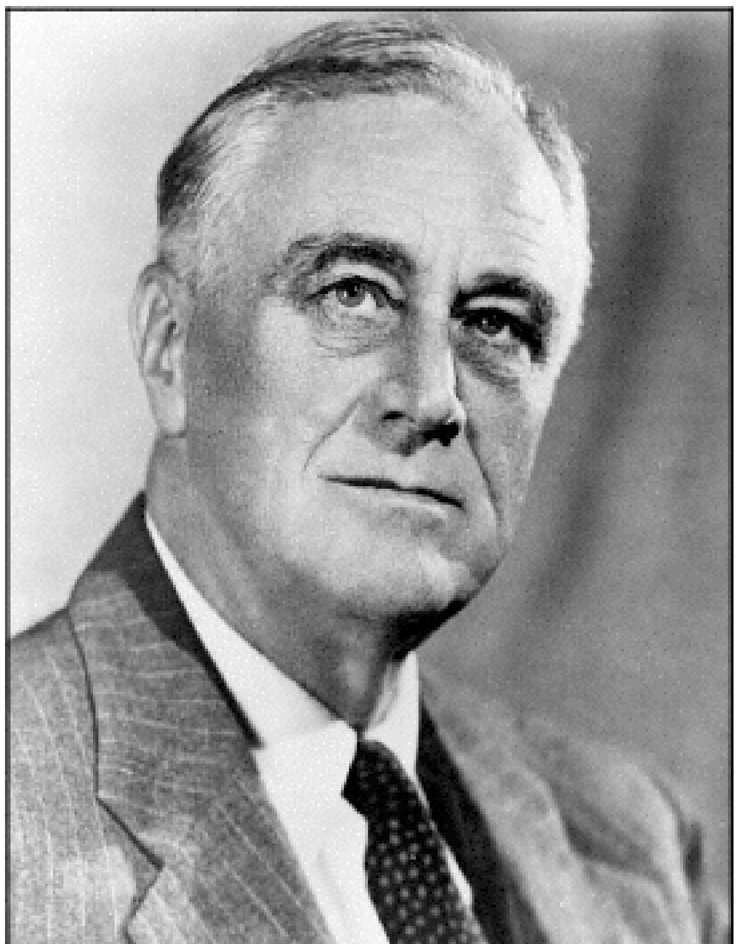


Figure 1. Photograph of President Franklin Delano Roosevelt taken in 1944 by Leon A Perskie. (By permission of Beatrice Perskie Foxman, Silver Springs, Maryland, USA.)

Table 2. Diagnostic probabilities of eight key symptoms in Roosevelt's paralytic illness appearing in Guillain–Bar poliomyelitis, tested by Bayesian analysis

FDR's case	*GBS (prior probability 0.51)*		*Poliomyelitis (pri*
	Symptom probability	*Posterior probability*	*Symptom probability*
Paralysis ascends for 10–13 days	0.70	**0.36**	0.02
Facial paralysis	0.50	**0.26**	0.02
Bladder/bowel dysfunction for 14 days	0.50	**0.26**	0.05
Numbness/dysaesthesia	0.50	**0.26**	<0.01
No meningismus	0.99	**0.50**	0.10
Fever	<0.01	<0.01	0.90
Descending recovery from paralysis	0.70	**0.36**	0.02
Permanent paralysis	0.15	0.08	0.50

The derivation of the estimates of prior probabilities (relative frequencies of the diseases in FDR's age range probabilities (the chance that a clinical feature occurred in a disease) of poliomyelitis and GBS is given in the considerations''. Posterior probabilities (the probability that FDR's symptoms were due to a disease) are the symptom probabilities. Greater posterior probabilities are in bold type.

References

Differential geometry

Amari, S.-I. (1985). *Differential-Geometric Methods in Statistics.* Springer-Verlag, New York.

Barndorff-Nielsen, O.E., Cox, D.R. and Reid, N. (1986). The role of differential geometry in statistical theory. *Int. Stat. Rev.* **54**, 83–96.

Barndorff-Nielsen, O.E. and Cox, D.R. (1994) *Inference and Asymptotics*. (Ch. 5) Chapman & Hall/CRC, Boca Raton.

Barndorff-Nielsen, O.E. and Cox, D.R. (1989) *Asymptotic Techniques*. Chapman & Hall/CRC, Boca Raton.

Efron, B. (1975). Defining the curvature of a statistical model. (with discussion). *Ann. Statist.* **6**, 362–376.

Murray, M.W. and Rice, J.O. (1993). *Differential Geometry and Statistics*. Chapman & Hall/CRC, Boca Raton.

Category theory

McCullagh, P. (2000) Invariance and factorial models. (with discussion) *J. R. Statist. Soc.* B **62**, 209–256.

McCullagh, P. (2002) What is a statistical model? (with discussion) *Ann. Statist.* **30**, 1225–1310.

`www.stat.uchicago.edu/p̃mcc`

Gröbner bases

Diaconis, P. and Sturmfels, B. (1998) Algebraic algorithms for sampling from conditional distributions. *Ann. Statist.* **26**, 363–397.

Pistone, G. and Wynn, H.P. (1996). Generalized confounding with Gröbner bases. *Biometrika* **83**, 653–666.

Pistone, G., Wynn, H.P., and Riccomagno, E. (2001). *Algebraic Statistics.* Chapman & Hall/CRC, Boca Raton.

Hoskan, S. (2003). `www.dima.unige.it/SMID/grostatvi`

Examples

Dobra, A., Fienberg, S.E. and Trottini, M. (2003). Assessing the risk of disclosure of confidential categorical data. Bayesian Statistics 7, Bernardo et al. (Eds). to appear. `www.stat.cmu.edu/-fienberg`

Tibshirani, R. and Redelmeier, D. A. (1997) Cellular telephones and motor-vehicle collisions: some variations on matched pairs analysis. *Canad. J. Statist.* **25**, 581–591.

Tukey

Brillinger, D.R. (2002). John W. Tukey: his life and professional contributions. *Ann. Statist.* **30**, 1535–1575.

Others

Lauritzen, S. L. (1996). *Graphical Models.* Clarendon, Oxford.

Andersen, P.K. , Borgan, O., Gill. R.D. and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Barndorff-Nielsen, O.E., Gill, R.D. and Jupp, P.E. (2003). Quantum statistics. *J. R. Statist. Soc.*, to appear (with discussion).

`www.math.mcgill.ca/keith/`

`www-stat.stanford.edu/-donoho` and `www-stat.stanford.edu/-wavelab`