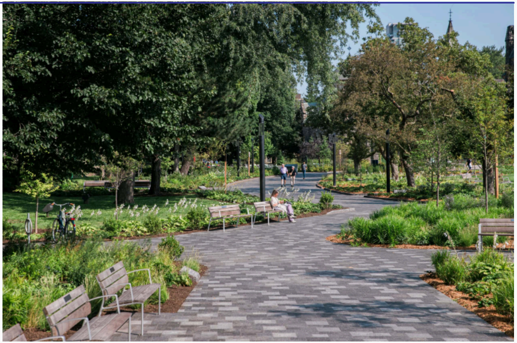
Models and likelihood

Nancy Reid
University of Toronto

September 30 2025



UNIVERSITY OF TORONTO
FACULTY OF ARTS & SCIENCE
Department of Statistical Sciences

Tattoos and melanoma

JOURNAL ARTICLE ACCEPTED MANUSCRIPT

Tattooing and risk of melanoma: a population-based case-control study in Utah

Rachel D McCarty, PhD, Britton Trabert, PhD, Lindsay J Collin, PhD, Morgan M Millar, PhD, David Kriebel, ScD, Laurie Grieshober, PhD, Mollie E Barnard, ScD, Jenna Sawatzki, BS, Marjorie Carter, MSPH, Valerie Yoder, BS ... Show more

JNCI: Journal of the National Cancer Institute, djaf235,
https://doi.org/10.1093/jnci/djaf235
Published: 21 August 2025 **Article history** ▾

Number of tattoo sessions					
No tattoo sessions	4980	(85%)	1027	(88%)	Ref
1	280	(5%)	79	(7%)	1.53 (1.16–2.00)
2-3	277	(5%)	37	(3%)	0.73 (0.50–1.04)
4 or more	298	(5%)	24	(2%)	0.44 (0.27–0.67)
Number of large tattoos					
No tattoo sessions	4980	(85%)	1027	(88%)	Ref
0	407	(7%)	80	(7%)	1.06 (0.81–1.38)
1-2	308	(5%)	*	(<5%)	0.93 (0.66–1.28)
3 or more	140	(2%)	*	(<5%)	0.26 (0.10–0.54)

*Censored to hide cell values <11
Note: Percentages indicated as "<" are not precise to prevent identification of study participants and may cause c
Adjusted for age (five-year groups), sex, race and ethnicity (Non-Hispanic White, Hispanic, all other racial and eth
four-year college degree or more), ever smoking (yes/no), physical activity in the past 30 days (yes/no), and body

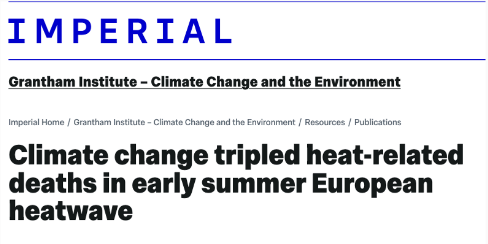
Scientific question: Do skin tattoos increase the risk of melanoma

observational data

Statistical model: "we computed odds ratios and 95% confidence intervals from logistic regression models"

"we created a DAG to identify potential confounders"

caution is warranted

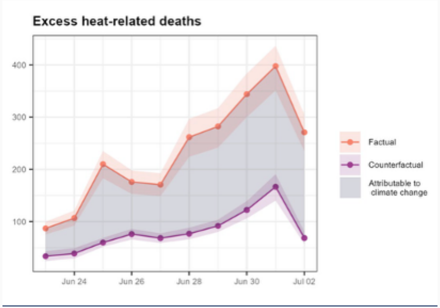


IMPERIAL

Grantham Institute – Climate Change and the Environment

Imperial Home / Grantham Institute – Climate Change and the Environment / Resources / Publications

Climate change tripled heat-related deaths in early summer European heatwave



Scientific question: Estimate the excess heat deaths attributable to climate change

simulated data

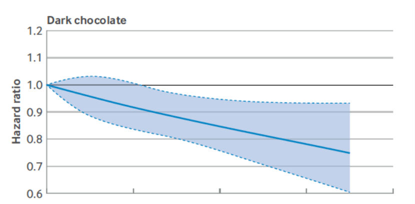
Statistical model: “We use established epidemiological models and the WWA climate-attribution framework”

extreme value distribution; regression

RESEARCH

Chocolate intake and risk of type 2 diabetes: prospective cohort studies

Binkai Liu,[1] Geng Zong,[2,3] Lu Zhu,[1] Yang Hu,[1] JoAnn E Manson,[4,5,6] Molin Wang,[4,5,7] Eric B Rimm,[1,4,5] Frank B Hu,[1,4,5] Qi Sun[1,4,5]



Scientific question: Does dark chocolate consumption affect the risk of Type 2 diabetes?

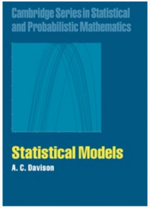
Statistical model: “We used Cox proportional hazards models to estimate the hazard ratios and corresponding 95% confidence intervals (CIs)”

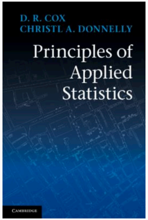
“a significant linear trend across four groups was observed”

Why these models?

- motivated by theory: economic, physical, ...
- motivated by design: randomized controlled trial, survey, regression discontinuity design, ...
- standard in the literature of that field — heat wave
- standard in the publications of that lab
- follow some prescription:
  - binary response — use logistic regression — tattoos
  - time to event — use PH model — chocolate
  - time series — use ARMA
  - repeated measures — use random effects
  - ...

Some guidance from the experts





- the key feature of a statistical model is that variability is represented using probability distributions
- the art of modelling lies in finding a balance that enables the questions at hand to be answered or new ones posed
- probability models as an aid to the interpretation of data
- perturbations of no intrinsic interest distort an otherwise exact measurement
- substantial natural variability in the phenomenon under study

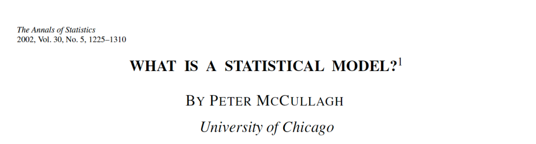
The role of parameters

- probability models very likely be parameterized
- thus defining a class of models $\{f(y;\theta); \theta \in \Theta\}$
- parameters may be finite- or infinite-dimensional — parametric vs nonparametric

- ideally one or more parameters represent key aspects of the model — for the application at hand

- other parameters complete the specification
- the meaning of various parameters varies with the application

- this sounds simpler than it is



The Annals of Statistics
2002, Vol. 30, No. 5, 1225–1310

WHAT IS A STATISTICAL MODEL?[1]

BY PETER MCCULLAGH

University of Chicago

e.g. Box-Cox $y^{\lambda} = x^{\mathrm{T}}\beta + \epsilon$

The likelihood function

- puts the emphasis on the model parameters: $L(\theta; \mathbf{y}) \propto f(\mathbf{y}; \theta) = \prod_{i=1}^{n} f(y_i; \theta)$

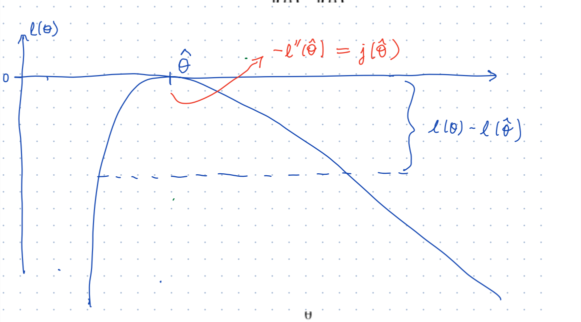
inverse problem

- provides a convenient way to compare parameter values

e.g. $L(\theta)/L(\hat{\theta})$

- provides reliable summary measures

$\ell(\theta; y) = \log L(\theta; y)$



- can be converted to a probability, given a prior probability for θ

Inference and asymptotics

$$(i)\ \ \ell(\theta) = \sum_{i=1}^{n} \log f(y_i;\theta \mid x_i), \quad (ii)\ \ \ell'(\theta) = \sum_{i=1}^{n} \nabla_\theta \log f(y_i;\theta \mid x_i), \quad (iii)\ \ \ell'(\hat\theta) = 0$$

Central Limit Theorem $\quad \frac{1}{\sqrt{n}}\ell'(\theta) \xrightarrow{d} N\{0, I_1(\theta)\}$ $\quad$ expected Fisher information per obs'n

$\implies$ MLE is approximately normally distributed $\quad J(\theta) = -\ell''(\theta)$

$$\hat\theta \mathrel{\dot\sim} N_p\{\theta, J^{-1}(\hat\theta)\}$$

$\implies$ LRT is approximately χ^2 distributed $\quad I(\theta) = \mathbb{E}_\theta\{J(\theta)\}$

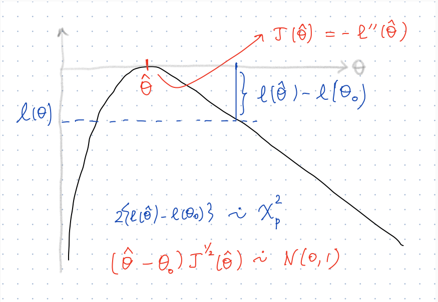
$$2\{\ell(\hat\theta) - \ell(\theta)\} \mathrel{\dot\sim} \chi^2_p$$

… Limit theory

Large-sample approximation:

$$\hat{\theta} \dot{\sim} N_p\{\theta, J^{-1}(\hat{\theta})\}, \qquad 2\{\ell(\hat{\theta}) - \ell(\theta)\} \dot{\sim} \chi^2_p$$

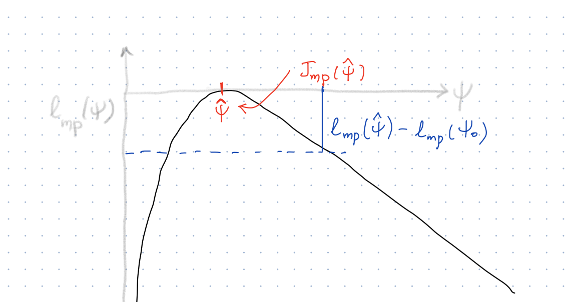
```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -3.079      0.987   -3.12   0.0018 **
aged1         -0.292      0.754   -0.39   0.6988
stage1         1.373      0.784    1.75   0.0799 .
grade1         0.872      0.816    1.07   0.2850
xray1          1.801      0.810    2.22   0.0263 *
acid1          1.684      0.791    2.13   0.0334 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 40.710  on 22  degrees of freedom
Residual deviance: 18.069  on 17  degrees of freedom
```



A bit too simple

- model $f(y;\theta)$, $\theta \in \mathbb{R}^p$
- $\theta = (\psi, \lambda)$ parameters of interest nuisance parameters
- results above used modified profile log-likelihood function

$$\ell_{\text{mp}}(\psi) = \ell(\psi, \hat{\lambda}_\psi) - \frac{1}{2}\log|J_{\lambda\lambda}(\psi, \hat{\lambda}_\psi)|$$



What can go wrong?

- distribution approximations might be poor — likelihood skewed; extremes more relevant
- too many parameters — $p \sim n^{\alpha}, \quad p/n \to C, \quad p/n \to \infty$
- irregular parameter space — $\alpha f(y;\theta_1) + (1-\alpha) f(y;\theta_2), \quad 0 \le \alpha \le 1$
- computational intractability — $L(\theta,\tau;y) = \int_{\mathbb{R}^k} f(y \mid z;\theta) f(z;\tau) dz$
- model is misspecified — true $Y \sim m(y), \quad f(\cdot;\theta) \neq m(\cdot)$ $\forall \theta$

1. Classical

Cox 1961,2; Huber 1967; White 1982

- true model $m(\boldsymbol{y})$ fitted model $f(\boldsymbol{y};\theta)$

$\boldsymbol{y} = (y_1, \ldots, y_n)$

$\ell(\theta;\boldsymbol{y}) \equiv \log f(\boldsymbol{y};\theta)$

- maximum likelihood estimator $\widehat{\theta}$

$\widehat{\theta} \equiv \arg\sup_\theta \ell(\theta;\boldsymbol{y})$

- $\widehat{\theta}$ converges to the "closest true value"

KL-divergence

$$\theta_m^o = \arg\min_\theta \int m(\boldsymbol{y}) \log\{\frac{m(\boldsymbol{y})}{f(\boldsymbol{y};\theta)}\} d\boldsymbol{y}$$

- $\widehat{\theta}$ has asymptotic normal distribution, but is not fully efficient

"sandwich variance"

$$\text{a.var.}\,(\widehat{\theta}) = G^{-1}(\theta_m^o), \qquad G(\theta) = J(\theta)I^{-1}(\theta)J(\theta)$$

$I = \text{var}_m(\ell'),\ J = \mathbb{E}_m(-\ell'')$

- change the inference goal, proceed more or less as usual

"we used robust standard errors "

2. More flexible inference functions

Composite likelihood

- true model $m(\boldsymbol{y}_i) = f(\boldsymbol{y}_i; \theta), \boldsymbol{y}_i \in \mathbb{R}^d$ fitted model $\prod_{A \in \mathcal{A}} f(y_{iA}; \theta)$ subsets A

- Example: pairwise likelihood $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$

$$L_{pair}(\theta; \boldsymbol{y}) = \prod_{i=1}^{n} \prod_{s \neq t} f_2(y_{is}, y_{it}; \theta)$$

- Example: $AR(1)$ likelihood $\boldsymbol{y} = (y_1, \ldots, y_n)$

$$L_{cond}(\theta; \boldsymbol{y}) = \prod_{i=1}^{n} f(y_i \mid y_{i-1}; \theta)$$

interpretation of θ

- Example: pseudo-likelihood in spatial models   condition on near neighbours; Besag 74

... More flexible inference functions

Quasi-likelihood and generalized estimating equations

$$g\{\mathbb{E}(y_i \mid \boldsymbol{x}_i)\} = g(\mu_i) = \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \quad \mathrm{var}(y_i \mid \boldsymbol{x}_i) = \sigma^2 V(\mu_i)$$

- estimating equation for $\boldsymbol{\beta}$

full distribution unspecified

$$\sum_{i=1}^{n} \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \frac{(y_i - \mu_i)}{V(\mu_i)} = \mathbf{0}$$

column vector

Quadratic inference functions

Qu, Lindsay, Li 2000; Hector 2023

- replace $V^{-1}(\mu_i)$ above with an expansion in basis functions
- apply generalized method of moments

3. More flexible models

- identify one or more parameters of interest
- use a highly flexible specification for other aspects of the model

- Example: proportional hazards regression

instantaneous failure rate

$$h(t; x, \beta) = h_0(t)\exp(x^{\mathrm{T}}\beta)$$

- Example: empirical likelihood

$T(F)$ to be specified; e.g. $\mathbb{E}_F(Y_i)$

$$\max_F L(F; \boldsymbol{y}), \text{ subject to } T(F) = \theta$$

$L(F; \boldsymbol{y}) = \prod_{i=1}^n F(y_i)$

- Example: semi-parametric regression

$$\mathbb{E}(y \mid T, x) = \psi T + \omega(x)$$

- when does parameter of interest have a stable interpretation

model assumption

Example: exponential matched pairs

Battey & Cox 2020

- independent exponential pairs (y_{1i}, y_{2i}), $\quad i = 1, \ldots, n$ — $n + 1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- ψ common parameter of interest $\quad \gamma_i \quad$ pair-specific nuisance parameters

- likelihood function

$$L(\psi, \gamma; \boldsymbol{y}) \propto \prod_{i=1}^{n} \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\}$$

- possibilities for eliminating nuisance parameters
  - profile (concentrated) likelihood — maximize over γ
  - marginal likelihood: $f(\boldsymbol{t}; \psi) = \prod_{i=1}^{n} f(t_i; \psi)$ — $t_i = y_{1i}/y_{2i}$
  - random effects $\gamma_i \sim g(\cdot; \boldsymbol{\lambda})$ — more efficient, if ...

- independent exponential pairs (y_{1i}, y_{2i}), $\quad i = 1, \ldots, n$ — $n + 1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- random effects: $\gamma_i \sim \text{Gamma}(\alpha, \beta)$ — λ = (shape, rate)
- likelihood function

$$L(\psi, \alpha, \beta; \boldsymbol{y}) \propto \prod_{i=1}^{n} \int \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\} g(\gamma_i; \alpha, \beta) d\gamma_i$$

- orthogonality:

$$\mathbb{E}_{gamma}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha}\right\} = 0, \quad \mathbb{E}_{gamma}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta}\right\} = 0$$

- even better

$$\mathbb{E}_{m}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha}\right\} = 0, \quad \mathbb{E}_{m}\left\{-\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta}\right\} = 0$$

any random effects distribution

- independent exponential pairs $(y_{1i}, y_{2i}), \quad i = 1, \dots, n$ — $n+1$ parameters
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively
- random effects: $\gamma_i \sim \text{Gamma}(\alpha, \beta)$ — λ = (shape, rate)
- likelihood function

$$L(\psi, \alpha, \beta; \boldsymbol{y}) \propto \prod_{i=1}^{n} \int \gamma_i^2 \exp\{-\gamma_i(\frac{y_{1i}}{\psi} + \psi y_{2i})\} g(\gamma_i; \alpha, \beta) d\gamma_i$$

- even better

$$\mathbb{E}_m \left\{ -\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\alpha} \right\} = 0, \quad \mathbb{E}_m \left\{ -\frac{\partial^2 \log L(\psi, \alpha, \beta)}{\partial\psi\partial\beta} \right\} = 0$$

- and

$$\hat{\psi} \xrightarrow{p} \psi$$

any random effects distribution

PNAS RESEARCH ARTICLE | STATISTICS OPEN ACCESS

On the role of parameterization in models with a misspecified nuisance component

Heather S. Battey[a,2,1] and Nancy Reid[b,2,1]

Contributed by Nancy Reid; received February 8, 2024; accepted July 23, 2024; reviewed by Emmanuel J. Candès and Edward I. George

August 30, 2024 | 121 (36) e2402736121

Significance

Overview

- parameter of interest ψ is well-defined
- model with nuisance parameters may be misspecified — random effects
- when can we recover the true value of ψ
- does parameter orthogonality play a role?

- yes, it does, but may be difficult to verify directly — requires moments under the true model
- models based on groups satisfy this orthogonality
- with particular symmetry in the parametrization

- most natural examples seem to involve misspecified random effects

- has some links to Neyman orthogonality in doubly-debiased machine learning

Another look at matched pairs

- independent exponential pairs $(y_{1i}, y_{2i}), \quad i = 1, \ldots, n$
- rate parameters γ_i/ψ and $\gamma_i\psi$, respectively

- $t_i = y_{1i}/y_{2i}, \quad i = 1, \ldots, n$

- marginal likelihood

$$L(\psi; \boldsymbol{t}) = \prod_{i=1}^{n} \frac{\psi^2}{(1 + \psi^2 t_i)^2}$$

- inference can proceed as in classical setting

- is this also robust to model misspecification?
- for example, if $y_{1i} \sim \text{Exp}(\alpha_i + \beta), \quad y_{2i} \sim \text{Exp}(\alpha_i - \beta)$

wip

What can go wrong?

- the distributional approximations might be poor — likelihood skewed; extremes more relevant
- too many parameters — $p \sim n^{\alpha}, \quad p/n \to C, \quad p/n \to \infty$
- irregular parameter space — $pf(y;\theta_1) + (1-p)f(y;\theta_2), \quad 0 \le p \le 1$
- computational intractability — $L(\theta,\tau;y) = \int_{\mathbb{R}^k} f(y \mid z;\theta) f(z;\tau) dz$
- model is misspecified — true $Y \sim m(y), \quad f(\cdot;\theta) \neq m(\cdot) \qquad \forall\theta$

Likelihood methods, $p = O(n)$

- data $\quad y = (y_1, \ldots, y_n)$
- model $\quad f(y; \theta), \quad \theta \in \mathbb{R}^p; \quad$ or $\quad f(y \mid x; \beta)$ $\quad$ e.g. semi-par reg
- parameter of interest and nuisance parameters $\quad \theta = (\psi, \lambda)$
- low-dimensional $\quad$ high-dimensional
- for example factorial and fractional factorial designs $\quad$ e.g. design matrix X is orthogonal
- for example adjustments to profile log-likelihood $\quad$ e.g. $\hat{\sigma}^2 = \frac{RSS}{n} \longrightarrow \tilde{\sigma}^2 = \frac{RSS}{n-p}$

… likelihood methods, $p = O(n)$

- log-likelihood function $\ell(\theta; y) = \log f(y; \theta), \quad \theta \in \mathbb{R}^p, \quad y \in \mathbb{R}^n$
- profile log-likelihood function $\ell_{\text{prof}}(\psi; y) = \ell(\hat\theta_\psi) = \ell(\psi, \hat\lambda_\psi) \qquad \theta = (\psi, \lambda)$
- good enough if p fixed, $n \to \infty$

$$w = 2\{\ell_{\text{prof}}(\hat\psi) - \ell_{\text{prof}}(\psi)\} \xrightarrow{d} \chi^2_1, \qquad r = \pm w^{1/2} \xrightarrow{d} N(0,1)$$

- fails if $p = p_n$:

$$w \xrightarrow{d} \frac{\sigma_*^2}{\lambda_*}\chi^2_1$$

Sur, Chen, Candès 2019; logistic regression, $\psi = \beta_j$

- (σ_*, λ_*) characterized as the solution of two equations the optimization path

also depends on $\lim_{n\to\infty} p_n/n$

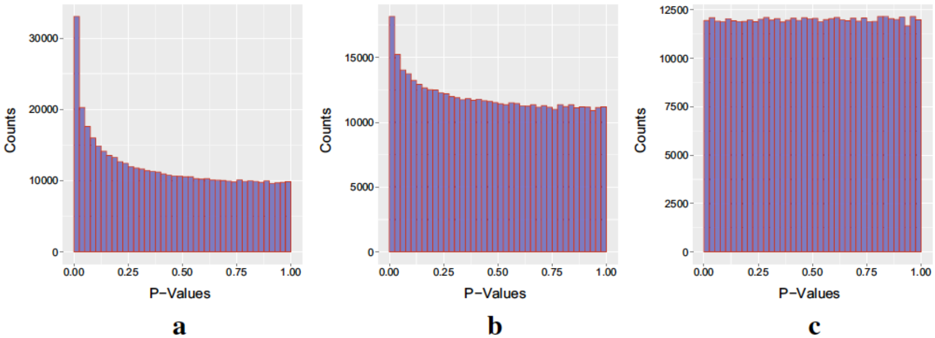


Fig. 1 Histogram of p-values for logistic regression under i.i.d. Gaussian design, when $\boldsymbol{\beta} = \mathbf{0}$, $n = 4000$, $p = 1200$, and $\kappa = 0.3$: **a** classically computed p-values; **b** Bartlett-corrected p-values; **c** adjusted p-values by comparing the LLR to the rescaled chi square $\alpha(\kappa)\chi_1^2$ (27)

Improvements to likelihood

1. adjust the profile log-likelihood function for estimation of nuisance parameters

   - $\ell_{\text{prof}}(\psi) = \ell(\psi, \hat\lambda_\psi) \quad \longrightarrow \quad \ell_{\text{mp}}(\psi) = \ell(\psi, \hat\lambda_\psi) - \frac{1}{2}\log|j_{\lambda\lambda}(\psi, \hat\lambda_\psi)|$ $\quad j_{\lambda\lambda}$: Fisher info
   - can lead to improved inference in finite samples

     e.g. Kosmidis & Firth 2019 *Bka* for logistic regression

     e.g. Sartori 2003 *Bka* for stratified models

2. adjust the log-likelihood ratio statistic

$$w = 2\{\ell_{\text{prof}}(\hat\psi) - \ell_{\text{prof}}(\psi)\} \dot\sim \chi^2_1$$

   - or its signed square root $\quad r = \text{sign}(\hat\psi - \psi)[2\{\ell_{\text{prof}}(\hat\psi) - \ell_{\text{prof}}(\psi)\}]^{1/2} \dot\sim N(0,1) + O_p(n^{-1/2})$

$$r^* = r + r_{\text{nuisance}} + r_{\text{info}}, \qquad r^* \sim N(0,1) + O_p(n^{-3/2})$$

- $$r^* = r + r_{\text{nuisance}} + r_{\text{info}} \sim N(0,1)$$

$p = O(n^{\alpha}), \alpha < 0.5$

- Tang & R Theorem 1:

$$r_{\text{nuisance}} = O_p(p^{3/2}/n^{1/2}), \qquad \text{can be as small as} \quad O_p(p/n^{1/2})$$

- Tang & R Theorem 2:

$$r_{\text{info}} = O_p(p/n^{1/2}), \qquad \text{can be as small as} \quad O_p(1/n^{1/2})$$

- $$r_{\text{nuisance}} \simeq \frac{1}{r}\log\left\{\frac{|j_{\lambda\lambda}(\hat\psi,\hat\lambda)|^{1/2}}{|j_{\lambda\lambda}(\psi,\hat\lambda_\psi)|^{1/2}}\right\}, \quad r_{\text{info}} \simeq \frac{1}{r}\log\left(\frac{t}{r}\right), \qquad t = (\hat\psi - \psi)/\hat\sigma$$

A different model and estimator

- $\hat{\beta}(\rho) = \arg\min \frac{1}{n}\sum_{i=1}^{n} \rho(y_i - x_i^{\mathrm{T}}\beta)$
- coordinate-wise asymptotic normality

$$max_j d_{TV}\left\{\mathcal{L}\left(\frac{\hat{\beta}_j - \beta_j^*}{\sqrt{\mathrm{var}(\hat{\beta}_j)}}\right), N(0,1)\right\} = o(1)$$

- “For instance for least-squares, standard degrees of freedom adjustments effectively take care of many dimensionality-related problems”
- ?perhaps HOA adjustments for nuisance parameters (= ‘standard degrees of freedom adjustments’) can be as effective as using $p/n \to \kappa$ asymptotics? when? why not?

Split likelihood ratio

- Assume $y = (y_1, \ldots, y_{2n})$ i.i.d. $f(y; \theta)$
- Create two subsamples $(y_{(1)}, y_{(2)}$ each of length n
- Use $y_{(2)}$ to estimate θ with $\hat{\theta}_{(2)}$
- Use $y_{(1)}$ to create a likelihood ratio

$$T_n(\theta; y) = \frac{L(\hat{\theta}_{(2)}; y)}{L(\theta; y_{(1)})}$$

- a universal confidence set for θ:

$$C(\theta; y) = \{\theta \in \Theta : T_n(\theta; y) \leq \frac{1}{\alpha}\}$$

Shi & Drton, 2025; Strieder & Drton, 2022

Overview

- likelihood function serves as a basis for inference
- seems inevitable if probability models are to be used
- many variations seem to be needed when the models get more complicated
- marginal (or conditional) likelihood, higher-order corrections to likelihood
- regularization, new distribution theory, different asymptotic analysis, computational simplifications, simulations, model selection, ...
- what guides solutions for new problems, complex models, scientific questions

Foundations

What are the foundations of statistics?

- Statistics needs a healthy interplay between theory and applications
- theory meaning foundations, rather than theoretical analysis of specific techniques
- must be continually tested against new applications
- "the practical application of general theorems is a different art from their establishment by mathematical proof" — Fisher 1958 SMRW

International Statistical Review

International Statistical Review (2015), 83, 2, 293–308 doi:10.1111/insr.12067

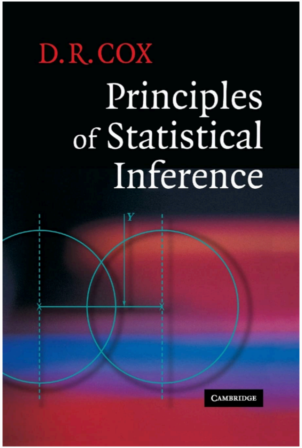
On Some Principles of Statistical Inference

Nancy Reid[1] and David R. Cox[2]

... What are the foundations of statistics?

- probability, analysis, applied mathematics — modelling
- Bayes, Neyman, Fisher — approaches to inference
- nature of uncertainty — epistemic, empirical
- nature of induction — belief functions, inferential models
- interpretation of p-values, confidence regions, credibility intervals, likelihood ratios
- role of sufficiency, ancillarity, conditioning, asymptotic theory
- sparsity, causality, assumption-free/lean inference, stability, prediction, decisions

… What are the foundations of statistics?



I'm fairly cautious about the impact of the book in that it really is very cryptic indeed on key issues but we will see. In particular quite apart from the Bayesian stuff I have essentially discarded (not rejected) the Neyman-Pearson machinery in favour of Fisher's original approach and I am sure this is the right route.

Thanks to AWF Edwards

What use are foundations?

- provide a rigorous basis for the development of techniques
- provide a common language for particular classes of problems
- help to clarify the nature of uncertainty in scientific conclusions
- highlight aspects of data analysis which are likely to raise difficult issues
- suggest strategies for tackling highly complex problems
- avoid ‘re-inventing the wheel’ for each new application

Likelihood and Its Extensions

Nancy Reid, Cristiano Varin, and Grace Y. Yi

Thank you

CANSSI INCASS ONTARIO