

# Models and likelihood

---

Nancy Reid  
University of Toronto

October 29 2025

**UF** Liberal Arts *and* Sciences

**STATISTICS**



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 Grieshaber, 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 ethn  
 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

are melanoma cases more likely to have tattoos

case-control study

**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

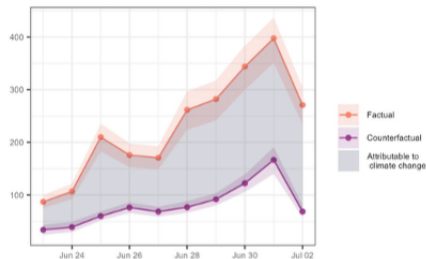
# IMPERIAL

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

Excess heat-related deaths



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

simulated data

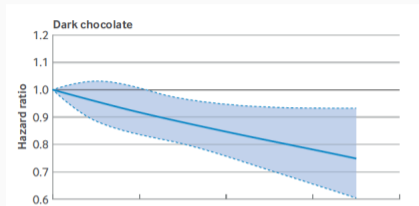
**Statistical model:** “We use established epidemiological models and the WWA climate-attribution framework”

extreme value distribution; regression

## RESEARCH

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

Binkai Liu,<sup>1</sup> Geng Zong,<sup>2,3</sup> Lu Zhu,<sup>1</sup> Yang Hu,<sup>1</sup> JoAnn E Manson,<sup>4,5,6</sup> Molin Wang,<sup>4,5,7</sup>  
Eric B Rimm,<sup>1,4,5</sup> Frank B Hu,<sup>1,4,5</sup> Qj Sun<sup>1,4,5</sup>



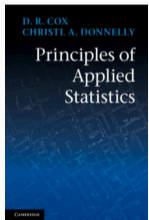
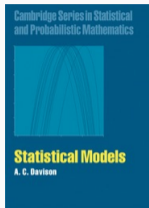
**Scientific question:** Does dark chocolate consumption affect the risk of Type 2 diabetes?

**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, case-control study, observational data, 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
  - ...



- 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 to aid the interpretation of data
- perturbations of no intrinsic interest distort an otherwise exact measurement physical models
- substantial natural variability in the phenomenon under study human subjects

# The role of parameters

- probability models very likely be parameterized
- thus defining a class of models
- parameters may be finite- or infinite-dimensional

$$\{f(y; \theta); \theta \in \Theta\}$$

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?<sup>1</sup>

BY PETER McCULLAGH

*University of Chicago*

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

# The likelihood function

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

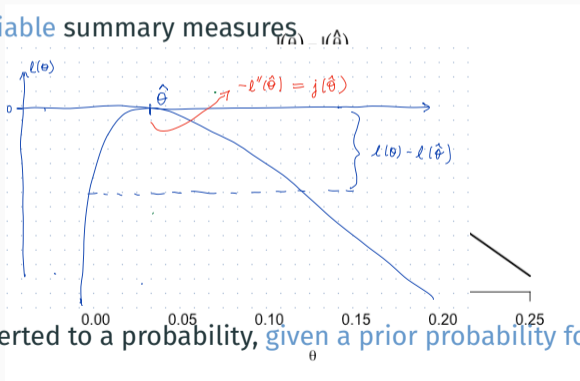
inverse problem

- provides a convenient way to compare parameter values

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

- provides **reliable** summary measures

$\ell(\theta; \mathbf{y}) = \log L(\theta; \mathbf{y})$



- can be converted to a probability, given a prior probability for  $\theta$

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

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

$\implies$  score statistic is approximately normally distributed

$$I^{-1/2}(\theta) \ell'(\theta) \sim N(\mathbf{0}, Id)$$

$\implies$  MLE is approximately normally distributed

$$J(\theta) = -\ell''(\theta)$$

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

$\implies$  LRT is approximately  $\chi^2$  distributed

$$I(\theta) = \mathbb{E}_{\theta}\{J(\theta)\}$$

$$2\{\ell(\hat{\theta}) - \ell(\theta)\} \sim \chi_p^2$$

Large-sample approximation:

$$\hat{\theta} \sim N_p\{\theta, J^{-1}(\hat{\theta})\}, \quad 2\{\ell(\hat{\theta}) - \ell(\theta)\} \sim \chi_p^2$$

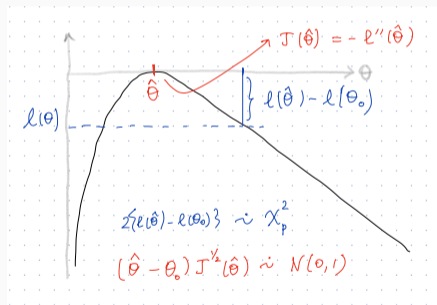
Coefficients:

	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)

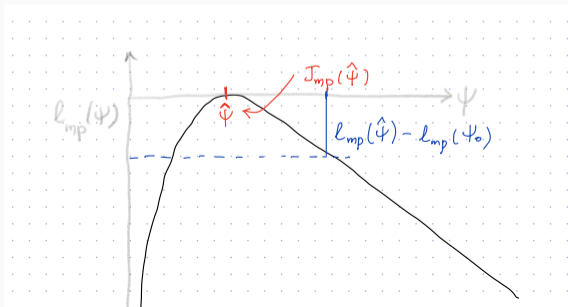
Null deviance: 40.710 on 22 degrees of freedom  
Residual deviance: 18.069 on 17 degrees of freedom



# 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
- too many parameters
- irregular parameter space
- computational intractability
- model is misspecified

likelihood skewed;  
extremes more relevant

$$p \sim n^\alpha, \quad p/n \rightarrow C, \quad p/n \rightarrow \infty$$

$$\alpha f(\mathbf{y}; \theta_1) + (1 - \alpha)f(\mathbf{y}; \theta_2), \quad 0 \leq \alpha \leq 1$$

$$L(\theta, \tau; \mathbf{y}) = \int_{\mathbb{R}^k} f(\mathbf{y} | \mathbf{z}; \theta) f(\mathbf{z}; \tau) d\mathbf{z}$$

$$\text{true } Y \sim m(\mathbf{y}), \quad f(\cdot; \theta) \neq m(\cdot)$$

$\forall \theta$

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

$$\mathbf{y} = (y_1, \dots, y_n)$$

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

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

$$\hat{\theta} \equiv \arg \sup_{\theta} \ell(\theta; \mathbf{y})$$

- $\hat{\theta}$  converges to the “closest true value”

KL-divergence

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

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

“sandwich variance”

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

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

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

“we used robust standard errors”

## 2. More flexible inference functions

### Composite likelihood

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

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

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

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

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

interpretation of  $\theta$

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

Quasi-likelihood and **generalized estimating equations**

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

- 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

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

- Example: empirical likelihood

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

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

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

- Example: semi-parametric regression

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

- **when does parameter of interest have a stable interpretation**

model assumption

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

- likelihood function

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

- possibilities for eliminating nuisance parameters
  - profile (concentrated) likelihood
  - marginal likelihood:  $f(\mathbf{t}; \psi) = \prod_{i=1}^n f(t_i; \psi)$
  - **random effects**  $\gamma_i \sim g(\cdot; \boldsymbol{\lambda})$

maximize over  $\gamma$   
 $t_i = y_{1i}/y_{2i}$   
 more efficient, if ...

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

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

- **orthogonality**:

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

- **even better**

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

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

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

- even better

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

- **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<sup>a,2,1</sup> and Nancy Reid  <sup>b,2,1</sup>

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

- **parameter of interest**  $\psi$  is well-defined
- model with nuisance parameters may be misspecified random effects
- when can we recover the true value of  $\psi$
- 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  $(y_{1i}, y_{2i})$ ,  $i = 1, \dots, n$
- rate parameters  $\gamma_i/\psi$  and  $\gamma_i\psi$ , respectively
- $t_i = y_{1i}/y_{2i}$ ,  $i = 1, \dots, n$

- marginal likelihood

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

- inference can proceed as in classical setting
- is this also robust to model misspecification?
- for example, if  $y_{1i} \sim \text{Exp}(\alpha_i + \beta)$ ,  $y_{2i} \sim \text{Exp}(\alpha_i - \beta)$

wip

## What can go wrong?

- the distributional approximations might be poor
- too many parameters
- irregular parameter space
- computational intractability
- model is misspecified

likelihood skewed;  
extremes more relevant

$$p \sim n^\alpha, \quad p/n \rightarrow C, \quad p/n \rightarrow \infty$$

$$pf(y; \theta_1) + (1 - p)f(y; \theta_2), \quad 0 \leq p \leq 1$$

$$L(\theta, \tau; \mathbf{y}) = \int_{\mathbb{R}^k} f(\mathbf{y} | \mathbf{z}; \theta) f(\mathbf{z}; \tau) d\mathbf{z}$$

$$\text{true } Y \sim m(\mathbf{y}), \quad f(\cdot; \theta) \neq m(\cdot) \quad \forall \theta$$

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

## ... likelihood methods, $p = O(n)$

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

$$W = 2\{\ell_{\text{prof}}(\hat{\boldsymbol{\psi}}) - \ell_{\text{prof}}(\boldsymbol{\psi})\} \xrightarrow{d} \chi_1^2, \quad \mathbf{r} = \pm W^{1/2} \xrightarrow{d} N(\mathbf{0}, \mathbf{1})$$

- fails if  $p = p_n$ :

$$W \xrightarrow{d} \frac{\sigma_*^2}{\lambda_*} \chi_1^2$$

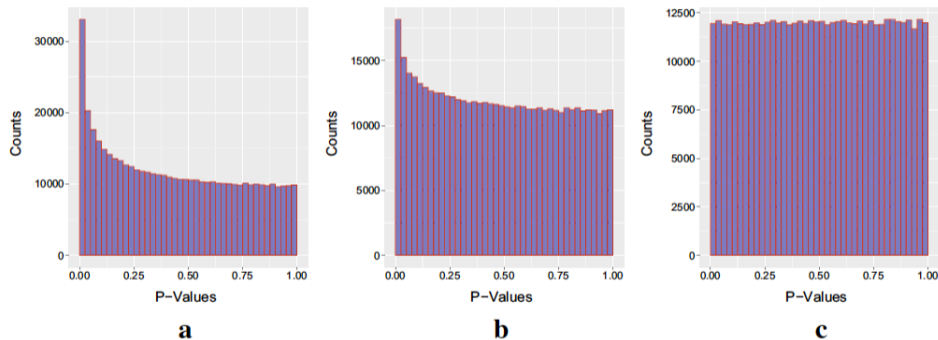
Sur, Chen, Candès 2019; logistic regression,  $\boldsymbol{\psi} = \boldsymbol{\beta}_j$

- $(\sigma_*, \lambda_*)$  characterized as the solution of two equations the optimization path

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

490

P. Sur et al.



**Fig. 1** Histogram of  $p$ -values for logistic regression under i.i.d. Gaussian design, when  $\beta = \mathbf{0}$ ,  $n = 4000$ ,  $p = 1200$ , 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)

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

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

$$W = 2\{\ell_{\text{prof}}(\hat{\psi}) - \ell_{\text{prof}}(\psi)\} \sim \chi_1^2$$

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

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

•

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

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

• Tang &amp; R Theorem 1:

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

• Tang &amp; R Theorem 2:

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

•

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

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

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

- a **universal** confidence set for  $\theta$ :

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

Shi & Drton, 2025; Strieder & Drton, 2022

- 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



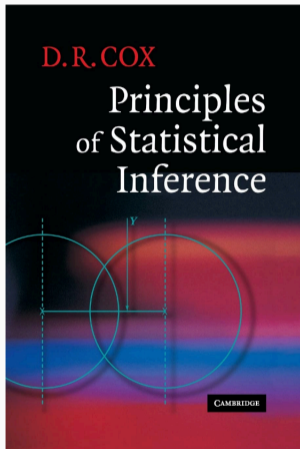
- Statistics needs a healthy interplay between theory and applications
- theory meaning **foundations**, rather than theoretical analysis of specific techniques
- must be continually tested against new applications
- “the practical application of general theorems is a different art from their establishment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

Fisher 1958 SMRW



## ... 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



*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!

UF

Liberal Arts *and* Sciences

**STATISTICS**

