Models and parameters

Inference under model misspecification

Nancy Reid
University of Toronto

August 6 2024

Statistical Science
2020, Vol. 35, No. 2, 308–320
https://doi.org/10.1214/19-STS734

A Conversation with Grace Wahba

Douglas Nychka, Ping Ma and Douglas Bates

Abstract. Grace Wahba (née Goldsmith, born August 3, 1934), I. J. Schoenberg-Hilldale Professor of Statistics at the University of Wisconsin-Madison (Emerita), is a pioneer in methods for smoothing noisy data. Her research combines theoretical analysis, computation and methodology motivated by innovative scientific applications. Best known for the development of generalized cross-validation (GCV), the connection between splines and Bayesian posterior estimates, and "Wahba's problem," she has developed methods with applications in demographic studies, machine learning, DNA microarrays, risk modeling, medical imaging and climate prediction.

Grace grew up in the Washington, DC area and New Jersey, and graduated

Outline

1. Examples: a haphazard selection
2. Models and parameters
3. Some approaches to misspecification
4. Example
5. Formalization
6. Discussion

Examples: a haphazard selection

Clim. Past, 20, 1387–1399, 2024
https://doi.org/10.5194/cp-20-1387-2024

Climate of the Past
Open Access

600 years of wine must quality and April to August temperatures in western Europe 1420–2019

Christian Pfister[1], Stefan Brönnimann[2], Andres Altwegg[3], Rudolf Brázdil[4], Laurent Litzenburger[5], Daniele Lorusso[6], and Thomas Pliemon[7]

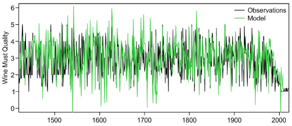


Figure 4. Observed series of wine quality (average; black) from 1420 to 2019 and series obtained with a statistical model calibrated in 1781–1800 (green). The model is explained in Sect. 3.

Scientific question: Can historical records of wine quality be used as temperature proxies?

observational data

Statistical model: “we used a statistical [linear regression] model for wine quality based on local temperature and precipitation”

yes, if used carefully

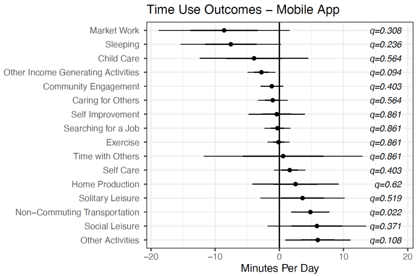
NBER WORKING PAPER SERIES

THE EMPLOYMENT EFFECTS OF A GUARANTEED INCOME:
EXPERIMENTAL EVIDENCE FROM TWO U.S. STATES

Eva Vivalt
Elizabeth Rhodes
Alexander W. Bartik
David E. Broockman
Sarah Miller

Working Paper 32719
http://www.nber.org/papers/w32719

Figure 5: Time Use Results: Mobile App



Scientific questions: Does guaranteed income supplement affect labor market measures?

randomized controlled trial

Statistical model: $Y_i = \alpha + \beta Treated_i + \gamma^{\mathrm{T}} X_i + \epsilon_i$

"support for both sides of this debate"

JAMA Oncology | Original Investigation

Bilateral Mastectomy and Breast Cancer Mortality

Vasily Giannakeas, PhD, MPH; David W. Lim, MDCM, MEd, PhD; Steven A. Narod, MD

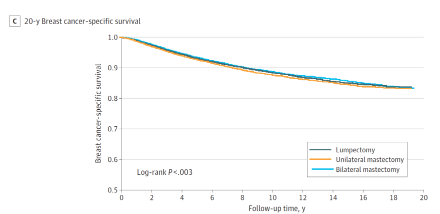
IMPORTANCE The benefit of bilateral mastectomy for women with unilateral breast cancer in terms of deaths from breast cancer has not been shown.

OBJECTIVES To estimate the 20-year cumulative risk of breast cancer mortality among women with stage 0 to stage III unilateral breast cancer according to the type of initial surgery performed.

DESIGN, SETTINGS, AND PARTICIPANTS This cohort study used the Surveillance, Epidemiology, and End Results (SEER) Program registry database to identify women with unilateral breast cancer (invasive and ductal carcinoma in situ) who were diagnosed from 2000 to 2019. Three closely matched cohorts of equal size were generated using 1:1:1 matching according to surgical approach. The cohorts were followed up for 20 years for contralateral breast cancer and for breast cancer mortality. The analysis compared the 20-year cumulative risk of breast cancer mortality for women treated with lumpectomy vs unilateral mastectomy vs bilateral mastectomy. Data were analyzed from October 2023 to February 2024.

EXPOSURES Type of breast surgery performed (lumpectomy, unilateral mastectomy, or bilateral mastectomy).

+ Editorial
+ Supplemental content



Scientific question: Does bilateral mastectomy for unilateral breast cancer improve 20-year survival?

matched case-cohort study

Statistical model: “We used the Kaplan-Meier method to estimate survival”

“preemptive surgery did not appear to reduce the risk of death”

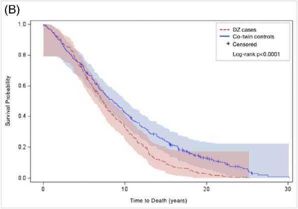

DOI: 10.1002/alz.13553

Alzheimer's & Dementia®
THE JOURNAL OF THE ALZHEIMER'S ASSOCIATION

RESEARCH ARTICLE

Dementia and mortality in older adults: A twin study

Jung Yun Jang[1] | Christopher R. Beam[2,3] | Ida K. Karlsson[4] | Nancy L. Pedersen[2,4] | Margaret Gatz[4,5]



Scientific question: Relationship between dementia and mortality

observational study of discordant twin pairs

Statistical model: multi-level Cox regression with random effects

"genetic variance contributes to the association between dementia risk and mortality"

Article https://doi.org/10.1038/s41550-023-01983-1

Variability of extragalactic X-ray jets on kiloparsec scales

Received: 17 May 2022

Accepted: 27 April 2023

Published online: 29 May 2023

Eileen T. Meyer[1], Aamil Shaik[1], Yanbo Tang[2], Nancy Reid[3], Karthik Reddy[1,4], Peter Breiding[5], Markos Georganopoulos[1], Marco Chiaberge[5,6], Eric Perlman[7], Devon Clautice[7], William Sparks[8,9], Nat DeNigris[1,10] & Max Trevor[1,11]

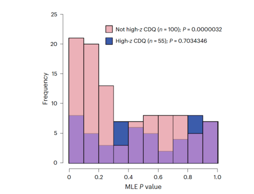


Fig. 3 | Histogram of the single-region *P* values from the directional test, not adjusted for multiple comparison. In pink, the subset of sources that

Yanbo Tang

Scientific question: Are observations of X-ray jets consistent with current theory?

observational data

Statistical model: compare background and sources measurements using Poisson:

$$x_i \sim Po(a_i \beta_i), \quad y_i \sim Po(b_i \beta_i + b_i f_i \mu_i)$$

$$H : \mu_i \equiv 0$$

"variability in the X-ray emission is not compatible with proposed mechanism"

Models and parameters

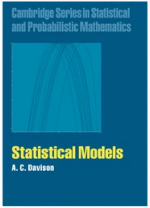
Why these models?

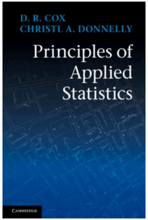
- standard in the literature of that field — income
- standard in the publications of that lab

breast cancer

- follow some prescription:
  - binary response — use logistic regression
  - time to event — use PH model
  - time series — use ARMA — wine
  - repeated measures — use random effects — Alzheimer's twin study
  - ...
- motivated by theory: economic, physical, ... — *X*-ray jets

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

Statistical Science
1990, Vol. 5, No. 2, 160–168

Model Specification: The Views of Fisher and Neyman, and Later Developments

E. L. Lehmann

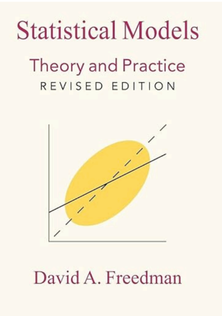
Role of Models in Statistical Analysis

D. R. Cox

Abstract. A number of distinct roles are identified for probability models used in the analysis of data. Examples are outlined. Some general issues arising in the formulation of such models are discussed.

empirical, or predictive models, contrasted with explanatory models

indirect models



The emphasis throughout is on the connection
– or lack of connection –
between the models and the real phenomena.

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 role of the data

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

WHAT IS A STATISTICAL MODEL?[1]

BY PETER MCCULLAGH
University of Chicago

Some approaches to misspecification

1. Classical

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

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

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

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

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

- $\widehat{\theta}$ 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"

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

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

- 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 $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$

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

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

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

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

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

- estimating equation for β

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 — here β
- use a highly flexible specification form for other aspects of the model

- Example: proportional hazards regression — instantaneous failure rate

$$\lambda(t; x, \beta) = \lambda_0(t)\exp(x^{\mathrm{T}}\beta)$$

- Example: empirical likelihood — $T(F)$ to be specified; e.g. $\mathrm{E}_F(Y_i)$

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

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

- Example: semi-parametric regression

$$\mathrm{E}\,(y \mid T, x) = \beta T + \omega(x)$$

- does parameter of interest have a stable interpretation — model assumption

- Possible model $\mathrm{E}\,(y \mid T, x) = \beta\, T + \omega(x)$ binary treatment T

- Define an estimand of interest limit of E-estimator, Robins et al 92

$$\frac{\mathrm{E}[\pi(x)\{1-\pi(x)\}\{\mathrm{E}(y \mid T=1, x) - \mathrm{E}(y \mid T=0, x)\}]}{\mathrm{E}[\pi(x)\{1-\pi(x)\}]}$$

propensity score $\pi(x) = \mathrm{pr}(T=1 \mid x)$

- reduces to β under this model
- is a meaningful quantity when the model is incorrect e.g. interaction between T and x

- more generally, given a link function g:

$$g\{\mathrm{E}(y \mid T, x)\} = \beta T + \omega(x)$$

$$\frac{\mathrm{E}\,(\pi(x)\{1-\pi(x)\}[g\{\mathrm{E}(y \mid T=1, x)\} - g\{\mathrm{E}(y \mid T=0, x)\}])}{\mathrm{E}[\pi(x)\{1-\pi(x)\}]}$$

- the proposed estimand is ‘parsimonious’: more complicated models not allowed to permit more complicated estimands Daniel
- models, even used only as tools, may implicitly affect the meaning of our estimands Didelez
- marginal assessment of treatment effect may be more relevant for policy Ding
- for a parameter of interest to be stable over uncertainty in other aspects of the model, is some version of orthogonality required/useful Battey
- estimands as nonparametric projections Hines & Diaz-Ordaz
- ...

Example

Exponential Matched Pairs

- survival times for n matched pairs (y_{1i}, y_{2i})

- random assignment of pair members to treatment/control
- nuisance parameters describing the pairs $\lambda_1, \ldots, \lambda_n$
- parameter of interest is the treatment effect

- model y_{1i} exponential with rate λ_i/ψ
  y_{2i} exponential with rate $\lambda_i\psi$
- ψ common parameter of interest; λ_i pair-specific nuisance parameters

- possible approaches to inference for ψ
  - profile likelihood: maximize over nuisance parameters
  - marginal likelihood: distribution of y_{1i}/y_{2i} is free of nuisance parameters
  - integrated likelihood: assume a distribution for λ_i random effects model

... Exponential matched pairs

Battey & Cox 2020

- model y_{1i} exponential with rate λ_i/ψ
  y_{2i} exponential with rate $\lambda_i\,\psi$

- random effects model: $\lambda_i \sim \text{Gamma}(\alpha, \beta)$ — shape, rate

- integrated likelihood

$$L(\psi, \alpha, \beta; \boldsymbol{y}) = \int f(\boldsymbol{y}; \psi, \boldsymbol{\lambda}) g(\boldsymbol{\lambda}; \alpha, \beta) d\boldsymbol{\lambda}$$

- in the integrated model, ψ is orthogonal to (α, β) — w.r.t. Fisher information
- even better: this is the case for any random effects distribution — not just Gamma

- conclude MLE $\widehat{\psi} \xrightarrow{p} \psi$ even if random effects model is misspecified

could be inefficient

can this be generalized?

Aside: details

1. for pair (y_{1i}, y_{2i})

$$L(\psi, \alpha, \beta; y_{1i}, y_{2i}) = \frac{\Gamma(\alpha+2)}{\Gamma(\alpha)} \frac{\beta^{\alpha}}{(y_{1i}/\psi + \psi y_{2i} + \beta)^{\alpha+2}}$$

2. orthog

$$\frac{\partial^2}{\partial\psi\partial\alpha} \log L(\psi, \alpha, \beta; y_{1i}, y_{2i}) = \frac{y_{2i} - y_{1i}/\psi^2}{y_{2i}\psi + y_{1i}/\psi + \beta}$$

3.

$$\mathrm{E}_m \left\{ \frac{y_{2i} - y_{1i}/\psi^2}{y_{2i}\psi + y_{1i}/\psi + \beta} \right\} = 0$$

4. for Gamma random effects, but also for any random effects distribution

interpretation of α, β when not Gamma

Formalization

Towards formalization

- true model $m(\boldsymbol{y})$ with parameter ψ and true value ψ_*
- fitted model $f(\boldsymbol{y};\psi,\lambda)$ same parameter of interest, (many) nuisance parameters

  interpretation of ψ is stable
- maximum likelihood estimates $(\hat{\psi},\hat{\lambda}) \xrightarrow{p} (\psi_m^0,\lambda_m^0)$

$$\mathrm{E}_m\{\partial\ell(\psi_m^0,\lambda_m^0)/\partial(\psi,\lambda)\} = 0$$

- assume no value of $\lambda \in \Lambda$ gives back $m(\cdot)$ misspecified
- Does $\widehat{\psi} \xrightarrow{p} \psi_*$? need $\mathrm{E}_m\{\partial\ell(\psi_*,\lambda_m^0)/\partial\psi\} = 0$ (1) λ_m^0 unknown
- can be easier to show $\mathrm{E}_m\{\partial\ell(\psi_*,\lambda)/\partial\psi\} = 0 \quad \forall\lambda$ (2)
- Result 1: (1) $\equiv$ (2) $\iff$ ψ_* is m-orthogonal to Λ

- Result 1: (1) $\equiv$ (2) $\iff$ ψ_* is *m*-orthogonal to Λ — hard $\equiv$ easy

- Definition *m*-orthogonal — $\psi_* \perp_m \Lambda$

$$\forall\lambda \quad \mathrm{E}_m\left\{\frac{\partial^2\ell(\psi,\lambda)}{\partial\psi\partial\lambda}\right\} = 0 \quad (2)$$

- But, $\widehat{\psi}$ can be consistent without this requirement

- Result 2: A weaker requirement — still too strong

$$\text{If } I^{\psi\psi}g_\psi + I^{\psi\lambda}g_\lambda = 0, \quad \forall\,\lambda, \text{ then } \quad \psi_m^0 = \psi_*$$

$I = \mathrm{E}_m\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\}; \quad g = \mathrm{E}_m\{\partial\ell(\theta)/\partial\theta\};$ partitioned

Parameter orthogonality

- we can often establish parameter orthogonality in the assumed model $f(\boldsymbol{y};\psi,\lambda)$
- all expectations with respect to this assumed model
- this is not usually the same as m-orthogonality in the true model $m(\boldsymbol{y};\psi)$
- Result 3 — a special case
  If the assumed log-likelihood function is linear in sufficient statistics S,
  and
  $$\mathrm{E}_m(S_j) = \mathrm{E}_{(\psi,\lambda)}(S_j),$$
- then assumed-model orthogonality $\Longrightarrow$ m-orthogonality — Sartori et al., 2010

Parameter Symmetry

- Example: matched exponential pairs $\mathrm{E}(Y_{1i}) = \psi/\lambda_i; \quad \mathrm{E}(y_{2i}) = 1/(\psi\lambda_i)$
- detailed calculation established $\hat{\psi} \xrightarrow{p} \psi_*$ under misspecification   using orthogonality
- why did this work?
- the parameter of interest enters symmetrically
- the calculations repeatedly use a change of variables to y_{1i}/y_{2i} and $y_{1i}y_{2i}$
- how to generalize this observation?

… Parameter Symmetry

- from earlier results, want *m*-orthogonal parametrization

$\ell = \log L$

$$\mathrm{E}_m\{-\partial^2 \ell(\psi, \lambda)/\partial\psi\partial\lambda^{\mathrm{T}}\} = 0$$

or at least at ψ_*

- we don't know the true model *m*, so can't check this
- the exponential matched pairs example is a group model

scale group

- their parametrization ensures cancellation of terms

- Result 4: If the joint distribution of (Y_1, Y_2) is parametrized ψ-symmetrically, and this parametrization induces anti-symmetry on the ψ-score function, then

$$\psi^* \perp_m \Lambda, \quad \mathrm{E}_m\{\partial\ell(\psi_*, \lambda)/\partial\psi\} = 0$$

- this in turn implies

$$\widehat{\psi} \xrightarrow{p} \psi_*.$$

… Formalization and Parameter Symmetry

- Result 4: If the joint distribution of (Y_1, Y_2) is parametrized ψ-symmetrically, and this parametrization induces anti-symmetry on the ψ-score function, then

$$\psi^* \perp_m \Lambda, \quad \mathrm{E}_m\{\partial \ell(\psi_*, \lambda)/\partial \psi\} = 0$$

- this in turn implies

$$\widehat{\psi} \xrightarrow{p} \psi_*.$$

- Example: Scale family

$g_\psi \in$ scale group

$$f_{Y_1}(y_1; \lambda\psi) = f_U(y_1/\psi; \lambda)(1/\psi),$$
$$f_{Y_2}(y_2; \lambda/\psi) = f_U(y_2\psi; \lambda)\psi,$$
$$U, \overset{d}{=} g_\psi^{-1} Y_1 \overset{d}{=} g_\psi Y_2$$

- Example: Location family

$g_\psi \in$ location group

$$f_{Y_1}(y_1; \lambda + \psi) = f_U(y_1 - \psi; \lambda),$$
$$f_{Y_2}(y_2; \lambda - \psi) = f_U(y_2 + \psi; \lambda)$$

Aside: definitions

- joint distribution of Y_1, Y_2 parametrized ψ-symmetrically:
  - p_1 and p_2 are measures for a transformation model on G

$$p(gy; g\lambda)d(gy) = p(y; \lambda)dy, \qquad g \in G, y \in \mathcal{Y}, \lambda \in \Lambda$$

  - **and** group action g depends only on ψ
  - **and** p_1 and p_2 are on the same λ-orbit: $\forall u \in \mathcal{Y}, p_1(gu; g\lambda)d(gu) = p_2(g^{-1}u; g^{-1}\lambda)d(g^{-1}u)$

- if this parametrization induces anti-symmetry on the ψ-score function
  - log-likelihood function

$$\ell(\psi; \lambda, y_1, y_2) = \log f_1(y_1; g_\psi \lambda) + \log f_2(y_2; g_\psi^{-1}\lambda)$$

  - as a function of u:

$$\ell(\psi; \lambda, u_1, u_2), \qquad u_1 = g_\psi^{-1} y_1, \quad u_2 = g_\psi y_2$$

  - anti-symmetry:

$$\partial\ell(\psi; \lambda, u_1, u_2)/\partial\psi = -\partial\ell(\psi; \lambda, u_2, u_1)/\partial\psi$$

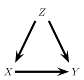
- then $\widehat{\psi} \xrightarrow{p} \psi_*$

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 — E_m
- models based on groups satisfy this orthogonality
- with particular parameter structure

- most natural examples seem to involve misspecified random effects — GLM disp
- another example is marginal structural model in a 'frugal parameterization'
- propensity score is the nuisance; other aspects correspond to ψ — Evans & Didelez (2024)
- E&D model has a parameter space cut, hence orthogonal



Discussion

Tentative conclusions, further work

- Results above only establish consistency
- asymptotic variance is much more difficult — although estimating it might be okay

- in the matched pairs examples, nuisance parameters treated as arbitrary constants can be eliminated by transformation to conditional or marginal distributions
- effectively assuming an arbitrary (nonparametric) mixing distribution
- less efficient when the random effects model is correct

- orthogonality under assumed model $\mathrm{E}_\theta\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\} = 0$ — $\theta = (\psi, \lambda)$
- *m*-orthogonality under true model $\mathrm{E}_m\{-\partial^2\ell(\theta)/\partial\theta\partial\theta^{\mathrm{T}}\} = 0$
- connection to Neyman orthogonality? — decorrelated score

$$\partial\ell(\psi,\lambda)/\partial\psi - w^{\mathrm{T}}\partial\ell(\psi,\lambda)/\partial\lambda, \quad w = I_{\psi\lambda}I_{\lambda\lambda}^{-1}$$

- extension to general estimating equations important in 2-debiased ML

Chernozhukov et al 2018, Ning et al 2017, Jorgensen & Knudsen 2004

Summary

- Result 1: orthogonal parameters lead to consistent estimate
- Result 2: slightly weaker condition than orthogonality — but hard to check
- Result 3: linearity in sufficient statistics $\implies$ orthogonality
- Result 4: certain symmetries of parametrization also $\implies$ orthogonality
- hence consistency

What's old is new again

Statistical Science
2011, Vol. 26, No. 3, 388–402
DOI: 10.1214/11-STS361

Misspecifying the Shape of a Random Effects Distribution: Why Getting It Wrong May Not Matter

Charles E. McCulloch and John M. Neuhaus

ELSEVIER
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 5142–5154
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Robustness of the linear mixed model to misspecified error distribution

Hélène Jacqmin-Gadda[a,b,*], Solenne Sibillot[a,b], Cécile Proust[a,b], Jean-Michel Molina[c], Rodolphe Thiébaut[a,b]

[a] *Institut National de la Santé et de la Recherche Médicale, Equipe de biostatistique E0338, Bordeaux, France*
[b] *Université Victor Segalen Bordeaux II, Bordeaux, France*
[c] *Département de maladies infectieuses, Assistance Publique Hôpitaux de Paris, Hôpital Saint Louis, Paris, France*

Available online 27 June 2006

Biometrika (2001), **88**, 4, *pp.* 973–985

Misspecified maximum likelihood estimates and generalised linear mixed models

By PATRICK J. HEAGERTY and BRENDA F. KURLAND
Department of Biostatistics, University of Washington, Seattle, Washington 98195, U.S.A.
heagerty@u.washington.edu kurland@u.washington.edu

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

WHAT IS A STATISTICAL MODEL?[1]

By Peter McCullagh
University of Chicago

References i

Battey, H.S. & Cox, D.R. (2020). High dimensional nuisance parameters: an example from parametric survival analysis. *Information Geometry* **3** 119–148. matched pairs exp

Battey, H.S. & Reid, N. (2024). On the role of parametrization in models with a misspecified nuisance component. arxiv *PNAS*, to appear.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B **36**, 192–236.

Chernozhukov, V. et al. (2018). Double debiased machine learning. *Econometrics Journal* **21**, C1–C68. doi: 10.1111/ectj.12097 Neyman orthogonality

Cox, D.R. (1961). Tests of separate families of hypotheses. *4th Berkeley Symposium* **1**, 105–123.

Cox, D.R. (1962). Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B **24**, 406–424.

References ii

Cox, D.R. (1990). Role of models in statistical analysis. *Statist. Sci.* **5**, 169–174.

Cox, D.R. and Donnelly, C.A. (2011). *Principles of Applied Statistics*. Cambridge University Press.

Davison, A.C. (2003). *Statistical Models*. Cambridge University Press.

Evans, R.J. & V.Didelez (2024). Parameterizing and simulating from causal models (with discussion). *J. R. Statist. Soc. B* **86**, to appear. frugal parametrization

Freedman, D.A. (2005). *Statistical Models*. Cambridge University Press.

Hector, E. (2023). Fused mean structure learning in data integration with dependence. *Canad. J. Statist.*

Huber, P.J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. *5th Berkeley Symposium* **1**, 221–233.

References iii

Jorgensen, B. & Knudsen, S.J. (2004). Parameter orthogonality and bias adjustment for estimating functions. *Scand. J. Statist.* **31**, 93–114.

Lehmann, E.L. (1990). Model specification: the views of Fisher and Neyman, and later developments. *Statist. Sci.* **5**, 160–168.

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

Ning, Y. & Liu, H. (2017). A general theory of hypothesis tests and confidence regions for sparse high-dimensional models. *Annals of Statistics* **45**, 158–195. decorrelated score

Qu, A.m Lindsay, B.G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika* **87**, 823–836.

Sartori, N., Severini, T.A. & Marras, E. (2010). An alternative specification of generalized linear mixed models. *Comp. Stat. Data. Anal.* **54**, 575–584. https://arxiv.org/abs/2402.05708

References iv

Vansteelandt, S. & Dukes, O. (2022). Assumption-lean inference for generalized linear models (with discussion). *J. R. Statist. Soc. B* **84**, 657–685. focus on estimand

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.