UCL statistics seminar

Induced replication and the assessment of models

Heather Battey
Department of Mathematics, Imperial College London

June 4, 2026

Includes work from *arXiv:2507.10373* with Daniel Garcia Rasines and Yanbo Tang, and *arXiv:2603.27718* with Nancy Reid.

Statistical models

Textbook definition:

> Parametrised family of probability distributions $\{P_\theta : \theta \in \Theta\}$ to which the "true" distribution P_{θ^*} belongs. Data are realisations of random variables from P_{θ^*}.

Independence or dependence between random variables is part of the model.

Parametric vs. nonpara vs. semipara reflected in Θ.

What is a (sensible) statistical model?

Subjective claim: sensible statistical models are idealisations, stressing commonalities between individuals, usually attributing unexplained differences to randomness.

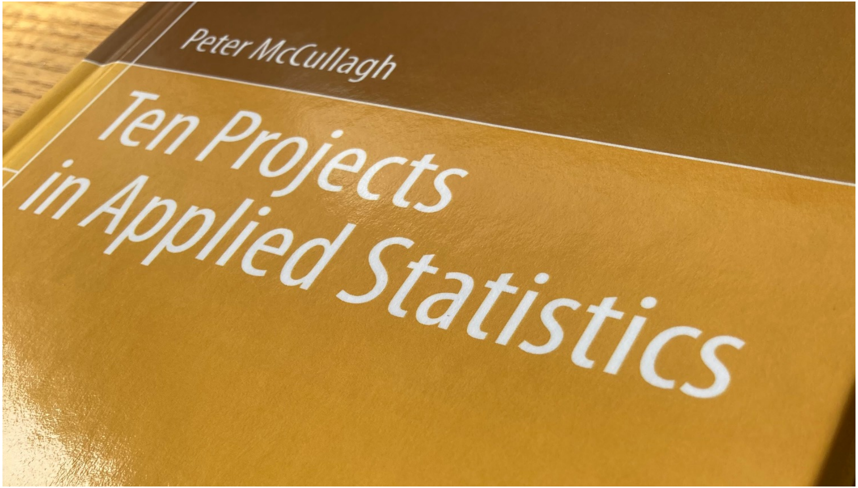
Purpose: insight/understanding; extrapolation of conclusions to new individuals.

Commonalities are encapsulated in stable parameters, e.g. treatment effects.

"individuals": more precisely, observational or experimental units.

What is a (sensible) statistical model?

A careful formalisation of the foundations of modelling can be found in chapters 11 and 14 of . . . .


Peter McCullagh
Ten Projects in Applied Statistics

Separation of information in parametric models

Information in the sample separates into:

- That relevant for inference on parameters of a given model, provisionally assumed true.
  Minimal sufficient statistic . . . or a notional part of it, a "maximal co-ancillary statistic".
- That relevant for assessment of the model.
  Everything else. How should this information be extracted?

Model assessment

Foundational “Fisherian” ideas due to Fisher (1922, 1934, etc.) and Bartlett (1937) apply in idealised parametric models.

No similar foundations for semiparametric and high-d models.

Model assessment: this talk

Part I: review of parametric foundations from a geometric perspective.

Part II: attempt to formulate foundations in the vein of Fisher/Bartlett for semiparametric and high-d models.

Inferential separations: parametric abstraction

Notional idealised separations

Sufficiency/co-sufficiency separation: $Y \cong (S, Q(S))$.

Ancillary/co-ancillary separation: $S \cong (A, C(A))$.

$Q(S)$ and A carry information for model assessment.

$C(A)$ relevant for inference on θ. Needs a separate talk[†].

[†] If interested, see: Battey (2024). Maximal co-ancillarity and maximal co-sufficiency. *Inf. Geom*, 7, 355–369, or slides on website.

Sufficiency/co-sufficiency separation

Let $d < n$ be the dimension of the minimal sufficient statistic.

Notional idealised separation: $Y \cong (S, Q(S))$.

The observed value $s^o = s(y^o)$ leaves $n - d$ degrees of freedom for variation of y consistent with the constraint $s(y) = s^o$.

Think of $Q(s^o)$ as having a distribution on the co-sufficient manifold

$$\mathcal{Q}(s^o) = \{y \in \mathbb{R}^n : s(y) = s^o\} \subset \mathbb{R}^n.$$

Calibration

$Q(s^o)$ has conditional distribution of Y given $S = s^o$.

Lives on co-sufficient manifold

$$\mathcal{Q}(s^o) = \{y \in \mathbb{R}^n : s(y) = s^o\} \subset \mathbb{R}^n.$$

By construction, $y^o \in \mathcal{Q}(s^o)$. If y^o extreme when calibrated against distribution of $Q(s^o)$, this casts doubt on model.

Inferential separation: parametric example

Example: normal linear regression

- Postulated model: $Y \sim N(X\beta, \sigma^2 I_n)$, $\beta \in \mathbb{R}^{d_\beta}$.
- Observed outcome: $y^o \in \mathbb{R}^n$.
- $X\beta \in \mathcal{X} \subset \mathbb{R}^n$; $\dim(\mathcal{X}) = d_\beta$.
- Residual vector $\hat{\varepsilon} \in \mathcal{X}^\perp \subset \mathbb{R}^n$; $\dim(\mathcal{X}^\perp) = n - d_\beta$.
- The n degrees of freedom for variation of y^o separate.

Example: normal linear regression

- Postulated model: $Y \sim N(X\beta, \sigma^2 I_n)$, $\beta \in \mathbb{R}^{d_\beta}$.
- Observed outcome: $y^o \in \mathbb{R}^n$.
- $X\beta \in \mathcal{X} \subset \mathbb{R}^n$; $\dim(\mathcal{X}) = d_\beta$.
- Residual vector $\hat{\varepsilon} \in \mathcal{X}^\perp \subset \mathbb{R}^n$; $\dim(\mathcal{X}^\perp) = n - d_\beta$.
- The n degrees of freedom for variation of y^o separate.

  First separation: $\mathcal{X} \oplus \mathcal{X}^\perp = \mathbb{R}^n$.

  But 1 deg. freedom from $\mathcal{X}^\perp$ needed for estimation of σ^2.

Example: normal linear regression

- Postulated model: $Y \sim N(X\beta, \sigma^2 I_n)$, $\beta \in \mathbb{R}^{d_\beta}$.
- Observed outcome: $y^o \in \mathbb{R}^n$.
- $X\beta \in \mathcal{X} \subset \mathbb{R}^n$; $\dim(\mathcal{X}) = d_\beta$.
- Residual vector $\hat{\varepsilon} \in \mathcal{X}^\perp \subset \mathbb{R}^n$; $\dim(\mathcal{X}^\perp) = n - d_\beta$.
- The n degrees of freedom for variation of y^o separate.

  First separation: $\mathcal{X} \oplus \mathcal{X}^\perp = \mathbb{R}^n$.

  But 1 deg. freedom from $\mathcal{X}^\perp$ needed for estimation of σ^2.

Example: normal linear regression

- Postulated model: $Y \sim N(X\beta, \sigma^2 I_n)$, $\beta \in \mathbb{R}^{d_\beta}$.
- Observed outcome: $y^o \in \mathbb{R}^n$.
- $X\beta \in \mathcal{X} \subset \mathbb{R}^n$; $\dim(\mathcal{X}) = d_\beta$.
- Residual vector $\hat{\varepsilon} \in \mathcal{X}^\perp \subset \mathbb{R}^n$; $\dim(\mathcal{X}^\perp) = n - d_\beta$.
- The n degrees of freedom for variation of y^o separate.

  Sufficient statistic $s^o = (X^{\mathrm{T}} y^o, \hat{\varepsilon}^{\mathrm{T}} \hat{\varepsilon})$: uses $d_\beta + 1$ d.f.

  Leaves $n - (d_\beta + 1)$ d.f. for model assessment.

  How can these be isolated without loss or redundancy?

Linear regression information partition

Sufficient statistic $s^o = (X^{\mathrm{T}} y^o, \hat{\varepsilon}^{\mathrm{T}} \hat{\varepsilon})$: uses $d_\beta + 1$ deg. freedom.

Isolate $n - (d_\beta + 1)$ deg. freedom for model assessment...

Basis for $\mathcal{X}^\perp$, e.g. the $(n - d_\beta)$ orthonormal eigenvectors v_j of $I - X(X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}}$ with non-zero eigenvals: $V = (v_1, \ldots, v_{n-d_\beta})$.

$V V^{\mathrm{T}} y^o \in \mathcal{X}^\perp \subset \mathbb{R}^n$, $V^{\mathrm{T}} y^o \in \mathbb{R}^{n-d_\beta}$.

$$q^o := \frac{V^{\mathrm{T}} y^o}{\| V^{\mathrm{T}} y^o \|} \in \mathcal{S}^{n-d_\beta}.$$

Co-sufficient information

Sufficient statistic $S = (X^{\mathrm{T}}Y, \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon})$. It can be shown that

$$Q := \frac{V^{\mathrm{T}}Y}{\|V^{\mathrm{T}}Y\|} \in \mathcal{S}^{n-d_\beta}$$

is independent of $\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}$ under postulated model; Q is co-sufficient.

In general, it is the *information* that is sufficient. In present context, the co-sufficient information corresponds to a statistic.

Information partition and degrees of freedom

$VV^{\mathrm{T}}y^o \in \mathcal{X}^\perp \subset \mathbb{R}^n$, $V^{\mathrm{T}}y^o \in \mathbb{R}^{n-d_\beta}$; $\|V^{\mathrm{T}}y^o\|^2 = \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}$.

$$q^o := \frac{V^{\mathrm{T}}y^o}{\|V^{\mathrm{T}}y^o\|} \in \mathcal{S}^{n-d_\beta}.$$

There are $n - d_\beta - 1$ independent angles $a_1^o, \ldots, a_{n-d_\beta-1}^o$ specifying position of q^o on unit sphere. Information partition:

$$\begin{array}{ccccccc} & & (X^{\mathrm{T}}y^o, & & \hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}, & & a_1^o, \ldots, a_{n-d_\beta-1}^o) \\ n & = & d_\beta & + & 1 & + & n-(d_\beta+1) \qquad \checkmark \end{array}$$

Assessment of the mean model in Gaussian linear regression

(a) If σ^2 is known, $\|V^{\mathrm{T}} y^o\|^2 = \hat{\varepsilon}^{\mathrm{T}} \hat{\varepsilon}$ is available for model assessment.

(b) Otherwise, only the information in q^o, or more compactly in $a_1^o, \ldots, a_{n-(d_\beta+1)}^o$, is available for model assessment.

Logic and calibration

Suppose for a potential "contradiction at level α" that the postulated model is true, then

- $\hat{\varepsilon}^{\mathrm{T}}\hat{\varepsilon}$ is a realisation of $R^2 \sim \sigma^2\chi^2_{n-d_\beta}$.
- q^o is a realisation of $Q \sim \mathrm{Unif}(\mathcal{S}^{n-d_\beta})$; Q independent of R^2.

If q^o is extreme when calibrated against α-level "tail" of distribution of Q, this casts doubt on adequacy of model.

There are some complications. More later...

Analogous partitions for semiparametric models?

Example 1: semipara multiplicative exponential

Example 1: postulated model

$(Y_{j1}, Y_{j0})_{j=1}^{m}$: outcomes on treated and untreated units in pair $j = 1, \ldots, m$.

Consider time-to-event outcomes. Heterogeneity across pairs: e.g.

$$Y_{j1} \sim \exp(\gamma_j \psi), \quad Y_{j0} \sim \exp(\gamma_j).$$

Parameter space grows with sample: not a parametric model in classical sense.

More general than typical semiparametric formulation. Why?

With covariates, could have taken $\gamma_j = h(x_j)$ for unknown function $h \in \mathcal{H}$.

Example 1 continued: co-sufficient information

Battey and Cox (2020, §5): suppose model true with $\psi = \psi_0$.

Then $S_j(\psi_0) = Y_{j0} + \psi_0 Y_{j1} \sim \text{Gamma}(2, \gamma_j)$ is sufficient for γ_j.

Co-sufficient information: variation in Y_{j1} remaining after conditioning on $S_j(\psi_0) = s_j^o(\psi_0)$. Calculation shows that

$$f_{Y_{j1}|S_j(\psi_0)}(y_{j1}|s_j^o(\psi_0)) = \psi_0/s_j^o(\psi_0), \text{ i.e.}$$

Y_{j1} is conditionally uniformly distributed on $[0, s_j^o(\psi_0)/\psi_0]$. Equivalently

$$U_j(\psi_0) := Y_{j1}\psi_0/s_j^o(\psi_0) \sim \text{Unif}[0,1], \quad j = 1, \dots, m.$$

Example 1 continued: model assessment

Y_{j1} is conditionally uniformly distributed on $[0, s_j^o(\psi_0)/\psi_0]$. Equivalently

$$U_j(\psi_0) := Y_{j1}\psi_0/s_j^o(\psi_0) \sim \text{Unif}[0,1], \quad j = 1, \ldots, m. \tag{1}$$

Holds if and only if model true with $\psi = \psi_0$.

If $U_1(\psi_0), \ldots, U_m(\psi_0)$ incompatible with (1): refutes model or ψ_0.

Example 1 continued: calibration

Combine evidence using, e.g. Fisher's (1925) method.

Two-sided p-value: $2\min\{p(\psi_0), 1 - p(\psi_0)\}$

$$p(\psi_0) := G_{2m}(-2\sum_{j=1}^{m} \log U_j(\psi_0))$$

where $G_{2m}(\cdot)$ is distribution function of χ^2_{2m}. Confidence set:

$$\mathcal{C}(\alpha) := \left\{\psi_0 \in \Psi : 2\min\{p(\psi_0), 1 - p(\psi_0)\} > \alpha\right\}$$

If $\mathcal{C}(\alpha)$ empty: casts doubt on model.

Example 1 continued: more power to reject model

Under model assumption, there exists a $\psi \in \Psi$ such that true distribution of $Z_j = Y_{j1}/Y_{j0}$ has density function

$$f_Z(z;\psi) = \frac{\psi}{(1+\psi z)^2}, \quad j = 1, \ldots, m.$$

No dependence on γ_j: MLE $\hat{\psi} \to_p \psi^*$ (true value) under model assumption $\Rightarrow U_j(\hat{\psi}) \rightsquigarrow U_j(\psi^*)$, $m \to \infty$.

Example 2: semipara time-dependent Poisson process

Example 2: postulated model

$(Y_{i1}, \ldots, Y_{iM_i})$ event times for individual i in fixed interval $(0, t_0)$.

Time-dependent Poisson process with intensity function

$$\lambda_i(t) = e^{\gamma_i + \beta t}, \qquad \text{(Cox, 1955)}.$$

Heterogeneity via γ_i: more general than typical semipara formulation.

Example 2 continued: co-sufficient information

Jointly sufficient statistics for (γ_i, β) are $(M_i, \sum_{j=1}^{M_i} Y_{ij})$.

Conditional density function of $Y_{i1}, \ldots, Y_{iM_i}$ given $M_i = m_i$ is

$$f(y_{i1}, \ldots, y_{im_i} \mid m_i) = \frac{m_i! \beta^{m_i} \exp(\beta \sum_{j=1}^{m_i} y_{ij})}{(e^{\beta t_0} - 1)^{m_i}}$$

No dependence on γ_i. If β were known, co-sufficient information would be contained in the variation permissible in $Y_{i1}, \ldots, Y_{im_i}$ when $M_i = m_i$.

Example 2 continued: model assessment

$f(y_{i1}, \dots, y_{im_i} \mid m_i)$ depends on $y_{i1}, \dots, y_{im_i}$ through $s_i := \sum_{j=1}^{m_i} y_{ij}$.

Calculate $F_{S_i \mid M_i}$. Then

$$U_i(\beta_0) := F_{S_i \mid M_i}(S_i \mid m_i; \beta_0) \sim \mathrm{Unif}[0, 1], \quad i = 1, \dots, n. \tag{2}$$

Holds if and only if model true with $\beta = \beta_0$.

If $U_1(\beta_0), \dots, U_n(\beta_0)$ incompatible with (2): refutes model or β_0.

Example 2 continued: calibration

Combine evidence using, e.g. Fisher's (1925) method.

Two-sided p-value: $2\min\{p(\beta_0), 1 - p(\beta_0)\}$

$$p(\beta_0) := G_{2n}(-2\sum_{i=1}^{n} \log U_i(\beta_0))$$

where $G_{2n}(\cdot)$ is distribution function of χ^2_{2n}. Confidence set:

$$\mathcal{C}(\alpha) := \left\{\beta_0 \in \mathcal{B} : 2\min\{p(\beta_0), 1 - p(\beta_0)\} > \alpha\right\}$$

If $\mathcal{C}(\alpha)$ empty: casts doubt on model.

Example 2 continued: more power to reject model

Under model assumption, can estimate true β^* via conditional lik.

$$L(\beta) := \prod_{i=1}^{n} f(y_{i1}, \ldots, y_{im_i} \mid m_i) = \prod_{i=1}^{n} \frac{m_i! \beta^{m_i} \exp(\beta \sum_{j=1}^{m_i} y_{ij})}{(e^{\beta t_0} - 1)^{m_i}}.$$

No dependence on γ_i: MLE $\hat{\beta} \to_p \beta^*$ under model assumption $\Rightarrow$ $U_i(\hat{\beta}) \rightsquigarrow U_i(\beta^*)$, $n \to \infty$.

Interest-parameter dependent co-sufficiency

- Fix interest parameter.
- Identify sufficient statistic for nuisance component.
- Condition to isolate co-sufficient information.
- Same info to estimate interest param, but MLE collapses this to dimension of parameter $\implies$ info remaining.

Generalisation: conditional co-sufficiency

Example 3: semipara proportional hazards model (in brief)

Proportional hazards model. Hazard function:

$$h(y; x) = h_0(y) \exp(x_i^{\mathrm{T}} \beta).$$

Baseline hazard function $h_0(y)$ completely unspecified.

Assess PH model using a notion of β-dependent conditional co-sufficiency.

Can convert to uniform RVs as in other examples (details omitted).

Another structure: interest-parameter-dependent ancillarity

Example 1 again

Example 1 from a different perspective: $Z_j = Y_{j1}/Y_{j0}$ have distn. function

$$F_Z(z;\psi^*) = \frac{\psi^* z}{1+\psi^* z}.$$

under correct model formulation. At a postulated ψ_0, this function at Z_j is

$$U_j(\psi_0) = \frac{\psi_0 Z_j}{1+\psi_0 Z_j} = \frac{Y_{j1}\psi_0}{S_j(\psi_0)}, \quad S_j(\psi_0) = Y_{j0} + Y_{j1}\psi_0.$$

Recovers the previous analysis based on ψ-dependent co-sufficiency.

Highlights a different structure: Z_j is ancillary for γ_j.

Extracting general principles

In the original formulations, there was no replication, e.g.

- Matched pair: between-pair heterogeneity reflected in γ_j; within-pair heterogeneity due to treatment effect.
- Inhomogeneous Poisson process: between-individual heterogeneity reflected in γ_i; within-individual heterogeneity due to time-dependence.

But there was an **inducable replication** under the postulated model.

Violation of model $\Longrightarrow$ no replication induced: data refutes false model.

Usual vs. illustrations above

Usual semiparametric:

- Fit model including e.g. $\hat{\gamma}_i = \hat{h}(x_i)$, where h belongs to a function class. Typically requires cross-validation for tuning $\hat{h}$.
- Assess prediction performance on a separate sample.

Parametric co-sufficient: a kind of prediction performance within sample.

Illustrations above: attempt to generalise to semiparametric models by inducing replication. Several context-specific routes.

Example 4: post-reduction inference.

Inducement of replication through carefully-injected noise

Example 4: low-dimensional background

$$\begin{aligned} Y = X\beta + \varepsilon, \quad \varepsilon &= (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}}, \quad \varepsilon_i \sim N(0,1), \quad && n = 100, \\ \beta &= (1, 1, 0, \ldots, 0)^{\mathrm{T}} \in \mathbb{R}^p, && p = 25. \end{aligned}$$

Large correlation b/w signal and noise variables $\Longrightarrow$ difficult to detect true model.

Comprehensive model: $[p] := \{1, \ldots, p\}$; true model: $\mathcal{S} = \{1, 2\}$.

In simulation: lasso (with cross-validated tuning) selects a single model: $\{2\}$.

A likelihood-ratio test of each low-dimensional submodel $\mathcal{S}_m \subset \{1, \ldots, p\}$ against $[p]$ declared $\{2\}, \{1, 2\}, \{2, 3\}$ and $\{1, 2, 3\}$ as statistically indistinguishable from $[p]$.

Confidence set of models: $\mathcal{M} = \{\{2\}, \{1, 2\}, \{2, 3\}, \{1, 2, 3\}\}$

Example 4: low-dimensional formulation

- Full set of variables: $[p] = \{1, \ldots, p\}$. True model: $\mathcal{E}^* \subset [p]$. $\#$ of possible models $= 2^p$.
- Sparsity $\implies$ consider models with fewer variables, i.e. models of size s or less, of which there are $\bar{m} = \sum_{v=1}^{s} \binom{p}{v}$.
- For every possible model $\mathcal{E}_m$, $m = 1, \ldots, \bar{m}$, test $\mathcal{E}_m \subset [p]$ against $[p]$ by a likelihood ratio test. Confidence set of models:

$$\mathcal{M}_\alpha = \{\text{all models not rejected by a LR test of size } \alpha\}$$

- By standard arguments $\mathrm{pr}(\mathcal{E}^* \in \mathcal{M}_\alpha) \to 1 - \alpha$.

Example 4: high-dimensional complications

- Likelihood ratio test is infeasible when $p > n$.

  Natural resolution: eliminate variables that appear to have no effect. Conservative reduction procedures come with theoretical guarantees.

- But: the reduced set $\hat{\mathcal{E}}$ has been selected in the light of the data. Fits the data better than an arbitrary model of the same size.

  Implication: a likelihood ratio test of any subset of variables $\mathcal{A} \subset \hat{\mathcal{E}}$ rejects $\mathcal{A}$ too often in hypothetical repeated use. Thus

$$\lim_{n\to\infty} \mathrm{pr}(\mathcal{E}^* \in \mathcal{M}_\alpha) \ll 1 - \alpha$$

Example 4: minimal sufficient statistic

$$\text{True model}: Y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 I_n).$$

Recall from beginning of talk: minimal sufficient statistic:

$$S = (X^{\mathrm{T}} Y, \hat{\varepsilon}^{\mathrm{T}} \hat{\varepsilon}) \cong (\hat{\beta}, \hat{\varepsilon}^{\mathrm{T}} \hat{\varepsilon}).$$

Use co-sufficient information Q to construct a confidence set of models.

Example 4: implications

- $Q \sim \mathrm{Unif}(\mathcal{S}^{n-d_\beta})$ holds iff we project onto $\mathcal{X}^\perp$.
  If postulated model is wrong, we project on the wrong space.
- Force a contradiction: for every subset of variables $\mathcal{E}_m \subset \hat{\mathcal{E}}$, project on the corresponding $\mathcal{X}_m^\perp$, and calculate the corresponding $Q^{(m)}$. If uniformity is violated, that is evidence against the model.
- This avoids any comparison of fit relative to the larger model $\hat{\mathcal{E}}$.
- Problem: from one observation of $Q^{(m)}$ there is virtually no power to reject a false model $\mathcal{E}_m$.

Example 4: synthetic replication

- In the vein of Rasines and Young (2023), generate pseudo-replicates of Y: Let L be an $n \times (k-1)$ matrix of independent $N(0,1)$ entries. There exists a $k \times k$ deterministic matrix Γ such that

$$[\tilde{Y}^{(1)} \cdots \tilde{Y}^{(k)}] = [Y \; L]\Gamma \overset{\text{indep}}{\sim} N(\mu, k\sigma^2 I_n).$$

- These yield independent pseudo-replicates

$$\tilde{Q}^{(1)}, \dots, \tilde{Q}^{(k)} \sim \text{Unif}(\mathcal{S}^{n-d_\beta})$$

if and only if the postulated model is true.

Example 4: calibration

- Inner products between m disjoint pairs from $\tilde{Q}^{(1)}, \ldots, \tilde{Q}^{(k)}$ are independent cosine angles, $Z_1, \ldots, Z_m$, say.
- For a true postulated model, these have density function (Fisher, 1915)

$$f_Z(z) = \frac{\Gamma((n - d_\beta)/2)}{\sqrt{\pi}\Gamma\{(n - d_\beta - 1)/2\}}(1 - z^2)^{(n-d_\beta-3)/2}, \quad -1 < z < 1.$$

- Construct $U_j = F(Z_j)$ and combine in the usual way.

Example 4: theoretical guarantees

- An α-level confidence set of models $\mathcal{M}$ is all low-dimensional subsets of variables in $\hat{\mathcal{E}}$ for which a test of uniformity of $\tilde{Q}^{(1)}, \ldots, \tilde{Q}^{(k)}$ does not reject at level α.
- Provided that $\mathcal{E}^* \subset \hat{\mathcal{E}}$, i.e. all variables in the true model survive reduction, $\mathrm{pr}(\mathcal{E}^* \in \mathcal{M}) = 1 - \alpha$.

Summary

- Cross-validation in standard semiparametric practice blurs boundary between estimation and model assessment.
- Examples show that in some contexts, a semiparametric generalisation of the Fisherian suff/co-suff. and ancil/co-ancil. separations are feasible and fruitful.
- Replaces out-of-sample prediction performance by a within-sample analogue.

Some references

- Background/parametric abstraction from:
  - Bartlett (1937). Properties of sufficiency and statistical tests. *Proceedings A*, 160, 268–282.
  - Battey (2024). Maximal co-ancillarity and maximal co-sufficiency. *Information Geometry*, 7, 355–369.
- Primary content based on:
  - Battey and Reid (2026). Induced replication and the assessment of models. *arXiv:2603.27718*.
  - Battey, Rasines and Tang (2026). Post-reduction inference for confidence sets of models. *Biometrika*, to appear.
- Early inspiration from:
  - Battey and Cox (2018). Large numbers of explanatory variables: a probabilistic assessment. *Proceedings A*, 474, 20170631.
  - Battey and Cox (2020). High-dimensional nuisance parameters: an example from parametric survival analysis. *Information Geometry*, 3, 119–148.
- Also cited:
  - Fisher (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika*, 10, 507–521.
  - Cox (1955). Some statistical methods connected with series of events (with discussion). *J. R. Statist. Soc. B*, 17, 129–164.
  - Rasines and Young (2023). Splitting strategies for post-selection inference. *Biometrika*, 110, 597–614.

The end