Causal Inference Isn't What You Think It Is

Philip Dawid

University of Cambridge

25 June 2020

Overview

You may think that statistical causal inference is about inferring causation. You may think that it can not be tackled with standard statistical tools, but requires additional structure, such as counterfactual reasoning, potential responses or graphical representations. I shall try to disabuse you of such woolly misconceptions by locating statistical causality firmly within the scope of traditional statistical decision theory. From this viewpoint, the enterprise of "statistical causality" could fruitfully be rebranded as "assisted decision making".

Statistics 101

Simple 2-sample experiment

- $N = n_0 + n_1$ individuals randomly sampled from relevant population [headache sufferers]
- n_0 randomly assigned to inactive control $(c, 0)$ [chalk]
  - responses $Y_{0j} : j = 1, \ldots, n_0$ [log-duration of headache]
- n_1 randomly assigned to active treatment $(t, 1)$ [aspirin]
  - responses $Y_{1j} : j = 1, \ldots, n_1$
- Model: $Y_{ij} \sim P_i = \mathcal{N}(\mu_i, \sigma^2)$, all independently

Effect of treatment? Compare P_1 and P_0 – *e.g.*, $\mu_1 - \mu_0$

- Purely distributional comparison
- Inference by *e.g.* 2-sample t-test.

Question: What about "potential responses?"

Potential Response Approach

Conceive of each unit u having two pre-existing potential responses:

- Y_{u0} if given chalk
- Y_{u1} if given aspirin
- pairs (Y_{u0}, Y_{u1}) from common bivariate distribution
- "individual causal effect": $\mathrm{ICE}_u := Y_{u1} - Y_{u0}$
- We can observe at most one of Y_{u1}, Y_{u0} — for treatment actually taken — the other then being "missing data"
  - so ICE_u is never observable
  - cannot estimate dependence/correlation ρ between Y_0 and Y_1
- Fundamental Problem of Causal Inference ?
  - or just of this approach to causal inference ??

Problems with ICE

We *can* estimate the *average causal effect*:

$$\begin{aligned} \mathrm{ACE} := \mathsf{E}(\mathrm{ICE}) &= \mathsf{E}(Y_1 - Y_0) \\ &= \mathsf{E}(Y_1) - \mathsf{E}(Y_0) \end{aligned}$$

- unaffected by unknowable dependence between Y_1 and Y_0
- could estimate by $\overline{Y}_1 - \overline{Y}_0$

However, consider the *individual ratio effect*, $\mathrm{IRE} := Y_1/Y_0$

- We can *not* estimate $\mathrm{ARE} := \mathsf{E}(\mathrm{IRE})$
- — since this involves the unknowable dependence between Y_1 and Y_0.

For the same reason we cannot estimate the "effect of treatment on the observed treated", $\mathsf{E}(\mathrm{ICE} \mid Y_1 = y)$.

Neyman's null hypothesis

Let $\widetilde{Y}_j$ $(j = 0, 1)$ be the average of Y_{uj} over all N units in the experiment

- unobservable

Neyman (1935) interpreted "No effect of treatment" as:

$$\widetilde{Y}_0 = \widetilde{Y}_1 \qquad (\widetilde{\mathrm{ICE}} = 0)$$

and showed that, for this null hypothesis, the usual (t, F) test

- is unbiased in a simple randomised experiment
- but not when the logic is extended to more complex designs, *e.g.* Latin Square
- unless we assume treatment-unit additivity:
  - $\mathrm{ICE}_u = Y_{u1} - Y_{u0}$ the same for all units
- requires $\rho = 1$ — an unknowable condition

Fisher's null hypothesis

???

- Neyman's null hypothesis depends crucially on which units are in the experiment
- So not sensible...
- Fisher: "the hypothesis to be tested was... that differences of treatment made no difference to the yields"
- $Y_{u0} = Y_{u1}$, all u $\quad (\mathrm{ICE}_u \equiv 0)$???

But any performable test can only target weaker properties, *e.g.*:

- all observed values are exchangeable (permutation test)
- all observed values are from a common distribution
  - *e.g.*, $\mathcal{N}(\mu, \sigma^2)$ $\quad (t$-test of $\mu_0 = \mu_1)$

These purely distributional hypotheses are perfectly good interpretations of "no treatment effect"

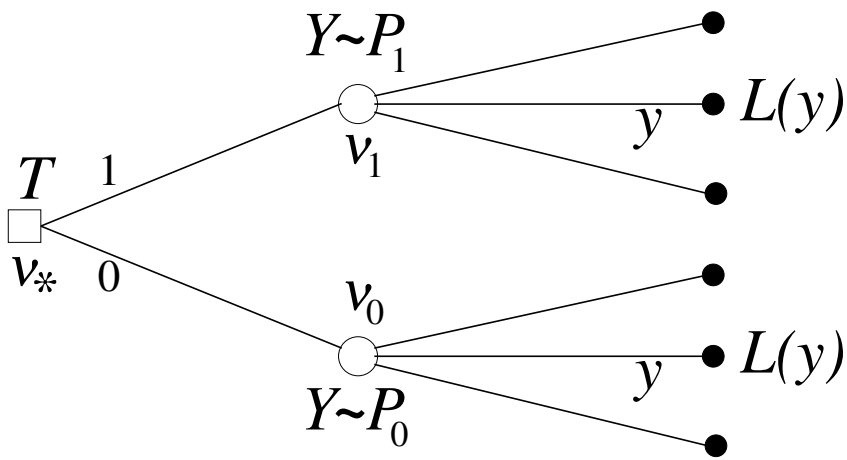
The potential outcome machinery has not added (indeed, has subtracted!) value

An Alternative Approach: Decision Theory

I have a headache. Should I take aspirin?

I have to compare the hypothetical future consequences of my two available decisions

I assess $Y \sim P_i$ if I were to take decision i.



- Take aspirin if $\mathrm{E}_{Y \sim P_1} L(Y) \leq \mathrm{E}_{Y \sim P_0} L(Y)$
- All that is need to solve any such decision problem is the pair of "hypothetical distributions", P_0 and P_1, for response Y

Causal Inference

- What then is "causal inference"?
  - Learning about required P_0, P_1 with assistance of available data (observational or – in this case – experimental)
- If I do take aspirin, I will become like (exchangeable with) those in the treated group.
  - then my response Y would be distributed as $Y \sim P_1 = \mathcal{N}(\mu_1, \sigma^2)$
- If I don't, I will become like (exchangeable with) those in the control group.
  - then my response Y would be distributed as $Y \sim P_0 = \mathcal{N}(\mu_0, \sigma^2)$

I can learn P_0, P_1 from the data. Then I have all I need to solve my decision problem:

- Take aspirin if $\mathrm{E}_{Y \sim P_1} L(Y) \leq \mathrm{E}_{Y \sim P_0} L(Y)$, where $L(y)$ is a suitable loss function
- If $L(y)$ is linear, take aspirin if the average causal effect $\mu_1 - \mu_0$ is negative

Some comments

- No use for potential responses. We have two hypothetical distributions for a single variable Y, not one joint distribution for a pair of potential variables $(Y_0.Y_1)$.
  - In particular, no need to consider the (unknowable) dependence between Y_0 and Y_1
- No need for counterfactual logic: "What would have happened to a treated study individual, if she had not been treated?"
- No determinism/predetermination: response Y can develop stochastically, even after application of treatment
- No missing data
  - so no "fundamental problem of causal inference"
- Average causal effect is a difference of expectations, $\mathrm{E}_{P_1}(Y) - \mathrm{E}_{P_0}(Y)$, not the expectation of a difference, $\mathrm{E}(Y_1 - Y_0)$
- No new principles/formulation/notation needed, only basic Fisher and standard decision theory. We can solve our problem without weighing it down with unnecessary baggage

Observational Study

Suppose now that assignment of treatments to study subjects was not done experimentally. When can we still use the data as before?

This would require post-treatment exchangeability:

- If I were to decide on treatment t, I would become exchangeable with those in the treatment group
- if I were to decide on treatment c, I would become exchangeable with those in the control group

NB: These can only both hold when (prior to treatment application), the treatment and control groups are exchangeable with each other

- so we would have to be "comparing like with like"
- ignorable treatment assignment

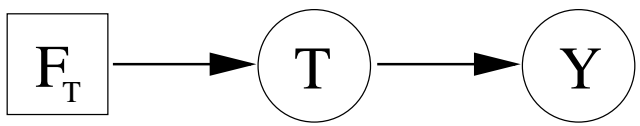
Variables and Regimes

- Binary treatment variable T
- Response variable Y
- Non-stochastic regime indicator F_T:
  - $F_T = 0$: (hypothetically) take treatment 0 ($\Rightarrow T = 0$)
  - $F_T = 1$: (hypothetically) take treatment 1 ($\Rightarrow T = 1$)
  - $F_T = \emptyset$: "Nature" chooses T (random)
- I am interested in comparing regimes $F_T = 1$ and $F_T = 0$
- I have data from $F_T = \emptyset$
- I should like to use the data to assist me with my decision problem
- So I will need to make (and justify) some connexions between the different regimes

Ignorability

The simplest case is when I can assume ignorability ("like an experiment"):

- The distribution of Y is the same in regime $F_T = 1$ [resp., 0] as in regime $F_T = \emptyset$, conditional on $T = 1$ [resp., 0].
- Y is independent of F_T, given $T = 1$ [resp., 0]
- $Y \perp\!\!\!\perp F_T \mid T$



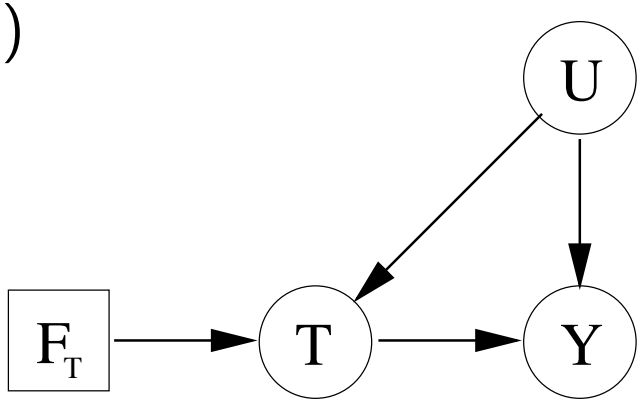
(If not, we have *confounding*)

Conditional ignorability ("no unobserved confounding")

When ignorability fails, it might be restored if we condition on additional information.

Call an observed variable U a *sufficient covariate* if:

- $U \perp\!\!\!\perp F_T$
- $Y \perp\!\!\!\perp F_T \mid (T, U)$



Can then estimate ACE by ("back door formula"):

$$\mathrm{ACE} = \mathrm{E}(\mathrm{SCE}_U \mid F_T = \emptyset)$$

where we define the specific causal effect (relative to U) as

$$\mathrm{SCE}_U := \mathrm{E}(Y \mid U, T = 1, F_T = \emptyset) - \mathrm{E}(Y \mid U, T = 0, F_T = \emptyset).$$

Not possible if U unobserved—confounding

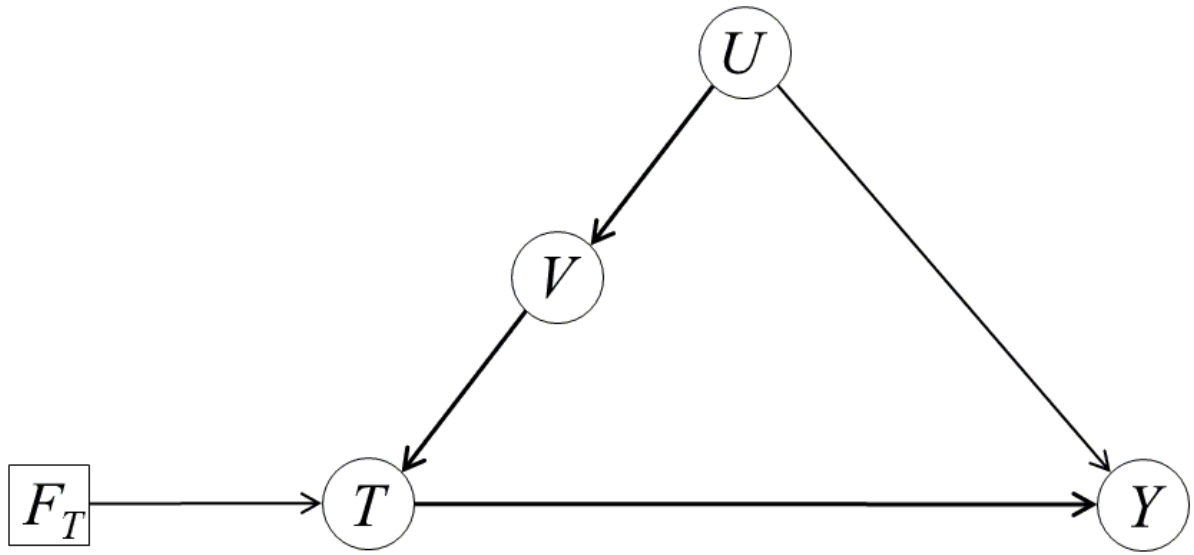
So such a variable U is an *unconfounder*.

Propensity score

U a sufficient covariate, V a function of U

V will be a sufficient covariate if

$$T \perp\!\!\!\perp U \mid (V, F_T = \emptyset)$$



- In the observational regime, choice of treatment depends on U only through V
- Does not involve response variable Y

Propensity score

$$T \perp\!\!\!\perp U \mid (V, F_T = \emptyset)$$

Equivalently,

$$U \perp\!\!\!\perp T \mid (V, F_T = \emptyset)$$

- Observational distribution of U, given V the same for both treatments
- V is a *balancing score* for U
- V is a *sufficient statistic* for comparing hypotheses $T = 1$, $T = 0$ for data U
- Minimal such V is likelihood ratio:

$$p(U \mid T = 1)/p(U \mid T = 0)$$

- Equivalently, posterior probability $p(T = 1 \mid U)$
  - single variable $\in [0, 1]$
  - *propensity score*

Effect of treatment on the treatable

Suppose I consider that, if I had been in the study, I would have been deemed "suitable for treatment" (*i.e.*, exchangeable with those for whom $T = 1$). Should I take the treatment?

Let $S = 1$ [0] denote "suitable [unsuitable] for treatment". In the study, $T = S$. I have $S = 1$ but have choice over T.

I should thus consider

$$\mathrm{ETT} := \mathsf{E}(Y \mid S = 1, T = 1) - \mathsf{E}(Y \mid S = 1, T = 0)$$

I can estimate first term from treated group in the study — but have no data directly relevant to second term.

Effect of treatment on the treatable

We can show:

$$\mathrm{ETT} = \frac{\mathsf{E}(Y \mid F_T = \emptyset) - \mathsf{E}(Y \mid F_T = 0)}{\Pr(T = 1 \mid F_T = \emptyset)}.$$

— estimable if I have also observed Y in an experimental control group.

Alternatively, if in the data we can observe a sufficient covariate U, conditional on which both $T = 0$ and $T = 1$ are possible, we can compute

$$\mathrm{ETT} = \mathsf{E}(\mathrm{SCE}_U \mid T = 1, F_T = \emptyset).$$

Compare the PR definition:

$$\begin{aligned} \mathrm{ETT} &= \mathsf{E}(\mathrm{ICE} \mid T = 1) \\ &= \mathsf{E}(Y_1 - Y_0 \mid T = 1) \end{aligned}$$

All above variants yield same value.

NB: Does not use information on outcome when treated

Sequential decisions

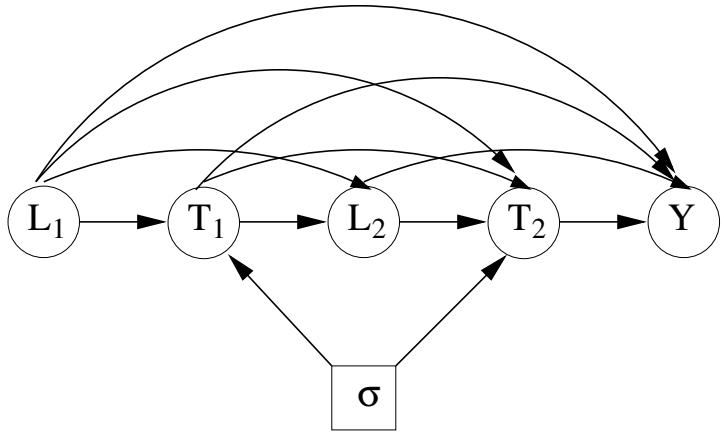
Variables arising in sequence: L_1—T_1—L_2—T_2—Y

- L_1 pre-treatment variable
- T_1 first treatment — in light of L_1
- L_2 response to T_1
- T_2 second treatment — in light of L_1, T_1, L_2
- Y response

Let s be a contemplated (possiby randomised) *strategy* for choosing T_1 and T_2, each in the light of previous values. We want to evaluate s, using observational data.

- Introduce non-stochastic regime indicator σ, values s and o (observation).
- Can we evaluate $\mathsf{E}(Y \mid \sigma = s)$ from properties of the observational regime $\sigma = o$?

Sequential ignorability



Assume

$$
\begin{aligned}
L_1 &\perp\!\!\!\perp \sigma \\
L_2 &\perp\!\!\!\perp \sigma \,|\, (L_1, T_1) \\
Y &\perp\!\!\!\perp \sigma \,|\, (L_1, T_1, L_2, T_2).
\end{aligned}
$$

Robins's G-formula for estimating $\mathrm{E}(Y \,|\, \sigma = s)$ from the observational data follows directly.

G-formula

$$
\begin{aligned}
p(l_1, t_1, l_2, t_2, y \,|\, \sigma = s) \;=\; & p(l_1 \,|\, \sigma = s) \\
& \times p(t_1 \,|\, l_1, \sigma = s) \\
& \times p(l_2 \,|\, l_1, t_1, \sigma = s) \\
& \times p(t_2 \,|\, l_1, t_1, l_2, \sigma = s) \\
& \times p(y \,|\, l_1, t_1, l_2, t_2, \sigma = s) \\
\\
= \; & p(l_1 \,|\, \sigma = o) \\
& \times p(t_1 \,|\, l_1, \sigma = s) \\
& \times p(l_2 \,|\, l_1, t_1, \sigma = o) \\
& \times p(t_2 \,|\, l_1, t_1, l_2, \sigma = s) \\
& \times p(y \,|\, l_1, t_1, l_2, t_2, \sigma = o).
\end{aligned}
$$

Now marginalise to find $p(y \,|\, \sigma = s)$.

Compare potential response approach

Conceive of potential versions of all variables under either regime:

$$\Pi^s := (L_1^s, T_1^s, L_2^s, T_2^s, Y^s)$$

$$\Pi^o := (L_1^o, T_1^o, L_2^o, T_2^o, Y^o)$$

—all with a joint distribution.

Principal assumptions/constraints:

Consistency

- $L_1^o = L_1^s$
- $T_1^o = s(L_1^o) \Rightarrow L_2^o = L_2^s$.
- $T_2^o = s(L_1^o, T_1^o, L_2^o) \Rightarrow Y^o = Y^s$.

Sequential ignorability

- $T_1^o \perp\!\!\!\perp (L_2^s, Y^s) \mid L_1^o$
- $T_2^o \perp\!\!\!\perp Y^s \mid L_1^o = l_1, T_1^o = s(l_1), L_2^o = l_2$

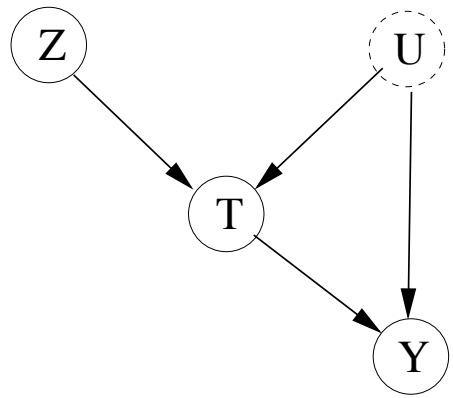
Instrumental variable

Suppose I can not observe U, but there is a variable Z such that:

1. Z affects T
2. Z affects the outcome Y only through T
3. Z does not share common causes with the outcome Y ("no confounding of the effect of Z on Y").

Observational conditional independence properties:

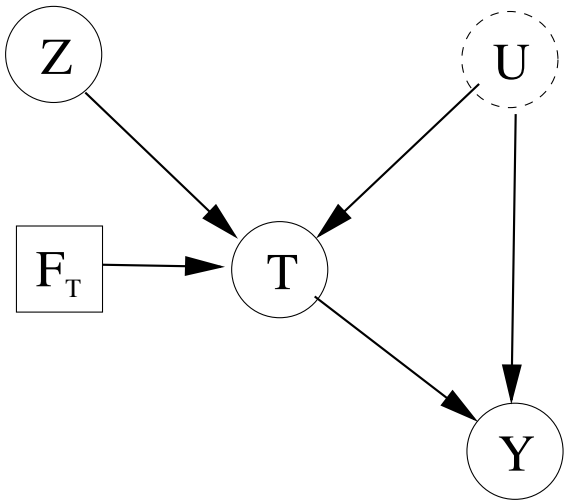
$$T \not\perp\!\!\!\perp Z$$
$$U \perp\!\!\!\perp Z$$
$$Y \perp\!\!\!\perp Z \mid (T, U)$$



—no causal content!

Instrumental variable

Add regime indicator F_T for causal content:



$$U \perp\!\!\!\perp (Z, F_T)$$
$$Y \perp\!\!\!\perp (Z, F_T) \mid (T, U).$$

Linear model

Assume

$$\mathrm{E}(Y \mid T, U, [Z, F_T]) = W + \beta\, T$$

for some function W of U.

Let $w_0 = \mathrm{E}(W)$, $= \mathrm{E}(W \mid Z, F_T)$ since $U \perp\!\!\!\perp (Z, F_T)$

Then

$$\mathrm{E}(Y \mid F_T = t) = w_0 + \beta\, t,$$

- β can be interpreted causally.
- Want to identify β.

Assumptions imply

$$\mathrm{E}(Y \mid Z, F_T = \emptyset) = w_0 + \beta\, \mathrm{E}(T \mid Z, F_T = \emptyset).$$

Then $\beta = \mathrm{Cov}_\emptyset(Y, Z)/\mathrm{Cov}_\emptyset(T, Z)$ can be estimated by the ratio of the coefficients of Z in the sample linear regressions of Y on Z and of T on Z.

Wrap-up

- (EoC) Causal Inference is really about assisted decision making
- Traditional statistical tools are adequate for this
- Arguments using potential responses are:
  - unnecessary
  - obscure
  - complex
  - potentially misleading
- Next time you are tempted to do such an analysis, think about doing so more directly

THANK YOU!

References

Constantinou, P. and Dawid, A. P. (2017). Extended conditional independence and applications in causal inference.
Annals of Statistics, **45**, 2618–53

Dawid, A. P. (2000). Causal inference without counterfactuals (with Discussion).
Journal of the American Statistical Association, **95**, 407–48

Dawid, A. P. (2002). Influence diagrams for causal modelling and inference.
International Statistical Review, **70**, 161–89.
Corrigenda, *ibid*., 437

Dawid, A. P. (2007a). Counterfactuals, hypotheticals and potential responses: A philosophical examination of statistical causality.
In *Causality and Probability in the Sciences*, Texts in Philosophy, Vol. 5, (ed. F. Russo and J. Williamson), pp. 503–32. College Publications, London

Dawid, A. P. (2007b). Fundamentals of statistical causality.
Research Report 279, Department of Statistical Science, University College London.
94 pp.
https://www.ucl.ac.uk/drupal/site_statistics/sites/statistics/files/migrated-files/rr279.pdf

Dawid, A. P. (2012). The decision-theoretic approach to causal inference.
In *Causality: Statistical Perspectives and Applications*, (ed. C. Berzuini, A. P. Dawid, and L. Bernardinelli), chapter 4, pp. 25–42. J. Wiley & Sons, Chichester, UK

Dawid, A. P. (2015). Statistical causality from a decision-theoretic perspective.
Annual Review of Statistics and its Application, **2**, 273–303.
DOI:10.1146/annurev-statistics-010814-020105

Dawid, A. P. (2020). Decision-theoretic foundations for statistical causality.
`arXiv:2004.12493`

Dawid, A. P. and Constantinou, P. (2014). A formal treatment of sequential ignorability.
Statistics in Biosciences, **6**, 166–88

Dawid, A. P. and Didelez, V. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview.
Statistical Surveys, **4**, 184–231

Geneletti, S. and Dawid, A. P. (2011). Defining and identifying the effect of treatment on the treated.
In *Causality in the Sciences*, (ed. P. M. Illari, F. Russo, and J. Williamson), pp. 728–49. Oxford University Press

Guo, H. and Dawid, A. P. (2010). Sufficient covariates and linear propensity analysis.
Journal of Machine Learning Research Workshop and Conference Proceedings, **9**, 281–8.
Proceedings of the Thirteenth International Workshop on Artificial Intelligence and Statistics (AISTATS) 2010, Chia Laguna, Sardinia, Italy, May 13–15, 2010, edited by Y. W. Teh and D. M. Titterington.
http://jmlr.csail.mit.edu/proceedings/papers/v9/guo10a/guo10a.pdf