Objective Bayesianism from a philosophical perspective

Jon Williamson

The Statistics Wars and their Casualties, 20th May 2021

1 Standard Bayesianism

1. Bayesian probability is epistemic.

   The rational degrees of belief of an agent.

   $B_E(A) \stackrel{\text{df}}{=}$ rational degree of belief in A, supposing only E.

2. Conditional beliefs are conditional probabilities.

   CBCP. There is a prior probability function $P_\varnothing$ such that $B_E(A) = P_\varnothing(A|E)$ for all A and E.

3. Current degrees of belief are obtained by conditionalising on current evidence.

   Conditionalisation. On evidence E, believe A to degree $B_E(A) = P_\varnothing(A|E)$.

   All degrees of belief are captured by a single prior probability function $P_\varnothing$.
   Evidence must be expressible in the agent's language (the domain of $P_\varnothing$).
   New evidence cannot shift degree of belief away from 0 or 1.

Subjectivism. Any prior is rationally permissible.

Objectivism. Only certain priors are appropriate.

EG The uniform prior.

The philosophical perspective

A divergence has taken place:

Bayesian statistics. Bayesianism for predictive inference, model selection etc.

> Common in statistics and many of the sciences.
>
> The domain is often continuous.
>
> Both subjectivism and objectivism have a strong presence.

Bayesian philosophy. Bayesianism as a theory of rational strength of belief.

> Often applied to generate an account of confirmation.
>
> Common in formal epistemology and philosophy of science.
>
> The domain is often finite, or captured by sentences of a logical language.
>
> Almost all philosophers are subjectivists.
>
> Almost all philosophers advocate **direct inference**.

2 Direct inference

A direct inference principle says that one should directly calibrate one's credences to non-epistemic probabilities (generic frequencies or single-case chances), insofar as one has evidence of them.

Bayesian statisticians tend not to appeal to direct inference principles because they:

- Often doubt the existence of non-epistemic probabilities (following Bruno de Finetti).
- See Bayesianism as a rival to frequentist statistical inference—'statistics wars'.
- Note that certain priors lead to convergence to frequencies, as data comes in.

In contrast, many philosophers:

- Accept non-epistemic probabilities.
- See Bayesianism as being concerned with belief, and so not a rival.
- See asymptotic convergence as merely contingent and long-run.

Calibration to non-epistemic probabilities is rationally required in the short run:

- If you're reliably informed that there is a 17% chance that Cheesewright gets a cough in the next year, then you ought to believe now that he gets a cough to degree 0.17.

Philosophers are also concerned with induction.

Two problems of induction:

Inductive Logic. Is there a viable logic of inductive inference?

- IE What inductive inferences ought we make?

Inductive Justification. Is there a viable justification of inductive inference?

- IE Can we convince detractors that we ought to make these inductive inferences?

Direct inference might help with these questions.

Example: 17 of a random sample of a hundred 21-year-olds develop a cough (E).

How confident should one be that Cheesewright, who is 21, gets a cough (A)?

That the sample frequency is 0.17 is evidence that the chance/frequency is ≈ 0.17.

Use this to constrain $P_{\varnothing}$: $P_{\varnothing}(A|E) \approx 0.17$.

- ✓ This provides a kind of logic of induction.
- ✓ If a detractor can be convinced of the merits of direct inference, then it might provide a kind of inductive justification.

We normally perform direct inference without thinking about it.

Even kea can perform a kind of direct inference (Bastos and Taylor, 2020a,b).

But we need a principled approach to deal with complex situations.

Two kinds of direct inference principle:

Reference Class Principle. Set degrees of belief to frequencies in reference classes—§3.

Principal Principle. Set degrees of belief to single-case chances—§4.

Unfortunately there are problems for each in the standard Bayesian framework . . .

3 The Principle of the Narrowest Reference Class

Hans Reichenbach:

> We then proceed by considering the narrowest class for which reliable statistics can be compiled. If we are confronted by two overlapping classes, we shall choose their common class. Thus, if a man is 21 years old and has tuberculosis, we shall regard the class of persons of 21 who have tuberculosis. (Reichenbach, 1935, p. 374.)

PNRC. $B_E(\alpha(c)) = x$ if E determines that the frequency $P^*_\rho(\alpha) = x$ and determines that ρ is the unique narrowest reference class containing c for which $P^*_\rho(\alpha)$ is available, and contains no more pertinent information.

PNRC is intuitive and widely endorsed.

For example,

A: Cheesewright gets a cough.

R: Cheesewright is 21.

S: Cheesewright has tuberculosis.

X: the frequency of 21-year-olds getting a cough is 0.17.

Y: the frequency of 21-year-olds with TB getting a cough is 0.97.

Then PNRC implies the following:

1. $P_{\varnothing}(A|XR) = 0.17$
2. $P_{\varnothing}(A|YRS) = 0.97$
3. $P_{\varnothing}(A|XYR) = 0.17$
4. $P_{\varnothing}(A|XYRS) = 0.97$
5. $P_{\varnothing}(A|XYR\bar{S}) = 0.17$.

The problem is that these assignments are inconsistent.

So the standard Bayesian framework struggles to accommodate PNRC.

4 The Principal Principle

Lewis (1980) uses chances rather than frequencies to constrain prior probabilities:

Principal Principle. $P_{\varnothing}(A|XE) = x$, where X says that the chance at time t of proposition A is x and E is any proposition that is compatible with X and admissible at time t.

✓ This is immune to the above problems with PNRC.

There are no reference classes and chances are unconditional.

✗ However, it does have its own share of problems . . .

A: it will rain tomorrow in Abergwyngregyn.

X: the present chance of A is 0.7.

PP implies:

6: $P_{\emptyset}(A|XE) = 0.7.$

F: Fred's fibrosarcoma will recur.

E provides evidence against F that is less compelling than the chance of F.

Then the following is rationally permissible:

7: $P_{\emptyset}(F|XE) = 0.3.$

Since evidence for F is less compelling than the chance of F, it should be permissible that:

8: $P_{\emptyset}(A|XE(A \longleftrightarrow F)) > 0.5.$

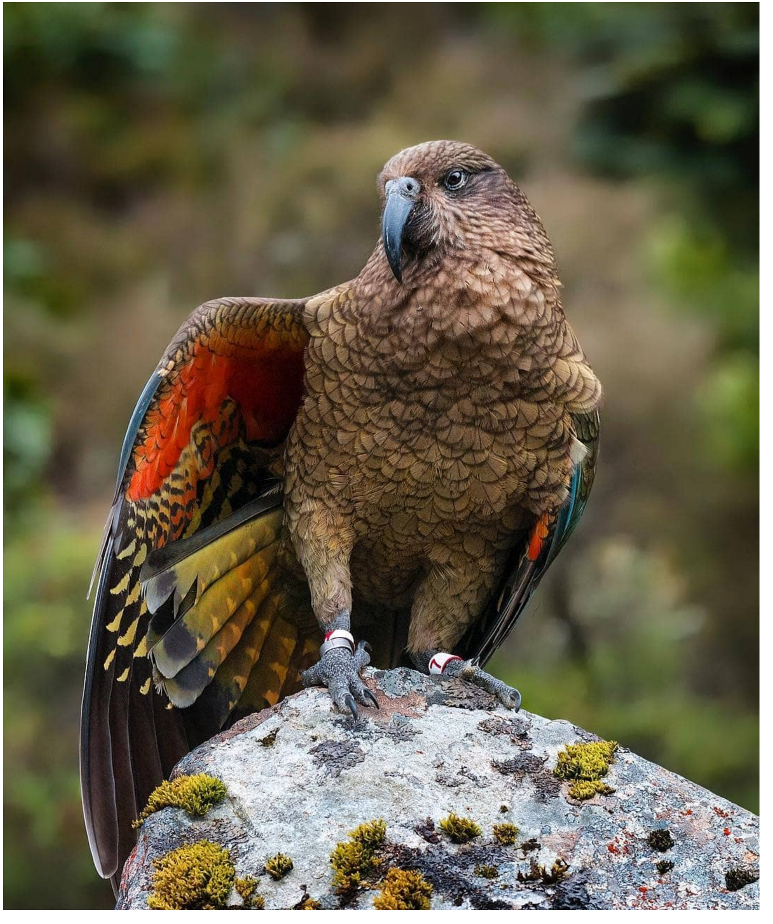
The problem is that 6–8 are inconsistent.

So the standard Bayesian framework struggles to accommodate the Principal Principle.

(Wallmann, C. and Williamson, J. (2020). The Principal Principle and subjective Bayesianism. *European Journal for the Philosophy of Science*, 10(1):3.)

Contents

5 Objective Bayesianism



There are two variants of objective Bayesianism

The usual variant sits within the standard Bayesian statistics framework.

It holds that only certain priors $P_\emptyset$ are rationally permissible.

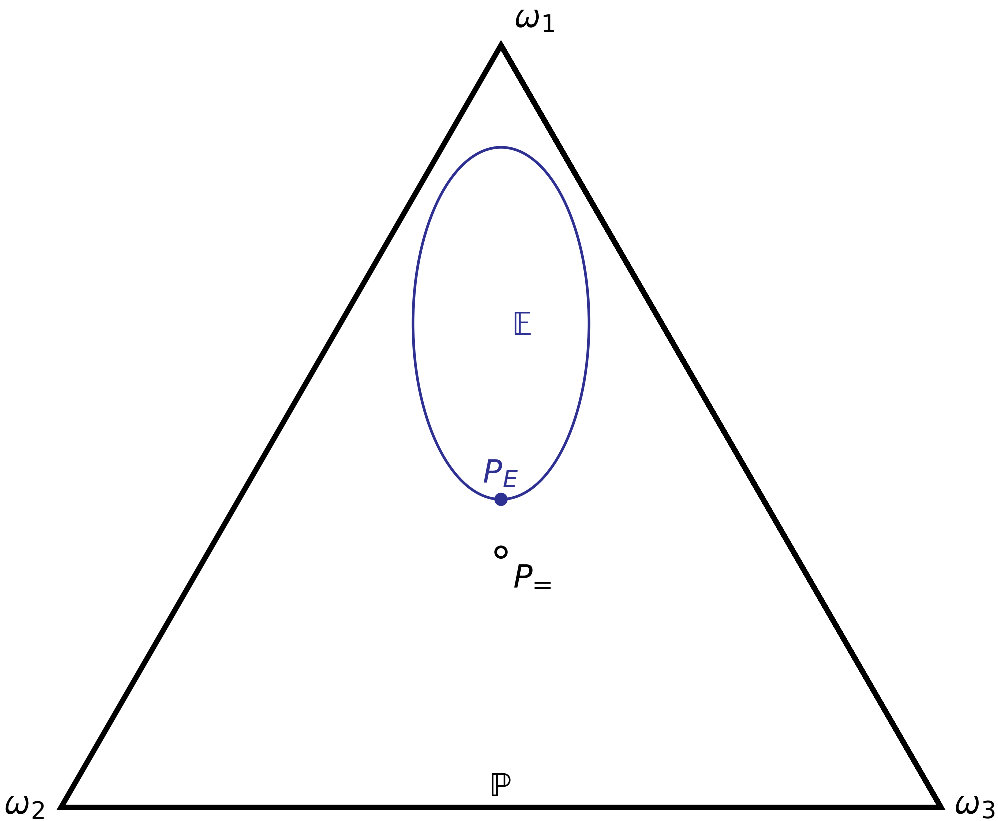
Jaynes (1957), physics and engineering: Maximum Entropy Principle.

Jeffreys (1939), statistics: other 'objective' or 'default' priors.

A key desideratum is that the prior should be uniquely determined by the problem formulation.

No use of direct inference.

Maximum Entropy Principle. P_E should satisfy constraints imposed by E and maximise entropy, $-\sum_{\omega\in\Omega} P(\omega)\log P(\omega)$



I argue for a non-standard version of Bayesianism.

It does **not** take conditional beliefs to be conditional probabilities.

✗ **CBCP.** There is a probability function $P_\varnothing$ such that $B_E(A) = P_\varnothing(A|E)$ for all A and E.

But it does take conditional beliefs to be probabilities.

CBP. For any E, there is a probability function P_E such that $B_E(A) = P_E(A)$ for all A.

∴ Conditionalisation is not a core principle.

It appeals to the Maximum Entropy Principle to say how beliefs should track evidence.

This usually agrees with Conditionalisation, but not always.

✓ One can revise beliefs away from 0 and 1, when faced with unexpected evidence.

✓ No need to assume evidence is expressible in the domain of the probability function.

This version does not require that there is a unique rationally permissible belief function.

Beliefs can depend on the agent's language, utilities and goals, in addition to her explicit evidence.

In addition there is sometimes room for subjectivity.

✓ Arguably it handles subjectivity better:

In the standard approach, if x and y are both rationally permissible for A, you need to choose one, x say. Learning something unrelated, you need to conditionalise and suddenly y is no longer permissible.

With the Maximum Entropy Principle, both x and y are permissible at both stages.

This version appeals to direct inference.

Frequency Calibration. If, according to E, the frequency $P^*_\rho(\alpha) \in X$, and ρ is the unique narrowest reference class containing c with respect to which E determines non-trivial bounds on the frequency of α, and E includes no more pertinent information, then $P_E(\alpha(c)) \in \langle X \rangle$, the convex hull of X.

Chance Calibration. If, according to current evidence E, the current chance function P^* lies in the set $\mathbb{P}^*$ of probability functions, then $P_E \in \langle \mathbb{P}^* \rangle$, the convex hull of $\mathbb{P}^*$.

Back to the problems with direct inference

By denying CBCP, we no longer over-constrain a single prior $P_{\varnothing}$.

The applications of PNRC are consistent:

1′. $P_{XR}(A) = 0.17$

2′. $P_{YRS}(A) = 0.97$

3′. $P_{XYR}(A) = 0.17$

4′. $P_{XYRS}(A) = 0.97$

5′. $P_{XYR\bar{S}}(A) = 0.17$.

Each of these consequences constrains a different probability function.

The assignments relating to the Principal Principle are also consistent:

6′. $P_{XE}(A) = 0.7$.

7′. $P_{XE}(F) = 0.3$.

8′. $P_{XE(A \leftrightarrow F)}(A) > 0.5$.

Direct inference and induction

Initially, $P_R(A) = 0.5$, say.

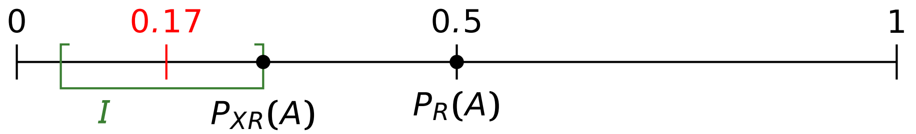
Then a sample of reference class ρ yields proportion 0.17 for α.

You're prepared to establish just that the frequency $P^*_\rho(\alpha)$ is in confidence interval I.

So your evidence changes from R to XR where X is $P^*_\rho(\alpha) \in I$.

Frequency Calibration requires that $P_{XR}(A) \in I$.

The Maximum Entropy Principle requires that $P_{XR}(A)$ is as equivocal as possible.



The confidence level can be determined from your utilities (Williamson, 2021).

This procedure integrates:

- Bayesian decision theory—finding the confidence level;
- Classical statistical inference—determining the confidence interval;
- Direct inference—using the interval to constrain your degree of belief;
- The Maximum Entropy Principle—selecting an appropriate value in the interval.

The story so far

We saw that there was a divergence in standard Bayesianism:

> Bayesian statistics: subjective or objective.
>
> Bayesian philosophy: mainly subjective; direct inference.

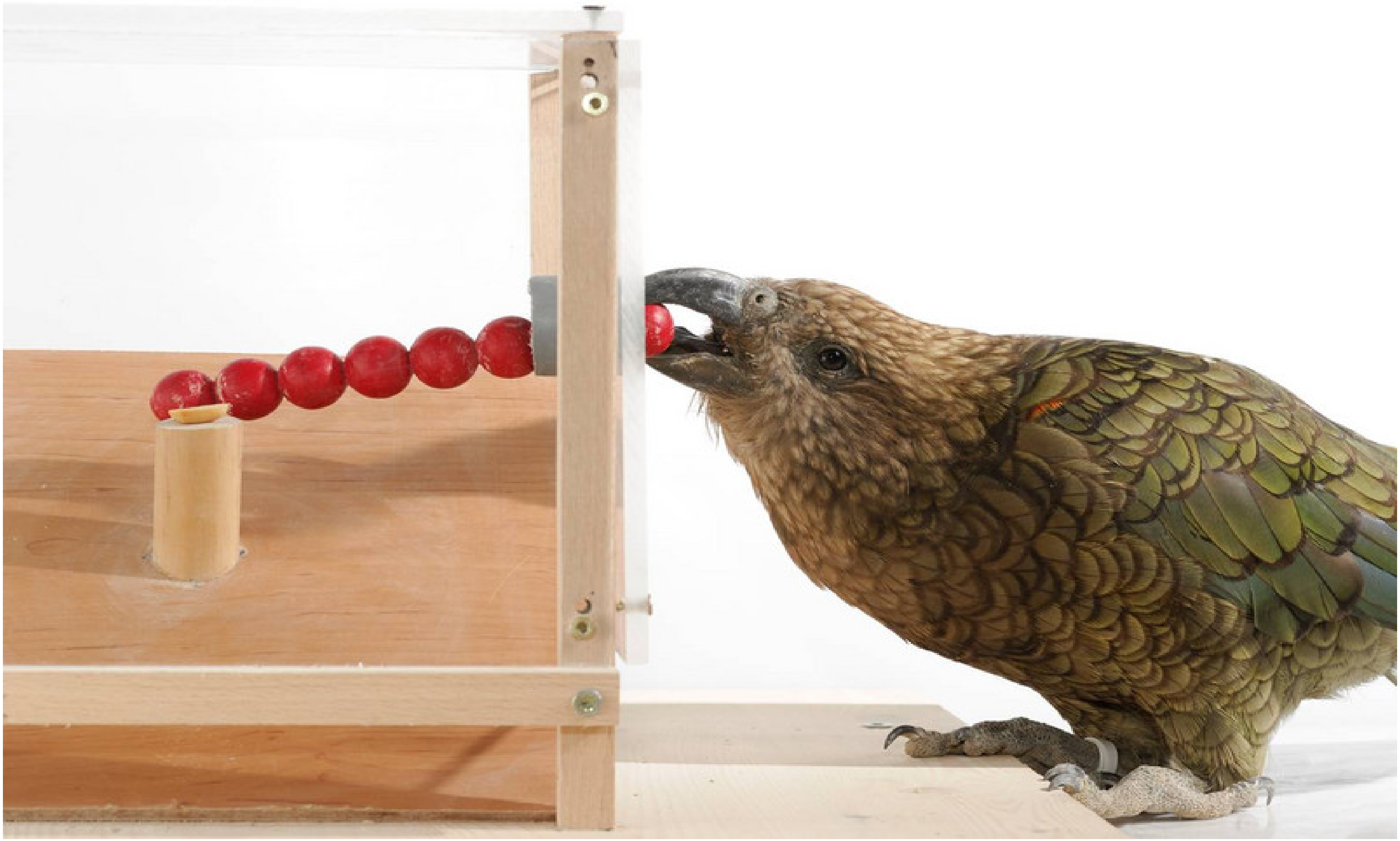
But there is a tension in the latter position:

> Standard Bayesianism struggles to accommodate direct inference.

So I've argued for a non-standard version of Bayesianism:

> No CBCP; no conditionalisation norm; objective Bayesianism; uniqueness not required.

6 Application to medicine



Although these developments are mainly philosophical, they can be applied.

Systems biology studies systems of molecules and their causal interactions within the cell.

Using data-intensive functional genomics techniques.

EG Transcriptomics, metabolomics, proteomics.

Systems medicine applies systems biology to medicine, with two goals:

Practical. For diagnosis, prognosis and treatment.

IE Causal discovery.

Theoretical. For discovering pathophysiological mechanisms.

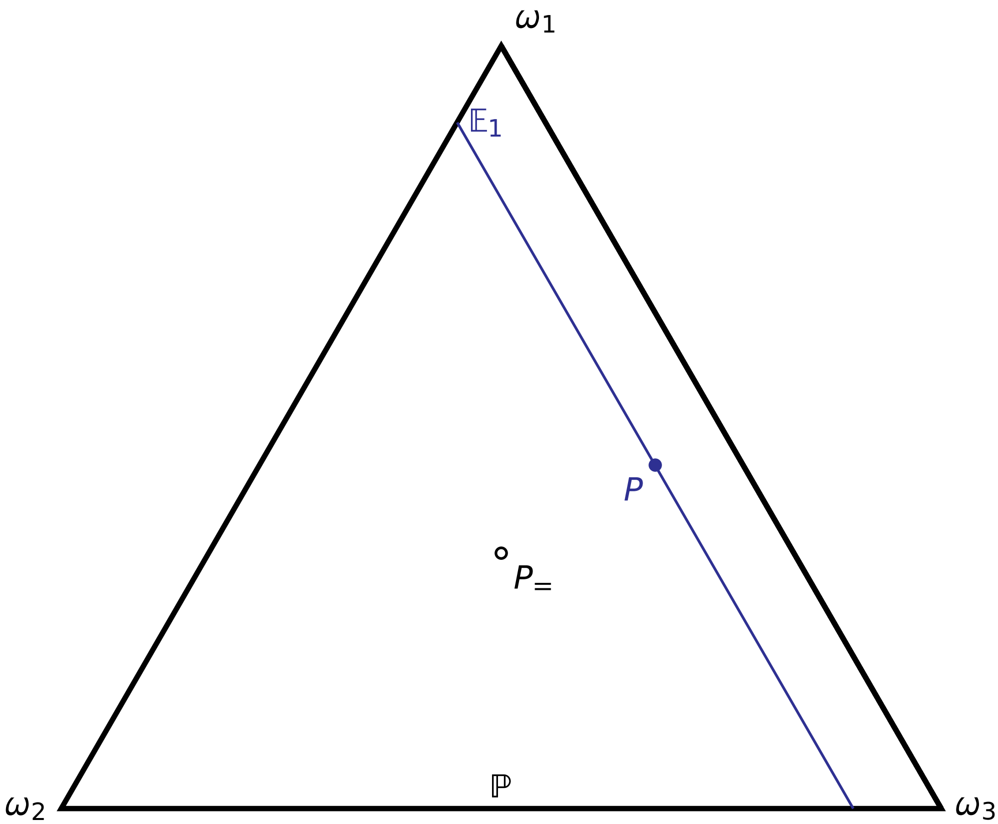
Systems medicine includes higher-level clinical and environmental data and mechanisms.

Key problem: Few variables are measured by more than one dataset.

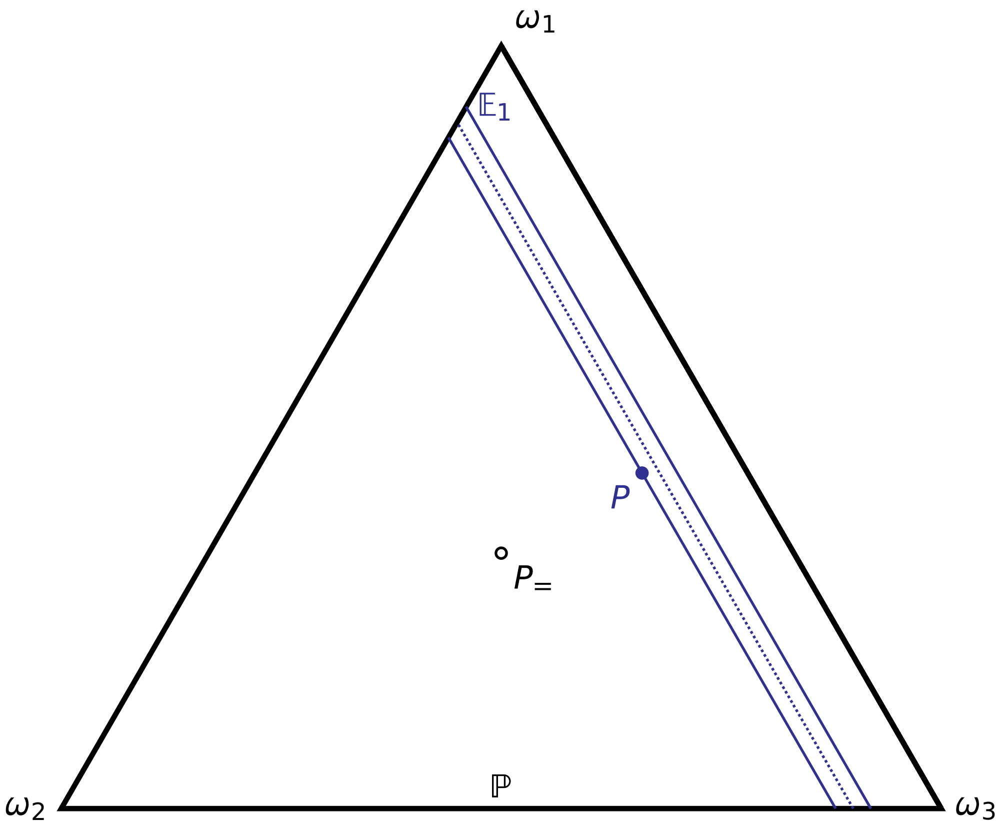
How can one integrate the datasets into a model that involves all variables of interest?

Suggestion: apply objective Bayesianism (non-standard version).

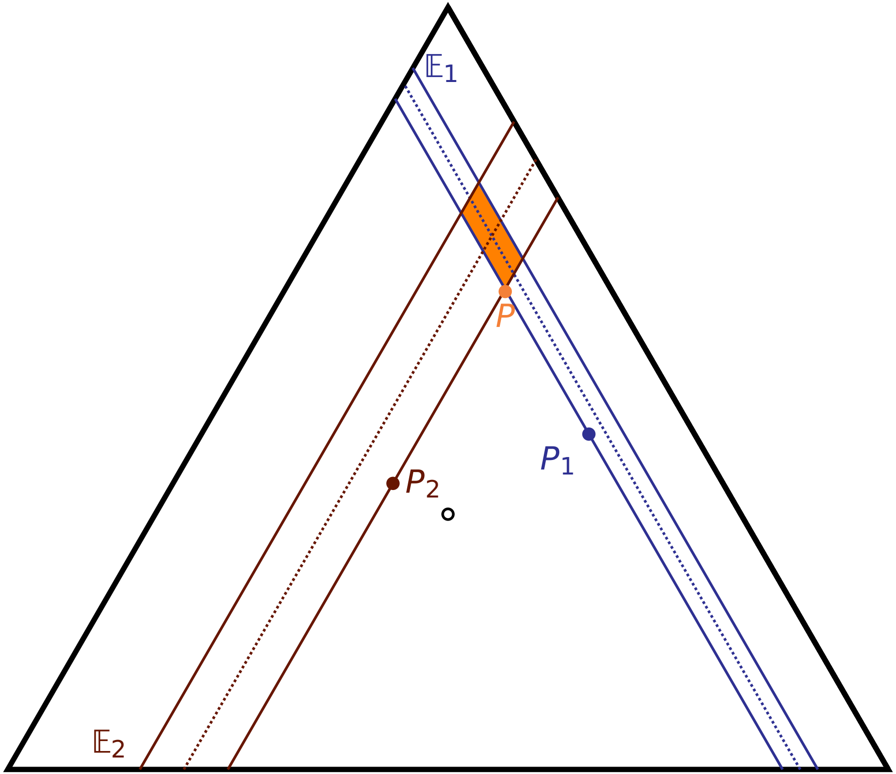
Single dataset, point estimate:



Single dataset, confidence region:



The objective Bayesian approach to data integration:



Objective Bayesian net

A Bayesian net that represents the objective Bayesian probability function.

- This is a model that can include all the measured variables.

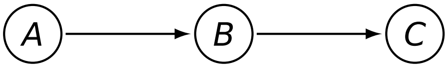
Dataset 1:

	A	B	C
Jill	✓	✗	✓
Keith	✓	✗	✗
Linda	✗	✗	✓
...	...	...	...

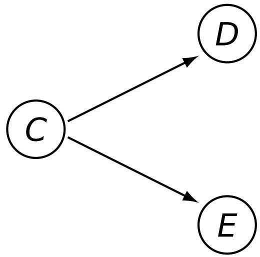
Dataset 2:

	C	D	E
Jim	✗	✗	✓
Kirsty	✓	✓	✓
Lionel	✗	✓	✗
...	...	...	...

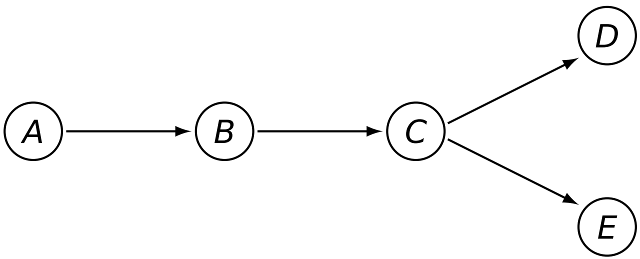
Bayesian net 1:



Bayesian net 2:



Objective Bayesian net:



Example: breast cancer prognosis.

The problem is to decide how to treat a breast cancer patient

Some treatments have very harsh side-effects

$\therefore$ the higher the probability of recurrence, the more aggressive the treatment.

(Nagl, S., Williams, M., and Williamson, J. (2008). Objective Bayesian nets for systems modelling and prognosis in breast cancer. In Holmes, D. and Jain, L., editors, *Innovations in Bayesian networks: theory and applications*, pages 131–167. Springer, Berlin.)

Evidence from:

- clinical datasets
- genomic datasets
- scientific papers (causal relations, mechanisms, and dependencies)
- experts (causal relationships and mechanisms)
- medical informatics systems (ontological relationships, logical relationships)

Clinical dataset, SEER study,

3 million patients in the US from 1975–2003;

of these 4731 were breast cancer patients.

Variables: Age, Tumour size (mm), Grade (1–3), HR Status (Oestrogen/Progesterone receptors), Lymph Node Tumours, Surgery, Radiotherapy, Survival (months), Status (alive/dead).

Age	T Size	Grade	HR	LN	Surgery	Radiotherapy	Survival	Status
70–74	22	2	1	1	1	1	37	1
45–49	8	1	1	0	2	1	41	1
...	...	...	...	...	...	...	...	...

Genomic data, Progenetix dataset

502 cases.

1p31	1p32	1p34	2q32	3q26	4q35	5q14	7p11	8q23	20p13	Xp11	Xq13
0	0	0	1	-1	0	0	1	0	0	0	-1
0	0	1	1	0	0	0	-1	-1	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...

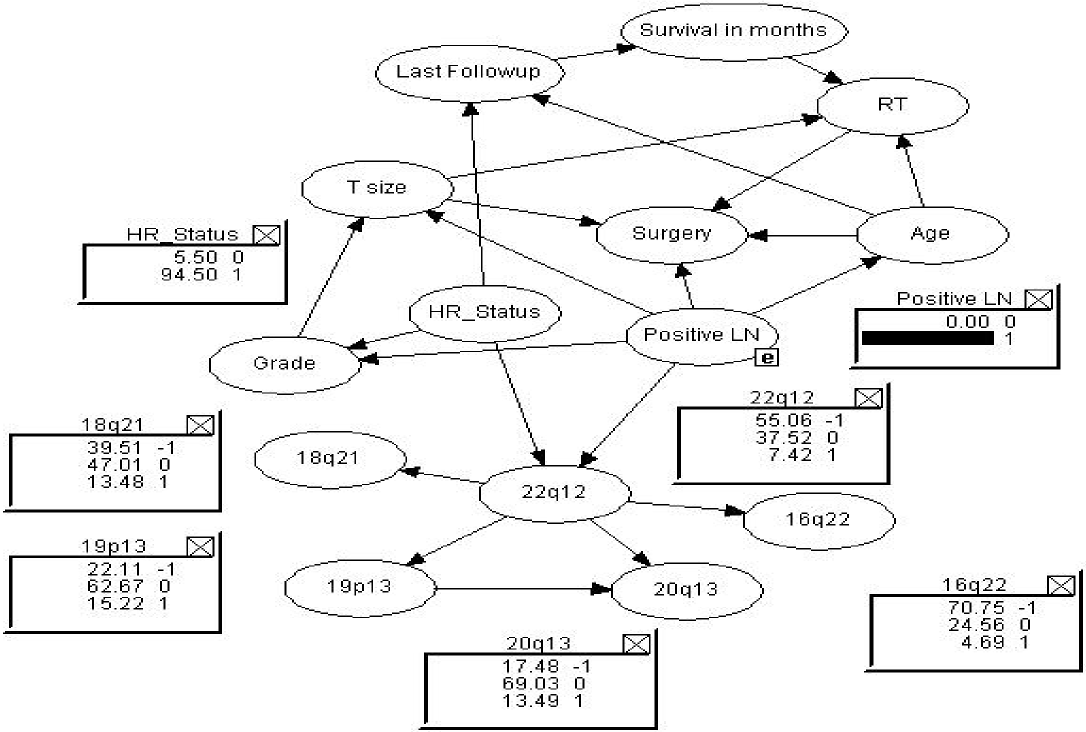
A further genomic dataset (119 cases with clinical annotation) from the Progenetix database:

Lymph Nodes	1q22	1q25	1q32	1q42	7q36	8p21	8p23	8q13	8q21	8q24
0	1	1	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...

A published study (Fridlyand et al., 2006),

causal and quantitative information involving HR_status and 22q12

The objective Bayesian net:



7 Conclusion

Bayesian philosophers are (rightly) keen to incorporate direct inference.

But standard Bayesianism struggles to accommodate both PNRC and the Principal Principle.

Non-standard objective Bayesianism does not face these problems.

It can appeal to classical estimation methods.

This yields an account of the logic of induction.

This approach can be applied to medicine.

EG To systems medicine.

NB Direct inference plays a crucial role.

The Maximum Entropy Principle enables the use of Bayesian net models.

- Objective Bayesian nets.

Arguably, any statistical matching application **needs** these foundations:

The model represents a rational basis for action, not an estimate of frequency/chance.

Matching the data distributions requires direct inference.

Acknowledgements: Christian Wallmann, the Leverhulme Trust, DFG.

Bibliography

Bastos, A. P. M. and Taylor, A. H. (2020a). Addendum: Kea show three signatures of domain-general statistical inference. *Nature Communications*, 11(2908):1–2.

Bastos, A. P. M. and Taylor, A. H. (2020b). Kea show three signatures of domain-general statistical inference. *Nature Communications*, 11(828):1–8.

Fridlyand, J., Snijders, A., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B., Jain, A., McLennan, J., Ziegler, J., Chin, K., Devries, S., Feiler, H., Gray, J., Waldman, F., Pinkel, D., and Albertson, D. (2006). Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer*, 6:96.

Jaynes, E. T. (1957). Information theory and statistical mechanics. *The Physical Review*, 106(4):620–630.

Jeffreys, H. (1939). *Theory of Probability*. Clarendon Press, Oxford, third (1961) edition.

Lewis, D. K. (1980). A subjectivist's guide to objective chance. In *Philosophical papers*, volume 2, pages 83–132. Oxford University Press (1986), Oxford. With postscripts.

Nagl, S., Williams, M., and Williamson, J. (2008). Objective Bayesian nets for systems modelling and prognosis in breast cancer. In Holmes, D. and Jain, L., editors, *Innovations in Bayesian networks: theory and applications*, pages 131–167. Springer, Berlin.

Reichenbach, H. (1935). *The theory of probability: an inquiry into the logical and mathematical foundations of the calculus of probability*. University of California Press, Berkeley and Los Angeles, second (1949) edition. Trans. Ernest H. Hutten and Maria Reichenbach.

Wallmann, C. and Williamson, J. (2020). The Principal Principle and subjective Bayesianism. *European Journal for the Philosophy of Science*, 10(1):3.

Williamson, J. (2021). A Bayesian account of establishing. *British Journal for the Philosophy of Science*, in press. Doi 10.1086/714798.