Comparing models

Topics in Cognitive Modelling

John Lee, Chris Lucas
School of Informatics
University of Edinburgh
{jlee,clucas2}@inf.ed.ac.uk

How do can we compare models?

What makes one model or theory better than another?

- Explanatory completeness
- Predictive accuracy
- Being understandable

How do can we compare models?

What makes one model or theory better than another?

- Explanatory completeness
- Predictive accuracy
- Being understandable

Explanatory completeness

Generality

A good model accurately explains many results

- Fits data from many experiments
- Captures qualitatively different phenomena

Precision

A good model is precise

- Specific predictions, less wiggle room

Generality

E.g., for physical forces and particles:

Electricity

↓

Classical electromagnetism (+ magnetism)

↓

Quantum electrodynamics (+ quantum phenomena)

↓

Standard model (+ nuclear forces)

↓

"Theory of everything"

(gravity, dark matter, dark energy …)

Precision

Beware vagueness!

- "*Stuff happens*" is a hypothesis, but vague one.
- Better: "*X is related to Y.*"
- Better: "*As X increases, Y will decrease.*"
- Better: "*As X increases, Y will decrease according to the following function …*"

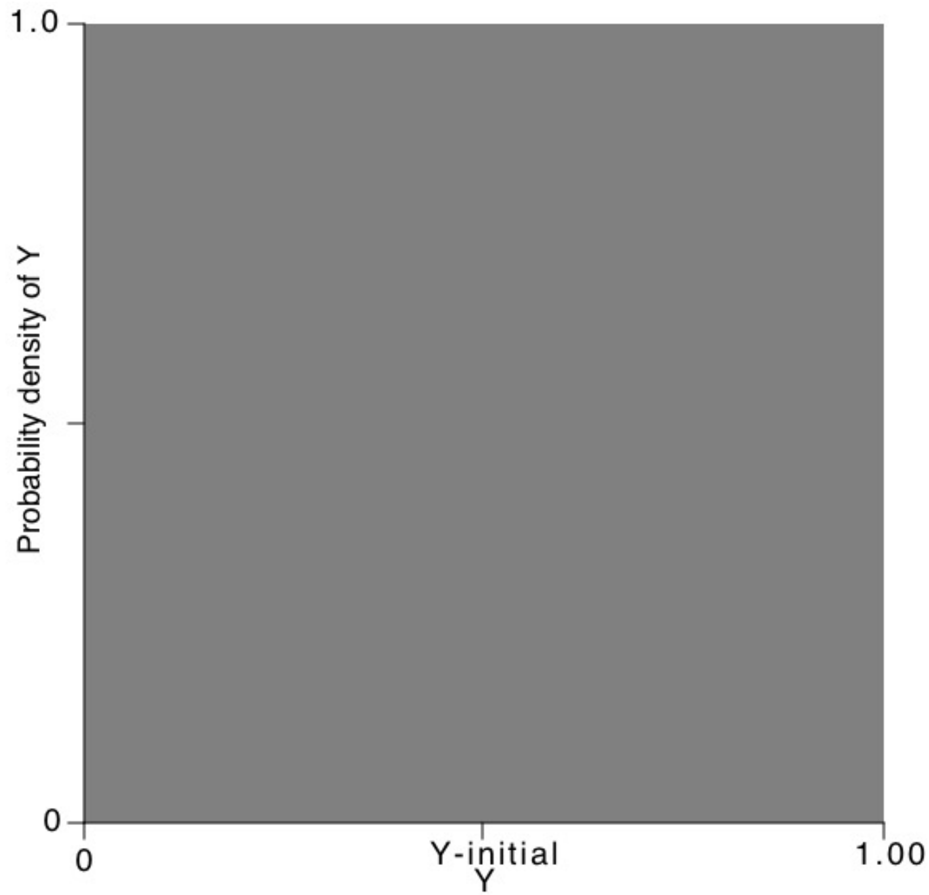
Probability theory lets us be precise about precision:

$$P(\mathbf{m}\text{odel}|\mathbf{d}\text{ata}) \propto P(m)P(d|m)$$

Precision

Suppose X increases. What do our different hypotheses say about Y?

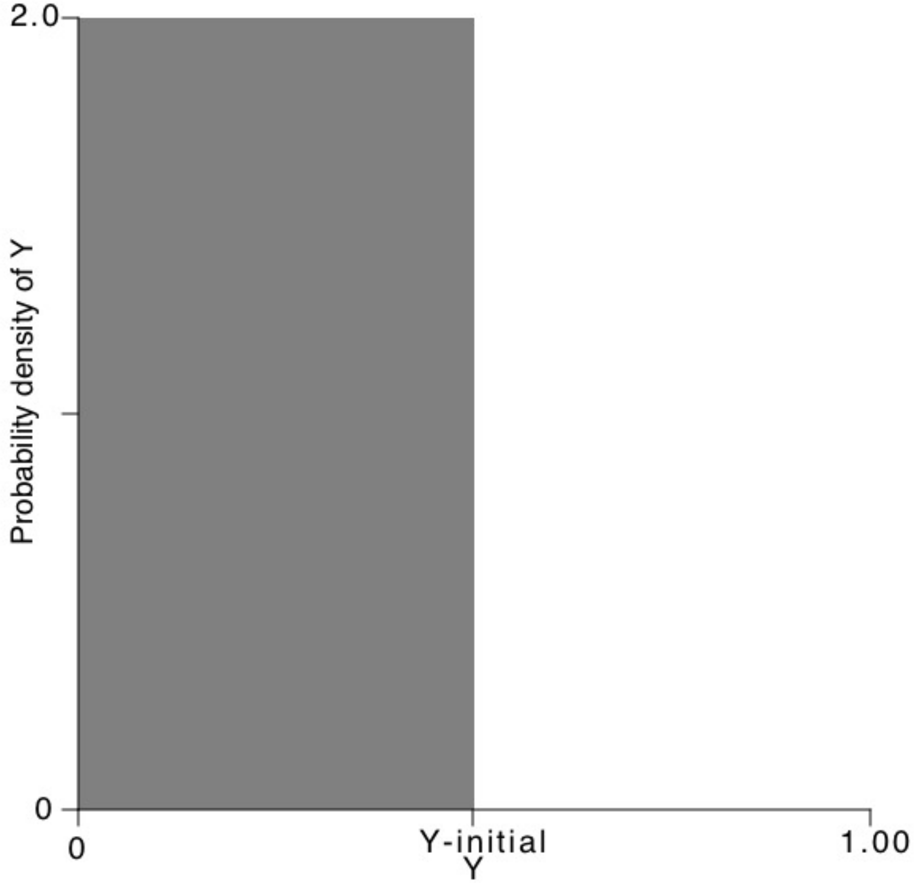
"Stuff happens"



Precision

Suppose X increases. What do our different hypotheses say about Y?

"As X increases, Y will decrease."



Precision

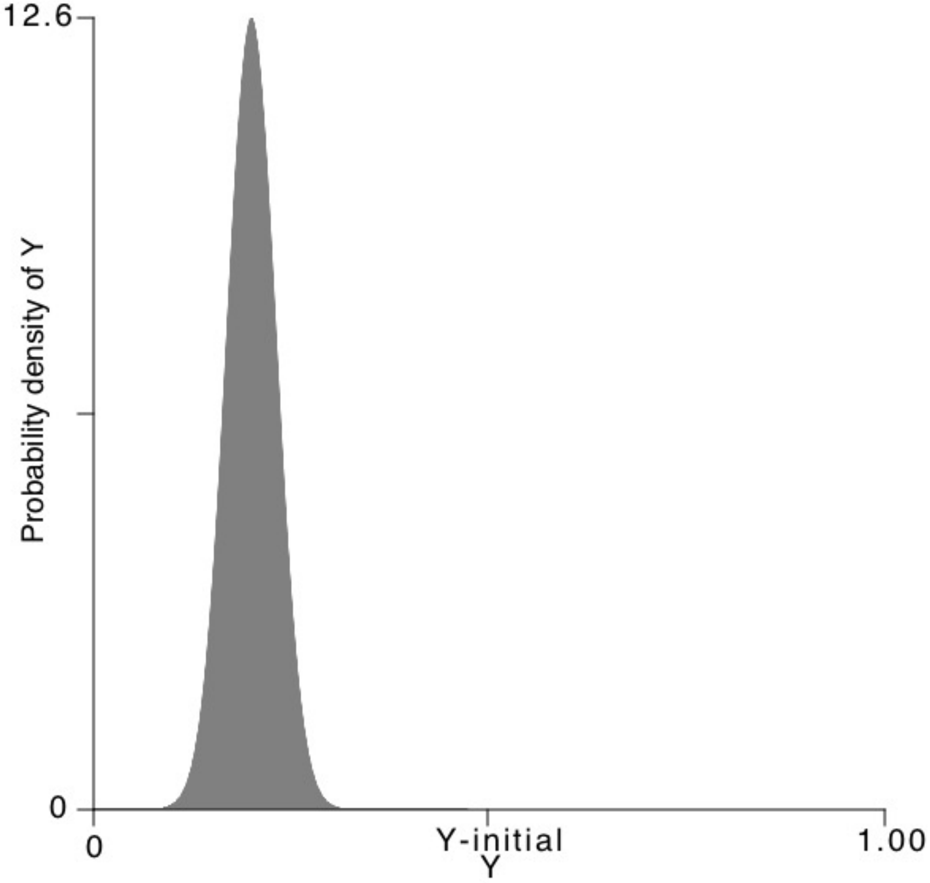
Suppose X increases. What do our different hypotheses say about Y?

“As X increases,

Y will decrease

according to the

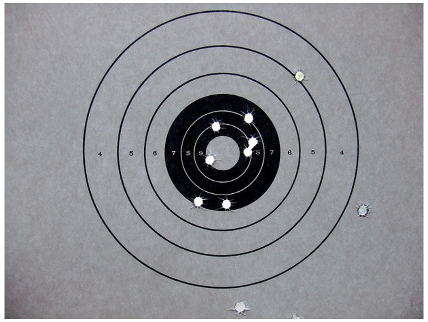
following function…”



Explanatory completeness

Beware the limits of post-hoc explanations!

- The Texas sharpshooter fallacy
  - A.K.A., Don't just test on your training data



- "My model predicts where people shoot – you just need to specify the bullseye-location parameter for each person!"

Explanatory completeness

- We don’t want models that just explain data after the fact!
- Rather, we want models that do well on the enormous variety of cases we haven’t yet seen.

That is, predictive accuracy.

How do can we compare models?

What makes one model or theory better than another?

- Explanatory completeness
- Predictive accuracy
- Being understandable

Predictive accuracy

Straightforward in principle:

1. Make predictions
2. Collect data
3. Evaluate model
4. Publish results

Predictive accuracy

Difficult in practice:

1. Publication bias
2. |old data| >> |new data|
3. Choosing criteria/loss functions
4. Free parameters

Predictive accuracy

Can we estimate predictive accuracy using old data?

- Cross-validation
- “Information criteria”

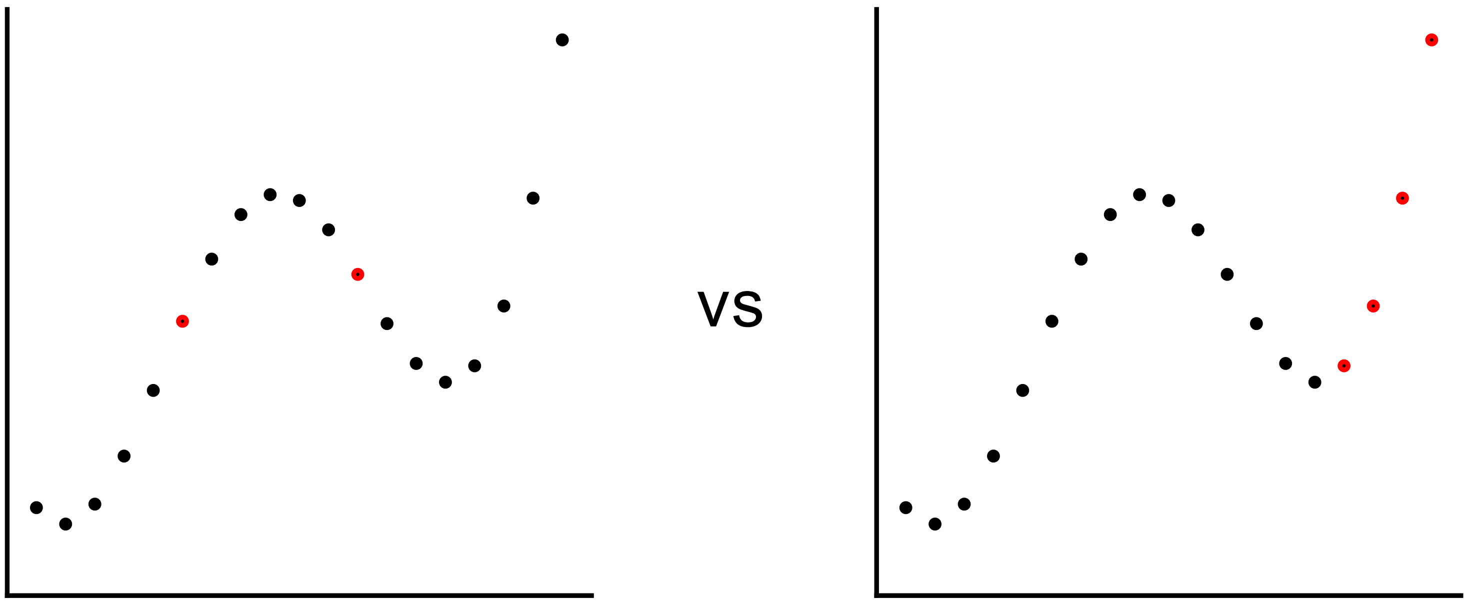
Cross-validation

1. Partition the data into training and validation sets
2. Fit the model on the training data
3. Get the probability* of the validation data under the fitted model.
4. Repeat steps for non-overlapping validation sets until all of the data have been covered.

Cross-validation

Issues:

- Can be computationally expensive
- Are cross-validation test sets like new cases?



Information criteria

Lower scores are better; generally

score = badness of fit + complexity penalty.

Most common badness of fit = -log(P(D|M,θ_{MLE}))

i.e., negative log likelihood of data given model, using likelihood-maximising parameters θ_{MLE}.

Perfect fit, e.g., P(D|M,θ_{MLE})=1 → badness of fit=0.

Information criteria

Different criteria vary by their complexity terms and goals:

Name	Goal	Fit term	Complexity term
Akaike **IC**	Find model with best hold-1-out cross validation accuracy[1,2]	$-2*\log(P(D\|M,\theta_{MLE}))$	2*k (k = # of params)
Bayesian **IC** (misnomer)	Find model with highest probability[1,2,3]	$-2*\log(P(D\|M,\theta_{MLE}))$	k*log(n) (n = # data points)
Watanabe-**A**kaike **IC**	Like AIC, but applies more generally	-log(P(D|M))[4]	Effective # params See (Wantanabe, 2010)

[1] Asymptotically

[2] If models are of a particular type (exponential family)

[3] If the true generating model is among those being tested

[4] Requires integrating over θ

(See also DIC, RIC)

Information criteria

Issues:

- Assumptions often aren't true
  - Sometimes a model is insensitive to a parameter or parameters are partially redundant
  - Sometimes a single parameter hides enormous flexibility
  - Sometimes parameters are hidden
- Criteria with weaker assumptions are sometimes intractable to compute (e.g., WAIC)

How do can we compare models?

What makes one model or theory better than other?

- Explanatory completeness
- Predictive accuracy
- Being understandable

Being understandable

- Part of a model's value is as a foundation for other models and theories.
- If we want to *understand* human cognition, then incomprehensible models aren't useful.
- One criterion: can a sophisticated person implement the model from a description?

Conclusions

Models are better when they're more

- General
- Precise
- Predictively accurate
- Parsimonious
- Comprehensible

Some of these notions can be expressed formally, e.g., using probability theory.

They should complement, rather than replace, your intuitions about how plausible, useful, or reasonable a model is.

References and further reading

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, *24*(6), 997-1016.

Jeffreys, W. H., & Berger, J. O. (1992). Ockham's Razor and Bayesian analysis. *American Scientist*, *80*(1), 64–72.

Shepard, R. N. (1987). Towards a universal law of generalization for psychological science. *Science*, *237*, 1317–1323.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *The Journal of Machine Learning Research*, *11*, 3571–3594.