Necessary Changes in the Philosophy and Practice of Probability & Statistics

William M. Briggs

Statistician to the Stars!

.

matt@wmbriggs.com

What is probability?

All men are mortal

Socrates is man

Socrates is mortal

Just half of Martians are mortal

Socrates is Martian

Socrates is mortal

Pr(A) *does not exist*!

Pr(A|evidence) might exist

A *does not* “have” a distribution

Distributions do not exist

If Pr(A) = limiting relative frequency, then no probability can ever be known.

If Pr(A|evidence) is subjective, then

$$\mathrm{Pr}(x = 7 | x + y = 12) = 1$$

if I say so.

Interocitors can take states $s_1, \ldots, s_p$

This is an interocitor

This interocitor is in state s_j

$\Pr(s_j|\text{Interocitors can...}) = 1/p$ :: *No symmetry*!

Jack said he saw a whole bunch of guys

There were 12 guys

$$\Pr(12|\text{Jack said...}) = \textit{not too unlikely}.$$

*Be*cause

Cause :: form + material + mechanism + direction :: essence + power

$$\text{Pr}(\text{Y}|\text{cause or determine}) = 1$$

$$y = \tan(\theta) \cdot x - g(2v_0^2 \cos^2\theta)^{-1} \cdot x^2$$

$$\text{Pr}(y|xgv_o\theta) \in \{0, 1\}$$

Chance or randomness are not ontic, thus powerless. *No probability model is causal* (including QM). Every potential must be made actual by something actual (including QM).

We have $\Pr(Y|X)$, where X is that information we think or assume is probative of Y—meaning we think X is related to the causal path of Y. If not, *pain*.

Hypothesis testing? We cannot derive from $\Pr(\mathsf{Y}|\mathsf{X}) = p$ that Y. *Probability is not decision*!

P-value $= \Pr(\text{larger ad hoc stat}|\mathsf{M}_\ominus, x, \theta_s = 0)$, which is *no way* related to $\Pr(\theta_s = 0|x, \mathsf{M}_\ominus)$.

$\Pr(\text{larger ad hoc stat}|\mathsf{M}_\ominus, x, \theta_s \neq 0)$ may be lower!

Models

Bayes is not important: probabaility is.

A parameterized model M relates X to Y probabilistically, e.g. $\mu = \beta_0 + \beta_1 x$ where μ is central parameter of normal used to characterize uncertainty in some y. "Priors" a real distraction: start finite!

With rare exceptions, parameters are of no interest to man nor beast.

$\hat{\mathsf{Y}} = f(\mathsf{X}, \widehat{\theta}(\mathsf{M}_\theta))$ ignores uncertainty, and makes a decision.

$\Pr(\theta|$data, $\mathsf{M}_\theta)$ only about unobservable parameters.

We want this:

$\Pr(Y|\text{new } X, \text{data}, M)$, where the data are old values of Y and X, and M are the arguments that led to a (parameterized) model; the parameters having been integrated out.

This—and only this—captures the *full* uncertainty, given M. *Prediction!*

Every model—neural net, statistical, machine learning, artificial intelligence, anything—can fit into the Pr(Y|XDM) schema. What differentiates them is usually a matter of *ad hoc* complexity and form—and a building in of decision.

Demystifying "learning"

ANNs, GANs, Deep this-and-thats, etc. $=$ parameterized non-linear regressions

Learning $=$ estimating parameters

Extracting features $= f(\text{input data})$

There is no such thing as *unsupervised learning.*

Every algorithm does *exactly what it is designed to do*, and therefore gives *correct* results—conditional on the algorithm.

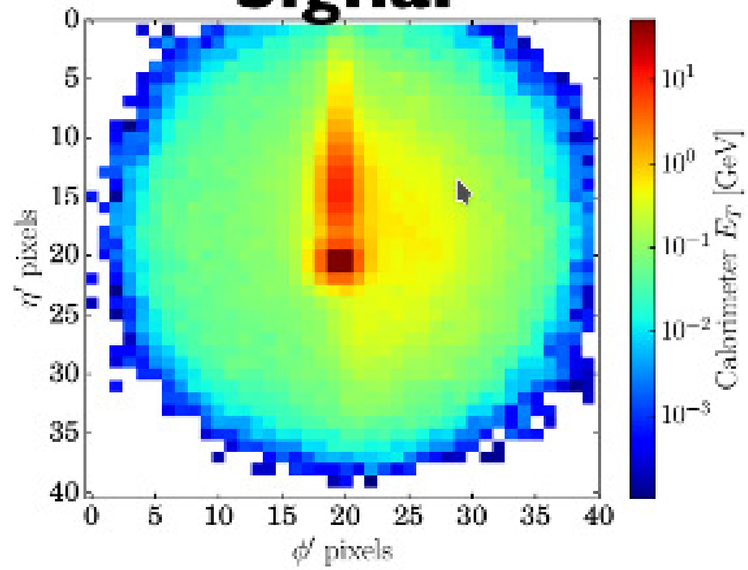
Not all probability is quantitative, and not all algorithms live in machines.

Monte Carlo — The place to lose your money, and your way.

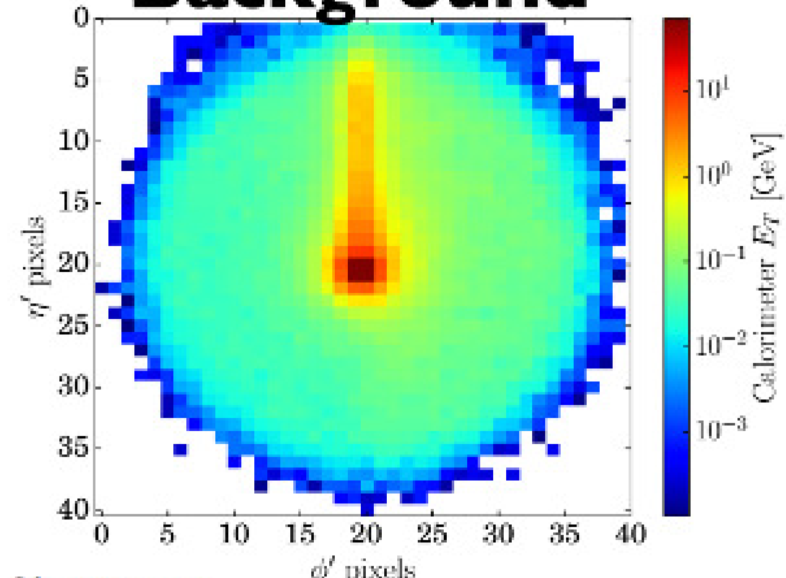
Jaynes: "It appears to be a quite general principle that, whenever there is a randomized way of doing something, then there is a nonrandomized way that delivers better performance but requires more thought."

Disadvantage of DNN approach: choosing network architecture is a bit of voodoo This seems to work though

Signal



Background



Kasieczka

Image D with possible signal + background

$$\Pr(d_{ij}|\mathsf{M}_{\mathsf{B}}) \sim \text{Poisson}(\lambda_B)$$

$$\Pr(d_{ij}|\mathsf{M}_{\mathsf{S+B}}) \sim \text{Poisson}(\lambda_S + \lambda_B)$$

$$\Pr(d_{ij}|\mathsf{M}_{\mathsf{S+B}}, \mathsf{M}_{\mathsf{B}}) = p\mathsf{P}(\lambda_B) + (1-p)\mathsf{P}(\lambda_S + \lambda_B)$$

$\Pr(\mathsf{M}_{\mathsf{S+B}}|d_{ij})$ Guglielmetti et al., 2002, Mon. Not. R. Astron. Soc

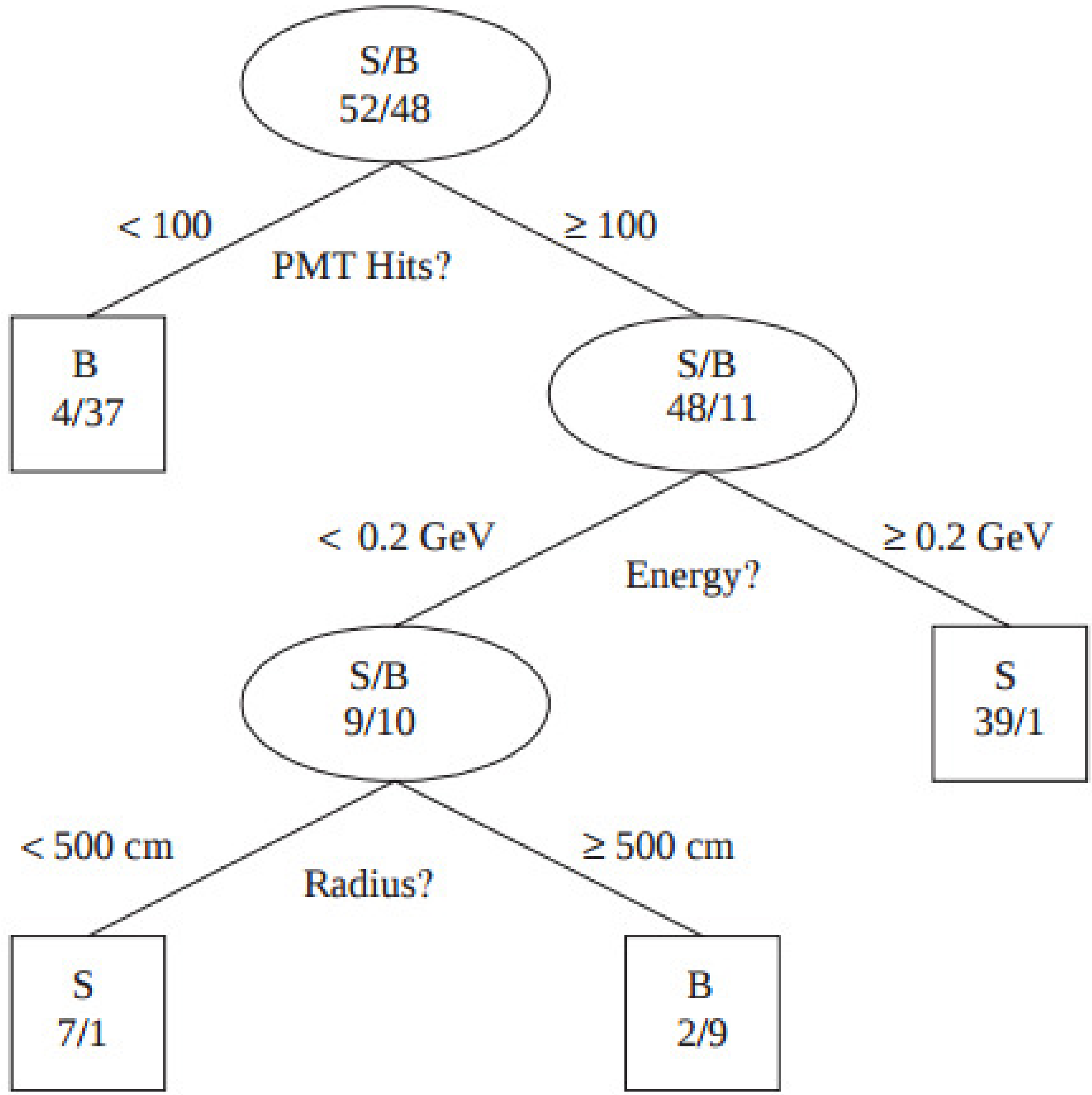


Fig. 1. Schematic of a decision tree. **S** for signal, **B** for background. Terminal nodes (called leaves) are shown in boxes. If signal events are dominant in one leaf, then this leaf is signal leaf; otherwise, background leaf.

Roe et al.

Skill:

		Obs	
		S	B
Mod	S	3	5
	B	5	87

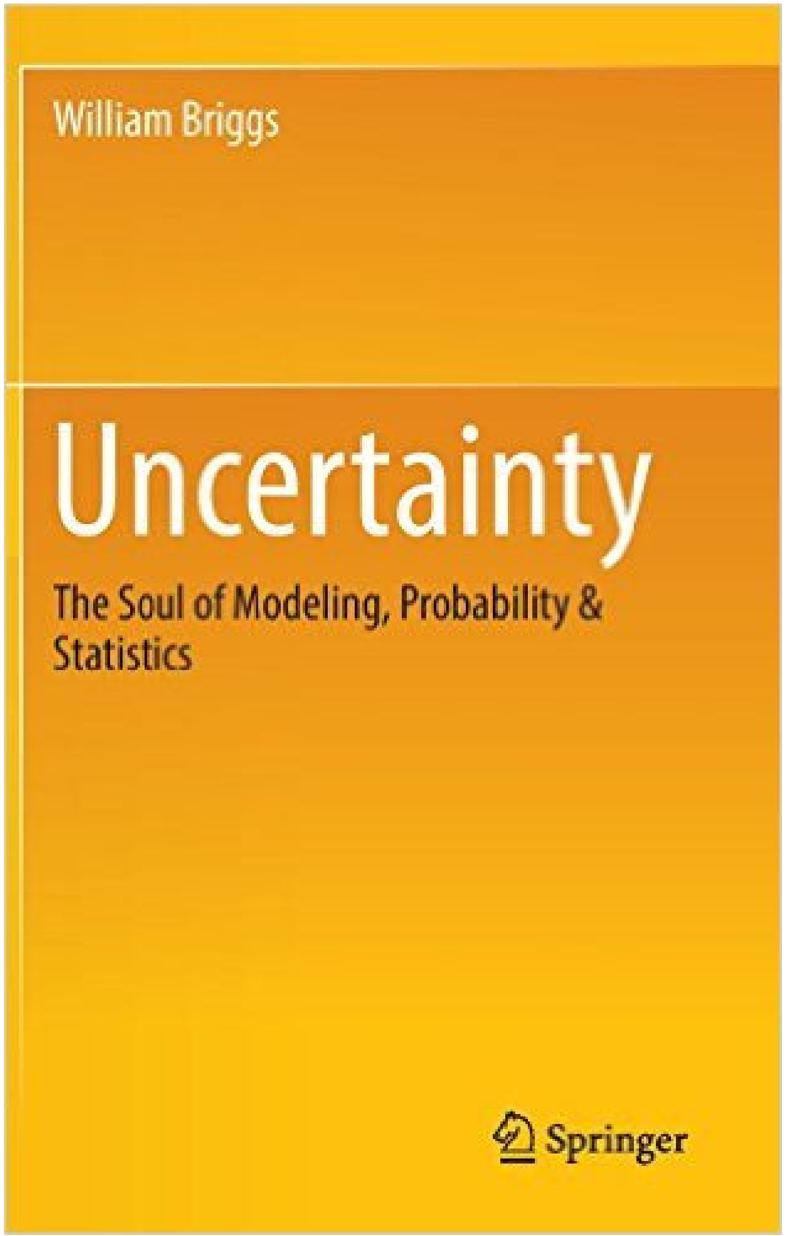
Super machine neural deep-learning boosting forest machine boasts 90% accuracy!

Skill and calibration curves, *not* ROC.

Model-based vs. verification-based uncertainty; verify "features".

All uncertainty carried through to the bitter end.

In the absence of knowledge of cause, all probabilistic models will classify imperfectly.



"This is not not a statistics text, it is not a treatise on philosophy of science or logic. This work is like nothing I have seen before, an excellent combination of the above, indeed the 'the soul of modeling, probability ...', presented with passion and accessible to everybody."

"It is a deep philosophical treatment of probability written in a plain language and without the interference of unnecessary math."