Derivation of an output pdf from Bayes' theorem and the principle of maximum entropy

Ignacio Lira, Clemens Elster,

Wolfgang Wöger and Maurice Cox

Catholic University of Chile

Physikalisch-Technische Bundesanstalt (PTB)

National Physical Laboratory (NPL)

Introduction

At present there exists a sound, consistent and accepted theory for the evaluation of measurement uncertainty

The theory provides a method for:

- unambiguously expressing the information that is available about the quantities that enter a measurement model

- propagating this information through the model in order to represent and communicate the resulting, generally incomplete, state of knowledge about the output quantity.

Basic elements:

- probability calculus

- Bayes' theorem

- principle of maximum entropy

Probability calculus

Measurement model:

$$\mathbf{X} \longrightarrow \boxed{f} \longrightarrow Y$$

N input quantities output quantity

$$g_{\mathbf{X}}(\boldsymbol{\xi}) \longrightarrow g_Y(\eta)$$

Product rule:

$$g_{Y,\mathbf{X}}(\eta,\boldsymbol{\xi}) = g_{Y|\mathbf{X}}(\eta|\boldsymbol{\xi})\,g_{\mathbf{X}}(\boldsymbol{\xi})$$

Marginalization:

$$g_Y(\eta) = \int\cdots\int g_{Y|\mathbf{X}}(\eta|\boldsymbol{\xi})\,g_{\mathbf{X}}(\boldsymbol{\xi})\,\mathrm{d}\xi_1\cdots\mathrm{d}\xi_N$$

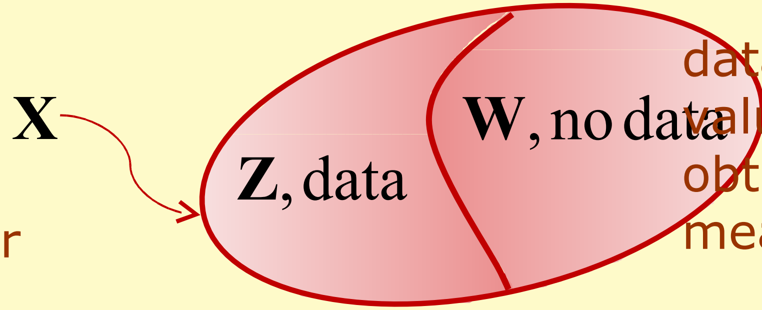
Some authors prefer to refer to this black box as the measurement equation, perhaps to avoid confusion with the term statistical model used by statisticians.

However, to the extent that the relation between the input and output quantities is usually derived from a theoretical construct representing the physical phenomena involved in the measurement procedure, the term measurement model seems quite appropriate.

Bayes' theorem

$g_{\mathbf{X}}(\xi) \longrightarrow g_{\mathbf{X}}(\xi|\text{information},\text{data})$

"prior" not a very good name, because in general this PDF may depend on information that becomes available only after having obtained the data



data to refer to given values or indications obtained by direct measurements

$$\underbrace{g_{\mathbf{Z},\mathbf{P}}(\mathbf{V},\boldsymbol{\pi}|\text{data},\text{info})}_{\text{posterior}} \propto \int \underbrace{l(\mathbf{V},\boldsymbol{\pi}|\text{data})}_{\text{likelihood}}\,\underbrace{g_{\mathbf{Z},\mathbf{P}}(\mathbf{V},\boldsymbol{\pi}|\text{no data},\text{info})}_{\text{prior}}\,\mathrm{d}\pi$$

information to refer to any other type of available knowledge about any or all quantities, for example, some of them being non-negative

P (with possible values π) is a set of one or more parameters from which the frequency distribution may depend. These parameters must be integrated out

depends on the frequency distribution from which the data are assumed to be drawn

Typical example

$$l(\zeta, z) \propto \frac{1}{\sqrt{\nu}} \exp\left(-\frac{(\zeta - z)^2}{2\nu}\right)$$

$$l(\zeta, z_1, \ldots, z_n) \propto \prod_i \frac{1}{\sqrt{\nu}} \exp\left(-\frac{(\zeta - z_i)^2}{2\nu}\right)$$

$$g_{Z,V}(\zeta, \nu) = g_Z(\zeta)\, g_V(\nu)$$

$$g_Z(\zeta \mid \text{no data, no info}) \propto 1$$

$$g_V(\nu \mid \text{no data, no info}) \propto \frac{1}{\nu}$$

Substitution of all these elements into Bayes' theorem and integration of the variance leads to

$$g_Z(\zeta \mid \text{data, no info}) \propto \int_0^\infty \frac{1}{\nu} \prod_i \frac{1}{\sqrt{\nu}} \exp\left(-\frac{(\zeta - z_i)^2}{2\nu}\right) \mathrm{d}\nu$$

which can be shown to be a scaled-and-shifted *t*-distribution

Note that in this case the variance takes the place of a (single) unknown parameter of the frequency distribution

Typically data are obtained, assuming a Gaussian process with unknown variance V

Respect to data that will be obtained, the global likelihood is just the product of the individual likelihoods

If we have no information about Z and V, so in this case non informative PDFs must be used

For a given datum z of Z, the likelihood is in this case a Gaussian function where ζ are the possible values of Z and ν are the possible values of V

The quantity Z and the parameter V can be considered as being independent, so their joint (prior) PDF factorizes into the product of their respective marginals

Because Z establishes the location of the frequency distribution its non informative prior is a constant

Instead, V is a scale parameter and so Jeffreys' prior is adequate

So where do we stand?

Recall that from the rules of probability calculus we found that the PDF for the output quantity is given by this multiple integral

$$g_Y(\eta) = \int \cdots \int g_{Y|\mathbf{X}}(\eta|\boldsymbol{\xi})\, g_{\mathbf{X}}(\boldsymbol{\xi})\, \mathrm{d}\xi_1 \cdots \mathrm{d}\xi_N$$

The second joint PDF depends on whether we have information and/or data about the input quantities

The likelihood is in this case not obtained through the concept of underlying frequency distribution. The reason is that if the values of the input quantities are assumed to be given, and thus as data, then the output quantity can only assume the value established by the measurement model

$$g_{\mathbf{X}}(\boldsymbol{\xi}|\text{data, no info})$$

$$g_{\mathbf{W}}(\boldsymbol{\zeta}|\text{no data, info})$$

For those quantities for which we have data but no information, we use Bayes' theorem with non informative priors

But what do we do with those quantities ***W*** for which we have information, but no data?

$$g_{\mathbf{Z}}(\boldsymbol{\zeta}|\text{data,info})$$

For those quantities for which we have data *and* information, we use Bayes' theorem with appropriate *informative* priors

$$g_{Y|\mathbf{X}}(\eta|\boldsymbol{\xi},\text{info}) \propto l(\eta|\boldsymbol{\xi})\, g_Y(\eta,\text{info})$$

: special application of Bayes' theorem

$$\delta(\eta - f(\boldsymbol{\xi}))$$

This means that the likelihood becomes a delta function

And with this PDF, in case previous knowledge about the input quantity exists?

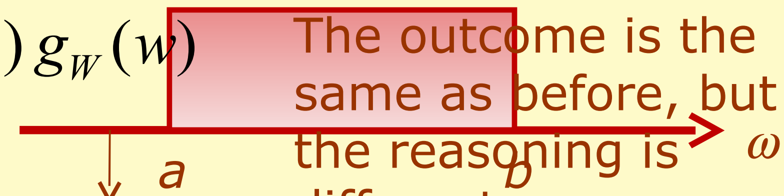
Principle of maximum entropy (PME)

Precisely, the principle of maximum entropy is used to construct PDFs for those quantities for which no measurement data are available, and thus for which Bayes' theorem cannot be used

It is interesting to comment that some authors derive this last result from Bayes' theorem taking the estimate and standard uncertainty as the "data"

$$g_W(\omega \mid w, u(w)) \propto l(\omega \mid w, u(w))\, g_W(w)$$

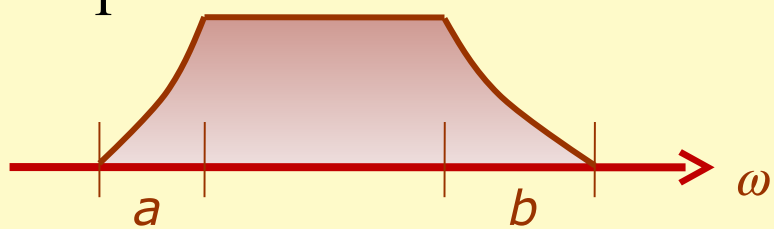
Typical example: $a < \omega < b$



$$\exp\left(-\frac{(\omega - w)^2}{2u^2(w)}\right)$$

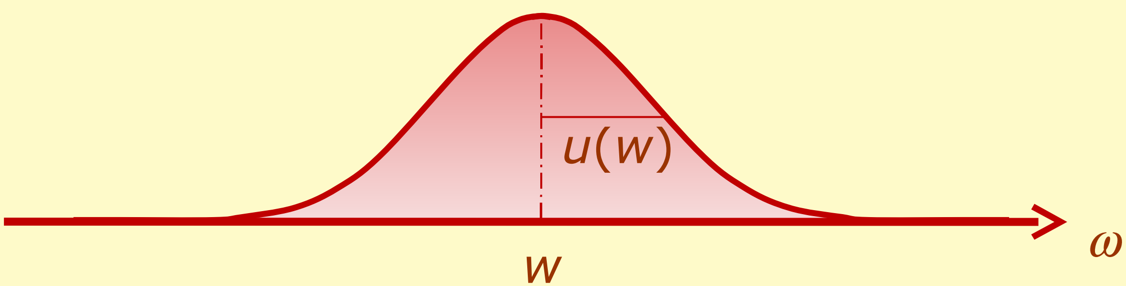
$$1$$

Variation: the limits themselves are known to be contained within given intervals



The outcome is the same as before, but the reasoning is different

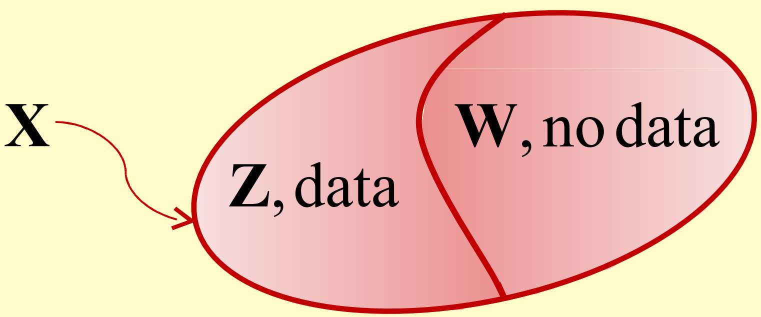
Typical example: w, $u(w)$ known



Summary

Probability calculus: $$g_Y(\eta) = \int\cdots\int g_{Y|\mathbf{X}}(\eta\,|\,\boldsymbol{\xi})\, g_{\mathbf{X}}(\boldsymbol{\xi})\,\mathrm{d}\boldsymbol{\xi}$$

$$g_Y(\eta\,|\,\text{data},\text{info}) = \int\cdots\int \delta\left(\eta - f(\boldsymbol{\xi})\right) g_{\mathbf{Z}}(\boldsymbol{\zeta})\, g_{\mathbf{W}}(\boldsymbol{\omega})\, g_Y(\eta)\,\mathrm{d}\boldsymbol{\zeta}\cdots\mathrm{d}\boldsymbol{\omega}$$



Bayes' theorem

PME

That all, folks!

In general, this PDF cannot be obtained analytically, but the Monte Carlo method in the Supplement 1 to the GUM allows for its numerical implementation