Handling model uncertainties by means of comparison densities

Sara Algeri

School of Statistics, University of Minnesota

CalStats Working Group Meeting

December 1, 2020.

The key ingredients

The skew-G density model (Mukhopadhyay, 2017)

Given a random variable X, let F and f be its cdf and pdf (or pmf) respectively. Let G a suitable cdf and let g be the respective pdf (or pmf). Then,

$$f(x) = g(x) \; d(G(x); G, F) \tag{1}$$

where

$$d(G(x); G, F) = \frac{f(x)}{g(x)} \quad \text{is called } \textbf{comparison density} \tag{2}$$

Note:

$d(G(x); G, F)$ it is not a density w.r.t. x but it is with respect to $G(x)$.
$\Rightarrow$ Set $u = G(x)$, $d(u; G, F)$ is a density w.r.t. u.

But what are g and f in practice?

It really depends on the astrophysical problem...

For instance...

- If we are trying to assess if the background model is correct:

$$\underbrace{f(x)}_{\text{true (unknown) bkg}} = \underbrace{g(x)}_{\text{postulated bkg}} \underbrace{d(G(x); G, F)}_{\text{comparison density}} \tag{3}$$

- If we are trying to detect the signal of a new source:

$$\underbrace{f(x)}_{\substack{\text{truth (may or may not}\\ \text{contain the signal)}}} = \underbrace{g(x)}_{\substack{\text{true (or estimated)}\\ \text{bkg model}}} \underbrace{d(G(x); G, F)}_{\text{comparison density}} \tag{4}$$

Modelling and estimation

If we represent the comparison density through a series of $T_j(x;G)$ orthonormal functions of $G(x)$ then

$$f(x) = g(x) \underbrace{\left\{1 + \sum_{j>0} \theta_j \; T_j(x;G)\right\}}_{d(G(x);G,F)} \tag{5}$$

To estimate it

- Truncate the series at m
- Estimate the θ_js with $\widehat{\theta}_j = \frac{1}{n}\sum_{i=1}^{n} T_j(x_i;G)$.

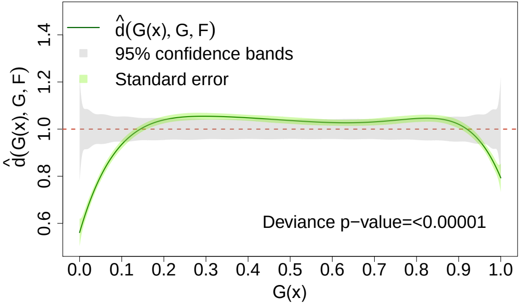
Inference

H_0 : d(G(x);G,F)=1 **vs** $H_1 : d(G(x); G, F) \neq 1$

... but $d(G(x); G, F) = 1 + \sum_{j>0} \theta_j T_j(x; G)$ **remember?**

$\Rightarrow d(G(x); G, F) = 1$ whenever $\theta_1 = \theta_2 = \cdots = 0$!

- Deviance test statistic: $D = \sum_{j=1}^{m} \widehat{\theta}_j^2 \xrightarrow{d} \chi_m^2$ (under H_0, as $n \to \infty$)
- Confidence bands:



References

- Algeri, S., 2020. Detecting new signals under background mismodeling. *Physical Review D 101, 015003.*
- Algeri, S. (2020+). Informative goodness-of-fit for multivariate distributions. *Under review. arXiv:2009.00503*
- Mukhopadhyay, S., 2017. Large-scale mode identification and data-driven sciences. *Electronic Journal of Statistics.*