DRAFT 0.0

Glen Cowan
30 April 2020

Ideas on how to include systematics in a search

Suppose that a search for a new signal has been set up based on a neural network that is designed, e.g., to distinguish a background-only model from a signal model such as clockwork gravity. Suppose the signal model has M parameters of interest $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_M)$ (in the case of clockwork gravity, $\boldsymbol{\mu} = (k, m_5)$), which for the background-only model takes on the special values $\boldsymbol{\mu}_0$.

An experiment results in data here symbolically $\mathbf{n} = (n_1, \ldots, n_N)$, e.g., a histogram of a kinematic variable such as a dilepton mass. After training the neural network results in a statistic $y(\mathbf{n})$, which we treat here as a fixed function. The distribution of y can be written $p(y|\boldsymbol{\mu}, \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_L)$ represent nuisance parameters, e.g., related to the electromagnetic energy resolution.

Suppose that the nuisance parameters $\boldsymbol{\theta}$ have nominal values $\tilde{\boldsymbol{\theta}}$ that are treated in the same way as measurements (they may indeed represent control measurements). Suppose we have a function $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ that represents the distribution of the $\tilde{\boldsymbol{\theta}}$ under assumption of a given $\boldsymbol{\theta}$. This could be e.g. a multivariate Gaussian centred about $\boldsymbol{\theta}$ with some covariance matrix V, or in simpler cases the components of $\tilde{\boldsymbol{\theta}}$ might be independent and so $p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta})$ could be a product of 1-D Gaussians. In any case we will treat the control measurements $\tilde{\boldsymbol{\theta}}$ and the statistic y as being statistically independent.

An experiment is thus characterized by a value of y and a value of $\tilde{\boldsymbol{\theta}}$. The likelihood function is given by

$$L(\boldsymbol{\mu}, \boldsymbol{\theta}) = p(y, \tilde{\boldsymbol{\theta}}|\boldsymbol{\mu}, \boldsymbol{\theta}) = p(y|\boldsymbol{\mu}, \boldsymbol{\theta}) p(\tilde{\boldsymbol{\theta}}|\boldsymbol{\theta}) \tag{1}$$

For now suppose that all of the ingredients on the right-hand side of Eq. (1) are known. For example, the distribution of y is determined using MC and its dependence on the nuisance parameters $\boldsymbol{\theta}$ has been found by interpolating distributions that are shifted by variations of the nuisance parameters (e.g., using HistFactory or similar).

To test a point in parameter space $\boldsymbol{\mu}$, including the special point $\boldsymbol{\mu}_0$ that represents the background-only model, we define as usual the profile likelihood ratio [1]

$$\lambda(\boldsymbol{\mu}) = \frac{L(\boldsymbol{\mu}, \hat{\hat{\boldsymbol{\theta}}})}{L(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\theta}})} , \tag{2}$$

where $\hat{\hat{\boldsymbol{\theta}}}$ represents the profiled (constrained) values of the nuisance parameters, i.e., the values of $\boldsymbol{\theta}$ that maximize the likelihood for the specific $\boldsymbol{\mu}$. Using $\lambda(\boldsymbol{\mu})$ we then define

$$t_{\boldsymbol{\mu}} = -2 \ln \lambda(\boldsymbol{\mu}) . \tag{3}$$

From Wilks' theorem we know that in the asymptotic limit and under regularity conditions [2] the statistic $t_{\boldsymbol{\mu}}$ follows a chi-square distribution for M degrees of freedom, where M is the number of parameters of interest, i.e., the number of components of $\boldsymbol{\mu}$ (2 for clockwork gravity). It remains to be investigated whether these regularity conditions are valid in the case of clockwork gravity.

Let us suppose that we can find the pdf $f(t_{\boldsymbol{\mu}}|\boldsymbol{\mu},\boldsymbol{\theta})$ either using the asymptotic form or from Monte Carlo simulation. Taking larger values of $t_{\boldsymbol{\mu}}$ to represent increased disagreement between the data and the hypothesis of $\boldsymbol{\mu}$, the p-value of the hypothesis is given by

$$p_{\boldsymbol{\mu}} = \int_{t_{\boldsymbol{\mu},\mathrm{obs}}}^{\infty} f(t_{\boldsymbol{\mu}}|\boldsymbol{\mu},\boldsymbol{\theta})\, dt_{\boldsymbol{\mu}} = 1 - F(t_{\boldsymbol{\mu},\mathrm{obs}}|\boldsymbol{\theta}) \ , \tag{4}$$

where $F(t_{\boldsymbol{\mu}}|\boldsymbol{\theta})$ is the cumulative distribution of $t_{\boldsymbol{\mu}}$. If the asymptotic chi-square distribution is valid, the this distribution is in fact independent of the nuisance parameters. Thus if one finds $p_{\boldsymbol{\mu}} < \alpha$ for some threshold α then $\boldsymbol{\mu}$ is rejected. If the dependence of the distribution on $\boldsymbol{\theta}$ cannot be neglected then one has various options, e.g., profile construction [3]. In any case the effect of the systematic uncertainties related to the nuisance parameters is taken into account by the procedure above.

References

[1] Glen Cowan, Kyle Cranmer, Eilam Gross and Ofer Vitells, Eur. Phys. J. C 71 (2011) 1554.

[2] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*, Ann. Math. Statist. **9** (1938) 60-62.

[3] K. Cranmer, in *Statistical Problems in Particle Physics, Astrophysics and Cosmology (PHYSTAT 05): Proceedings, Oxford, UK, September 12-15, 2005*, 112–123 (2005), [arXiv:physics/0511028].