Statistical Background Ensemble

This note discusses the strategy for estimating the spurious signal that is used by the LPX search for Non-Resonant phenomena in the dilepton (ee,$\mu\mu$) final states. This analysis uses a background estimation defined with a fit to data. This note describes how the spurious signal is estimated for this data-driven background.

Summary: This procedure is described in detail, however the essential idea is as follows: the spurious signal is measured from an ensemble of *plausible* background shapes that are informed by systematic uncertainties.

Challenge

We define our expected background using a PDF, $f(m_{\ell\ell})$, fit to a dataset $\vec{D}$ with dimension N_{data}. The full analysis procedure of fitting the dataset and obtaining a background estimate is described by the function $F(\vec{D})$, with a range of dimension N_{bkg} and domain of dimension N_{data}. The output of $F(\vec{D})$ is the expected background, given the provided dataset $\vec{D}$. The remainder of this procedure is agnostic with regard to the particular definition of $F(\vec{D})$. In the case of the Non-Resonant search, $F(\vec{D})$ is the fitting and extrapolation procedure, where N_{data} is the number of bins in the Control Region, and N_{bkg} is the single bin of the Signal Region.

A main uncertainty on $F(\vec{D})$ is the spurious signal σ^{spur}. We define σ^{spur} as the number of signal events measured via $F(\vec{D})$, given $\vec{D}$ is a background-only Asimov dataset. As with $F(\vec{D})$, we consider $\sigma^{\text{spur}}(\vec{D})$ to be a function of a potential background shape $\vec{D}$. To discuss this, let's define the potential background-only distributions on which $F(\vec{D})$ can operate:

1. $\vec{D}^{\text{MC}}$, the nominal simulated background, with enough events to mitigate statistical fluctuations.

2. $\vec{D}_i^{\text{syst}}$, a set of simulated systematic variations on the nominal background.

3. $\vec{D}^{\text{data}}$, the recorded dataset which will be used for the final background estimation. We consider this recorded data to be randomly sampled from a "true" PDF, $f^{\text{truth}}(m_{\ell\ell})$.

4. $\vec{D}^{\text{truth}}$, the Asimov dataset generated by $f^{\text{truth}}(m_{\ell\ell})$. In practice, this is not known.

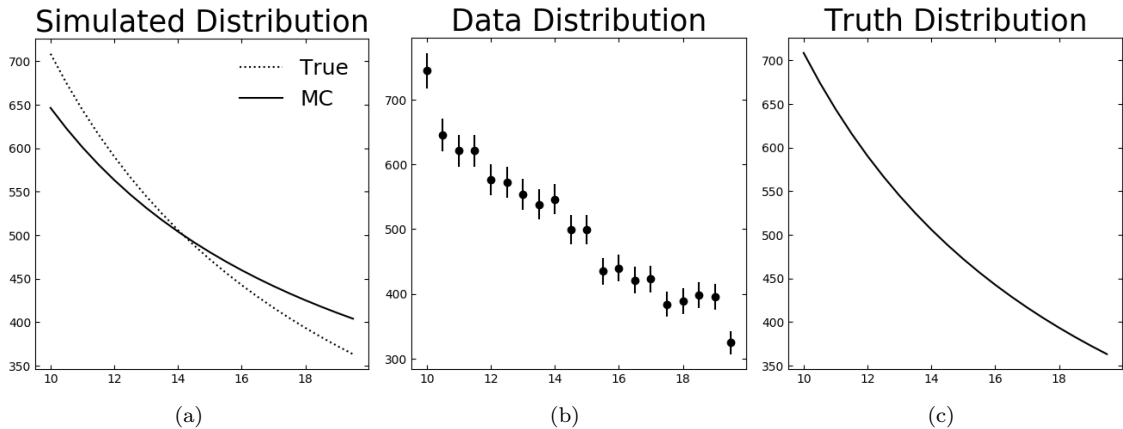For illustration purposes, examples of these are shown in figure 1.

Figure 1: Examples of $\vec{D}^{\text{MC}}$ (a), $\vec{D}^{\text{data}}$ (b), and $\vec{D}^{\text{truth}}$ (c). $\vec{D}^{\text{MC}}$ is shown to diverge from the true underlying distribution that has generated the data.

It is trivial to measure the σ^{spur} that corresponds to $\vec{D}^{\text{MC}}$, denoted $\sigma^{\text{spur}}(\vec{D}^{\text{MC}})$. This is usually done by computing $F(\vec{D}^{\text{MC}})$, and measuring the number of signal events spuriously recovered from the procedure. However if we want to measure $\sigma^{\text{spur}}(\vec{D}^{\text{data}})$ and use it in the analysis, we are faced with three main concerns:

1. $\vec{D}^{\mathrm{data}}$ may not be a background-only distribution, and we certainly do not want to assume that it is before making a measurement.

2. Due to limited statistics, $\vec{D}^{\mathrm{data}}$ is not an Asimov dataset, and therefore $F(\vec{D}^{\mathrm{data}})$ will be influenced by the statistical fluctuations of the recorded data.

3. The most fundamental issue is that the σ^{spur} measured from $F(\vec{D}^{\mathrm{data}})$ is essentially our signal measurement.

If we want to use $F(\vec{D}^{\mathrm{data}})$ for the expected background, we need an alternative way to estimate σ^{spur}. We cannot use σ^{spur} estimated from $F(\vec{D}^{\mathrm{data}})$ for the reasons listed above. We prefer not to simply use σ^{spur} estimated from $F(\vec{D}^{\mathrm{MC}})$, since this makes the assumption that $\vec{D}^{\mathrm{MC}}$ was sampled from a PDF matching $f^{\mathrm{truth}}(m_{\ell\ell})$. Instead, we would like to use σ^{spur} measured from $F(\vec{D}^{\mathrm{truth}})$. Since we do not know $\vec{D}^{\mathrm{truth}}$, we use the procedure described here to obtain our best estimate of $\vec{D}^{\mathrm{truth}}$ given our available prior knowledge.

Approach

The exact form of $\vec{D}^{\mathrm{truth}}$ is not known. While our best theoretical estimate of the exact shape of $\vec{D}^{\mathrm{truth}}$ may be $\vec{D}^{\mathrm{MC}}$, this does not encompass our all of our information. Since $\vec{D}^{\mathrm{truth}}$ is a N_{bkg} dimensional vector, we know that $\vec{D}^{\mathrm{truth}} \in \mathbb{R}^{N_{\mathrm{bkg}}}$. Likewise, $\vec{D}^{\mathrm{MC}} \in \mathbb{R}^{N_{\mathrm{bkg}}}$. Going further, we have a strong suspicion that $\vec{D}^{\mathrm{truth}}$ is "close" to $\vec{D}^{\mathrm{MC}}$ within $\mathbb{R}^{N_{\mathrm{bkg}}}$. One could consider the systematic uncertainties $\vec{D}_i^{\mathrm{syst}}$ as a description of how close $\vec{D}^{\mathrm{truth}}$ is to $\vec{D}^{\mathrm{MC}}$. Instead, we would like to use $\vec{D}_i^{\mathrm{syst}}$ to define a PDF $h(\mathbb{R}^{N_{\mathrm{bkg}}}) \mapsto \mathbb{R}$ corresponding to the estimated likelihood of finding $\vec{D}^{\mathrm{truth}}$ to be at a particular point in $\mathbb{R}^{N_{\mathrm{bkg}}}$. In the future, we would like to rework this Bayesian reasoning to a more frequentist one.

For each systematic variation $\vec{D}_i^{\mathrm{syst}}$, we define a nuisance parameter $\theta_i \in \mathbb{R}$. For simplicity, consider that θ_i are defined such that have 1σ confidence that the value of $\hat{\theta}_i$ measured on $\vec{D}^{\mathrm{truth}}$ is $\in [-1, 1]$, and $\theta_i = 0$ corresponds to $\vec{D}^{\mathrm{MC}}$. This allows us to define *Standard Gaussian* priors for each systematic, $G(\theta_i)$. (If the upward and downward variations are measured separately, $G(\theta_i)$ can still be a Standard Gaussian and the impact of $\theta_i > 0$ can be treated differently than $\theta_i < 0$.) Taken together, θ_i define a vector $\vec{\theta} \in \mathbb{R}^{N_{\mathrm{syst}}}$. This leads to a simplification wherein we now consider the dimensionally smaller $\vec{\theta} \in \mathbb{R}^{N_{\mathrm{syst}}}$ instead of the full $\vec{D} \in \mathbb{R}^{N_{\mathrm{bkg}}}$, as long as $N_{\mathrm{syst}} < N_{\mathrm{bkg}}$. The measured shapes of the systematics, $\{\vec{D}^{\mathrm{syst}}\}$, are used to map choice of $\vec{\theta}$ onto a full background shape: $g(\vec{\theta}|\{\vec{D}^{\mathrm{syst}}\}) \mapsto \mathbb{R}^{N_{\mathrm{bkg}}}$. An example of this map is shown later in Equation 7.

Each choice of $\vec{\theta} \in \mathbb{R}^{N_{\mathrm{syst}}}$ has a prior expected probability of:

$$p(\vec{\theta}) = \prod_{i=0}^{N_{\mathrm{syst}}} G(\theta_i). \tag{1}$$

We should also keep in mind that we are essentially assigning a prior probability to each choice of $\vec{D} \in \mathbb{R}^{N_{\mathrm{bkg}}}$:

$$p(\vec{D}) = \oint_{\mathbb{R}^{N_{\mathrm{syst}}}} p(\vec{\theta})\, \delta(g(\vec{\theta}_j) - \vec{D})\, d\vec{\theta} \tag{2}$$

This $p(\vec{D})$ is zero for many non-physical choices of $\vec{D}$, since these cannot be represented by a choice of nuisance parameters.

Now we can return to the goal of approximating $\sigma^{\mathrm{spur}}(\vec{D}^{\mathrm{truth}})$. As a substitute, we propose to calculate $\sigma^{\mathrm{spur}}(\vec{D})$ for every $\vec{D} \in \mathbb{R}^{N_{\mathrm{bkg}}}$. These are then weighted by the corresponding probability $p(\vec{D})$:

$$\Delta^D(\vec{D}) = p(\vec{D}) \times \sigma^{\mathrm{spur}}(\vec{D}). \tag{3}$$

The resulting distribution of Δ^D represents the spurious signal for every possible background shape $\vec{D}$, weighted by the likelihood that each the $\vec{D} = \vec{D}^{\mathrm{truth}}$. We consider the mean of this distribution, $\mathrm{E}[\Delta^D]$, to be our best estimate of $\sigma^{\mathrm{spur}}(\vec{D}^{\mathrm{truth}})$, and the width of this distribution, $\mathrm{var}[\Delta^D]$, to be our uncertainty on that estimate.

Every choice of $\vec{\theta}$ maps onto a $\vec{D}$. Therefore, we can measure the spurious signal as $\sigma^{\mathrm{spur}}(\vec{\theta}) = \sigma^{\mathrm{spur}}(g(\vec{\theta}))$, and define a quantity analogous to Equation 3:

$$\Delta^{\theta}(\vec{\theta}) = p(\vec{\theta}) \times \sigma^{\mathrm{spur}}(\vec{\theta}). \tag{4}$$

The value of $\mathrm{E}[\Delta^D]$ is measured over the volume of potential values of $\vec{D}$. The denominator of the mean is $\oint_{\mathbb{R}^{N_{\mathrm{bkg}}}} p(\vec{D}) = 1$. A couple of substitutions show that we can calculate $\mathrm{E}[\Delta^{\theta}]$ instead of $\mathrm{E}[\Delta^D]$.

$$
\begin{aligned}
\mathrm{E}[\Delta^{\mathrm{D}}] &= \oint_{\mathbb{R}^{N_{\mathrm{bkg}}}} p(\vec{D})\ \sigma^{\mathrm{spur}}(\vec{D})\ d\vec{D} \\
&= \oint_{\mathbb{R}^{N_{\mathrm{bkg}}}} \left(\oint_{\mathbb{R}^{N_{\mathrm{syst}}}} p(\vec{\theta})\ \delta(g(\vec{\theta}) - \vec{D})\ d\vec{\theta} \right)\ \sigma^{\mathrm{spur}}(\vec{D})\ d\vec{D} \\
&= \oint_{\mathbb{R}^{N_{\mathrm{syst}}}} p(\vec{\theta})\ \sigma^{\mathrm{spur}}(\vec{\theta})\ d\vec{\theta} \\
\mathrm{E}[\Delta^{\mathrm{D}}] &= \mathrm{E}[\Delta^{\theta}]
\end{aligned}
\tag{5}
$$

The value of $\mathrm{var}[\Delta^D]$ can be replaced with $\mathrm{var}[\Delta^{\theta}]$ using similar substitutions from Equation 5:

$$
\begin{aligned}
\mathrm{var}[\Delta^D] &= \sqrt{\mathrm{E}[(\Delta^D)^2] - (\mathrm{E}[\Delta^D])^2} \\
&= \sqrt{\mathrm{E}[(\Delta^{\theta})^2] - (\mathrm{E}[\Delta^{\theta}])^2} \\
\mathrm{var}[\Delta^D] &= \mathrm{var}[\Delta^{\theta}]
\end{aligned}
\tag{6}
$$

Based on Equations 5 and 6, we will measure $\mathrm{E}[\Delta^{\theta}]$ and $\mathrm{var}[\Delta^{\theta}]$ as an estimate of $\sigma^{\mathrm{spur}}(\vec{D}^{\mathrm{truth}})$.

Implementation

Our goal is to compute $\mathrm{E}[\Delta^{\theta}]$ and $\mathrm{var}[\Delta^{\theta}]$. Since Δ^{θ} is weighted by a probability $p(\vec{\theta})$, we take the straight forward approach of obtaining the distribution of Δ^{θ} via a Monte-Carlo random sampling of the possible values of $\vec{\theta} \in \mathbb{R}^{N_{\mathrm{syst}}}$. We choose values for $\vec{\theta}$ selected randomly based on the priors $G(\theta_i)$ for each nuisance parameter θ_i. Each choice of $\vec{\theta}$ corresponds to a "toy" dataset $\vec{D}$, where the probability of being selected corresponds to $p(\vec{\theta})$.

For each toy $\vec{\theta}_j$, we calculate $\sigma^{\mathrm{spur}}(\vec{\theta}_j)$. The distribution of $\sigma^{\mathrm{spur}}(\vec{\theta}_j)$ toy measurements, in the limit of many toys, approaches the distribution of Δ^{θ}. The enumerated procedure to build a distribution of $\sigma^{\mathrm{spur}}(\vec{\theta}_j)$ is as follows:

1. For each θ_i a random value is sampled from the PDF $G(\theta_i)$. This selects a point $\vec{\theta}_j$.

2. Using the information from $\vec{D}^{\mathrm{MC}}$ and $\{\vec{D}^{\mathrm{syst}}\}$, $\vec{\theta}_j$ determines a toy background shape $\vec{D}_j$. For example,

$$\vec{D}_j = \vec{D}^{\mathrm{MC}} + \sum_{i=0}^{N_{\mathrm{syst}}} (\vec{D}_i^{\mathrm{syst}} - \vec{D}^{\mathrm{MC}}) \times (\vec{\theta}_j)_i. \tag{7}$$

3. Compute $\sigma^{\mathrm{spur}}(\vec{D}_j)$, the spurious signal measured on the toy dataset $\vec{D}_j$.

Next, we measure the mean and standard deviation of the toy measurements, as an approximation of $\mathrm{E}[\Delta^{\theta}]$ and $\mathrm{var}[\Delta^{\theta}]$ respectively. We then use these to approximate $\sigma^{\mathrm{spur}}(\vec{D}^{\mathrm{truth}})$ with a corresponding uncertainty. This allows us to estimate our background from the data using $F(\vec{D}^{\mathrm{data}})$, and with this use a σ^{spur} that we believe corresponds to the underlying PDF that generated $\vec{D}^{\mathrm{data}}$.

Caveats

The following is a brief discussion about some of the complexities that arise when implementing this procedure.

The 1σ approximation

In this procedure, we represent our systematic uncertainties as prior PDFs. Ideally, these PDFs would be measured for many different confidence intervals. In practice, these are often measured as the $\pm 1\sigma$ intervals in which the systematic variation is expected. We then make the assumption that the $\pm n\sigma$ interval is n-times the size of the $\pm 1\sigma$ interval. In doing this, we are making the assumption that the shape of the prior PDF is Gaussian. In some cases this assumption may be unreasonable; for example if the -3σ p_T variation is negative.

Asymmetric systematics

Equation 7 shows one form of $g(\vec{\theta}|\{\vec{D}^{\mathrm{syst}}\})$, a map from $\vec{\theta}$ to a background shape. In practice, a more nuanced version should be used because systematics are measured in different ways. A subset of systematics with indices $\mathbb{A} \in \{1, 2, ..., N_{\mathrm{syst}}\}$ may be measured separately to produce upper, $\vec{D}_i^{+\mathrm{syst}}$, and lower, $\vec{D}_i^{-\mathrm{syst}}$, uncertainty bands. We still consider these to be described by one nuisance parameter θ_i with a Standard Gaussian prior:

$$p(\theta_i) = G(\theta_i).$$

To handle the asymmetry of $\vec{D}_i^{+\mathrm{syst}}$ and $\vec{D}_i^{-\mathrm{syst}}$, we modify the impact of θ_i on each toy (j) template $\vec{D}_j$. We are sculpting the impact of θ_i on the toy shape in order to preserve the symmetrical Gaussian prior. Equation 7 is expanded to treat symmetric and asymmetric systematics separately:

$$\vec{D}_j = \vec{D}^{\mathrm{MC}} + \sum_{i \in \mathbb{B}} (\vec{D}_i^{+\mathrm{syst}} - \vec{D}^{\mathrm{MC}}) \times (\vec{\theta}_j)_i + \sum_{i \in \mathbb{C}} (\vec{D}_i^{-\mathrm{syst}} - \vec{D}^{\mathrm{MC}}) \times (\vec{\theta}_j)_i + \sum_{i \in \mathbb{D}} (\vec{D}_i^{\mathrm{syst}} - \vec{D}^{\mathrm{MC}}) \times (\vec{\theta}_j)_i \quad (8)$$

$$\mathbb{B} = \{i \in \mathbb{A} | (\vec{\theta}_j)_i > 0\}$$
$$\mathbb{C} = \{i \in \mathbb{A} | (\vec{\theta}_j)_i < 0\}.$$
$$\mathbb{D} = \{i \notin \mathbb{A}\}$$

Here, symmetric systematics are treated as before, while asymmetric systematics use the uncertainty band corresponding to whether $(\vec{\theta}_j)_i$ is positive or negative. This is straightforwardly implemented with conditionals when building the toy template $\vec{D}_j$.

Envelope uncertainties

In the Non-Resonant search, the many dominant systematics (PDC_Choice, Bad-Muon-Veto, etc.) are measured as *envelopes*. These are statements that $\vec{D}^{\mathrm{truth}}$ falls within a particular range. This formulation of systematic uncertainty does not lend itself easily to a simple description of a prior on a corresponding nuisance parameter. One choice of the prior G' be a flat step function:

$$G'(\theta_i) = \begin{cases} 1/2 & \text{if } |\theta_i| \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

This choice does not reflect our belief that the nominal $\vec{D}^{\mathrm{MC}}$ is more likely than the extreme bounds of the envelope. Another choice of the prior would be a Gaussian, such as is appropriate for Gaussian distributed nuisance parameters. This leads to an enlarged impact from the envelope systematics, since now large deviations corresponding to $\theta_i > 1$ are allowed. This is in conflict with the statement of the envelope, that such shapes are excluded (though to what level is not clear). The third option is a compromise is a Gaussian prior restricted to the confines of the envelope:

$$G'(\theta_i) = \begin{cases} \frac{G(\theta_i)}{\mathrm{erf}(\sqrt{0.5})} & \text{if } |\theta_i| \leq 1 \\ 0 & \text{otherwise} \end{cases}.$$

While this choice of PDF is not ideal (as mentioned above, we would prefer to measure the PDF at a number of intervals when assessing the impact of the systematic), it is a compromise between our physical intuitions and the requirement that the allowed variation fall within the measured envelope.