The General Idea

Let $\vec{\mu}(\vec{\lambda}) = \{\mu_1(\vec{\lambda}), \mu_2(\vec{\lambda}), \dots \mu_\iota(\vec{\lambda}) \dots, \mu_k(\vec{\lambda})\}$ is the prediction of the mean number of events, which are expected to be observed in k bins of the relevant kinematical variables. Obviously this defines a histogram (of any number of dimensions) , e.g the theoretical expectations for k regions of the relevant kinematical variables (let us say: of energy, cosθ and L) will be a 3-d histogram of mxpxq = k bins. The predictions depend on several (e.g. ρ) parameter values, which are symbolized as $\vec{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_\rho\}$.

Let us assume that we want to study the sensitivity of estimating the values of $\vec{\lambda}$ in the case that the true values of the parameters are $\vec{\lambda} = \overrightarrow{\lambda^t}$.

1. Produce a histogram $\vec{n}(\vec{\lambda}) = \left\{n_1\left(\overrightarrow{\lambda^t}\right), n_2\left(\overrightarrow{\lambda^t}\right), \dots n_\iota\left(\overrightarrow{\lambda^t}\right) \dots, n_k\left(\overrightarrow{\lambda^t}\right)\right\}$ such as that each n_i has been chosen according to a Poissonian distribution with mean value $\mu_\iota\left(\overrightarrow{\lambda^t}\right)\left(i.e. \quad P(n_i) = \frac{\mu_i^{n_i} e^{-\mu_\iota}}{n_i!}\right)$. Hereafter, we call the $\vec{n}(\vec{\lambda})\ vector\ as$ "the pseudo-data".
   I. Form the likelihood for the pseudo-data as a function of the parameter values: $\mathcal{L}(\vec{\lambda}) = \prod_{i=1}^{k} \frac{\mu(\vec{\lambda})_i^{n(\overrightarrow{\lambda^t})_i} e^{-\mu(\vec{\lambda})_i}}{n(\overrightarrow{\lambda^t})_i!}$ (1)

      Note that the expected values (predictions) are functions of the parameter values whilst the pseudo-data, number of events, have been chosen for $\vec{\lambda} = \overrightarrow{\lambda^t}$.
   II. Minimize the function $Q(\vec{\lambda}) = -ln\left(\mathcal{L}(\vec{\lambda})\right)$ with respect to $\{\lambda_1, \lambda_2, \dots, \lambda_\rho\}$ and estimate the Q_{min}, $\hat{\vec{\lambda}} = \{\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_\rho\}$ and the corresponding covariant matrix
2. Repeat, many times, the step "1" by producing new pseudodata corresponding to the same value of $\vec{\lambda} = \overrightarrow{\lambda^t}$ and, each time, estimate the parameter values by minimizing eq. (1), as well as Q_{min} and the corresponding covariant matrix.
3. After step "2", we have several estimations, $\hat{\vec{\lambda}}_j = \{\lambda_1, \hat{\lambda}_2, \dots, \hat{\lambda}_\rho\}_j$, j=1,2,3,…,N_{exp}, as well as Q^j_{min} and the corresponding covariant matrices (V^j) each for every "pseudo-experiment" (i.e. set of pseudo-data).
4. The histogram (ρ-dimensions) of the $\hat{\vec{\lambda}}_j = \{\lambda_1, \hat{\lambda}_2, \dots, \hat{\lambda}_\rho\}_j$, j=1,2,3,…,N_{exp}, expresses the distribution of the estimated parameters (in principle are correlated, i.e the covariant matrix has non-zero the non-diagonal. elements). From these pseudo-experiments, we can set unbias and consistent limits and errors. Also we can make pool-distributions to verify that each of the fits described in 1-II are unbias.

Including the Background Contribution.

In the case that the expected number of events per bin comprises signal and background, we should have, next to the signal prediction $\vec{s}(\vec{\lambda}) = \{s_1(\vec{\lambda}), s_2(\vec{\lambda}), \dots s_\iota(\vec{\lambda}) \dots, s_k(\vec{\lambda})\}$, and

another prediction vector $\vec{b} = \{b_1, b_2, \dots b_\iota \dots, b_k\}$ to account for the background contribution. We choose the pseudo-data for each bin by selecting, according to Poissanian distributions:

a) an integer, $ns_\iota\left(\overrightarrow{\lambda^t}\right)$, for the signal (with mean value equal to $s_\iota\left(\overrightarrow{\lambda^t}\right)$) and

b) another integer, nb_ι, for the background (with mean value equal to b_ι) .

Then, the pseudo-data contain in the I[th] bin, $n_\iota\left(\overrightarrow{\lambda^t}\right) = ns_\iota\left(\overrightarrow{\lambda^t}\right) + nb_\iota$, events. Note that the background is independent of the parameter values but participates in all the minimizations because $\vec{\mu}(\vec{\lambda}) = \vec{s}(\vec{\lambda}) + \vec{b}$. We use eq. 1 for the likelihood function but we include background to both the predictions and the pseudo-data, as explained above.

Taking into account priors

Assuming that previous experiments or other theories provide some prior knowledge on the parameter values, we can modify the likelihood function definition in order to incorporate such a knowledge into the estimator. We start with the Bayesian statistics.

$$P(\vec{\lambda}|\vec{n}) = \frac{P(\vec{n}|\vec{\lambda})P(\vec{\lambda})}{P(\vec{n})} \qquad (2)$$

Where

$P(\vec{\lambda}|\vec{n})$: is the probability that the parameter values are $\vec{\lambda}$ given that the data are $\vec{n}$. This is what we want to find from the data and our statistical model.

$P(\vec{n}|\vec{\lambda})$: is the probability that the data are $\vec{n}$, given that the parameter values are $\vec{\lambda}$. This is the likelihood function of eq. 1

$P(\vec{\lambda})$: is the probability that the parameter values are $\vec{\lambda}$. This is the prior knowledge which can be expressed with the function, $f(\lambda_1, \lambda_2, \dots, \lambda_\rho)d\lambda_1 d\lambda_2 \dots, d\lambda_\rho$

$P(\vec{n})$: is the probability to observe the data for any value of the parameters. In principle $P(\vec{n}) = \left[\int \mathcal{L}(\vec{\lambda}) f(\lambda_1, \lambda_2, \dots, \lambda_\rho)d\lambda_1 d\lambda_2 \dots, d\lambda_\rho\right] dn_1 dn_2 \dots, dn_k$. But $P(\vec{n})$ is independent of the λ's, it is a common numerical factor for all the likelihood values and obviously does not affect the minimum of eq. 2.

We minimize the function $\mathcal{L}(\vec{\lambda}) f(\lambda_1, \lambda_2, \dots, \lambda_\rho) = \prod_{i=1}^{k} \frac{\mu(\vec{\lambda})_i^{n(\overrightarrow{\lambda^t})_i} e^{-\mu(\vec{\lambda})_i}}{n(\overrightarrow{\lambda^t})_i!} f(\lambda_1, \lambda_2, \dots, \lambda_\rho)$

Taking into Account Systematic Errors

Let us assume that there are several other parameters (e.g. energy scale, cross section normalization, etc) which affect the predictions and let us say that we know their joint probability function, $g(p_1, p_2, \dots, p_m)$. As an example, it can be a m-dimensional Gaussian with V_g covariant matrix and $\overline{p_1}, \overline{p_2}, \dots, \overline{p_m}$ the mean values. Then we write the likelihood with systematics as

$$L_{syst}(\lambda_1, \lambda_2, \dots, \lambda_\rho; p_1, p_2, \dots, p_m) = [\mathcal{L}(\vec{\lambda}, p_1, p_2, \dots, p_m) f(\lambda_1, \lambda_2, \dots, \lambda_\rho)] g(p_1, p_2, \dots, p_m) \quad (3)$$

Where

$$\mathcal{L}(\vec{\lambda}, p_1, p_2, \dots, p_m) = \prod_{i=1}^{k} \frac{\mu(\vec{\lambda}, p_1, p_2, \dots, p_m)_i^{n(\overrightarrow{\lambda^t})_i} e^{-\mu(\vec{\lambda} p_1, p_2, \dots, p_m,)_i}}{n\left(\overrightarrow{\lambda^t}\right)_i !}$$

We can then minimize for ρ+m parameters.

Another way is to write the likelihood as

$$L_{syst}(\lambda_1, \lambda_2, \dots, \lambda_\rho) = \int [\mathcal{L}(\vec{\lambda} p_1, p_2, \dots, p_m) f(\lambda_1, \lambda_2, \dots, \lambda_\rho)] g(p_1, p_2, \dots, p_m) \, dp_1 dp_2, \dots dp_m$$

And minimize only for the λ's. Unfortunately, the integration, for most of the cases, is difficult to be performed analytically and we must use numerical or statistical integration.