Draft v0.1
29 January, 2006

LIKELIHOOD IMPLEMENTATION IN TOY MC

1 Likelihood Definitions

The likelihood is defined as the product of event PDF's.

$$\mathcal{L} = \prod_{e=evts} \mathcal{P}(\vec{x_e}\,\vec{p}) \tag{1}$$

where $\vec{x_e}, \vec{p}$ are a vectors of event observables and likelihood function parameters respectively. The total PDF for an event is the sum of PDFs for each possible source contributing to the data sample weighted by their fractions, f_a in the samples:

$$\mathcal{P}(\vec{x}\,;\,\vec{p}) = \sum_{a=src} f_a \mathcal{P}_a(\vec{x_a}\,;\,\vec{p}) \tag{2}$$

As we will see below, the values of the event observables may depend on the assumed source.

Since there are several observables for each event, which may be correlated, the PDF ($\mathcal{P}$) is given above is the *joint* PDF of all the variables. The *Probability* section of the PDG Review [1] describes how to relate this joint PDF to the conditional PDF that is easy to implement mathematically. In the following we will assume that the first element of the observables vector, $\vec{x}$, corresponds to the "main" variable sensitive to mixing – *e.g.* the VPDL or L_{xy}. The joint PDF can then be written as:

$$\mathcal{P}_a(x_1, x_2, \cdots, x_n) = \mathcal{P}_a(x_1 \mid x_2, \cdots, x_n)\mathcal{P}_a(x_2)\cdots\mathcal{P}_a(x_n) \tag{3}$$

where

- $\mathcal{P}_a(x_1, x_2, \cdots, x_n)$ is the probability of *jointly* observing $x_1, x_2, \cdots, x_n$ in an event for source a.
  Normalized such that $\int dx_1 dx_2 \cdots dx_n\, \mathcal{P}_a(x_1, x_2, \cdots, x_n) = 1$
- $\mathcal{P}_a(x_1 \mid x_2, \cdots, x_n)$ is the probability of observing x_1 *given* $x_2, \cdots, x_n$ for source a.
  Normalized such that $\int dx_1\, \mathcal{P}_a(x_1 \mid x_2, \cdots, x_n) = 1$
- $\mathcal{P}_a(x_i)$ is the probability density of variable x_i for source a.
  Normalized such that $\int dx_i\, \mathcal{P}_a(x_i) = 1$

In the above we have assumed that the "nuisance" variables $x_2, \cdots, x_n$ are uncorrelated, howver, this can be avoided by continuing to construct conditional PDFs for the variables that are left.

2 Fit Results Histograms

To check the quality of the fit, it is useful to compare the distributions of data and the prediction of the fit vs. the main variable for samples tagged as "unmixed" and "mixed". This is a far more sensitive indicator of problems with the fit than the result for Δm_s.

To construct these histograms, the unmixed and mixed PDFs, at the best-fit parameters, $\vec{p}_0$, must be integrated over the secondary variables.

$$\begin{aligned} \mathcal{P}^{tag}(x_1\,;\,\vec{p}_0) &= \int dx_2 \cdots dx_n\, \mathcal{P}^{tag}(\vec{x}\,;\,\vec{p}) \\ &= \sum_a f_a \int dx_2 \cdots dx_n\, \mathcal{P}_a^{tag}(x_1 \mid x_2, \cdots, x_n\,;\,\vec{p}_0) \\ &\quad \times \mathcal{P}_a(x_2) \cdots \mathcal{P}_a(x_n) \end{aligned} \tag{4}$$

Integrals are approximated by sums in the code.

The histogram, $\mathcal{H}^{tag}(x_1)$, is then

$$\mathcal{H}^{tag}(x_1) = \int_{bin} dx_1 \mathcal{P}^{tag}(x_1\,;\,\vec{p}_0) \tag{5}$$

which can be approximated by the value at the bin center for small enough bins.

3 Variables Used in the Fit

Various observable sets can be used in the fit. The two sets implemented so far are the following.

3.1 Decay Length Based

Main Variable:

- L_{xy}: the 2D distance from the primary to the secondary vertex.

Secondary Variables:

- P_t^m: the measured $(D_s + \ell)$ transverse momentum in the event.
- σ_L: the measured error on L_{xy}.
- T: the mixing tag – +1=unmixed, -1=mixed
(note: in the code the convention 0=unmixed, 1=mixed is often used)
- $\mathcal{D}$: the dilution

3.2 VPDL Based

Main Variable:

- X_a: the VPDL (depends on the source assumed – see below)

Secondary Variables:

- σ_{xa}: the measured error on X_a.
- T: the mixing tag – +1=unmixed, -1=mixed
(note: in the code the convention 0=unmixed, 1=mixed is often used)
- $\mathcal{D}$: the dilution

3.3 Defining the VPDL Correctly

A generic VPDL and its error can be defined for any event as:

$$\begin{aligned} X &= M(B_s)\frac{\vec{L}_{xy}\cdot\vec{P}_t^m}{P_t^{m2}} \sim M(B_s)\frac{L_{xy}}{P_t^m} \\ \sigma_x &\sim M(B_s)\frac{\sigma_L}{P_t^m} \end{aligned} \tag{6}$$

However, this does not take into account the different masses of B-mesons. When used as a variable in the PDF for source a, the vpdl should be modified to:

$$\begin{aligned} X_a &= X\frac{M(B_a)}{M(B_s)} \\ \sigma_{xa} &= \sigma_x\frac{M(B_a)}{M(B_s)} \end{aligned} \tag{7}$$

unless a mass is not defined for this source ($c\bar{c}$, combinatoric background, etc.) in which case the original VPDL variables are used.

4 Likelihood Implementation

The likelihood as implemented in the code consists of several components.

1. Convolution over the K-factor.
$$K = \frac{P_t^m}{P_t(B_s)} \tag{8}$$
Note: K=1 for $c\bar{c}$, combinatoric background, etc.
2. Convolution over the resolution function, $\mathcal{R}(x^t - x^m;\sigma_x)$, assumed to be a gaussian with width given by the event error, where x^t, m are the true and measured values of the VPDL or L_{xy}.
3. Convolution over the efficiency function, ε, as a function of x^m.
4. The efficiency function normalization, $A(\vec{p}, K, x_2, \cdots, x_n)$, which depends on the values of the fit parameters. It is defined by the condition:
$$\begin{aligned} A_a(\vec{p}, K, x_2, \cdots, x_n) &= \int dx^m \varepsilon_a(x^m) \int dx^t \mathcal{R}(x^t - x^m;\sigma_x) \\ &\times [f_a^{unm}(x^t, \cdots) + f_a^{mix}(x^t, \cdots)] \end{aligned} \tag{9}$$
since the sum of the unmixed and mixed physics functions must be normalized to unity.

5. The physics function, $f_a^{tag}(x^t \mid x_2, \cdots, x_n\,;\,\vec{p})$

6. Histograms for the secondary variables representing their PDFs – $\mathcal{H}_a(v) \sim \mathcal{P}_a(v)$. These are normalized such that: $\sum_{i=bins} \mathcal{H}_a(v_i) = 1$

Convolution integrals are approximated as sums in the code.

4.1 Decay Length Based PDF

The joint PDF for each source using the decay length based variable set is given by:

$$\mathcal{P}_a(\vec{x}\,;\,\vec{p}) = \mathcal{P}_z(L_{xy} \mid P_t^m, \sigma_L, T, \mathcal{D}\,;\,\vec{p})\mathcal{H}_a(P_t^m)\;(\mathcal{H}(\sigma_L)) \qquad (10)$$

Note that:

- The $\mathcal{H}_a(T)$'s will be absorbed in the physics functions as constants: *e.g.* the factor of 1/2 in the unm/mix functions.
- $\mathcal{H}_a(\mathcal{D})$ is uniform in the range 0–1 for all sources, so it is does not affect the likelihood and is excluded, both in fits and histogram creation. *Note*: if this is not true, then some way of estimating the dilution PDFs for each source must be found.
- $\mathcal{H}(\sigma_L)$, the distribution of decay length errors is assumed to be independent of source. It thus factors out of the sum of the individual source PDFs and does not affect the fit. It is only included when plots are made of the predicted number of events vs. L_{xy} based on the fit results.

The conditional PDF is then:

$$\begin{aligned}\mathcal{P}_a(L_{xy}^m \mid P_t^m, \sigma_L, T, \mathcal{D}\,;\,\vec{p}) &= \sum_k \Delta K \mathcal{H}_a(K_k)\;\sum_i \Delta L^t \mathcal{R}(L_i^t - L^m; \sigma_L) \\ &\times \frac{\varepsilon_a(L^m)}{A_a(\vec{p}, K, \cdots)} f_a^T(L^t \mid K, P_t^m, \sigma_L, \mathcal{D}\,;\,\vec{p}) \qquad (11)\end{aligned}$$

4.2 VPDL Based PDF

The joint PDF for each source using the VPDL based variable set is given by:

$$\mathcal{P}_a(\vec{x}_a\,;\,\vec{p}) = \mathcal{P}_z(X_a \mid \sigma_{xa}, T, \mathcal{D}\,;\,\vec{p})\mathcal{H}_a(\sigma_{xa})) \qquad (12)$$

Note that:

- $\mathcal{H}_a(\sigma_{xa})$, the distribution of VPDL errors for source a, is required to be included in the likelihood definition used in the fit since the width of each source's gaussian distribution will be different.

The conditional PDF is then:

$$\begin{aligned}\mathcal{P}_a(X_a^m \mid \sigma_{xa}, T, \mathcal{D}\,;\,\vec{p}) &= \sum_k \Delta K \mathcal{H}_a(K_k)\;\sum_i \Delta X_a^t \mathcal{R}(X_{ai}^t - X_a^m; \sigma_{xa}) \\ &\times \frac{\varepsilon_a(X_a^m)}{A_a(\vec{p}, K, \cdots)} f_a^T(X_a^t \mid K, \sigma_{xa}, \mathcal{D}\,;\,\vec{p}) \qquad (13)\end{aligned}$$

4.3 PDFs for "Zero"-Lifetime Backgrounds

The PDFs for backgrounds arising from sources with true decay lengths that are small compared to the resolution – $c\bar{c}$, combinatoric backgrounds, etc. – are calculated differently than those for other sources. For these backgrounds, a K-factor cannot be computed, because there is no true $P_t(B)$ or $M(B)$. Thus the following definitions are used:

$$\begin{aligned} K &= 1 \\ X_{cc} &= M(B_s)\frac{L_{xy}}{P_t^m} \\ \sigma_{xcc} &= M(B_s)\frac{\sigma_L}{P_t^m} \end{aligned} \tag{14}$$

Convolution over the K-factor is not necessary and, because, the physics function is approximately a delta-function, $\delta(x^t = 0)$, convolution over the resolution function can also be avoided. (Note: even if this is not the case, the physics function for the 0-lifetime case can be constructed to include effects of resolution.) With the 0-lifetime physics function re-defined to include effects of resolution, the PDF for this source can be written as:

Decay Length Based

$$\mathcal{P}_{cc}(L_{xy}^m \mid P_t^m, \sigma_L, T, \mathcal{D}\ ;\ \vec{p}) = \frac{\varepsilon_{cc}(L^m)}{A_{cc}(\vec{p}, K, \cdots)} f_{cc}^T(L^m \mid P_t^m, \sigma_L, \mathcal{D}\ ;\ \vec{p}) \tag{15}$$

VPDL Based

$$\mathcal{P}_{cc}(X^m \mid \sigma_x, T, \mathcal{D}\ ;\ \vec{p}) = \frac{\varepsilon_{cc}(X^m)}{A_{cc}(\vec{p}, K, \cdots)} f_{cc}^T(X^m \mid \sigma_x, \mathcal{D}\ ;\ \vec{p}) \tag{16}$$

5 Physics Functions

The physics functions used in the code for various source classes are defined below. All use proper time, t_0, defined as:

$$\begin{aligned} ct_0 &= M(B_a)\frac{L_{xy}}{P_t^m}K \qquad \text{(decay length based)} \\ &= X_a K \qquad \text{(VPDL based)} \end{aligned} \tag{17}$$

The functions are also labeled by their tag, T:

$$\begin{aligned} T &= +1 \qquad \text{unmixed (unm)} \\ &= -1 \qquad \text{mixed (mix)} \end{aligned} \tag{18}$$

They are normalized such that:

$$\int dx_1[f_a^{unm}(x_1 \mid x_2, \cdots, x_n\ ;\ \vec{p}) + f_a^{mix}(x_1 \mid x_2, \cdots, x_n\ ;\ \vec{p})] = 1 \tag{19}$$

5.1 Mixing Sources

$$
\begin{aligned}
f_a^T &= \frac{N}{2}\exp(-t_0/\tau_a)\,[1 + T\mathcal{D}\cos(\Delta m_a t_0)] \\
N &= \frac{M_a K}{c\tau_a P_t^m} \quad \text{(decay length)} \\
N &= \frac{K}{c\tau_a} \quad \text{(VPDL)}
\end{aligned} \tag{20}
$$

Here we are using $\mathcal{P}(T = mix) = \mathcal{P}(T = unm) = 1/2$.

5.2 Long-Lifetime (non-mixing) Sources

$$
\begin{aligned}
f_a^T &= N\frac{[1 + T\mathcal{D}(1 - 2P_{mix})]}{2}\exp(-t_0/\tau_a) \\
N &= \frac{M_a K}{c\tau_a P_t^m} \quad \text{(decay length)} \\
N &= \frac{K}{c\tau_a} \quad \text{(VPDL)}
\end{aligned} \tag{21}
$$

In this case we usually have $\mathcal{P}(T = unm) = 1$ and $\mathcal{P}(T = mix) = 0$. Then, the probability that an event shows up in the mixed sample is just the mis-tag probability: $(1 - \mathcal{D})/2$.

5.3 Zero-Lifetime (non-mixing) Sources

$$
\begin{aligned}
f_{cc}^T &= \frac{[1 + T\mathcal{D}(1 - 2P_{mix})]}{2\sqrt{2\pi}\sigma_L}\exp(-\frac{L_{xy}^2}{2\sigma_L^2}) \quad \text{(decay length)} \\
&= \frac{[1 + T\mathcal{D}(1 - 2P_{mix})]}{2\sqrt{2\pi}\sigma_L}\exp(-\frac{X^2}{2\sigma_X^2}) \quad \text{(VPDL)}
\end{aligned} \tag{22}
$$

where $P_{mix} \equiv \mathcal{P}(T = mix)$ is the probability of the event to appear in the mixed sample: usually set to 0.5. Note also that the dilution cancels in these functions if $P_{mix} = 0.5$

References

[1] http://pdg.lbl.gov