Further analysis topics with maximum likelihood

Derek Glazier
University of Glasgow

Covered here

- Maximum Likelihood
- Acceptance Corrections
- Background Subtractions
- Bootstrap

Probability Distribution Functions

- The Probability Distribution Function for some model f depending on observables $x_i=\{x_0,x_1,...\}$ and parameters $\theta_j=\{\theta_0,\theta_1,...\}$:

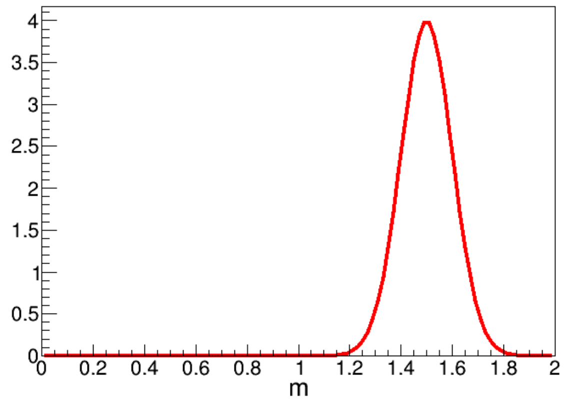
$$p(x_i:\theta_j) = \frac{f(x_i:\theta_j)}{\int f(x_i:\theta_j)dx}$$

i.e. its value over its integral

- Examples,

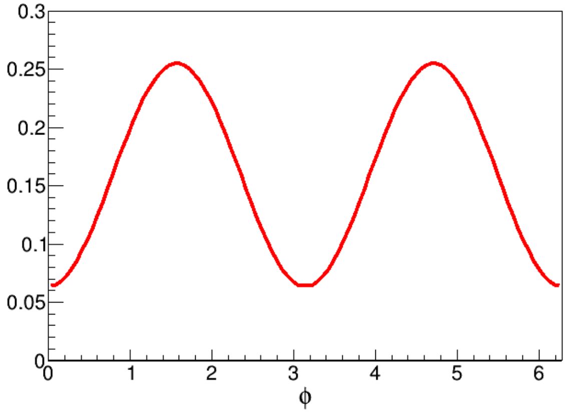
$$p(m:\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(m-\mu)^2}{2\sigma^2}\right)$$

Where $x_i = \{m\}$ and $\theta_i = \{\mu,\sigma\}$



$$p(\phi,P_\gamma:\Sigma) = \frac{1}{2\pi}(1-P_\gamma\Sigma\cos(2\phi))$$

Where $x_i = \{\phi,P_\gamma\}$ and $\theta_i = \{\Sigma\}$



Likelihood

May generalise to histograms but here we just consider event based distributions (histograms lose information)

- Product of probabilities of each data event

$$L(\theta_j) = \prod_{k=0}^{N} p(x_{i,k} : \theta_j)$$

Minimise this

$$\mathcal{L}(\theta_j) = -\ln(L(\theta_j)) = -\sum_{k=0}^{N} \ln[p(x_{i,k} : \theta_j)]$$

Better to compute - probabilities may be very small

- Extended likelihood allows for uncertainty in normalisation

$$L(\theta_j, Y) = \prod_{k=0}^{N} p(x_{i,k} : \theta_j) e^{-Y} \frac{Y^N}{N!}$$

Observed events N is a random variable from Poisson distribution of mean Y

$$\mathcal{L}(\theta_j, Y) = -\ln(L(\theta_j, Y)) = -\sum_{k=0}^{N} \ln[\frac{f(x_{i,k} : \theta_j)}{\int f(x_i : \theta_j) dx}] + Y - N \ln Y + \cancel{\ln N!}$$

Minimise this

$$\mathcal{L}(\theta_j, Y) = -\ln(L(\theta_j, Y)) = -\sum_{k=0}^{N} \ln \frac{f(x_{i,k} : \theta_j)}{\int f(x_i : \theta_j) dx} + Y - N \ln Y$$

May choose to cancel dashed cross terms in calculation

with $Y = \int f(x_i : \theta_j) dx_i$

Normalisation Integral

- To normalise our PDF we need to integrate the function over the observables x_i(e.g. mass, angles) at a fixed value for the parameters θ_j

$$\int f(x_i : \theta_j)dx$$

- This has to be recalculated everytime a parameter changes during minimisation

- This can be OK if analytical form is known

- If we need to numerically estimate it, this can be slower

- Estimation can be done via summing Monte-Carlo samples or a grid of values in low dimensions (e.g. a 1D histogram)

$$\int f(x_i : \theta_j)dx_i = \frac{V}{M}\sum_{s}^{M} f(x_{i,s} : \theta_j)$$

MC integration with M events
And volume V (product of x_i ranges)

Acceptance Corrections

- Include acceptance dependence in function

Physics model . Detector acceptance

$$f(x_i:\theta_j) = I(x_i:\theta_j).\eta(x_i)$$

- Function PDF can be approximated by

$$p(x_i:\theta_j) = \frac{I(x_i:\theta_j)\eta(x_i)}{\sum_s^M I(x_{i,s}:\theta_j)}$$

Normalisation Integral ~ Sum over *accepted* events from detector simulation Monte-Carlo.

Thus includes η

- Extended Likelihood is then

$$L(\theta_j,Y) = \prod_{k=0}^{N} p(x_{i,k}:\theta_j)e^{-Y}\frac{Y^N}{N!}$$

$$-\ln L(\theta_j,Y) = -\sum_{k=0}^{N}\ln[\frac{I(x_{i,k}:\theta_j)}{\sum_s^M I(x_{i,s}:\theta_j)}] + Y - N\ln Y - \cancel{\sum_k^N \ln\eta(x_{i,k})}$$

Sum over acceptance function does not depend on parameters θ_j

$$\mathcal{L}(\theta_j,Y) = -\ln L(\theta_j,Y) = -\sum_{k=0}^{N}\ln\frac{I(x_{i,k}:\theta_j)}{\sum_s^M I(x_{i,s}:\theta_j)} + Y - N\ln Y$$

Signal and Background

- Generally when fitting experimental data signal events will be contaminated by backgrounds

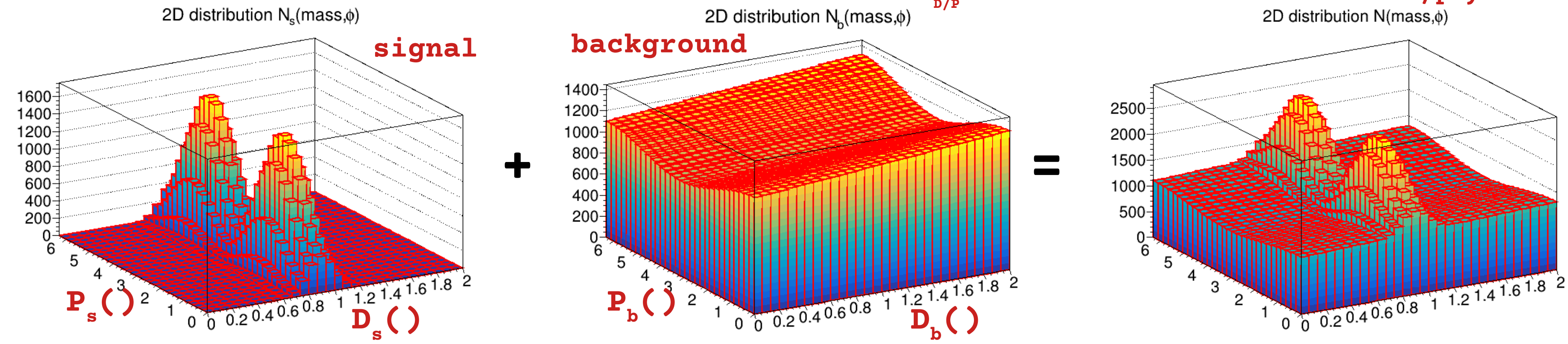
- Backgrounds will have different observable distributions than signal. They may not be fit-able with your signal model!

- We need additional discriminatory observables to disentangle signal

- If the distributions may be factorised, you can specify a joint function :

$$
\begin{aligned}
f\,(x_i : \theta_j) =& N_s.D_s(x_{Di} : \theta_{Di}).P_s(x_{Pi} : \theta_{Pi}) \\
&+ N_b.D_b(x_{Di} : \theta_{Di}).P_b(x_{Pi} : \theta_{Pi})
\end{aligned}
$$

$N_{s/B}$ = number signal/background events
$D_{s/B}$ = Sig/BG PDF dependent discriminating vars
$P_{s/B}$ = Sig/BG PDF dependent physics variables
$X_{D/P}$ = variables for discrimination/physics
$\theta_{D/P}$ = variables for discrimination/physics



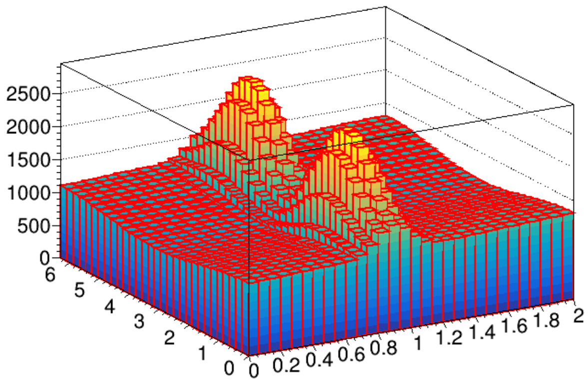
Example Signal and Background

D=Discriminatory P=Physics

- Full distribution :

$$N(m,\phi:\mu,\sigma,c_0,c_1,\Sigma,_b) = N_s.D_s(m:\mu,\sigma).P_s(\phi,P_\gamma:\Sigma) + N_b.D_b(m:c_0,c_1).P_b(\phi:A_B)$$

2D distribution N(mass,φ)



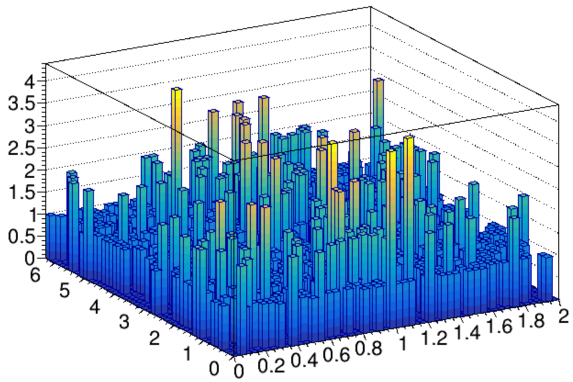
$$D_s(m:\mu,\sigma) = Gaussian(m:\mu,\sigma)$$
$$P_s(\phi,P_\gamma:\Sigma) = PhotonAsymmetry(\phi,P_\gamma:\Sigma)$$

Signal

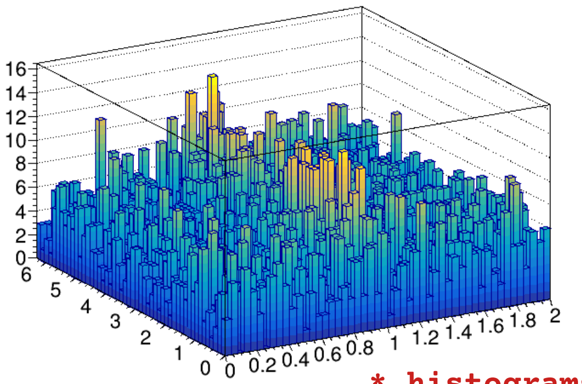
$$D_b(m:c_0,c_1) = \frac{c_0 + c_1.m}{\int(c_0+c_1.m)dm}$$
$$P_b(\phi:A_b) = \frac{1+A_b\cos(\phi)}{\int(1+A_b\cos(\phi))d\phi}$$

Background

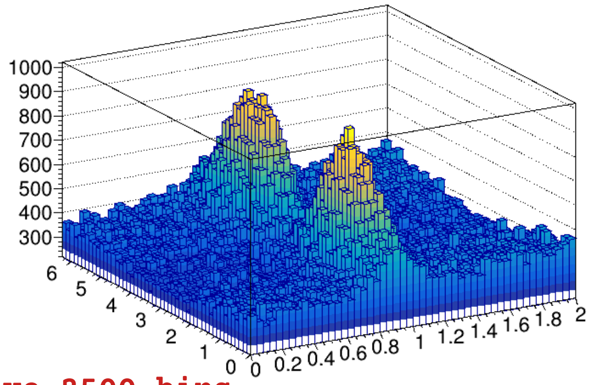
Sample of 1k events



Sample of 10k events



Sample of 1M events



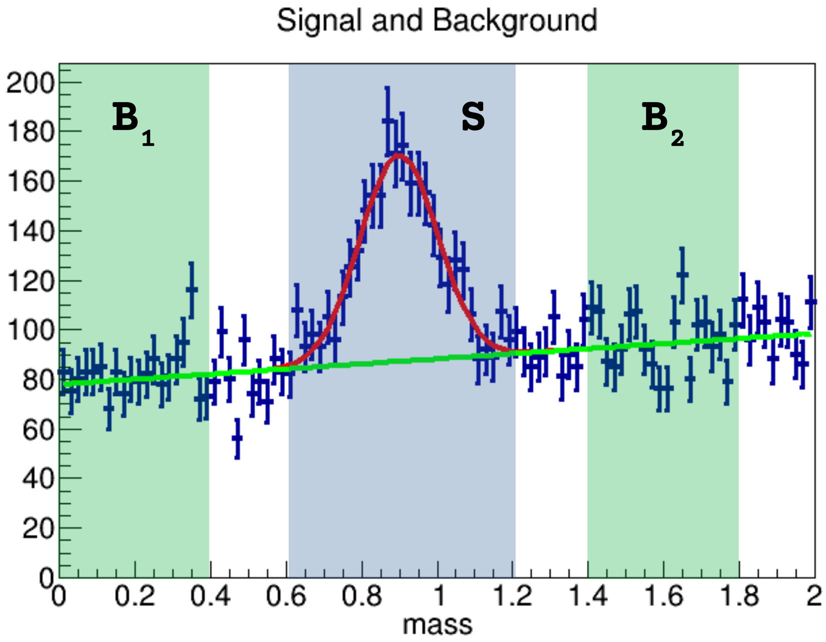
* histograms have 2500 bins

Background Subtraction : Sideband

- Sideband subtraction is useful when we can identify regions in our data of pure background, e.g. in discriminatory variable

- Then we can normalise these events to expected background in signal region and subtract via event(k) weights w_k

$$w_k = \begin{cases} 1 & \text{if } k \in S \\ \dfrac{-1 \int_{sgnl} D_b(x{:}\theta_j)dx}{\int_{side} D_b(x{:}\theta_j)dx} & \text{if } k \in B \\ 0 & \text{otherwise} \end{cases}$$



- In the case background PDF is flat the weight simplifies to the ratio of signal and background regions

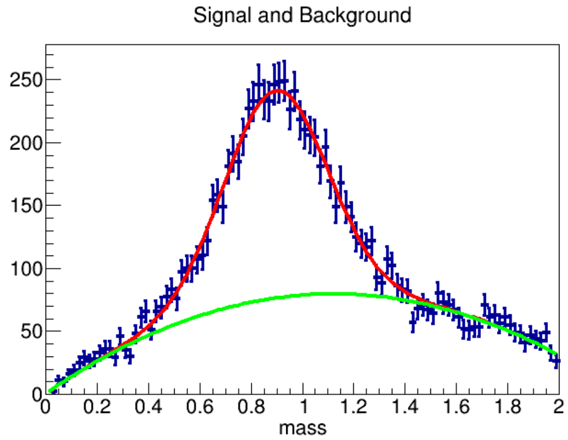
$$w_k = \frac{-1\Delta_{sgnl}}{\Delta_{side}} \text{ if } k \in B$$

- Weights can be used to fill histograms OR in likelihood event based fits

sPlots

- Can't always identify background only regions to subtract

- sPlots provide more general weights for disentangling signal and background "physics" distributions



- Can extract weights from Likelihood and combined PDF fitted to discriminatory observables in data for different event types t (e.g. signal,background1,...)

A preliminary minimisation Of * can be done to estimate θ_j^{best}

$$P_t(x_k) = p_t(x_{Di,k} : \theta_{Dj}^{best})$$

* $$\mathcal{L}(Y_t, \theta_j = \theta_j^{best}) = -\sum_k^{N_k} \ln\left[\sum_t^{N_t} Y_t.P_t(x_k)\right] + \sum_t^{N_t} Y_t$$

$$[V_{tt'}]^{-1} = \frac{\partial^2 \mathcal{L}(\theta_j^{best}, Y_t)}{\partial Y_t Y_{t'}} = \sum_k^{N_k} \frac{P_t(x_k) P_{t'}(x_k)}{\left(\sum_{t'}^{N_t} Y_{t'}.P_{t'}(x_k)\right)^2}$$

Calculated via fit routine or **Calculated from Sum over data**

$$w_{t,k} = \frac{\sum_{t'}^{N_t} V_{tt'}.P_{t'}(x_k)}{\sum_{t'}^{N_t} Y_{t'}.P_{t'}(x_k)}$$

Weights for event k to extract type t (e.g. t = signal)

Including Weights in Likelihood fits

- Given our event weights we now need to extract Physics parameters

- In principle this is straightforward :
  subtract the background contribution from the summation of log[PDF]

$$\mathcal{L}(\theta_j) = -\sum_k^{N_k} \boxed{w_k} \ln[p(x_{Pi,k} : \theta_{Pj})]$$

- Unfortunately our uncertainty estimates from the Likelihood are no longer valid

- An approximate correction for uncertainties may be applied using

$$\mathcal{L}(\theta_j) = -\boxed{\frac{\sum_k^{N_k} w_k}{\sum_k^{N_k} w_k^2}} \sum_k^{N_k} \boxed{w_k} \ln[p(x_{Pi,k} : \theta_{Pj})]$$

- However this is not always asymptotically correct (see references)

Bootstrapping Uncertainties

- Alternatively, one may try bootstrapping the data to generate distributions for our parameters

- These can give approximately the correct mean and variance for our parameters from the fit to the data

- Principle is to redo our fits many times (N_{boot})
  - For each fit we resample our data events from the original N_k
  - For each event we randomly choose an event from the original allowing for replacements,
    i.e. choose from the full N_k set for each event.
  - Now analyse the N_{boot} results for mean and variance of parameters
  - If multi-parameter fit we can produce covariances etc.

More complicated PDFs

- The beauty of ML is it naturally extends to problems with many observables and parameters
- Some Examples

K+Λ photoproduction

With recoil polarisations we have 7 parameters

and $\theta_{x,y,z}$ are the Λ decay angles

$$\begin{aligned} f(\Phi, \theta_x, \theta_y, \theta_z : P, \Sigma, T, O_x, O_z, C_x, C_z) \\ = \{1 + P\cos(\theta_y) \\ - P_\gamma \cos(2\Phi)(\Sigma + T\cos(\theta_y)) \\ - P_\gamma \sin(2\Phi)(O_x\cos(\theta_x) + O_z\cos(\theta_z)) \\ - P_\odot(C_x\cos(\theta_x) + C_z\cos(\theta_z))\} \end{aligned}$$

Two pion Partial Wave Analysis

$$\begin{aligned} f(\Omega, \Phi : \{[\ell]_m^{(\epsilon)}\}) = I^0(\Omega) - P_\gamma I^1(\Omega)\cos(2\Phi) \\ - P_\gamma I^2(\Omega)\sin(2\Phi), \end{aligned}$$

$$I^0(\Omega) = \sum_\epsilon \left\{|U^{(\epsilon)}(\Omega)|^2 + |\tilde{U}^{(\epsilon)}(\Omega)|^2\right\},$$

$$I^1(\Omega) = -2\sum_\epsilon \epsilon\Re\left\{U^{(\epsilon)}(\Omega)\left[\tilde{U}^{(\epsilon)}(\Omega)\right]^*\right\},$$

$$I^2(\Omega) = -2\sum_\epsilon \epsilon\Im\left\{U^{(\epsilon)}(\Omega)\left[\tilde{U}^{(\epsilon)}(\Omega)\right]^*\right\}.$$

Partial wave amplitudes **Spherical harmonics**

$$U^{(\epsilon)}(\Omega) = \sum_{\ell m}[\ell]_m^{(\epsilon)} Y_\ell^m(\Omega),$$

$$\tilde{U}^{(\epsilon)}(\Omega) = \sum_{\ell m}[\ell]_m^{(\epsilon)}\left[Y_\ell^m(\Omega)\right]^*.$$

Covered here

- Maximum Likelihood
- Acceptance Corrections
- Background Subtractions
- Bootstrap

Additional topics

- Toy MC
  - Validate your parameter extraction and Uncertainties via many simulated analysis
- Unfolding/deconvoluting resolutions
- Bayesian sampling techniques
  - e.g MCMC , nested sampling
  - Generate accurate samples of parameter posterior distributions
- Global versus local maxima

References

General Statistics

L. Lyons,
Statistics for Nuclear and Particle Physicists
Cambridge University Press, 1986.

Extended Maximum Likelihood

R. Barlow,
Nuclear Instruments and Methods A: 297, 496 (1990).

sPlots

M. Pivk and F. Le Diberder,
Nuclear Instruments and Methods A: 555, 356 (2005).

sPlots ++

H. Dembinski, M. Kenzie, C. Langenbruch, and M. Schmelling,
Nuclear Instruments and Methods A: 1040, 167270 (2022).

Asymptotically correct Uncertainties

C. Langenbruch,
The European Physical Journal C 82, 393 (2022).

Bootstrap

B. Efron, The Annals of Statistics 7, 1 (1979)