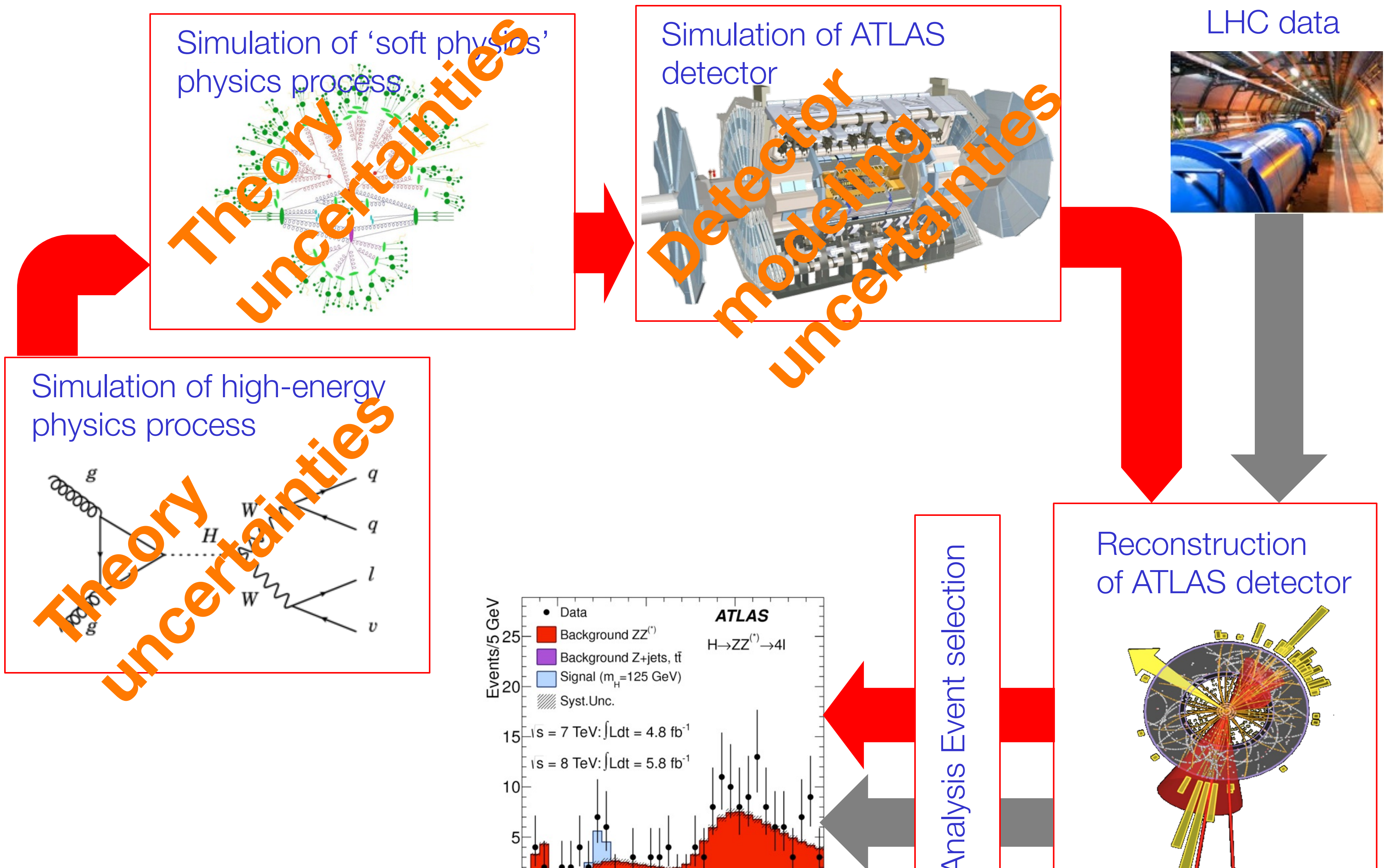
# Statistics

W. Verkerke

[Recap]

# The simulation workflow and origin of uncertainties

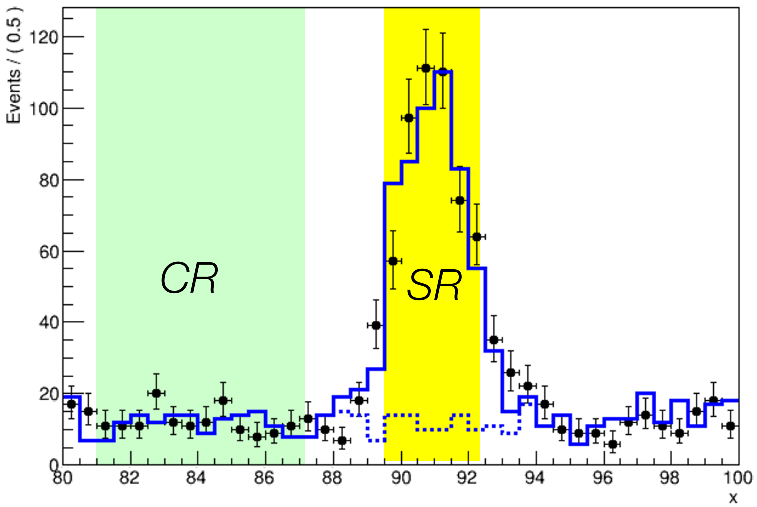




Wouter Verkerke, NIKHEF

# The sideband measurement

- Suppose your data in reality looks like this ➔




Can estimate level of background in the 'signal region' from event count in a 'control region' elsewhere in phase space

$$L_{SR}(s,b) = Poisson(N_{SR} \mid s+b)$$

$$L_{CR}(b) = Poisson(N_{CR} \mid \tilde{\tau} \cdot b)$$

NB: Define parameter 'b' to represents the amount of bkg is the SR.

Scale factor τ accounts for difference in size between SR and CR

*"Background uncertainty constrained from the data"*

- Full likelihood of the measurement ('simultaneous fit')

$$L_{full}(s,b) = Poisson(N_{SR} \mid s+b) \cdot Poisson(N_{CR} \mid \tilde{\tau} \cdot b)$$

[Recap]

# Generalizing the concept of the sideband measurement

- Background uncertainty from sideband clearly clearly not a 'systematic uncertainty'

$$L_{full}(s,b) = Poisson(N_{SR} | s+b) \cdot Poisson(N_{CR} | \tilde{\tau} \cdot b)$$

- Now consider scenario where *b* is not measured from a sideband, but is taken from MC simulation **with an 8% cross-section 'systematic' uncertainty**

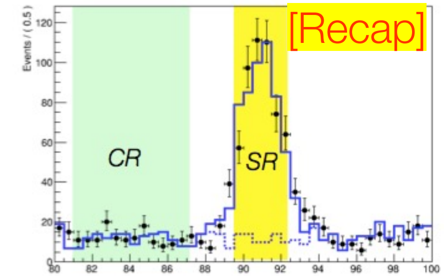
'Measured background rate by MC simulation'

$$L_{full}(s,b) = Poisson(N_{SR} | s+b) \cdot \underbrace{Gauss(\tilde{b} | b, 0.08)}_{\text{'\textbf{Subsidiary measurement}' of background rate}}$$

  - *We can model this in the same way, because the cross-section uncertainty is also (ultimately) the result of a measurement*

**Generalize: 'sideband' → 'subsidiary measurement'**

# Modeling a detector calibration uncertainty




$$L_{full}(s,b) = Poisson(N_{SR} \mid s+b) \cdot Gauss(\tilde{b} \mid b, 0.08)$$

- **Now consider a detector uncertainty**, e.g. jet energy scale calibration, which can affect the analysis acceptance in a non-trivial way (unlike the cross-section example)

Nominal calibration

Assumed calibration

Signal rate (our parameter of interest)

$$L(N,\tilde{\alpha} \mid s,\alpha) = Poisson(N \mid s + \underbrace{\tilde{b}(\alpha/\tilde{\alpha}) \cdot 2)}) \cdot \underbrace{Gauss(\tilde{\alpha} \mid \alpha, \sigma_\alpha)}$$

Observed event count

Nominal background expectation from MC (a constant), obtained with $a=\tilde{a}$

**Response function for JES uncertainty**
(a 1% JES change results in a 2% acceptance change)

Uncertainty on nominal calibration **(here 5%)**

“Subsidiary measurement”
Encodes ‘external knowledge’ on JES calibration

[Recap]

# Modeling a detector calibration uncertainty

- Simplify expression by renormalizing "subsidiary measurement"

$$Gauss(\tilde{\alpha} \mid \alpha, \sigma_\alpha)$$

Signal rate (our parameter of interest)

$$L(N \mid s, \alpha) = Poisson(N \mid s + \tilde{b}(1 + 0.1\alpha)) \cdot Gauss(0 \mid \alpha, 1)$$

Observed event count

Nominal background expectation from MC (a constant)

***Response function for normalized JES parameter***
[a unit change in α – a 5% JES change – still results in a 10% acceptance change]

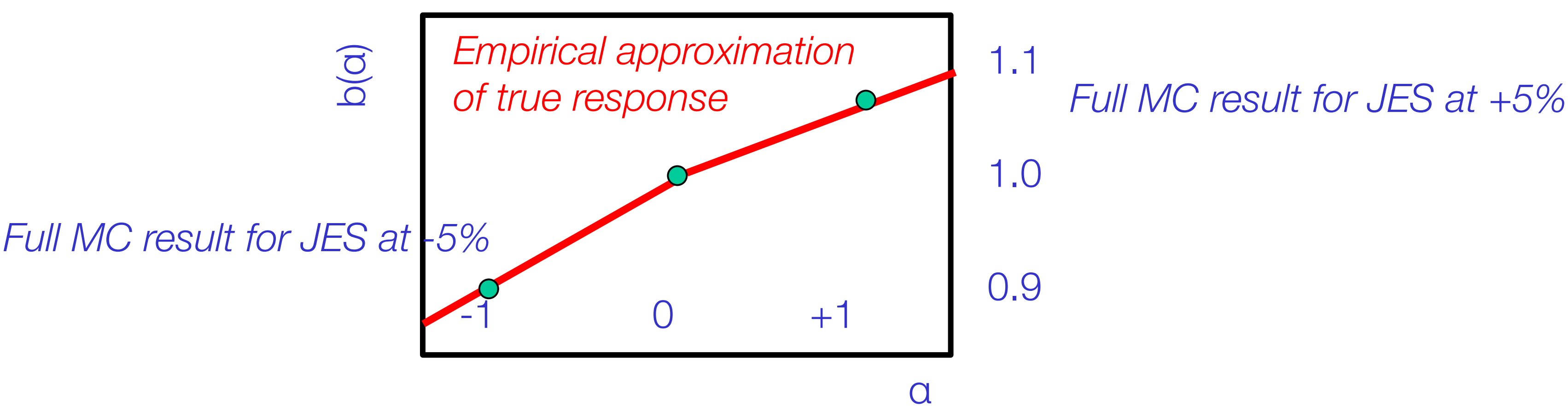
**"Normalized subsidiary measurement"**

***The scale of parameter α is now chosen such that values ±1 corresponds to the nominal uncertainty (in this example 5%)***

# The response function as empirical model of full simulation

$$L(N,0 \mid s,\alpha) = Poisson(N \mid s + \underbrace{b(\alpha)}) \cdot Gauss(0 \mid \alpha, 1)$$

- Note that the response function is generally not linear, but can in principle *always be determined by your full simulation chain*
  - But you cannot run your full simulation chain for any arbitrary 'systematic uncertainty variation' → Too much time consuming
  - Typically, run full MC chain for nominal and ±1σ variation of systematic uncertainty, and approximate response for other values of NP with interpolation
  - For example run at nominal JES and with JES shifted up and down by ±5%



[Recap]

# What is a systematic uncertainty?

- It is an uncertainty in the Likelihood of your physics measurement that is characterized deterministically, up to a set of parameters, of which the true value is unknown.
- A fully specified systematic uncertainty defines
  - 1: A set of one or more parameters
    of which the true value is unknown,
  - 2: A response model that describes the effect of those parameters on the measurement
    *(sampled from full simulation, and interpolation)*
  - 3: A subsidiary measurement of the parameters
    that constrains the values the parameters can take
    (implies a specific distribution: Gaussian *(default, CLT)*, Poisson *(low-stats counting)*, or otherwise)

# Modeling multiple systematic uncertainties

- Introduction of multiple systematic uncertainties presents no special issues
- Example JES uncertainty plus generator ISR uncertainty

$$L(N,0 \mid s,\alpha_{JES},\alpha_{ISR}) = P(N \mid s + b\underbrace{(1+0.1\alpha_{JES}+0.05\alpha_{ISR})}_{\text{Joint response function for both systematics}}) \cdot \underbrace{G(0 \mid \alpha_{JES},1) \cdot G(0 \mid \alpha_{ISR},1)}_{\text{One subsidiary measurement for each source of uncertainty}}$$

- A brief note on correlations
  - Word “correlations” often used sloppily – proper way is to think of correlations of parameter estimators. Likelihood defines parameters $\alpha_{JES}$, $\alpha_{ISR}$. The (ML) estimates of these are denoted $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$
  - The ML estimators of $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$ using the Likelihood of the subsidiary measurements are uncorrelated (since the product factorize in this example)
  - The ML estimators of $\hat{\alpha}_{JES},\hat{\alpha}_{ISR}$ using the full Likelihood may be correlated. This is due to physics modeling effects encoded in the joint response function

# Modeling systematic uncertainties in multiple channels

- Systematic effects that affect multiple measurements should be modeled coherently.
  - Example – Likelihood of two Poisson counting measurements

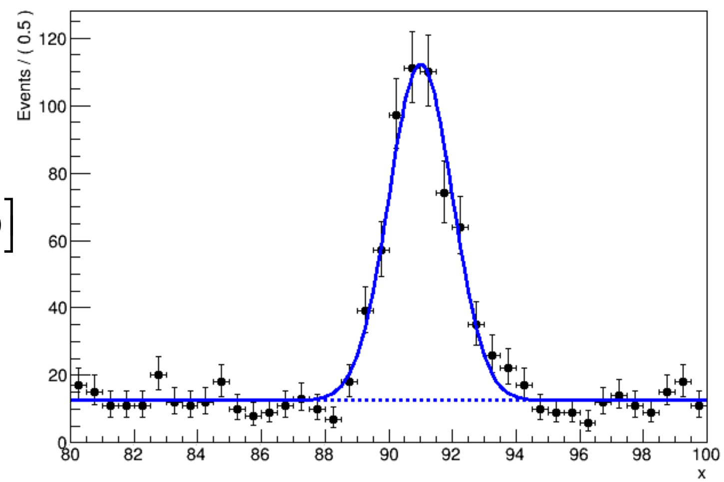
$$L(N_A, N_B \mid s, \alpha_{JES}) = P(N_A \mid s \cdot f_A + b_A \underbrace{(1+0.1\alpha_{JES})}_{\text{JES response function for channel A}}) \cdot P(N_B \mid s \cdot f_B + b_B \underbrace{(1-0.3\alpha_{JES})}_{\text{JES response function for channel B}}) \cdot \underbrace{G(0 \mid \alpha_{JES}, 1)}_{\text{JES subsidiary measurement}} \cdot$$

  - Effect of changing JES parameter $\alpha_{JES}$ coherently affects both measurement.
  - Magnitude and sign effect does not need to be same, this is dictated by the physics of the measurement

# Introducing response functions for shape uncertainties

- Modeling of systematic uncertainties in Likelihoods describing distributions follows the same procedure as for counting models
  - Example: Likelihood modeling distribution in a di-lepton invariant mass. POI is the signal strength μ

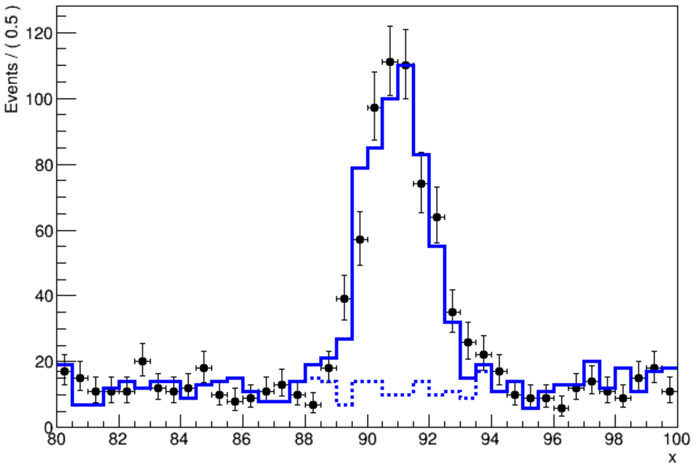
$$L(\vec{m}_{ll} \mid \mu) = \prod_i \left[ \mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91, 1) + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right]$$




- Consider a lepton energy scale systematic uncertainty that affects this measurement
  - The LES has been measured with a 1% precision
  - The effect of LES on $m_{ll}$ has been determined to a 2% shift for 1% LES change

$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[ \mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \cdot \underbrace{(1 + 2\alpha_{LES}, 1)}_{\text{Response function}} + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot \underbrace{Gauss(0 \mid \alpha_{LES}, 1)}_{\text{Subsidiary measurement}}$$

# Response modeling for distributions




- For a change in the **rate**, response modeling of histogram-shaped distribution is straightforward: **simply scale entire distribution**

$$L(\vec{N}\,|\,\mu) = \prod_i Poisson(N_i\,|\,\mu\tilde{s}_i + \tilde{b}_i)$$

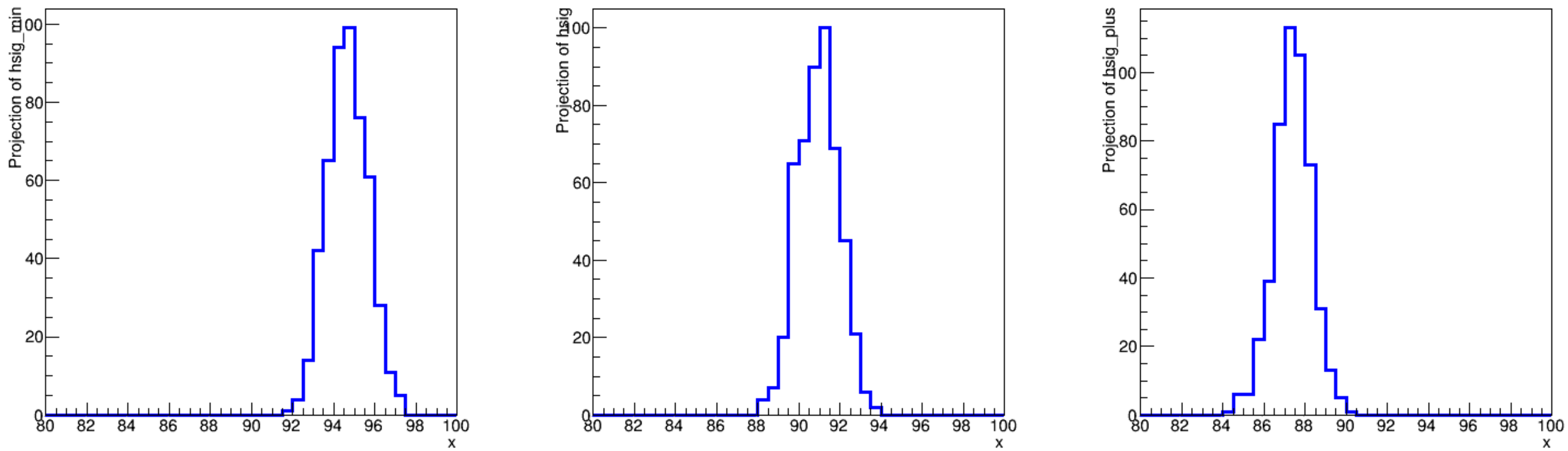
$$L(\vec{N}\,|\,\mu,\alpha) = \prod_i Poisson(N_i\,|\,\mu\tilde{s}_i \cdot \underbrace{(1+3.75\alpha)}_{\text{Response function for signal rate}} + \tilde{b}_i) \cdot \underbrace{Gauss(0\,|\,\alpha,1)}_{\text{Subsidiary measurement}}$$

- But what about a systematic uncertainty that shifts the mean, or affects the distribution in another way?

# Modeling of shape systematics in the likelihood

- Effect of *any* systematic uncertainty that affects the shape of a distribution can in principle be obtained from MC simulation chain
  - Obtain histogram templates for distributions at '+1σ' and '-1σ' settings of systematic effect

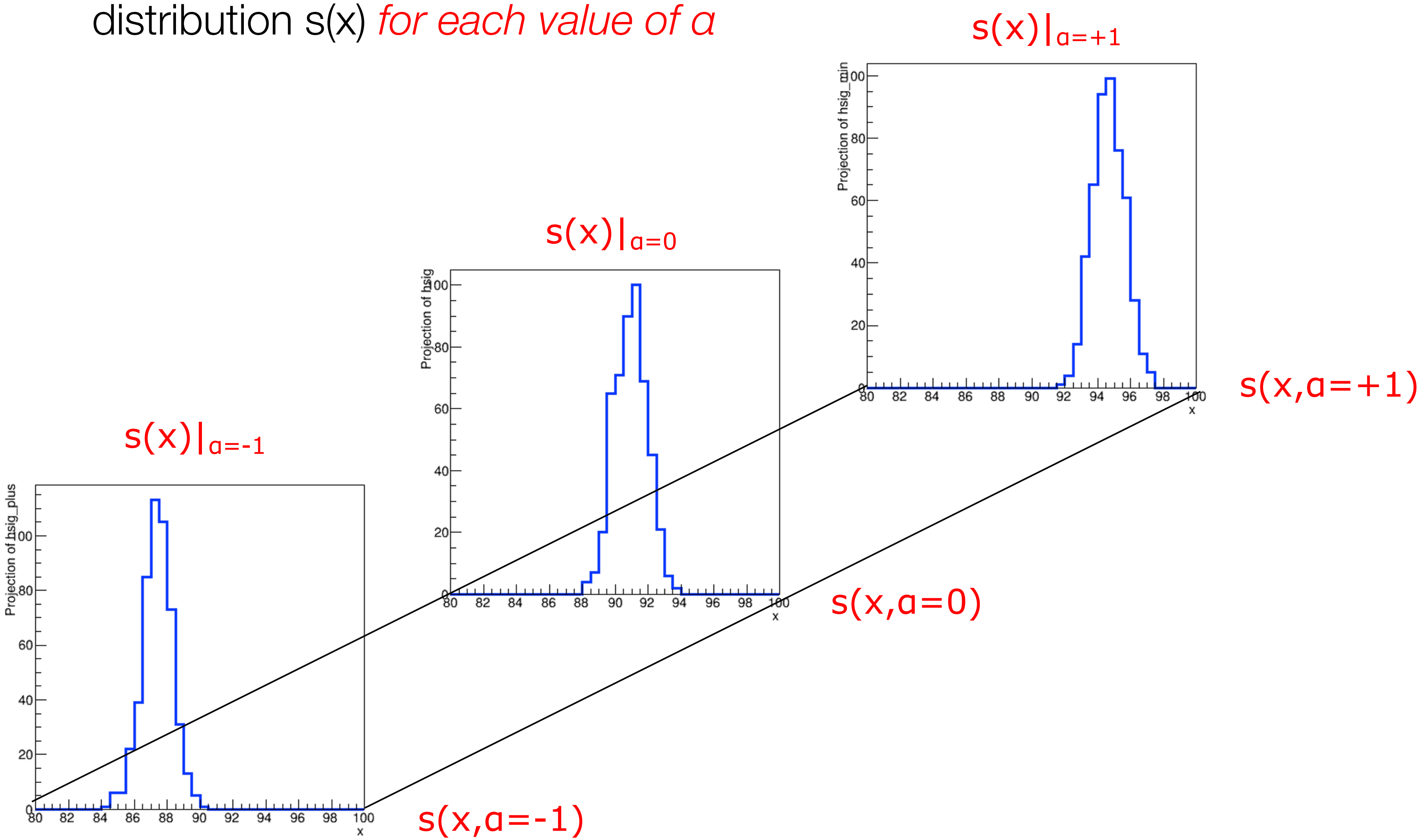
'-1σ' 'nominal' '+1σ'




- Challenge: **construct an empirical response function based on the interpolation of the shapes of these three templates**.

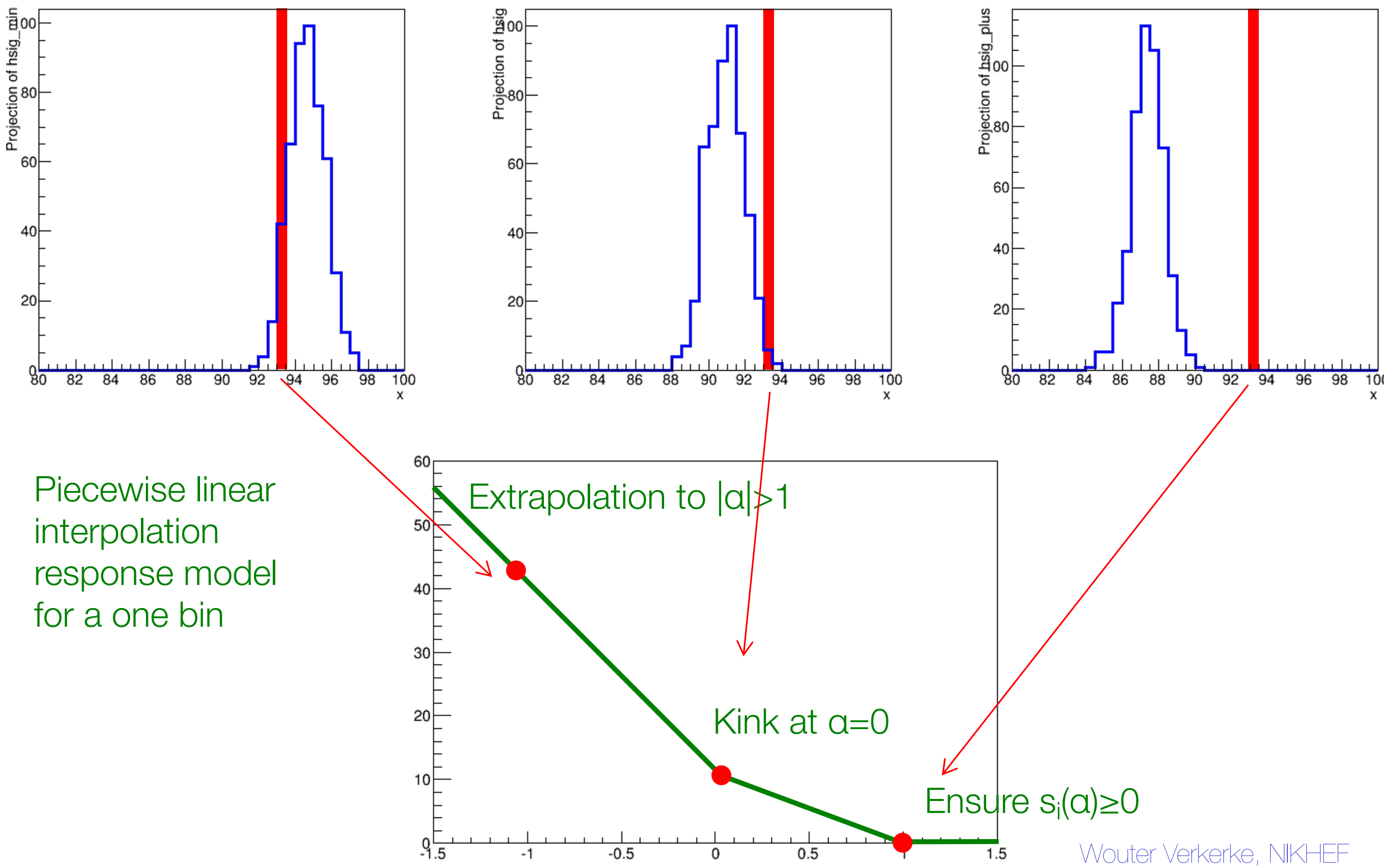
# Need to interpolate between template models

- Need to define 'morphing' algorithm to define distribution s(x) *for each value of α*

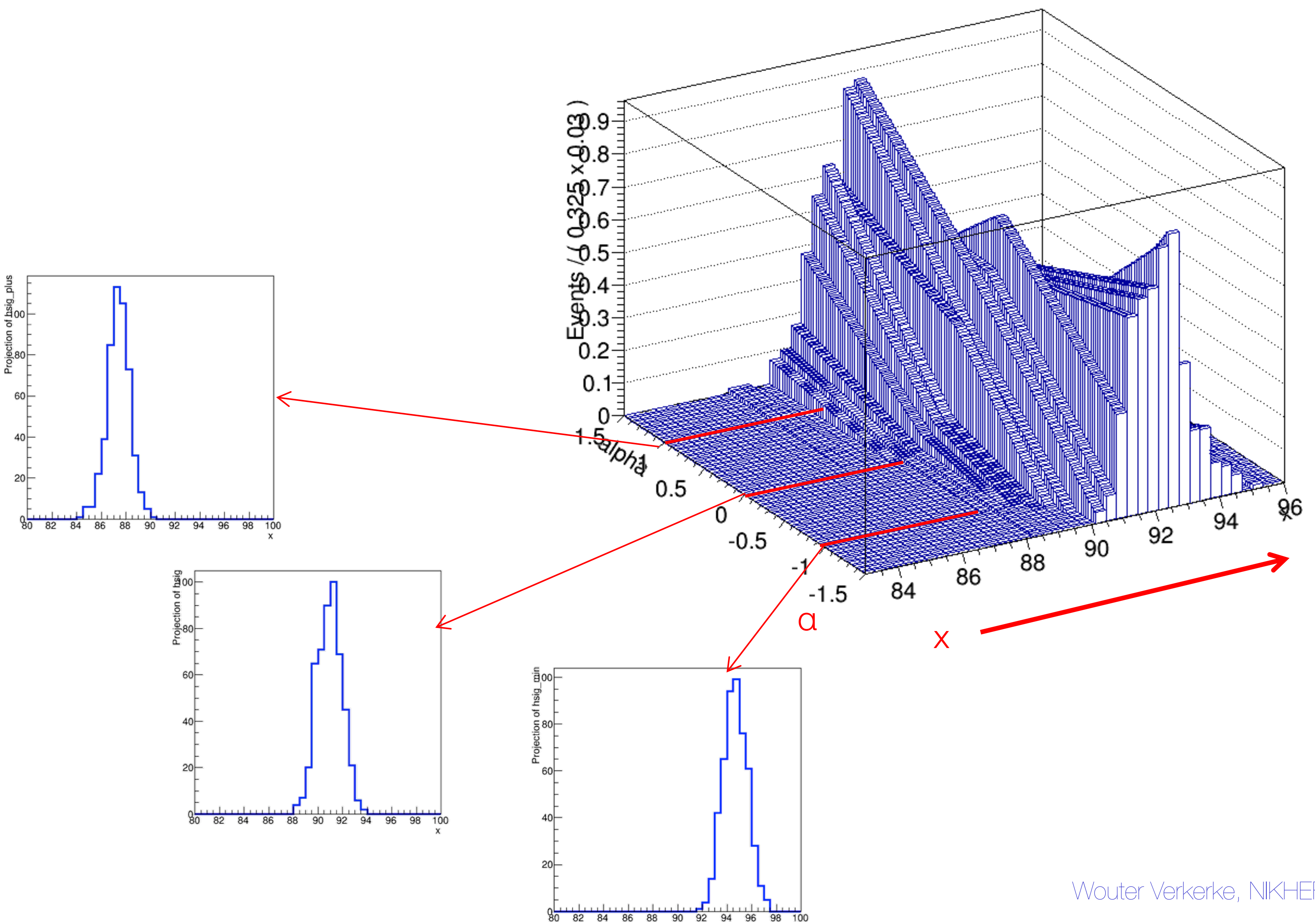


# Piecewise linear interpolation

- Simplest solution is piece-wise linear interpolation for each bin

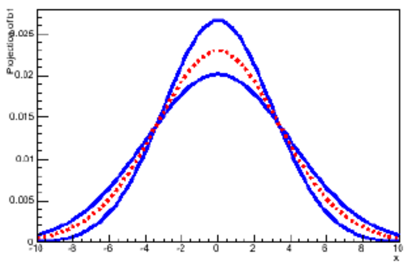
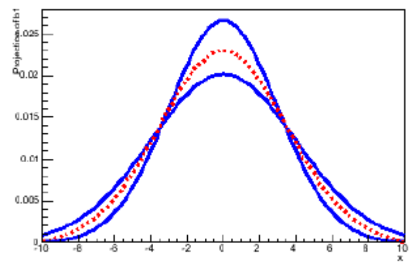
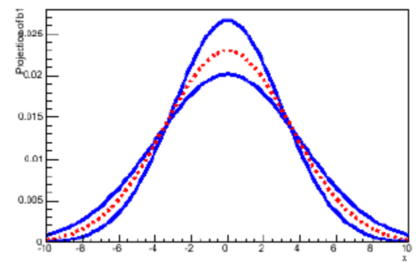
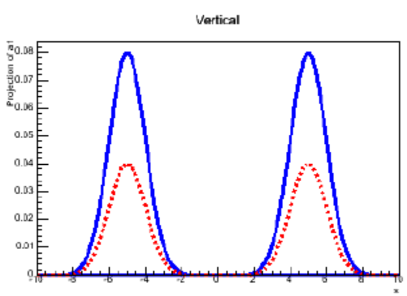
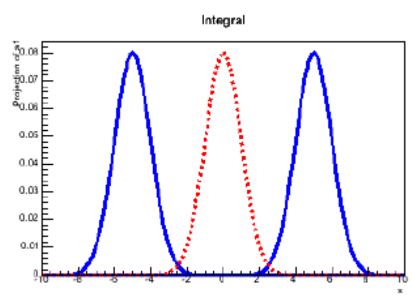
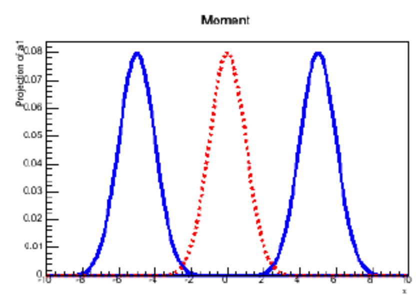
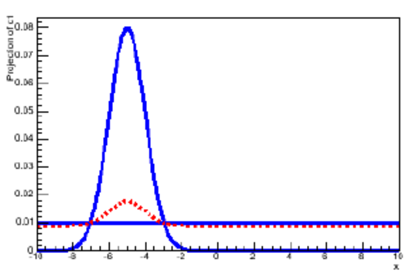
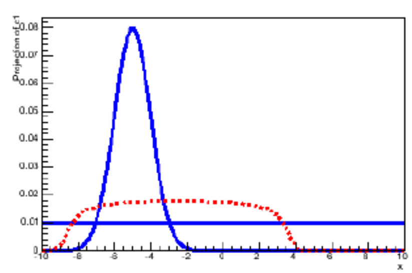
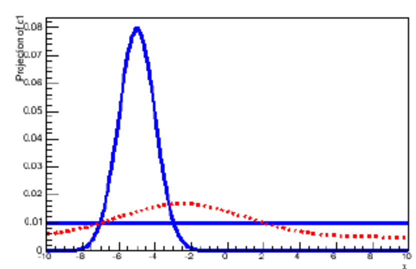


# Visualization of bin-by-bin linear interpolation of distribution



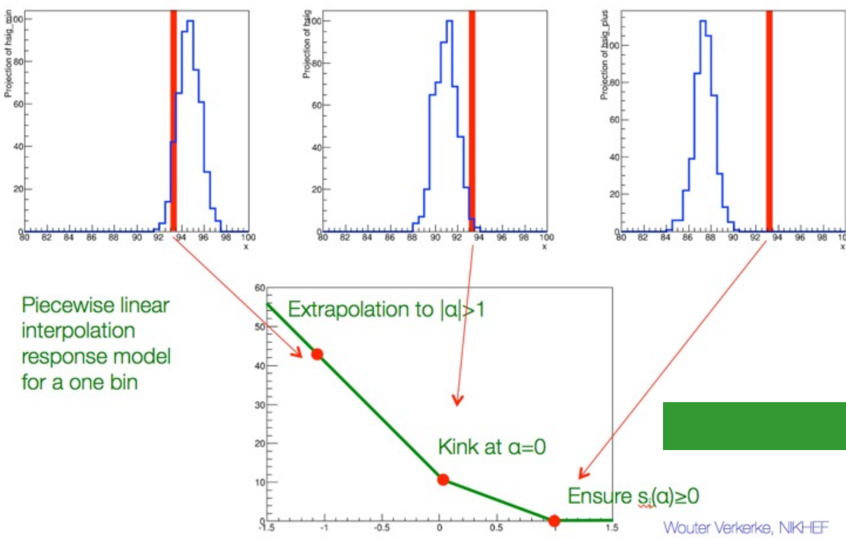

Wouter Verkerke, NIKHEF

# There are other morphing algorithms to choose from

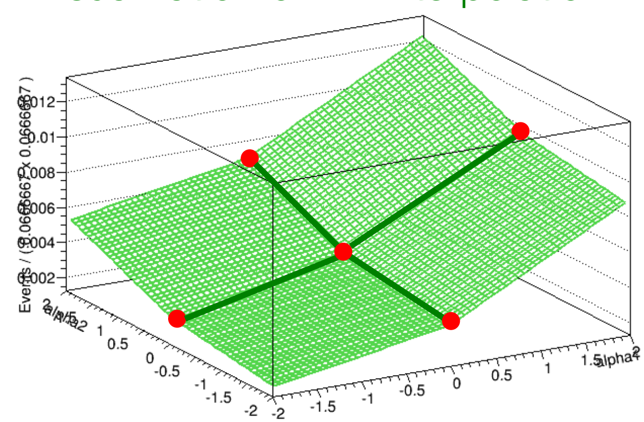
| | Vertical Morphing | Horizontal Morphing | Moment Morphing |
|---|---|---|---|
| Gaussian varying width |  |  |  |
| Gaussian varying mean |  |  |  |
| Gaussian to Uniform (this is conceptually ambigous!) |  |  |  |
| n-dimensional morphing? | ✓ | X | ✓ |

# Piece-wise interpolation for >1 nuisance parameter

- Concept of piece-wise linear interpolation can be trivially extended to apply to morphing of >1 nuisance parameter.
  - Difficult to visualize effect on full distribution, but easy to understand concept at the individual bin level

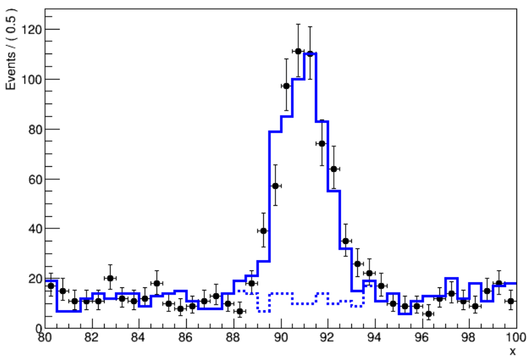


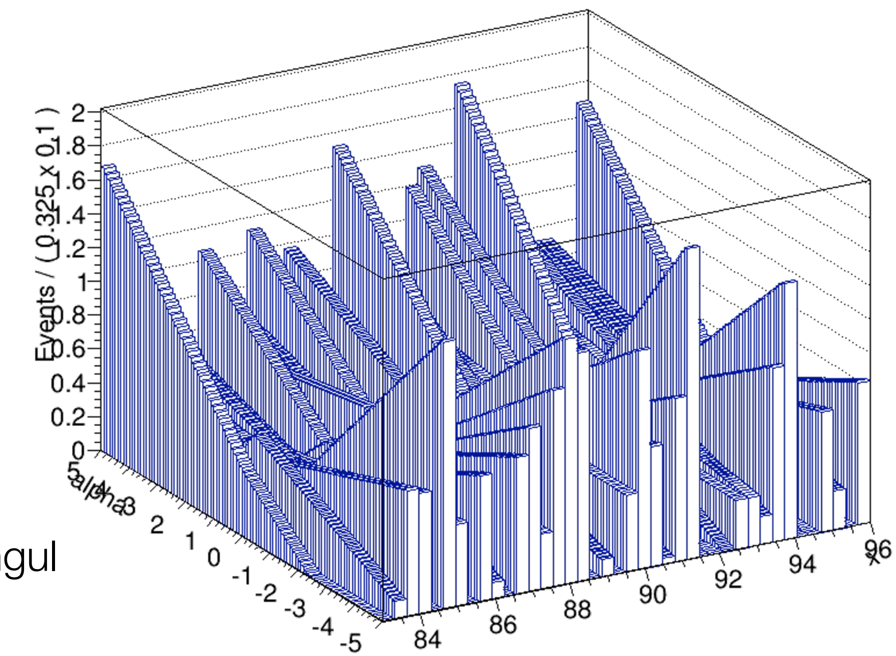

Visualization of 2D interpolation



# Shape, rate or no systematic?

- Be judicious with modeling of systematic with little or no significant change in shape (w.r.t MC template statistics)
  - Example morphing of a very subtle change in the background model
  - Is this a meaningful new degree of freedom in the likelihood model?

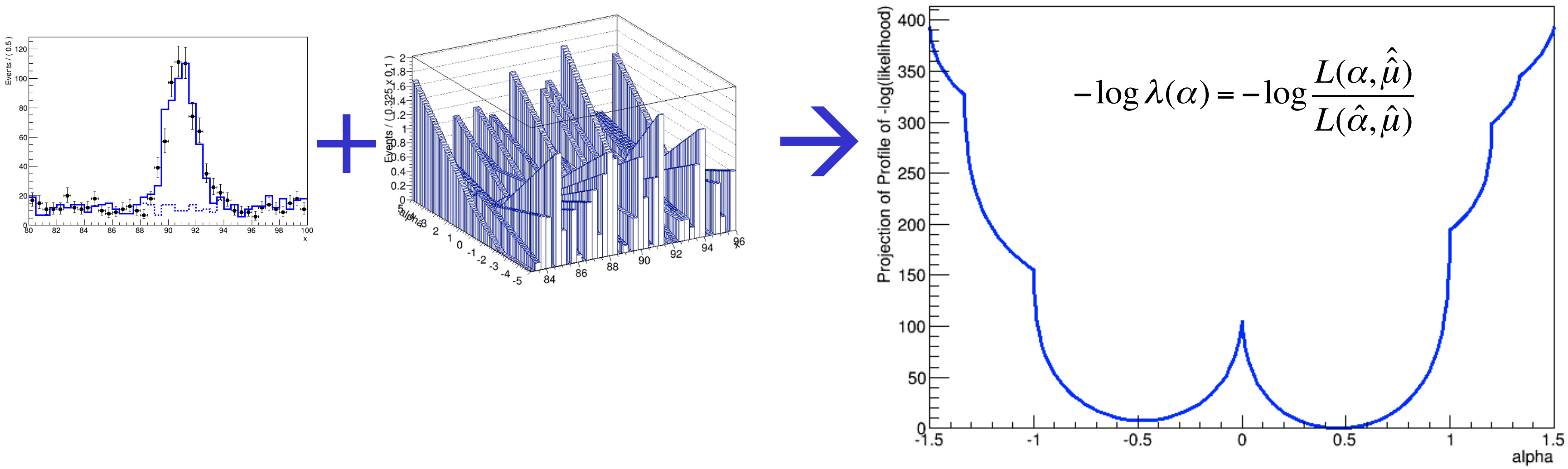



  - A χ2 or KS test between nominal and alternate template can help to decide if a shape uncertainty is meaningul
  - Most systematic uncertainties affect both rate and shape, but can make independent decision on modeling rate (which less likely to affect fit stability)

# Fit stability due to insignificant shape systematics

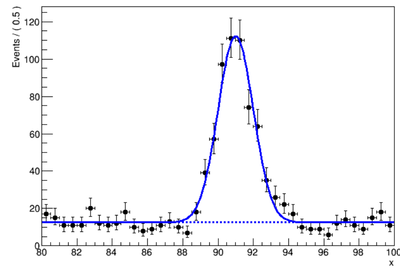
- Shape of profile likelihood in NP α clearly raises two points




- 1) Numerical minimization process will be ‘interesting’
- 2) MC statistical effects induce strongly defined minima that are fake
  - Because for this example all three templates were sampled from the same parent distribution (a uniform distribution)

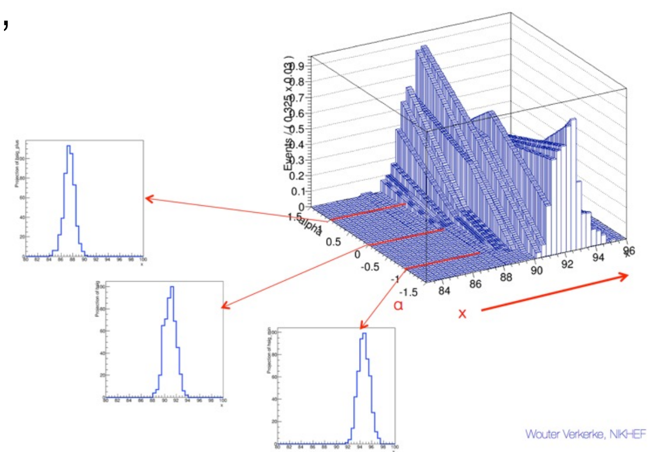
# Recap on shape systematics & template morphing

- Implementation of shape systematic in likelihoods modeling distributions conceptually no different that rate systematics in counting experiments




$$L(\vec{m}_{ll} \mid \mu, \alpha_{LES}) = \prod_i \left[ \mu \cdot \mathrm{Gauss}(m_{ll}^{(i)}, 91 \cdot (1 + 2\alpha_{LES}, 1) + (1-\mu) \cdot \mathrm{Uniform}(m_{ll}^{(i)}) \right] \cdot Gauss(0 \mid \alpha_{LES}, 1)$$

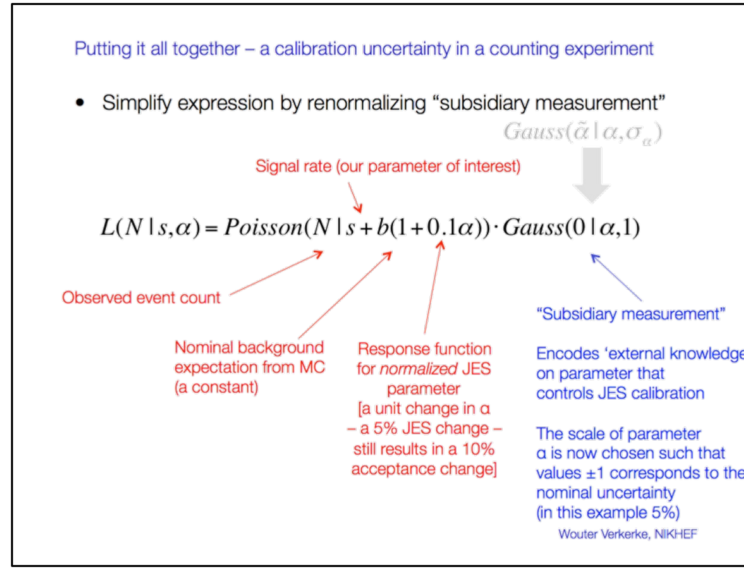
- For template modes obtained from MC simulation template provides a technical solution to implement response function
  - Simplest strategy piecewise linear interpolation, but only works well for small changes
  - Moment morphing better adapted to modeling of shifting distributions
  - Both algorithms extend to n-dimensional interpolation to model multiple systematic NPs in response function
  - Be judicious in modeling 'weak' systematics: MC systematic uncertainties will dominate likelihood




Wouter Verkerke, NIKHEF

# Example 1: counting expt

- Will now demonstrate how to construct a model for a counting experiment with a systematic uncertainty

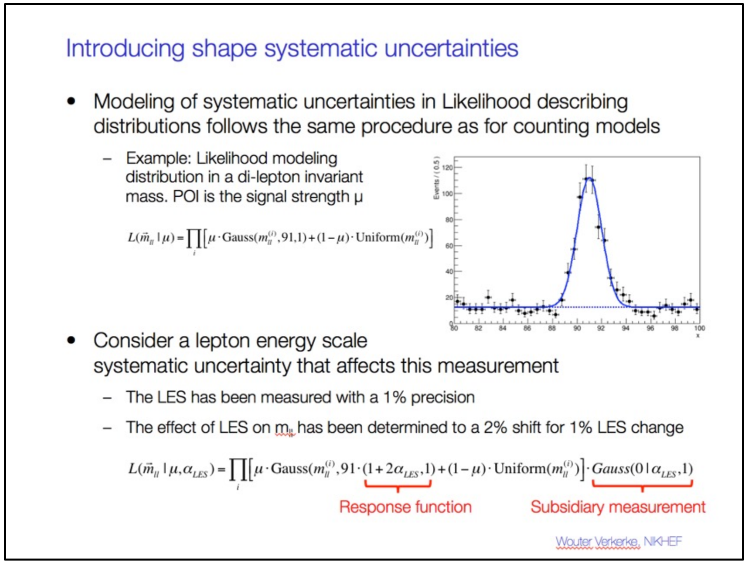



$$L(N \mid s,\alpha) = Poisson(N \mid s + b(1+0.1\alpha)) \cdot Gauss(0 \mid \alpha, 1)$$

```
// Subsidiary measurement of alpha
w.faxtory("Gaussian::subs(0,alpha[-5,5],1)") ;

// Response function mu(alpha)
w.factory("expr::mu('s+b(1+0.1*alpha)',s[20],b[20],alpha)") ;

// Main measurement
w.factory("Poisson::p(N[0,10000],mu)");

// Complete model Physics*Subsidiary
w.factory("PROD::model(p,subs)") ;
```

# Example 2: unbinned L with syst.

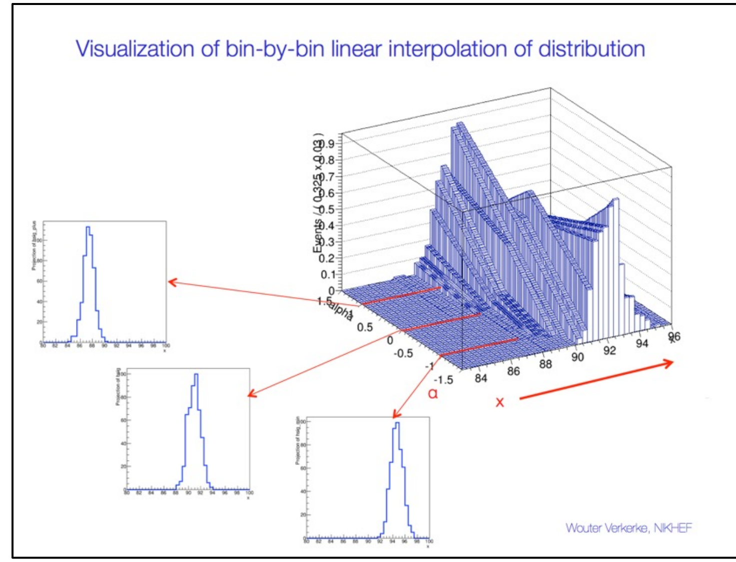
- Will now demonstrate how to code complete example of the unbinned profile likelihood of Section 5:




$$L(\vec{m}_{ll}\,|\,\mu,\alpha_{LES}) = \prod_i \left[\mu\cdot\mathrm{Gauss}(m_{ll}^{(i)},\underbrace{91\cdot(1+2\alpha_{LES}),1)+}(1-\mu)\cdot\mathrm{Uniform}(m_{ll}^{(i)})\right]\cdot\underbrace{Gauss(0\,|\,\alpha_{LES},1)}$$

```
// Subsidiary measurement of alpha
w.factory("Gaussian::subs(0,alpha[-5,5],1)");

// Response function m(alpha)
w.factory("expr::m_a("m*(1+2alpha)",m[91,80,100],alpha)") ;

// Signal model
w.factory("Gaussian::sig(x[80,100],m_a,s[1])")

// Complete model Physics(signal plus background)*Subsidiary
w.factory("PROD::model(SUM(mu[0,1]*sig,Uniform::bkg(x)),subs)") ;
```

# Example 3 : binned L with syst




- Example of template morphing systematic in a binned likelihood

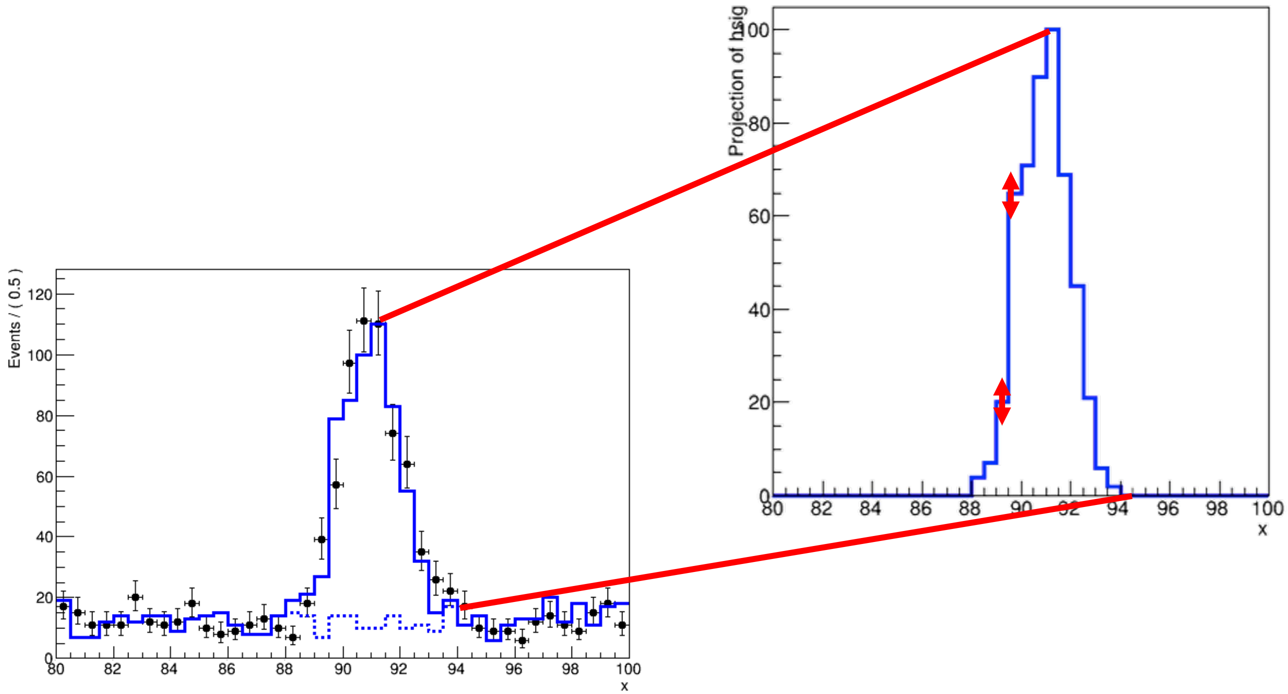
$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha > 0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha < 0 \end{cases}$$

$$L(\vec{N}\,|\,\alpha,\vec{s}^-,\vec{s}^0,\vec{s}^+) = \prod_{bins} P(N_i\,\underbrace{|\,s_i(\alpha,s_i^-,s_i^0,s_i^+))}_{}\cdot\underbrace{G(0\,|\,\alpha,1)}_{}$$

```
// Import template histograms in workspace
 w.import(hs_0,hs_p,hs_m) ;

 // Construct template models from histograms
 w.factory("HistFunc::s_0(x[80,100],hs_0)") ;
 w.factory("HistFunc::s_p(x,hs_p)") ;
 w.factory("HistFunc::s_m(x,hs_m)") ;

 // Construct morphing model
 w.factory("PiecewiseInterpolation::sig(s_0,s_,m,s_p,alpha[-5,5])") ;

 // Construct full model
 w.factory("PROD::model(ASUM(sig,bkg,f[0,1]),Gaussian(0,alpha,1))") ;
```

# Other uncertainties in MC shapes – finite MC statistics

- In practice, MC distributions used for template fits have finite statistics.

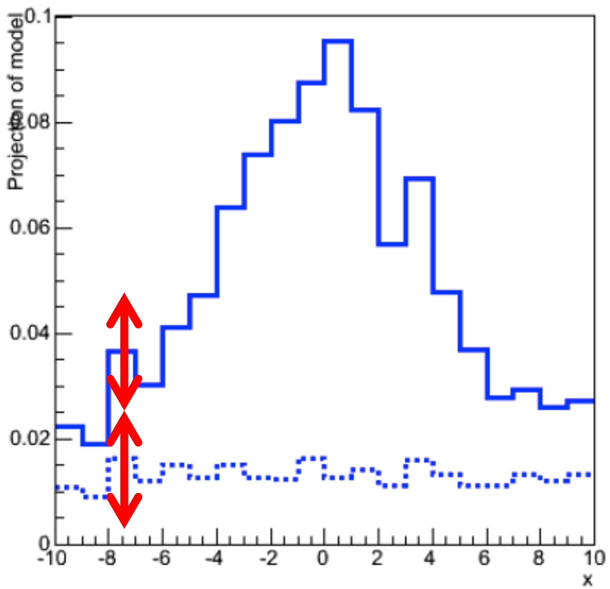



- Limited MC statistics represent an uncertainty on your model → how to model this effect in the Likelihood?

# Other uncertainties in MC shapes – finite MC statistics

- Modeling MC uncertainties: *each MC bin has a Poisson uncertainty*
- Thus, apply usual 'systematics modeling' prescription.
- For a single bin – exactly like original counting measurement

Fixed signal, bkg MC prediction




$$L_{bin-i}(\mu) = Poisson(N_i \mid \mu \cdot \tilde{s}_i + \tilde{b}_i)$$

Signal, bkg
MC nuisance params

$$L_{bin-i}(\mu, s_i, b_i) = Poisson(N_i \mid \mu \cdot s_i + b_i)$$

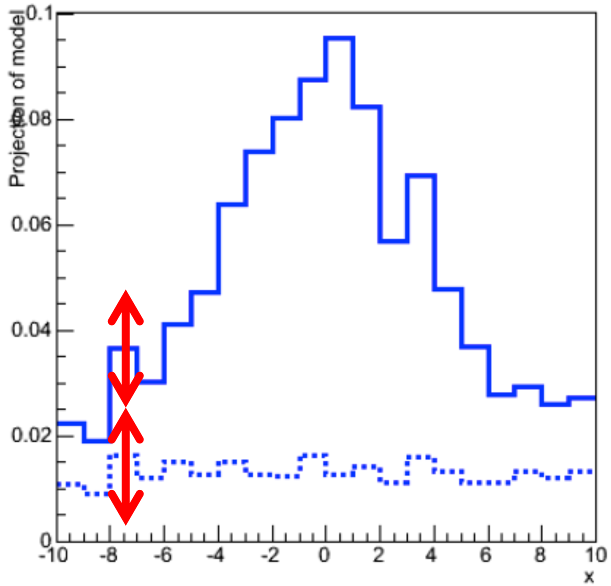
$$\cdot Poisson(N_i^{MC-s} \mid s_i)$$

$$\cdot Poisson(N_i^{MC-b} \mid b_i)$$

Subsidiary measurement for signal MC
('measures' MC prediction $s_i$ with Poisson uncertainty)

# Nuisance parameters for template statistics

- Repeat for all bins




$$L(\vec{N}\mid\mu)=\prod_{bins}P(N_i\mid\mu\cdot\tilde{s}_i+\tilde{b}_i)$$

Binned likelihood with rigid template

$$L(\vec{N}\mid\mu,\vec{s},\vec{b})=\prod_{bins}P(N_i\mid\underbrace{\mu\cdot s_i+b_i}_{\text{Response function w.r.t. } s,\, b \text{ as parameters}})\underbrace{\prod_{bins}P(\tilde{s}_i\mid s_i)\prod_{bins}P(\tilde{b}_i\mid b_i)}_{\text{2x } N_{bins} \text{ subsidiary measurements of } s,b \text{ from s\~{},b\~{}}}$$

Response function w.r.t. *s, b* as parameters

2x $N_{bins}$ subsidiary measurements of *s* ,*b* from s~,b~

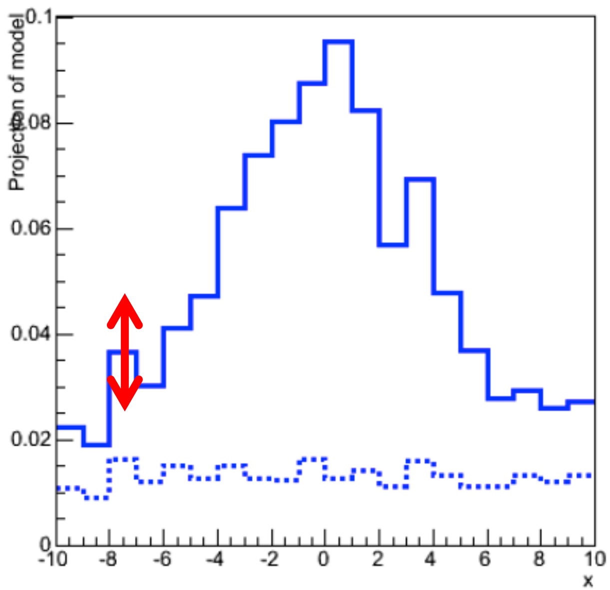
- Result: accurate model for MC statistical uncertainty, but lots of nuisance parameters (#samples x #bins)...

# Reducing the number NPs – Beeston-Barlow 'lite'

- Another approach that is being used is called 'BB' – lite
- Premise: effect of statistical fluctuations on sum of templates is dominant → Use one NP per bin instead of one NP per component per bin

'Beeston-Barlow'

$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,s_i + b_i) \prod_{bins} P(\tilde{s}_i\,|\,s_i) \prod_{bins} P(\tilde{b}_i\,|\,b_i)$$




'Beeston-Barlow lite '

$$L(\vec{N}\,|\,\vec{n}) = \prod_{bins} P(N_i\,|\,n_i) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i\,|\,n_i)$$

Response function w.r.t. *n* as parameters

Subsidiary measurements of *n* from s~+b~

$$L(\vec{N}\,|\,\vec{\gamma}) = \prod_{bins} P(N_i\,|\,\gamma_i(\tilde{s}_i + \tilde{b}_i)) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i\,|\,\gamma_i(\tilde{s}_i + \tilde{b}_i))$$

Normalized NP lite model (nominal value of all γ is 1)

## The interplay between shape systematics and MC systematics

- Best practice for template morphing models is to also include effect of MC systematics
- Note that that for every 'morphing systematic' there is an set of two templates that have their own (independent) MC statistical uncertainties.
  - A completely accurate should model MC stat uncertainties of all templates

$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha > 0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha < 0 \end{cases}$$

$$L(\vec{N}\,|\,\alpha,\vec{s}^{\,-},\vec{s}^{\,0},\vec{s}^{\,+}) = \prod_{bins} P(N_i\,|\,\underbrace{s_i(\alpha,s_i^-,s_i^0,s_i^+)}_{\text{Morphing response function}})\underbrace{\prod_{bins} P(\tilde{s}_i^-\,|\,s_i^-)\prod_{bins} P(\tilde{s}_i^0\,|\,s_i^0)\prod_{bins} P(\tilde{s}_i^+\,|\,s_i^+)}_{\text{Subsidiary measurements}}$$

- But has severe practical problems
  - Can only be done in 'full' Beeston-Barlow model, not in 'lite' mode, enormous number of NP models with only a handful of shape systematics…

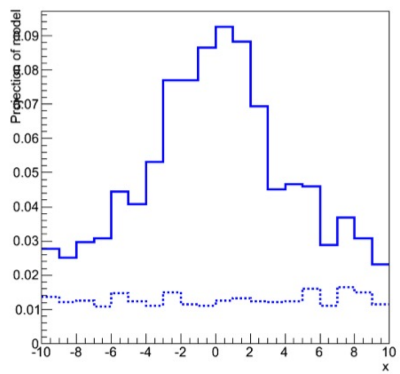
# The interplay between shape systematics and MC systematics

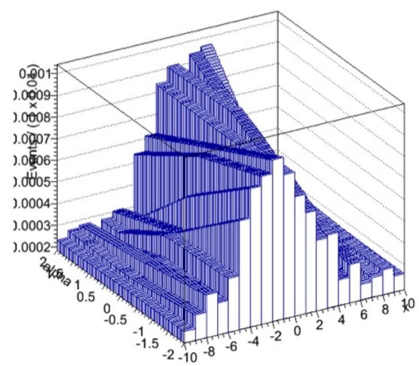
- Commonly chosen practical solution

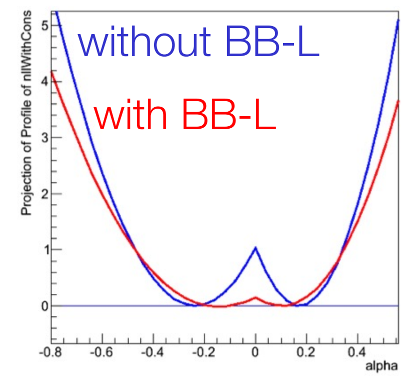
$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall \alpha > 0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall \alpha < 0 \end{cases}$$

$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} \underbrace{P(N_i\,|\,\gamma_i\cdot[s_i(\alpha,s_i^-,s_i^0,s_i^+)+b_i])}_{\text{Morphing \& MC response function}} \prod_{bins} \underbrace{P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i\cdot[\tilde{s}_i+\tilde{b}_i])G(0\,|\,\alpha,1)}_{\text{Subsidiary measurements}}$$

*Models relative MC rate uncertainty for each bin w.r.t the nominal MC yield, even if morphed total yield is slightly different*



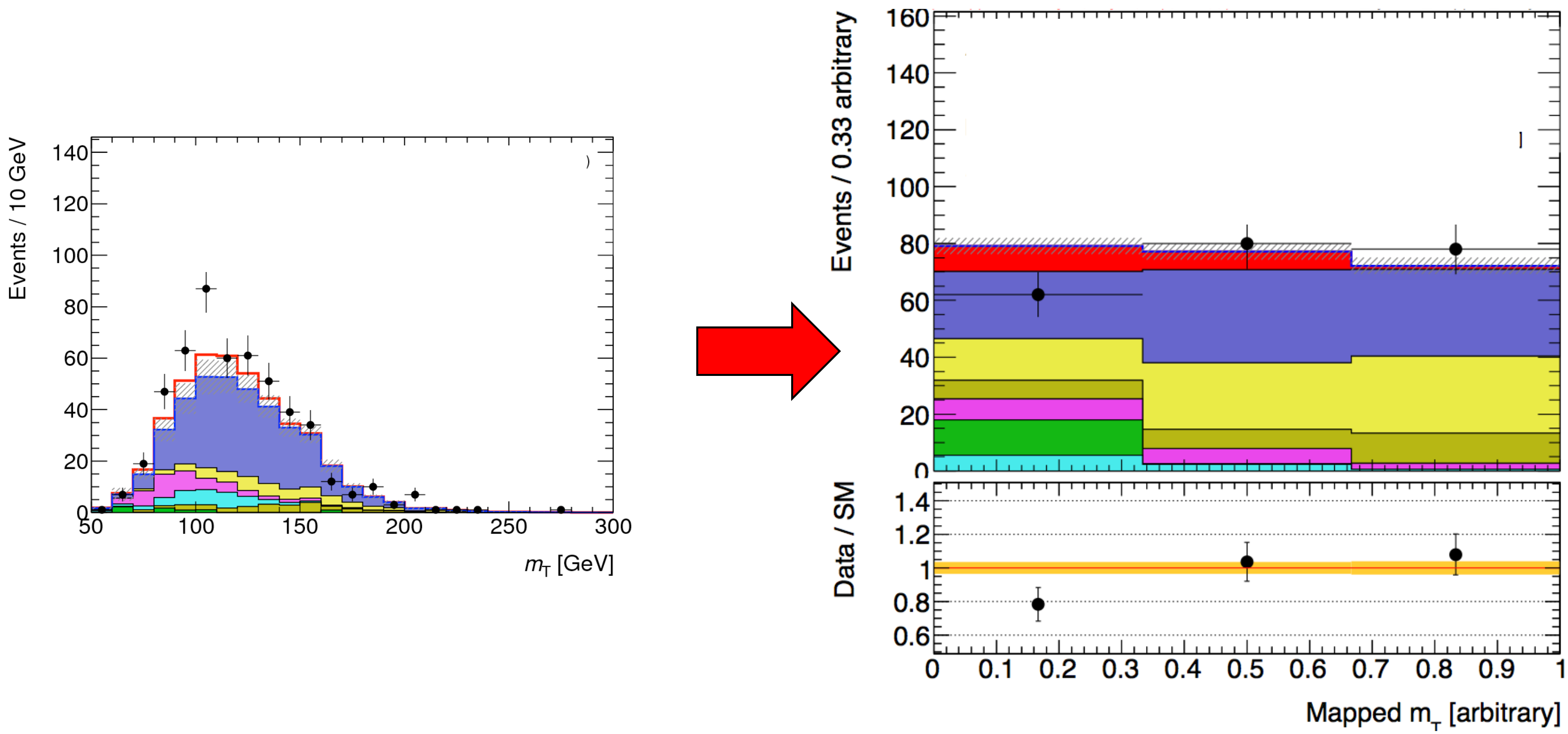




- Approximate MC template statistics already significantly improves influence of MC fluctuations on template morphing
  - Because ML fit can now 'reweight' contributions of each bin

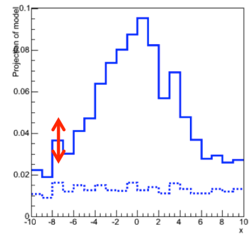
# Adapting binning to event density

- Effect of template statistics can also be controlled by rebinning data such all bins contain expected and observed events
  - For example choose binning such that expected background has a uniform distribution (as signals are usually small and/or uncertain they matter less)



# Example 4 – Beeston-Barlow light

- Beeston-Barlow-(lite) modeling of MC statistical uncertainties

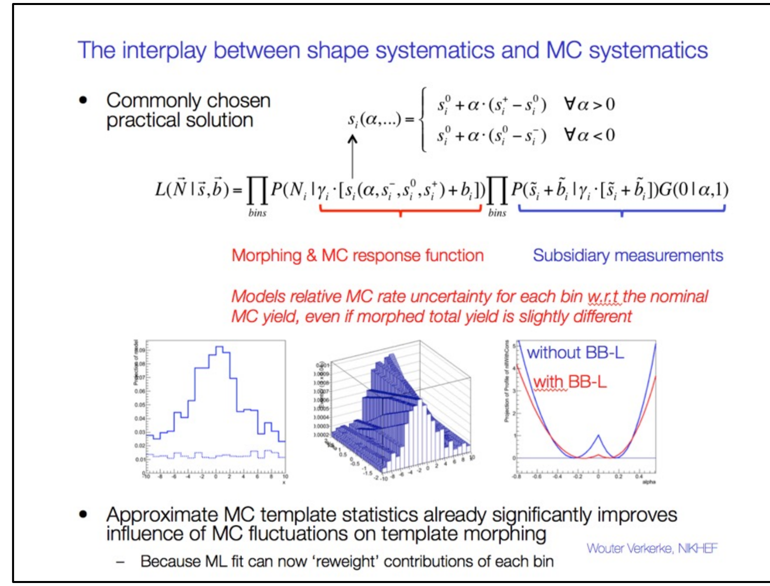
> Reducing the number NPs – Beeston-Barlow 'lite'
>
> - Another approach that is being used is called 'BB' – lite
> - Premise: effect of statistical fluctuations on sum of templates is dominant → Use one NP per bin instead of one NP per component per bin
>
> 'Beeston-Barlow'
>
> $$L(\vec{N}\mid\vec{s},\vec{b}) = \prod_{bins} P(N_i \mid s_i + b_i) \prod_{bins} P(\tilde{s}_i \mid s_i) \prod_{bins} P(\tilde{b}_i \mid b_i)$$
>
> 'Beeston-Barlow lite '
>
> $$L(\vec{N}\mid\vec{n}) = \prod_{bins} P(N_i \mid n_i) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i \mid n_i)$$
>
> Response function w.r.t. $n$ as parameters
>
> Subsidiary measurements of $n$ from s~+b~
>
> 
> 
>
> $$L(\vec{N}\mid\vec{\gamma}) = \prod_{bins} P(N_i \mid \gamma_i(\tilde{s}_i + \tilde{b}_i)) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i \mid \gamma_i(\tilde{s}_i + \tilde{b}_i))$$
>
> Normalized NP lite model (nominal value of all γ is 1)

$$L(\vec{N}\mid\vec{\gamma}) = \prod_{bins} P(N_i \mid \gamma_i(\tilde{s}_i + \tilde{b}_i)) \prod_{bins} P(\tilde{s}_i + \tilde{b}_i \mid \gamma_i(\tilde{s}_i + \tilde{b}_i))$$

```
// Import template histogram in workspace
 w.import(hs) ;

// Construct parametric template models from histograms
// implicitly creates vector of gamma parameters
 w.factory("ParamHistFunc::s(hs)") ;

 // Product of subsidiary measurement
 w.factory("HistConstraint::subs(s)") ;

 // Construct full model
 w.factory("PROD::model(s,subs)") ;
```

# Example 5 – BB-lite + morphing

- Template morphing model with Beeston-Barlow-lite MC statistical uncertainties



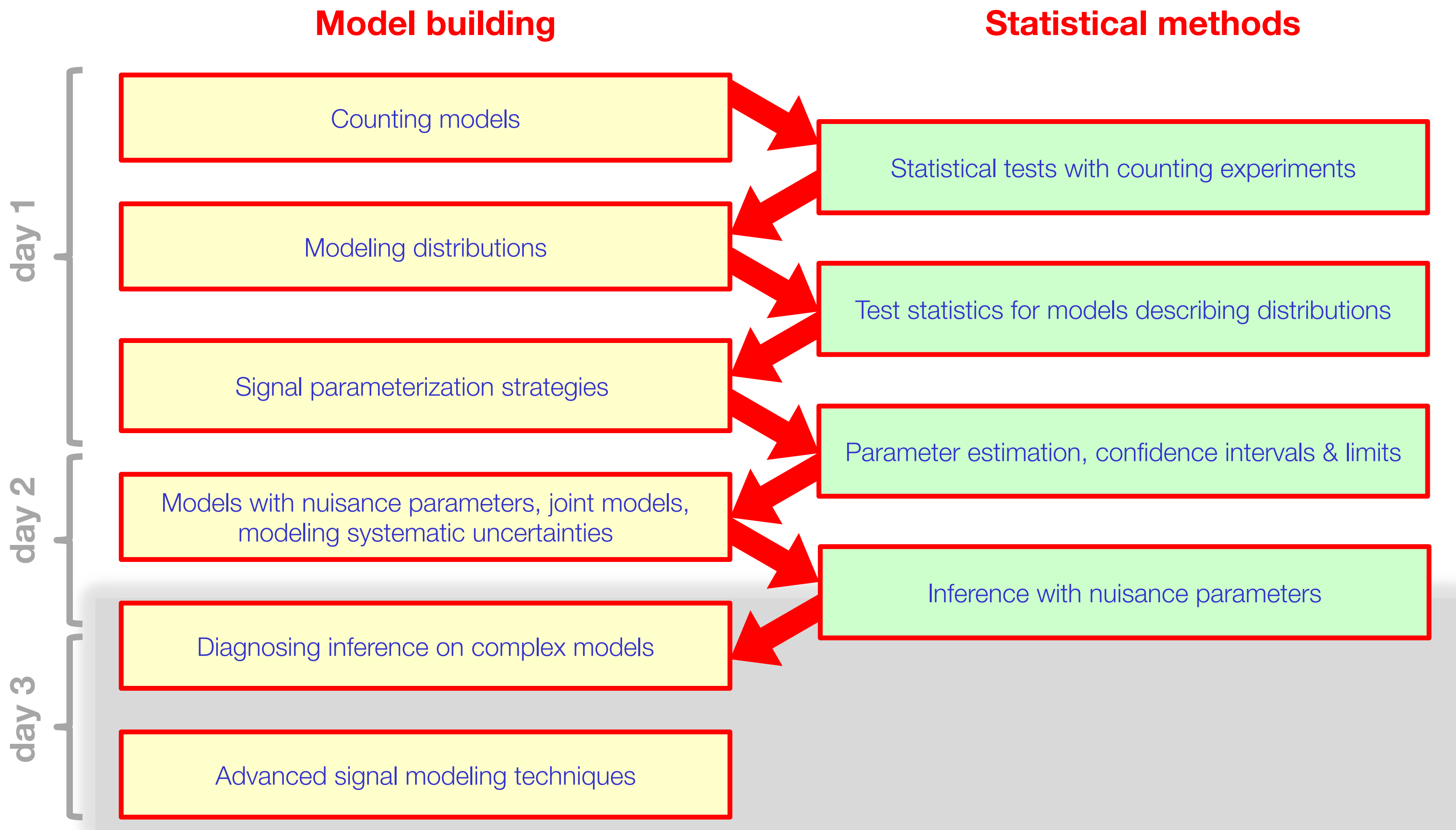

$$s_i(\alpha,\ldots) = \begin{cases} s_i^0 + \alpha\cdot(s_i^+ - s_i^0) & \forall\alpha>0 \\ s_i^0 + \alpha\cdot(s_i^0 - s_i^-) & \forall\alpha<0 \end{cases}$$

$$L(\vec{N}\,|\,\vec{s},\vec{b}) = \prod_{bins} P(N_i\,|\,\gamma_i\cdot[s_i(\alpha,s_i^-,s_i^0,s_i^+)+b_i])\prod_{bins} P(\tilde{s}_i+\tilde{b}_i\,|\,\gamma_i\cdot[\tilde{s}_i+\tilde{b}_i])G(0\,|\,\alpha,1)$$

```
// Import template histograms in workspace
 w.import(hs_0,hs_p,hs_m,hb) ;

 // Construct parametric template morphing signal model
 w.factory(“ParamHistFunc::s_p(hs_p)”) ;
 w.factory(“HistFunc::s_m(x,hs_m)”) ;
 w.factory(“HistFunc::s_0(x[80,100],hs_0)”) ;
 w.factory(“PiecewiseInterpolation::sig(s_0,s_,m,s_p,alpha[-5,5])”) ;

 // Construct parametric background model (sharing gamma’s with s_p)
 w.factory(“ParamHistFunc::bkg(hb,s_p)”) ;

 // Construct full model with BB-lite MC stats modeling
 w.factory(“PROD::model(ASUM(sig,bkg,f[0,1]),
            HistConstraint({s_0,bkg}),Gaussian(0,alpha,1))”) ;
```

# Roadmap of this course

- Start with basics, gradually build up to complexity



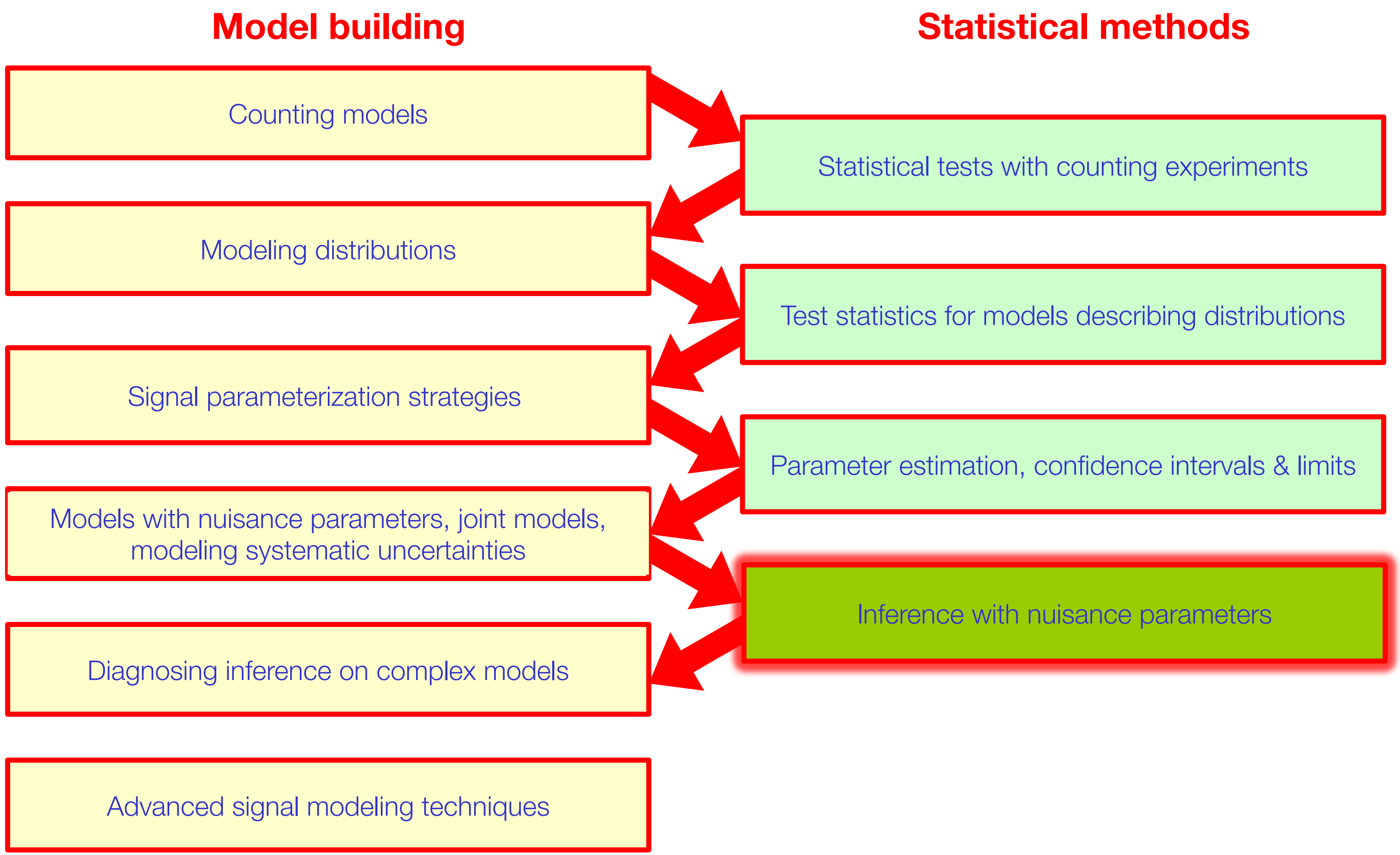
# Statistical methods 4

## Parameters of interest vs nuisance parameters, dealing with nuisance parameters in inference methods

Wouter Verkerke, NIKHEF

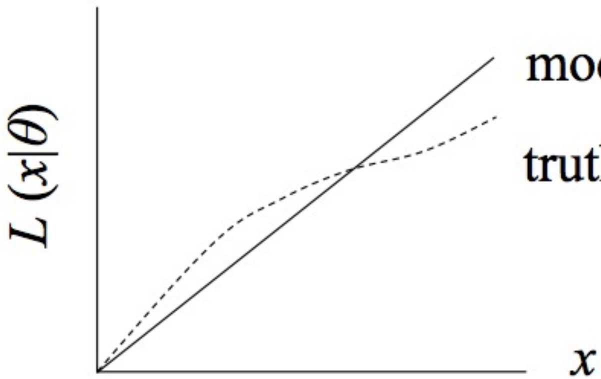
# Roadmap of this course

- Start with basics, gradually build up to complexity



# The statisticians view on nuisance parameters

- In general, our model of the data is not perfect




model: $L(x|\theta) = \theta x$

truth: $L(x|\theta) = \theta x + \alpha x^2 + \beta x^3 + \cdots$

- Can improve modeling by including additional adjustable parameters
- Goal: some point in the parameter space of the enlarged model should be "true"
- Presence of nuisance parameters decreases the sensitivity of the analysis of the parameter(s) of interest

# Treatment of nuisance parameters in variance estimation

- Maximum likelihood estimator of parameter variance is based on 2nd derivative of Likelihood
  - For multi-parameter problems this 2nd derivative is generalized by the **Hessian Matrix** of partial second derivatives

$$\hat{\sigma}(p)^2 = \hat{V}(p) = \left(\frac{d^2 \ln L}{d^2 p}\right)^{-1} \Rightarrow \hat{\sigma}(p_i)^2 = \hat{V}(p_{ii}) = \left(H^{-1}\right)_{ii}$$

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

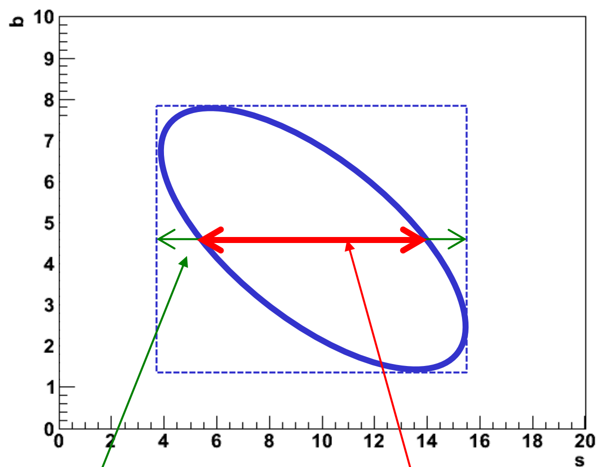
- For multi-parameter likelihoods estimate of covariance $V_{ij}$ of pair of 2 parameters in addition to variance of individual parameters
  - Usually re-expressed in terms dimensionless correlation coefficients ρ

$$V_{ij} = \rho_{ij}\sqrt{V_{ii}V_{jj}}$$

# Treatment of nuisance parameters in variance estimation

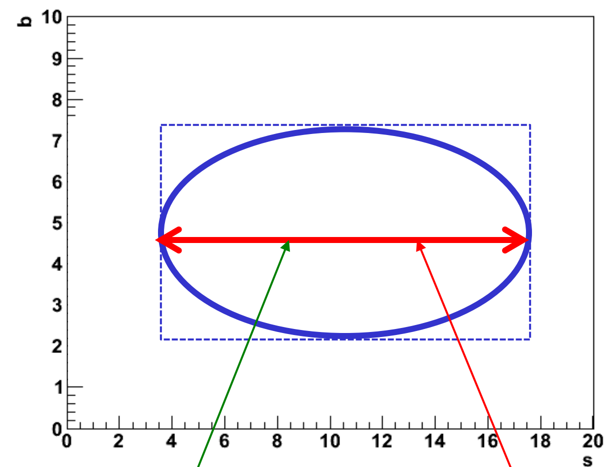
- Effect of NPs on variance estimates visualized

**Scenario 1**
Estimators of
POI and NP correlated
i.e. ρ(s,b)≠0




$$\hat{V}(s)\ from \begin{bmatrix} \frac{\partial^2 L}{\partial s^2} & \frac{\partial^2 L}{\partial s \partial b} \\ \frac{\partial^2 L}{\partial s \partial b} & \frac{\partial^2 L}{\partial b^2} \end{bmatrix}^{-1} \qquad \hat{V}(s)\ from \left[\frac{\partial^2 L}{\partial s^2}\right]^{-1}_{b=\hat{b}}$$

**Scenario 2**
Estimators of
POI and NP correlated
i.e. ρ(s,b)=0



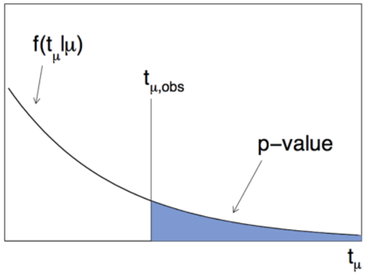

$$\hat{V}(s)\ from \begin{bmatrix} \frac{\partial^2 L}{\partial s^2} & \frac{\partial^2 L}{\partial s \partial b} \\ \frac{\partial^2 L}{\partial s \partial b} & \frac{\partial^2 L}{\partial b^2} \end{bmatrix}^{-1} \qquad \hat{V}(s)\ from \left[\frac{\partial^2 L}{\partial s^2}\right]^{-1}_{b=\hat{b}}$$

*Uncertainty on background increases uncertainty on signal*

# Treatment of NPs in hypothesis testing and conf. intervals

- We've covered frequentist hypothesis testing and interval calculation using likelihood ratios based on a likelihood with a single parameter (of interest) L(μ)
  - Result is p-value on hypothesis with given μ value, or
  - Result is a confidence interval $[\mu_-,\mu_+]$ with values of μ for which p-value is at or above a certain level (the confidence level)
- How do you do this with a likelihood L(μ,θ) where θ is a nuisance parameter?
  - With a test statistics $q_\mu$, we calculate p-value for hypothesis θ as

$$p_\mu = \int_{q_{\mu,obs}}^{\infty} f(q_\mu \mid \mu,\theta) dq_\mu$$




- But what values of θ do we use for $f(q_\mu|\mu,\theta)$?
  Fundamentally, we want to reject μ only if $p<\alpha$ for all θ
  → Exact confidence interval

# Hypothesis testing & conf. intervals with nuisance parameters

- The goal is that the parameter of interest should be covered at the stated confidence **for every value of the nuisance parameter**
- if there is *any value* of the nuisance parameter which makes the data consistent with the parameter of interest, that value of the POI should be considered:
  - e.g. don't claim discovery if any background scenario is compatible with data
- But: technically very challenging and significant problems with over-coverage
  - Example: how broadly should 'any background scenario' be defined? Should we include background scenarios that are clearly incompatible with the observed data?

# Example of over-coverage

- The 1958 thought expt of David R. Cox focused the issue:
  - Your procedure for weighing an object consists of flipping a coin to decide whether to use a weighing machine with a 10% error or one with a 1% error; and then measuring the weight.
- Then "surely" the error you quote for your measurement should reflect which weighing machine you actually used, and not the average error of the "whole space" of all measurements!
- But this is not how the classical frequentist confidence interval works!
  - Suppose weight=100, coin='1% error' Can you exclude weight=90 at 95% C.L?
  - No: because for 'coin=10% error' weight=90 cannot be excluded at 95% C.L.
- Solution: conditioning on observed data will make result more relevant (at expense of exact frequentist coverage)
  - Restricting whole space of probabilities to 'coin=1% error' only if that is observed allows to exclude weight=90 at 95% C.L.

# The profile likelihood construction as compromise

- For LHC the following prescription is used:

  NPs

  Given L(μ,θ)

  POI

  perform hypothesis test for each value of μ (the POI),

  using values of nuisance parameter(s) θ that best fit the data under the hypothesis μ

- Introduce the following notation

$$\hat{\hat{\theta}}(\mu)$$

M.L. estimate of θ for a given value of μ
(i.e. a conditional ML estimate)

- The resulting confidence interval will have exact coverage for the points $(\mu, \hat{\hat{\theta}}(\mu))$
  - Elsewhere it may overcover or undercover (but this can be checked)

# The profile likelihood ratio

- With this prescription we can construct the profile likelihood ratio as test statistic

Likelihood for given μ

Maximum Likelihood for given μ

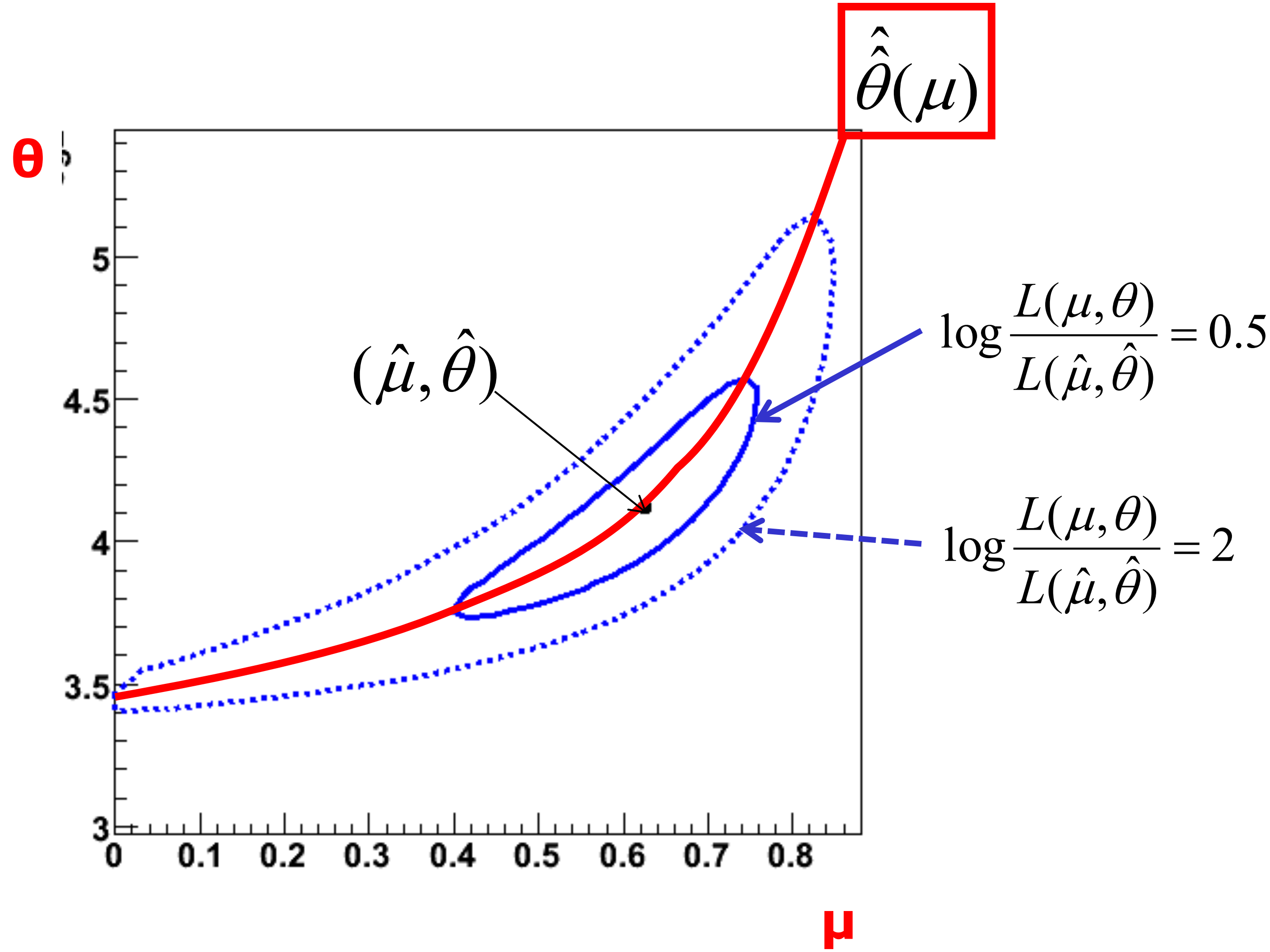
$$\lambda(\mu) = \frac{L(\mu)}{L(\hat{\mu})} \qquad \lambda(\mu) = \frac{L(\mu, \hat{\hat{\theta}}(\mu))}{L(\hat{\mu}, \hat{\theta})}$$

Maximum Likelihood

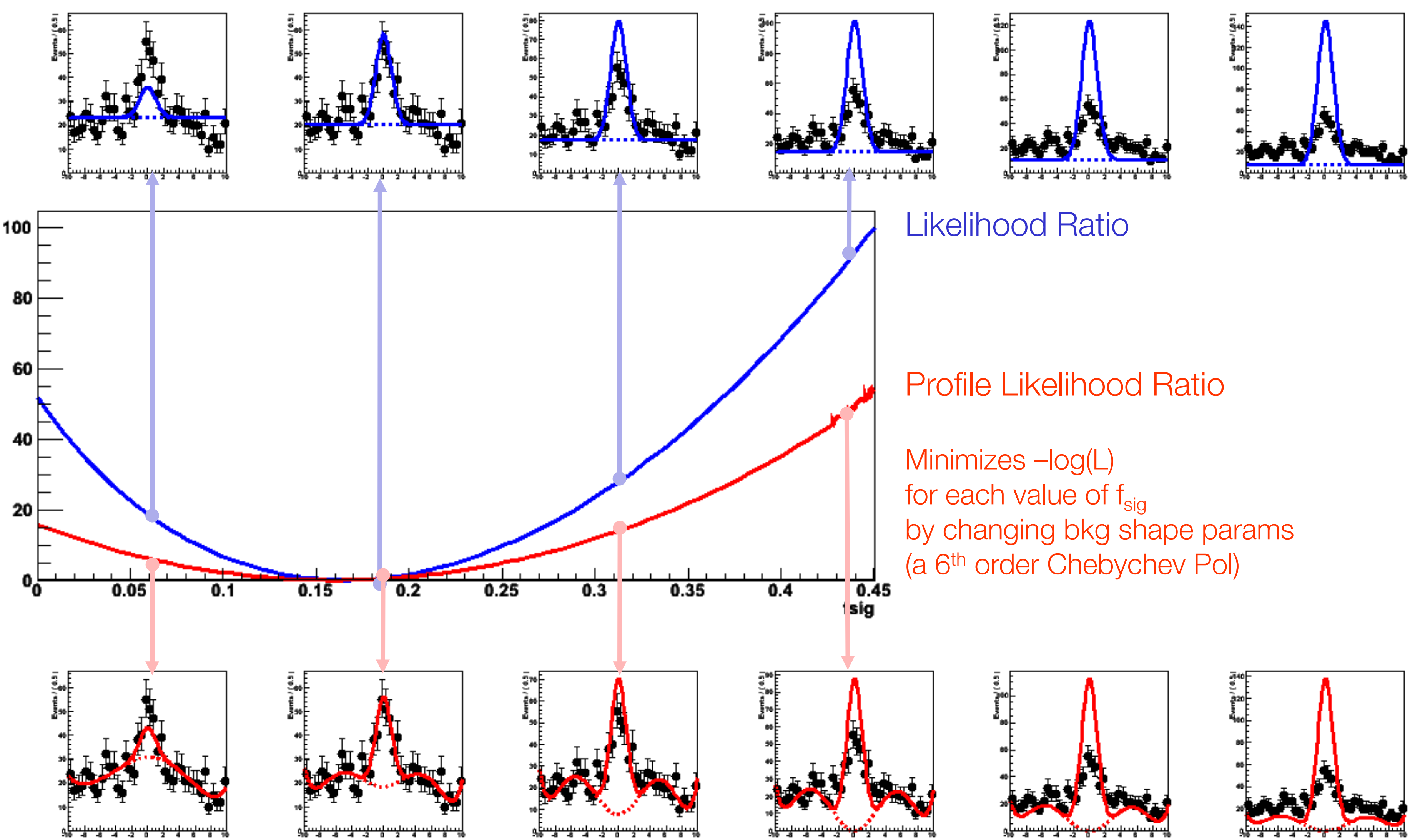
Maximum Likelihood

- NB: value profile likelihood ratio does *not* depend on θ

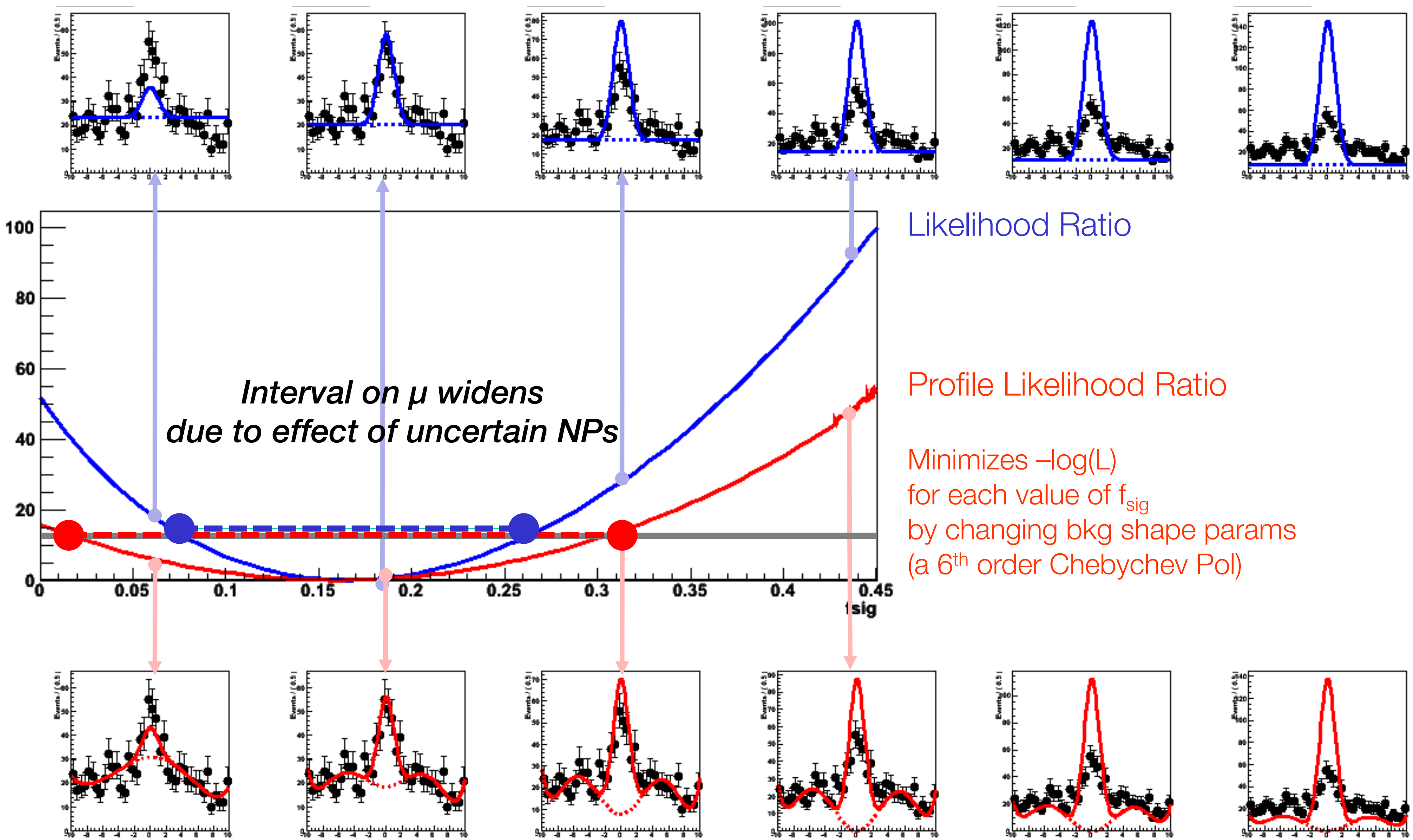
# Profiling illustration with one nuisance parameter



# Profile scan of a Gaussian plus Polynomial probability model



# Profile scan of a Gaussian plus Polynomial probability model




Likelihood Ratio

Profile Likelihood Ratio

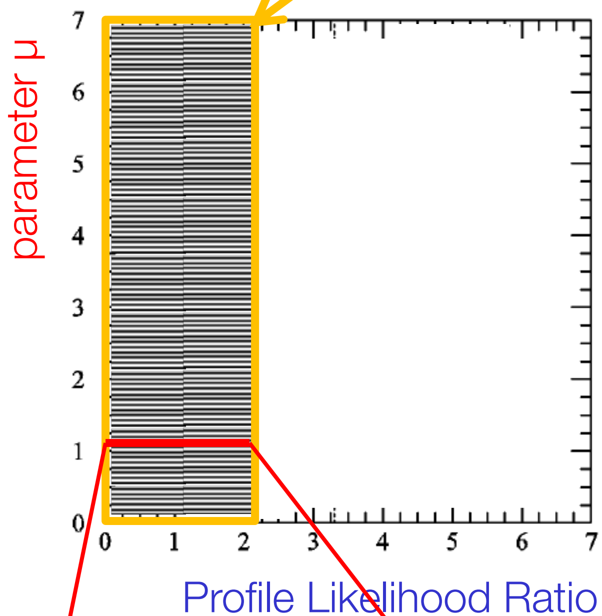
Minimizes –log(L)
for each value of $f_{sig}$
by changing bkg shape params
(a $6^{th}$ order Chebychev Pol)

# PLR Confidence interval vs MINOS

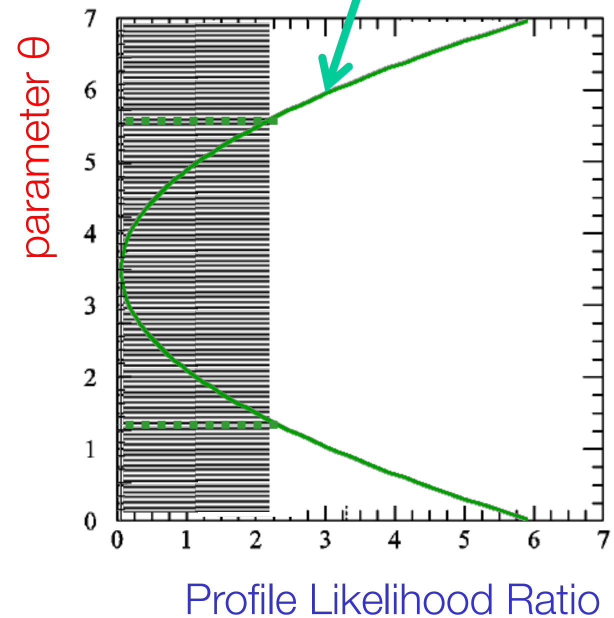
Confidence belt now range in PLR

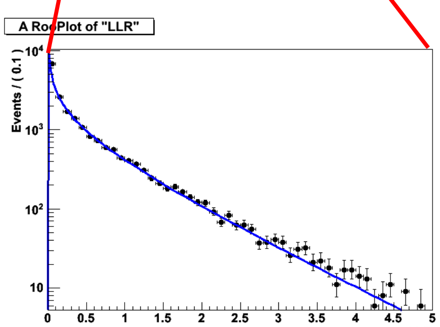
Measurement = $t_\mu(x_{obs},\mu)$ is now a function of μ

$t_\mu(x,\mu)$




$t_\mu(x,\mu)$

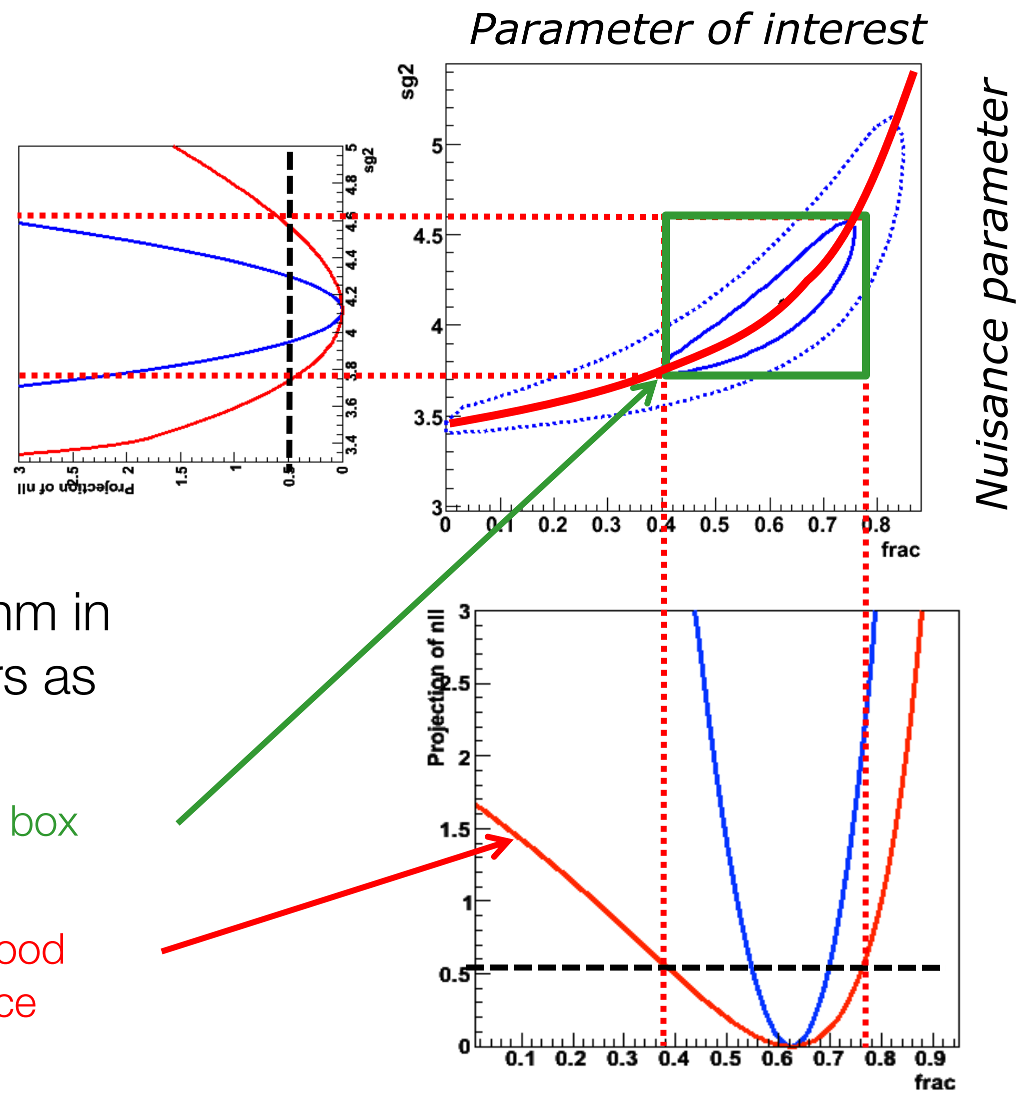



Asymptotically, distribution is identical for all μ

*NB: asymptotically, distribution is also independent of true values of θ*

# Link between MINOS errors and profile likelihood




- Note that MINOS algorithm in MINUIT gives same errors as Profile Likelihood Ratio
  - MINOS errors is bounding box around λ(s) contour
  - Profile Likelihood = Likelihood minimized w.r.t. all nuisance parameters

NB: Similar to graphical interpretation of variance estimators, but those always assume an elliptical contour from a perfectly parabolic likelihood

# Summary on NPs in confidence intervals

- Exact confidence intervals are difficult with nuisance parameters
  - Interval should cover for any value of nuisance parameters
  - Technically difficult and significant over-coverage common
- LHC solution Profile Likelihood ratio → Guaranteed coverage at *measured* values of nuisance parameters only
  - Technically replace likelihood ratio with profile likelihood ratio
  - Computationally more intensive (need to minimize likelihood w.r.t all nuisance parameters for each evaluation of the test statistic), but still very tractable
- Asymptotically confidence intervals constructed with profile likelihood ratio test statistics correspond to (MINOS) likelihood ratio intervals
  - As distribution of profile likelihood becomes asymptotically independent of θ, coverage for all values of θ restored

## Dealing with nuisance parameters in Bayesian intervals

- Elimination of nuisance parameters in Bayesian interval: Integrate over the full subspace of all nuisance parameters;

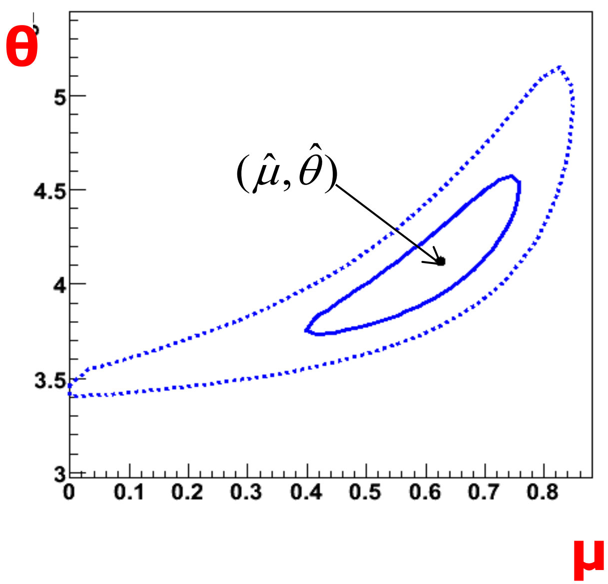
$$P(\mu \mid x) \propto L(x \mid \mu) \cdot \pi(\mu)$$

$$\Downarrow$$

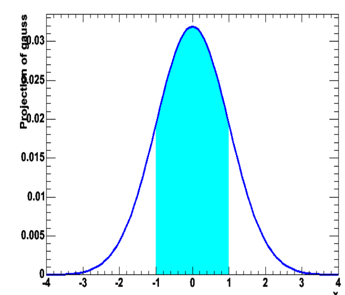
$$P(\mu \mid x) \propto \int \left( L(x \mid \mu, \vec{\theta}) \pi(\mu) \pi(\vec{\theta}) \right) d\vec{\theta}$$

- You are left with posterior pdf for μ

$\int$




$\times \, \pi(\mu, \theta) =$



Credible interval: area that integrates X% of posterior

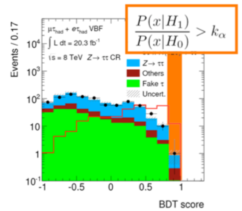
# Computational aspects of dealing with nuisance parameters

- Dealing with many nuisance parameters is computationally intensive in both Bayesian and (LHC) Frequentist approach
- Profile Likelihood approach
  - Computational challenge = ***Minimization*** *of likelihood w.r.t. all nuisance parameters for every point in the profile likelihood curve*
  - Minimization can be a difficult problem,
    e.g. if there are strong correlations, or multiple minima
- Bayesian approach
  - Computational challenge = ***Integration*** *of posterior density of all nuisance parameters*
  - Requires sampling of very potentially very large space.
  - Markov Chain MC and importance sampling techniques can help, but still very CPU consuming

# Nuisance parameters also impact event selection optimization!

## Choosing the 'best' high-signal region

- A common scenario for searches in a low-statistics regime is to perform a simplified analysis
  1. Train MVA to obtain discriminant D
  2. Apply a cut on D
  3. Perform only a counting analysis




- And a common question is then – what is the 'optimal cut on D'?
  – NB: the question arise due to choice for simplified
  If a *probability density model* is used for the analy
  'the full range of the discriminant'
  – To answer question a 'figure of merit' (FOM) mus
  the optimality of the selection. The ideal FOM for
  expected signal significance.

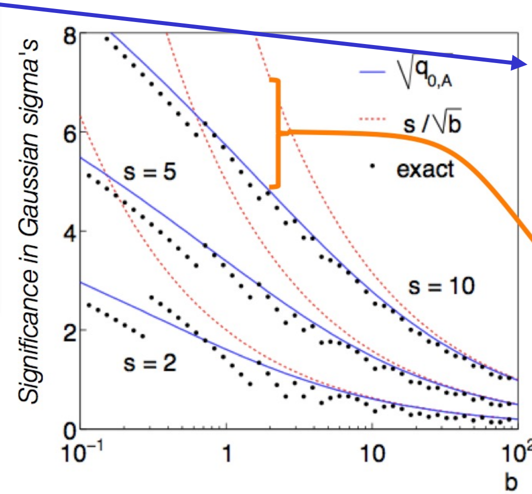
If the estimate of the background rate B is uncertain then

Figure of Merit $\sqrt{q_{0,A}}$ (and also $S/\sqrt{B}$)

*overestimate* counting model significance. Effect depends both on B and σ(B) → *can also effect location of optimum*

## Choosing the 'best' high-signal region

- The estimated significance assuming a Poisson process modeled by Poisson(N|S+B) is $\sqrt{2\left((s+b)\ln(1+s/b)-s\right)}$ .
- E.g. for 'discovery FOM' $s/\sqrt{b}$ illustration of approximation for s=2,5,10 and b in range [0.01-100] *shows significant deviations of $s/\sqrt{b}$ from actual significance at low b*



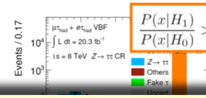

$$\sqrt{q_{0,A}} = \sqrt{2\left((s+b)\ln(1+s/b)-s\right)} \,.$$
$$= \frac{s}{\sqrt{b}}\left(1+\mathcal{O}(s/b)\right) \,.$$

Wouter Verkerke, NIKHEF

# Nuisance parameters also impact event selection optimization!

Choosing the ‘best’ high-signal region

- A common scenario for searches in a low-statistics regime is to perform a simplified analysis



$\frac{P(x|H_1)}{P(x|H_0)} > k_\alpha$

Can improve counting model significance estimate used as Figure of Merit *by including background uncertainty* (if known and sizable)

Approximate counting probability model with B uncertainty as

Poisson($N_{on}$|μS+B)**Poisson($N_{off}$|τB)**

NB: Assumes Poisson (not Gaussian) model for B uncertainty.
For x% fractional uncertainty on B choose

$$N_{off}=1/x^2 \quad \text{and} \quad \tau=N_{off}/B_{nom} \rightarrow \hat{B}=B_{nom}, \quad \sigma(\hat{B})=x\%$$

Signal significance for this model is analytically known in terms of the ‘Incomplete Beta funtion’

→ Easy to use implementation in ROOT (returns significance Z)

```
RooStats::NumberCountingUtils::BinomialObsZ(Double_t nObs,
                                            Double_t bExp, Double_t fracBUnc) ;
```

n

oisson process modeled

) .

f approximation for

s significant deviations of

$\sqrt{2\,((s+b)\ln(1+s/b)-s)}$ .

$\frac{s}{\sqrt{b}}\,(1+\underbrace{\mathcal{O}(s/b)}_{})$ .

b

# Summary of statistical treatment of nuisance parameters

- Each statistical method has an associated technique to propagate the effect of uncertain NPs on the estimate of the POI
  - Parameter estimation → Joint unconditional estimation
  - Variance estimation → Replace $d^2L/dp^2$ with Hessian matrix
  - Hypothesis tests & confidence intervals → Use profile likelihood ratio
  - Bayesian credible intervals → Integration ('Marginalization')

- Be sure to use the right procedure with the right method
  - Anytime you integrate a Likelihood you are a Bayesian
  - If you are minimizing the likelihood you are usually a Frequentist
  - If you sample something chances are you performing either a (Bayesian) Monte Carlo integral, or are doing glorified error propagation
- Answers can differ substantially between methods!
  - This is not always a problem, but can also be a consequence of a difference in the problem statement
- Don't forget large nuisance parameters in your event selection optimization