CAPITULO IV

FUNCIONES DE DENSIDAD DE PROBABILIDAD DE USO COMUN

Distribución	Uso(s) en altas energías	Otro(s) nombre(s)
Binomial	Tasa de decaimiento, eficiencias	Bernouilli
Poisson	Conteo de eventos	"ley de eventos raros"
Uniforme	Integración Monte-Carlo	
Exponencial	Vida media, tiempos de relajación	
Gaussiana	Resolución	Normal
Breit-Wigner	Resonancia	Cauchy, Lorentz
χ^2	"Goodness-of-fit"	"bondad de ajuste"

Lista no exhaustiva; otras de uso frecuente incluyen la distribución de Student, Galton o lognormal...

Ejemplo de una distribución discreta: Binomial (I)

Escenario con dos únicas realizaciones posibles: "éxito" y "fracaso", y con una probabilidad fija p de "éxito". Nos interesamos solamente en el número k de "éxitos" después de n intentos $0 \leq k \leq n$; (suponemos que la secuencia de intentos es irrelevante)
El número entero k sigue la distribución binomial $P(k; n, p)$:

$$P_{\text{binomial}}(k; n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} ,$$

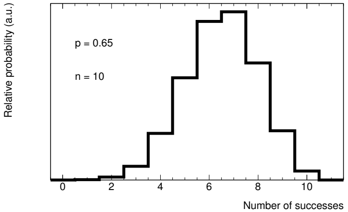
donde k es la variable aleatoria, mientras que n y p son parámetros.
Ejemplo típico: el número de eventos en una sub-categoría específica de eventos (p.e. una tasa de decaimiento)

Ejercicios :

- Verificar que la binomial está normalizada
- mostrar que la media y la varianza de una distr. binomial son

$$E[k] = \sum_{k=0}^{n} k P(k; n, p) = np ,$$
$$V[k] = np(1-p) .$$

- Sean dos binomiales con números de intentos n_1 y n_2 y de misma probabilidad p. Mostrar que la suma es también una binomial.



Teorema de la varianza de suma de binomiales:
La varianza de la suma de variables aleatorias que siguen distribuciones binomiales con probabilidades p_i diferentes,

$$k = \sum_i k_i \ ,$$

viene dada por

$$V[k] \quad = \quad n\overline{p}(1-\overline{p}) - ns^2 \ , \text{con } s^2 = \frac{1}{n}\sum_i (p_i - \overline{p})^2 \ ,$$

(donde $\overline{p}$ es el promedio de las probabilidades p_i), y es por tanto inferior o igual a la varianza de una variable binomial de probabilidad $\overline{p}$.
Ejercicio:
Verificar numéricamente la validez de este teorema, estimando con una simulación sencilla la varianza de la suma de dos variables aleatorias binomiales con probabilidades p_1 y p_2, y comparándola con la varianza de una variable aleatoria binomial con $p = (p_1 + p_2)/2$. Elegir varios valores de p_1 y p_2 que ilustren casos extremos del teorema.

Para la distribución binomial en el límite $n \to \infty$, $p \to 0$ (con $\lambda = np$ finito y no-nulo) la variable aleatoria k sigue la llamada distribución de Poisson $P(k;\lambda)$,

$$P_{\text{Poisson}}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!} ,$$

que tiene λ por único parámetro. For Poisson, the mean value and variance are the same:

$$E[k] = V[k] = \lambda .$$

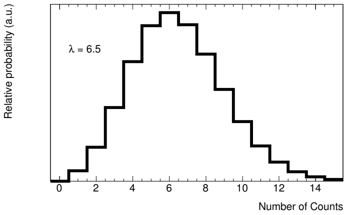
Esta distribución es también llamada “ley de los eventos raros” (debido al límite $p \to 0$), describe el número de observaciones en condiciones fijas si la tasa de ocurrencia es constante. Ejemplo: el número de decaimientos en un intervalo de tiempo fijo, para una fuente de actividad constante.

Ejercicios :

- mostrar que la media y la varianza de una Poisson son

$$E[k] = V[k] = \lambda .$$

- Sean dos distribuciones de Poisson con parámetros λ_1 y λ_2. Mostrar que la distribución de la suma es también una Poisson, con parámetro $\lambda = \lambda_1 + \lambda_2$.



Ejemplo de una distribución continua: Uniforme

Una variable aleatoria continua x, con densidad de probabilidad $P(x; a, b)$ constante y no-nula únicamente dentro de un intervalo finito $[a, b]$:

$$P_{\text{uniform}}(x; a, b) = \begin{cases} \frac{1}{b-a} & , \quad a \leq x \leq b \, , \\ 0 & , \quad x < a \text{ o } x > b \, . \end{cases}$$

Ejercicio : mostrar que para esta distribución uniforme, la media y la varianza son

$$E[x] = \frac{a+b}{2} \, , \; V[x] = \frac{(b-a)^2}{12} \, .$$

Uso cumún de la distribución uniforme : generación de distribuciones aleatorias

- "acepto-rechazo" para simular un proceso de probabilidad $p(x)$: utilizar dos variables aleatorias: x uniforme en el intervalo de interés, y x_0, uniforme en el intervalo $[0, 1]$. Aceptar el evento si $x_0 < p(x)$. Intuitivo y sencillo, pero usualmente poco eficaz.
- Transformada inversa: si se conoce la CDF de la distribución deseada, a partir de x uniforme en el intervalo $[0, 1]$, la variable aleatoria $z = CDF(x)$ sigue la distribución correspondiente.
- Transformación de Box-Muller: a partir de dos variables aleatorias x_1, x_2 uniformes en $[0, 1]$ e independientes, las variables z_1, z_2 dadas por

$$z_1 = \sqrt{-2 \ln x_1} \cos(2\pi x_2) \; , \; z_2 = \sqrt{-2 \ln x_1} \sin(2\pi x_2) \; ,$$

son independientes y siguen cada distribuciones Gaussianas, ambas de media nula y varianza unitaria.

Una variable aleatoria que sigue una densidad de probabilidad $P(x;\xi)$ dada por

$$P_{\text{exponential}}(x;\xi) = \begin{cases} \frac{1}{\xi}e^{-x/\xi} & , \quad x \geq 0 \, , \\ 0 & , \quad x < 0 \, , \end{cases}$$

y tiene por media y varianza

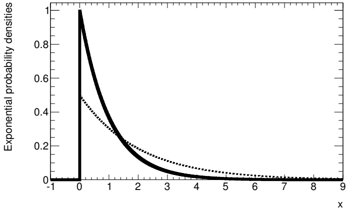
$$E[x] = \xi \, , \; V[x] = \xi^2 \, .$$

Ejercicio: evaluar la media y la varianza de una distribución exponencial truncada, que es no-nula solamente en un intervalo finito $a \leq x \leq b$. (Nota: verificar que la PDF que usen está correctamente normalizada)
La aplicación más común de esta distribución exponencial es la descripción de fenómenos independientes que se realizan a una tasa constante, como el tiempo de vida de una partícula inestable.

En vista del carácter auto-similar de la función exponencial:

$$P(t - t_0 | t > t_0) \;=\; P(t) \, ,$$

se dice que esta distribución es “sin memoria”.
La figura muestra dos PDF exponenciales, con parámetros $\xi = 1$ (línea sólida) y $\xi = 2$ (punteada).
Ejercicio: Obtener numéricamente una realización aleatoria de una distribución exponencial usando el método de la transformada inversa. Determinar el promedio y el RMS empíricos resultantes.

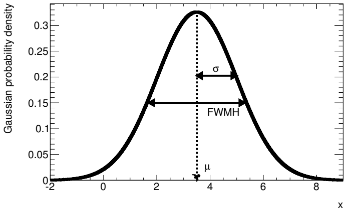


$$P_{\text{Gauss}}(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Para la distribución Normal (o Gaussiana), su media y varianza vienen dadas por

$$E[x] = \mu \ , \ V[x] = \sigma^2 .$$

Usamos deliberadamente los símbolos μ y σ tanto para los parámetros de la PDF Gaussiana como para su media y su varianza: la Gaussiana está caracterizada unívocamente por sus dos primeros momentos; $\mu_i = 0 \ \forall \ i > 2$.



La dispersión de una distribución a un solo máximo es en ocasiones caracterizada en términos de su *anchura a media altura* o FWHM (full width at half-maximum); para Gaussianas, la relación con la varianza es as $\mathrm{FWHM} = 2\sqrt{2\ln 2}\sigma \simeq 2{,}35\sigma$.
La Gaussiana es también la distribución límite para las binomiales y Poisson, en los límites a gran n y gran λ, respectivamente:

$$P_{\text{binomial}}(k; n\to\infty, p) \to P_{\text{Gauss}}(k; np, np(1-p)) ,$$
$$P_{\text{Poisson}}(k; \lambda\to\infty) \to P_{\text{Gauss}}\left(k; \lambda, \sqrt{\lambda}\right) .$$

Ejercicio:
Verificar que los momentos superiores de la Gaussiana son todos nulos.

Nota :, una corrección de continuidad es requerida: el rango de la Gaussiana se extiende a valores negativos, mientras que Binomial y Poisson están solamente definidas en el rango positivo.

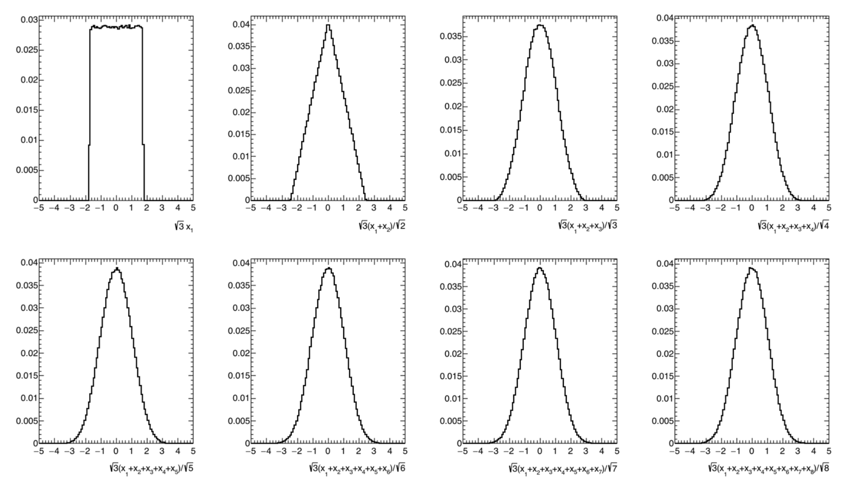
La distribución continua de mayor ubicuidad: la Gaussiana (II)

La primacía de la Gaussiana en términos de su relevancia conceptual y sus aplicaciones prácticas proviene en gran parte del *teorema del límite central* : si tenemos n variables aleatorias independendientes $\vec{x} = \{x_1, x_2, \ldots, x_n\}$, cada una con medias y varianzas μ_i y σ_i^2, la resultante $S(\vec{x})$ de sumarlas todas

$$S \;=\; \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{x_i - \mu_i}{\sigma_i} \;,$$

sigue una distribución que, en el límite de gran n, tiende a una distribución normal reducida.
(El caso particular $\mu = 0$, $\sigma = 1$ es llamado en ocasiones "normal reducida").
Por ello, una gran variedad de procesos, sujetos a múltiples fuentes independientes de incertidumbre, pueden describirse en buena aproximación con distribuciones Gaussianas, sin necesidad de conocer los detalles específicos de cada fuente de incertidumbre.
(ejemplo gráfico en la lámina siguiente, ilustrando la convergencia rápida del teorema del valor central para suma de distribuciones uniformes)
Ejercicio : verificar con una aplicación numérica la validez del teorema para otras distribuciones, por ejemplo sumas de exponenciales.

La distribución continua de mayor ubicuidad: la Gaussiana (III)



Ejemplo de una distribución continua: χ^2

$$P_{\chi^2}(x;n) = \begin{cases} \frac{x^{n/2-1}e^{-x/2}}{2^{n/2-1}\Gamma\left(\frac{n}{2}\right)} & , \quad x \geq 0 \text{ , } n \text{ entero} \\ 0 & , \quad x < 0 \text{ o } n \text{ no} - \text{entero} \text{ ,} \end{cases}$$

con único parámetro n, y donde $\Gamma(n/2)$ es la función Gamma. Su media y su varianza vienen dadas por

$$E[x] = n \text{ , } V[x] = 2n \text{ .}$$

La forma de la distribución de χ^2 depende de n, como se ve en la gráfica abajo que representa P_{χ^2} para $n = 1, \cdots, 6$.

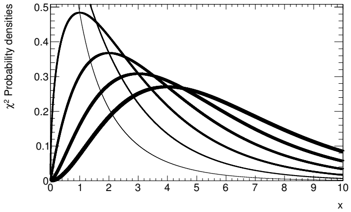
Definición importante: La función de χ^2 es la suma de cuadrados de n variables normales reducidas x_i, cada una de media μ_i y varianza σ_i^2:

$$\chi^2(x;n) = \sum_{i=1}^{n} \left(\frac{x_i - \mu_i}{\sigma_i}\right)^2 \text{ .}$$

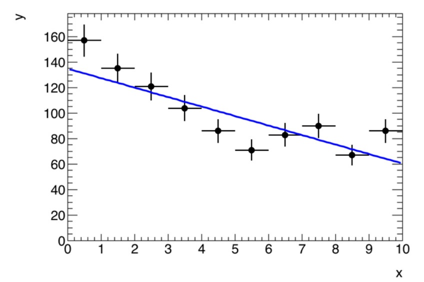
El parámetro n es también llamado "número de grados de libertad", y se refiere al comportamiento de los ajustes por el método de los cuadrados mínimos: cuando n_d puntos son utilizados para estimar n_p parámetros, el número correspondiente de grados de libertad es $n_d - n_p$.

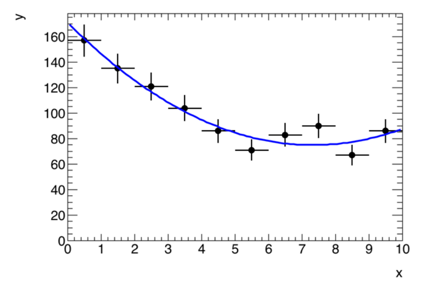
Si el modelo utilizado para ajustar es adecuado, los valores de los χ^2 obtenidos deben seguir la distribución P_{χ^2}.

Este criterio es llamado *goodness-of-fit* o "bondad del ajuste".



Tomemos una serie de 10 medidas de observables x e y, que muestran una clara anticorrelación. Suponemos que x es la variable *independiente*, y es la dependiente. Suponemos también que solamente las incertidumbres sobre y son pertinentes en este ejemplo.

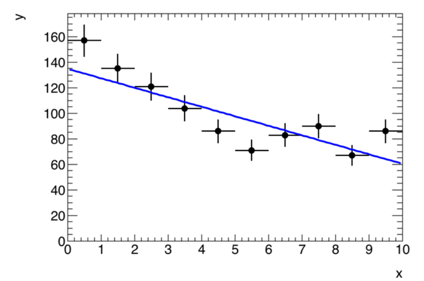


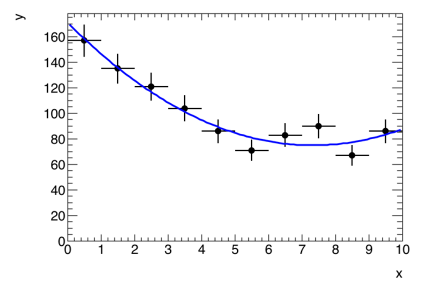


Discusión: ¿a qué corresponden las dos curvas azules?
Ejercicio: reproducir los ajustes mostrados en la lámina siguiente. Los valores precisos son: $y = [157; 135; 121; 104; 86; 71; 83; 90; 67; 86]$ y los errores en y son la raíz cuadrada de su valor.

Un ejemplo sencillo ("regresión") de un ajuste de χ^2

Tomemos una serie de 10 medidas de observables x e y, que muestran una clara anticorrelación. Suponemos que x es la variable *independiente*, y es la dependiente. Suponemos también que solamente las incertidumbres sobre y son pertinentes en este ejemplo.





Un ajuste a un polinomio de orden 1 da :
$\chi^2 = 23{,}1$ para 8 grados de libertad.

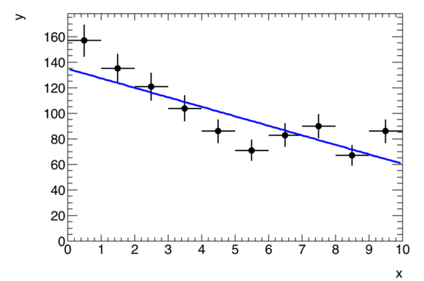
```
******************************************
Minimizer is Linear
Chi2                       =      23.1088
NDf                        =            8
p0                         =      134.861    +/-    7.02278
p1                         =     -7.43446    +/-    1.13342
******************************************
Prob(23.1,8) = 0.00322733
```

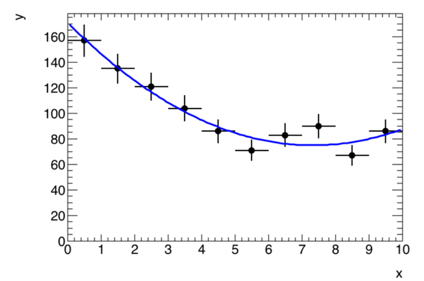
Mientras que un ajuste a un polinomio de orden 2 da :
$\chi^2 = 6{,}5$ para 7 grados de libertad.

```
******************************************
Minimizer is Linear
Chi2                       =      6.51604
NDf                        =            7
p0                         =      170.778    +/-    11.2722
p1                         =     -26.1437    +/-    4.73078
p2                         =      1.78313    +/-    0.43774
******************************************
Prob(6.5,7) = 0.480943
```

Un ejemplo sencillo ("regresión") de un ajuste de χ^2

Tomemos una serie de 10 medidas de observables x e y, que muestran una clara anticorrelación. Suponemos que x es la variable *independiente*, y es la dependiente. Suponemos también que solamente las incertidumbres sobre y son pertinentes en este ejemplo.





Comentarios:

- claramente, mientras más "flexible" sea la función utilizada, "mejor" será el ajuste
- pero, ¡ese es rara vez el objetivo! (claro, puedo imaginar algún contraejemplo...)
- idealmente, la decisión sobre la función a utilizar debe ser tomada, previamente, **sin mirar los datos**

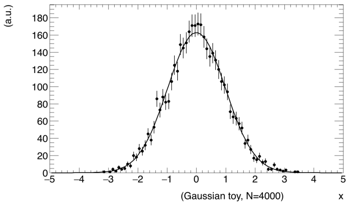
Otro ejemplo de ajuste de χ^2: un histograma

De manera general, la función de χ^2 que se minimiza es de esta forma:

$$\chi^2(\vec{\theta}) = \sum_{i=1}^{n} \left(\frac{y_i - f(x_i;\vec{\theta})}{\sigma_i} \right)^2 ,$$

donde $\vec{\theta} = \{\theta_1, \theta_2, \cdots\}$ son los parámetros que deben ser determinados de manera a minimizar el valor de χ^2.
En los ejemplos anteriores, teníamos

- p0 y p1 para el ajuste a un polinomio de orden uno
  - en este caso la minimización consiste en resolver un sistema lineal de 2 ecuaciones con dos incógnitas, la solución es analítica
- p0, p1 y p2 para el ajuste a un polinomio de orden dos



Otra situación usual es la de un histograma:

- aquí las incertidumbres asignadas a cada "bin" corresponden a la varianza en la altura del bin
- si se trata de conteos simples lo usual es asignar $\sigma_i = \sqrt{y_i}$
- cuidado sin embargo con los bines con pocas entradas...
- ¡y sobre todo con los bines vacíos!

Ejemplo de una distribución continua: Breit-Wigner (I)

También llamada distribución de Cauchy, o Lorentziana, la distribución

$$P_{\mathrm{BW}}(x;\Gamma,x_0) = \frac{1}{\pi}\frac{\Gamma/2}{(x-x_0)^2+\Gamma^2/4},$$

es a menudo utilizada para describir un proceso resonante (p.e. la masa invariante de productos del decaimiento de un estado intermedio resonante), con lo que x_0 y Γ son la masa y la anchura natural, que es su FWHM.
La BW es un ejemplo de una distribución *"fat-tailed"*:

- promedio empírico y RMS mal definidos (idem momentos superiores)
- sus varianzas aumentan con el tamaño de la muestra
- ejercicio : verificar con una simulación numérica que el RMS empírico de la Breit-Wigner aumenta (siguiendo una ley de potencia) con el rango en el que es estimado (ver gráfica en la lámina siguiente)
- un teorema curioso (que no demostraremos) el estimador de x_0 menos ineficaz es una media truncada sobre el 24 % central de la muestra : otras truncaciones son menos eficaces
- ejercicio opcional : verificar este teorema con una simulación numérica

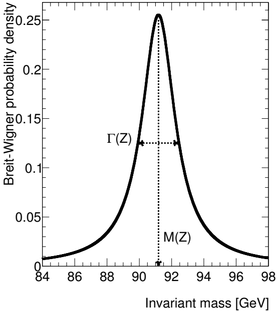
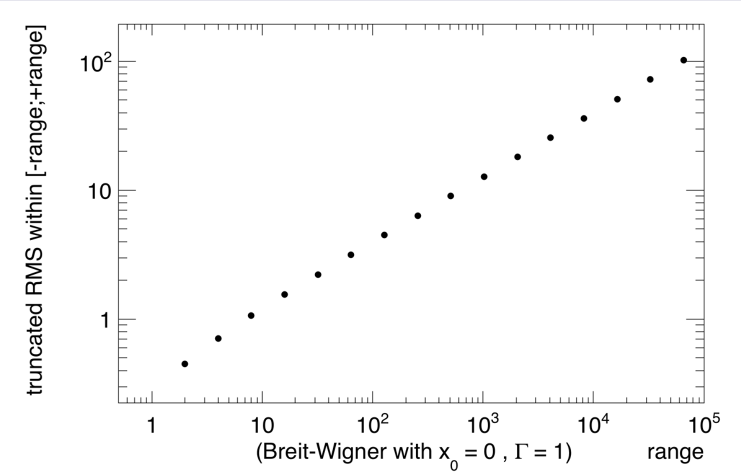


Gráfico : una Breit-Wigner con la masa y la anchura natural del bosón Z :
$M_Z = 91{,}1876 \pm 0{,}0021\ \mathrm{GeV}/c^2$, $\Gamma_Z = 2{,}4952 \pm 0{,}0023\ \mathrm{GeV}/c^2$.

Ejemplo de una distribución continua: Breit-Wigner (II)



Ejemplo de una distribución continua: Voigtiana (I)

La Voigtiana es la convolución de una Breit-Wigner con una Gaussiana,

$$P_{\text{Voigt}}\left(x; x_0, \Gamma, \sigma\right) \;=\; \int_{-\infty}^{+\infty} dx' P_{\text{Gauss}}\left(x'; 0, \sigma\right) P_{\text{BW}}\left(x - x'; x_0, \Gamma\right) \;\;,$$

- La Voigtiana es una distribución a tres parámetros : masa x_0, anchura natural Γ y resolución σ.
- Es un modelo de medidas de procesos resonantes, suponiendo que la resolución instrumental sea Gaussiana.
- No hay una forma analítica cerrada sencilla para la Voigtiana, pero hay implementaciones numéricas precisas y eficientes, p.e. la función miembro `TMath::Voigt` en ROOT, o la clase `RooVoigtian` en `RooFit`.
- Para valores de Γ y σ razonablemente similares, la FWHM de la Voigtiana se aproxima a una combinación en suma y en cuadratura :

$$\text{FWHM}_{\text{Voigt}} \;\simeq\; \left[(\Gamma/2) \oplus 2\sqrt{2\ln 2}\sigma\right] + \Gamma/2 \;.$$

Ejercicio :

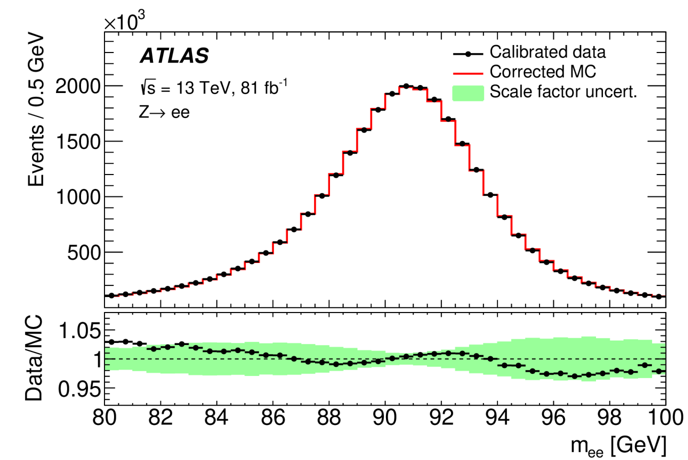
- La figura en la lámina siguiente muestra la distribución de masa invariante dielectrón alrededor de la masa del bosón Z, obtenida con datos ATLAS a 13 TeV.
- Suponiendo que la distribución puede ser descrita por una Voigtiana, determinar aproximadamente cuál es la resolución en masa dielectrón del detector ATLAS.
- La parte inferior de la figura muestra una comparación entre los datos y la simulación. Comentar.

$$\mathrm{FWHM}_{\mathrm{Voigt}} \simeq \Gamma/2 + \left[(\Gamma/2) \oplus 2\sqrt{2\ln 2}\sigma\right]$$

$M_Z = 91{,}1876 \pm 0{,}0021\ \mathrm{GeV}/c^2$
$\Gamma_Z = 2{,}4952 \pm 0{,}0023\ \mathrm{GeV}/c^2$
$\sigma(m_{ee}) = ?$
En palabras simplificadas :

$$\mathrm{FWHM}_{Z\to ee} \simeq \Gamma_Z \otimes \sigma_{\mathrm{exp.}}$$

- para la física, el parámetro de interés es Γ_Z
- pero lo que se mide es $\mathrm{FWHM}_{Z\to ee}$
- !'se requiere efectuar un "unfolding" o *desconvolución* de la resolución experimental $\sigma_{\mathrm{exp.}}$!



Otros ejemplos de distribuciones continuas

La "Gaussiana Bifurcada" : útil para describir distribuciones asimétricas

$$P_{\mathrm{BG}}\left(x;\mu,\sigma_L,\sigma_R\right) = \sqrt{\frac{2}{\pi}}\frac{1}{\sigma_L+\sigma_R}\times\begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma_L^2}} & , \quad x\leq\mu \\ e^{-\frac{(x-\mu)^2}{2\sigma_R^2}} & , \quad x\geq\mu \end{cases}$$

La "Crystal Ball" : en ocasiones utilizada para describir resoluciones con una componente de "fuga" (leakage)

$$P_{\mathrm{CB}}(x;\mu,\sigma,\alpha,n) = N\times\begin{cases} e^{-\frac{(x-\mu)^2}{2\sigma^2}} & , \quad \frac{x-\mu}{\sigma}\leq\alpha \\ \left(\frac{n}{n-\alpha^2+\alpha\sigma^{-1}(x-\mu)}\right)^n e^{\left(-\frac{\alpha^2}{2}\right)} & , \quad \frac{x-\mu}{\sigma}\geq\alpha \end{cases}$$

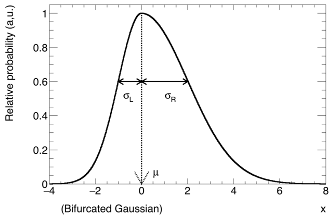
Funciones implementadas p.e. en `RooFit`

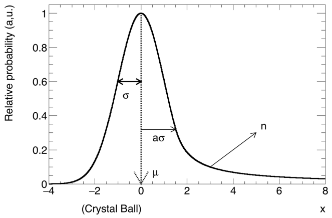
Interpretación "intuitiva" de los parámetros, pero

- los estimadores simples están sesgados
- correlaciones muy grandes entre ellos
- utilizar, pero con prudencia y ojo crítico

Ejercicio numérico : evaluar las matrices de correlación

Otras funciones similares : DSCB, ACB ...





Otro ejemplo proveniente de una convolución

Una situación usual es cuando una distribución de tipo exponencial "a nivel de la verdad" es observada "a nivel experimental" con una resolución σ comparable a su la escala característica ξ :

$$P_{\rm obs}(x;\xi,x_0,\sigma) = P_{\rm true}(x;\xi,x_0) \otimes P_{\rm Gauss}(x;0,\sigma)\ ,\ \ \text{con}\ \ P_{\rm true}(x;\xi,x_0) = \frac{1}{\xi}e^{-(x-x_0)/\xi}\ ,$$

que puede mostrarse fácilmente (¡ejercicio!) que tiene la forma siguiente:

$$P_{\rm obs}(x;\xi,x_0,\sigma) \propto e^{-\frac{\left(x-\frac{\sigma^2}{2\xi}\right)}{\xi}} \left[1+{\rm erf}\left(\frac{x-\sigma^2/\xi}{\sqrt{2}\sigma}\right)\right]\ ,{\rm erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x dz e^{-z^2}\ .$$

donde ${\rm erf}(x)$ es llamada "función de error".