

Probabilidad y estadística para la física experimental

José Ocariz

Université Paris Cité

ocariz@in2p3.fr

22 de febrero de 2023



Latin American alliance for
Capacity building in Advanced physics
LA-CoNGA physics



Cofinanciado por el
programa Erasmus+
de la Unión Europea





- ▶ El libro clásico de referencia (912 páginas) :
 - ▶ Stuart, K. Ord, S. Arnold, Kendall's Advanced theory of statistics Volume 2A : Classical Inference and the Linear Model, John Wiley & Sons, 2009
- ▶ Libros de estadísticas, escritos por físicos de partículas :
 - ▶ L. Lyons, Statistics for Nuclear and Particle Physics, Cambridge, 1986
 - ▶ G. Cowan, Statistical Data Analysis, Clarendon Press, Oxford 1998
(ver también http://www.p.rhu1.ac.uk/~cowan/stat_course.htm)
 - ▶ R.J. Barlow, A Guide to the Use of Statistical Methods in the Physical Sciences John Wiley & Sons, 1989
 - ▶ F. James, Statistical Methods in Experimental Physics, World Scientific, 2006
- ▶ El PDG también es una fuente conveniente para referencias rápidas :
 - ▶ 2020 Review of Particle Physics, P.A. Zyla et al. (Particle Data Group), Prog. Theor. Exp. Phys. 2020, 083C01 (2020) "Mathematical Tools" section
(ver también <https://pdg.lbl.gov/>)
- ▶ Las grandes colaboraciones internacionales tienen foros y grupos de trabajo, con muchos enlaces y referencias.



- ▶ Este curso se inspira ampliamente de mi experiencia personal como físico experimental de partículas
 - ▶ el lenguaje, las notaciones, etc... reflejan los usos y costumbres de mi área temática
 - ▶ otras comunidades tienen convenciones y definiciones diferentes!
 - ▶ en ocasiones indicaré diferencias en notaciones (cuando las conozco...)
- ▶ De la misma manera, las herramientas a las que estoy más acostumbrado provienen de la física de partículas
 - ▶ nunca he usado **R**
 - ▶ trabajo más a menudo con C++ que con python, pero ambos me van
 - ▶ mi herramienta cotidiana para análisis y visualización es **ROOT** (<https://root.cern/>)
 - ▶ existe una interfaz PyROOT que enlaza python con el C++ nativo de **ROOT**
 - ▶ para análisis estadísticos más elaborados (funciones de verosimilitud) utilizo el paquete RooFit
 - ▶ y en cuanto a Machine Learning, yo uso sobre todo **TMVA**, a Toolkit for MultiVariate data Analysis
 - ▶ nota : **TMVA** no incluye ciertas herramientas modernas tipo GAN, GNN, etc...
 - ▶ instalé TensorFlow en mi móvil, pero más por curiosidad que otra cosa...
 - ▶ pienso también probar con PyTorch, pero eso sigue en mi lista de tareas pendientes...
- ▶ Los ejercicios y tareas del curso pueden ser trabajados bajo forma de notebooks Jupyter
 - ▶ el uso de **ROOT** no es obligatorio (aunque por supuesto me sería más fácil a mí...)
 - ▶ me esfuerzo en que (buena parte de) los ejercicios no requieran funciones súper-específicas de HEP...
- ▶ Este documento retoma el material preparado para los cursos *LA-CoNGA physics* de los dos años opasados
 - ▶ Si bien ya cubre todos los capítulos del programa, sin duda ampliaré algunos segmentos a lo largo del curso



Este curso *no* es un curso de Machine Learning (aunque hablaremos de Machine Learning)

Este curso *no* es un curso sobre Teoría de Decisiones (que es una aplicación muy importante de la estadística)

Este curso *sí* trata de la inferencia científica, con el objetivo de pasar de lo intuitivo hacia una lógica más formal

- ▶ Cuando lees un artículo con un resultado así:

$$m_{\text{top}} = 173,34 \pm 0,27(\text{stat.}) \pm 0,71(\text{syst.})\text{GeV}$$

¿Cómo interpretas ese resultado?

- ▶ Cuando en otro artículo, te topas con una afirmación del estilo

Gluino and squark masses up to 1.5 TeV are excluded at 95 % C.L.

¿Cómo interpretas esa afirmación?

- ▶ Cuando en otro artículo, lees una frase del estilo

The most significant deviation with respect to the background-only hypothesis is observed for a mass of 19.35 GeV, corresponding to a local significance of $3,1\sigma$

¿Cómo interpretas esa frase?



CAPITULO I

PROBABILIDAD MATEMATICA



- ▶ ¿Qué es para tí la probabilidad? ¿Cómo la definirías?
- ▶ La probabilidad matemática es un concepto axiomático abstracto, desarrollado por Kolmogorov (1933) y otros
- ▶ La teoría de la probabilidad es el marco conceptual para el estudio de los procesos aleatorios
- ▶ Un proceso es llamado aleatorio si satisface dos condiciones :
 - ▶ su realización (un "evento") no puede ser predicha con total certeza ;
 - ▶ si el proceso se repite bajo las mismas condiciones, cada nueva realización puede ser diferente
- ▶ Es usual clasificar las fuentes de incertidumbre según su origen :
 - ▶ *reducibles* : errores en la la medición, p.e. limitaciones prácticas que en principio pueden ser mejoradas (mejores instrumentos, mejor control de las condiciones experimentales) ;
 - ▶ *cuasi-irreducibles* : errores aleatorios en la medición, como efectos térmicos o de turbulencia ;
 - ▶ *fundamentales* : cuando el proceso físico es intrínsecamente incierto (mecánica cuántica).

En física subatómica experimental, los tres tipos de incertidumbre deben ser considerados. Notar en particular :

- ▶ los eventos resultantes de colisiones de partículas son independientes, y son un ejemplo perfecto de procesos aleatorios de origen cuántico
- ▶ las partículas inestables obedecen probabilidades de desintegración descritas por la mecánica cuántica

Ejercicio : dar ejemplos de procesos físicos para cada una de las fuentes de incertidumbre mencionadas arriba.



Sea Ω el universo total de posibles realizaciones de un proceso aleatorio, y sean $X, Y \dots$ elementos de Ω . Una función de probabilidad \mathcal{P} se define como un mapa en los números reales :

$$\begin{aligned}\mathcal{P} : \{\Omega\} &\rightarrow [0 : 1] , \\ X &\rightarrow \mathcal{P}(X) .\end{aligned}$$

Ese mapeo debe satisfacer los siguientes axiomas :

$$\begin{aligned}\mathcal{P}(\Omega) &= 1 , \\ \text{si } X \cap Y &= \emptyset , \text{ entonces } \mathcal{P}(X \cup Y) = \mathcal{P}(X) + \mathcal{P}(Y) ,\end{aligned}$$

de los cuales se pueden derivar varias propiedades útiles, p.e. (donde \bar{X} es el complemento de X)

$$\begin{aligned}\mathcal{P}(\bar{X}) &= 1 - \mathcal{P}(X) , \\ \mathcal{P}(X \cup \bar{X}) &= 1 , \\ \mathcal{P}(\emptyset) &= 1 - \mathcal{P}(\Omega) = 0 , \\ \mathcal{P}(X \cup Y) &= \mathcal{P}(X) + \mathcal{P}(Y) - \mathcal{P}(X \cap Y) ,\end{aligned}$$



Probabilidad condicional, teorema de Bayes

La probabilidad condicional $\mathcal{P}(X | Y)$ se define como la probabilidad de X , dado Y

- ▶ equivale a restringir el universo Ω a la muestra Y .

El ejemplo más sencillo de probabilidad condicional es para realizaciones independientes :

- ▶ dos elementos X e Y son independientes (sus realizaciones no estén relacionadas en ninguna manera) si

$$\mathcal{P}(X \cap Y) = \mathcal{P}(X)\mathcal{P}(Y).$$

- ▶ por lo tanto, si X e Y son independientes, se satisface la condición

$$\mathcal{P}(X | Y) = \mathcal{P}(X)$$

El teorema de Bayes cubre el caso general : en vista de la relación $\mathcal{P}(X \cap Y) = \mathcal{P}(Y \cap X)$, se obtiene que

$$\mathcal{P}(X | Y) = \frac{\mathcal{P}(Y | X)\mathcal{P}(X)}{\mathcal{P}(Y)} .$$

Un corolario útil del teorema de Bayes : si Ω puede dividirse en un número de submuestras disjuntas X_i (una "partición"), entonces

$$\mathcal{P}(X | Y) = \frac{\mathcal{P}(Y | X)\mathcal{P}(X)}{\sum_i \mathcal{P}(Y | X_i)\mathcal{P}(X_i)} .$$



CAPITULO II

VARIABLES ALEATORIAS

FUNCIONES DE DENSIDAD DE PROBABILIDAD



Variables aleatorias, funciones de densidad de probabilidad (I)

El escenario más relevante para nosotros es cuando la realización de un proceso aleatorio se presenta en forma numérica (p.e. corresponde a una medición) : a cada elemento X corresponde una variable x (real o entera). Para x continuo, su función de densidad de probabilidad (PDF) $P(x)$ se define como :

$$\mathcal{P}(X \text{ en } [x, x + dx]) = P(x)dx ,$$

donde $P(x)$ es definida-positiva para todo de x , y satisface la condición de normalización

$$\int_{-\infty}^{+\infty} dx' P(x') = 1 .$$

Para x_i discreto, la definición es similar :

$$\begin{aligned} \mathcal{P}(X \text{ en } x_i) &= p_i , \\ \text{con } \sum_j p_j &= 1 \text{ y } p_k \geq 0 \forall k . \end{aligned}$$

Probabilidades finitas se obtienen por integración sobre un rango no-infinitesimal,

$$\mathcal{P}(a \leq X \leq b) = \int_a^b dx' P(x') .$$

Ejercicio: determinar el coeficiente de normalización de la función Gaussiana, $g(x; \mu, \sigma) = e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Y mejor aún: normalizar la Gaussiana en un intervalo truncado, con $a \leq x \leq b$.



En ocasiones es conveniente referirse a la función de densidad acumulativa (CDF) :

$$C(x) = \int_{-\infty}^x dx' P(x') ,$$

de modo que las probabilidades finitas corresponden a evaluar la CDF en los bordes del rango de interés :

$$\mathcal{P}(a \leq X \leq b) = \int_a^b dx' P(x') = C(b) - C(a) .$$

Una PDF no puede ser completamente arbitraria :

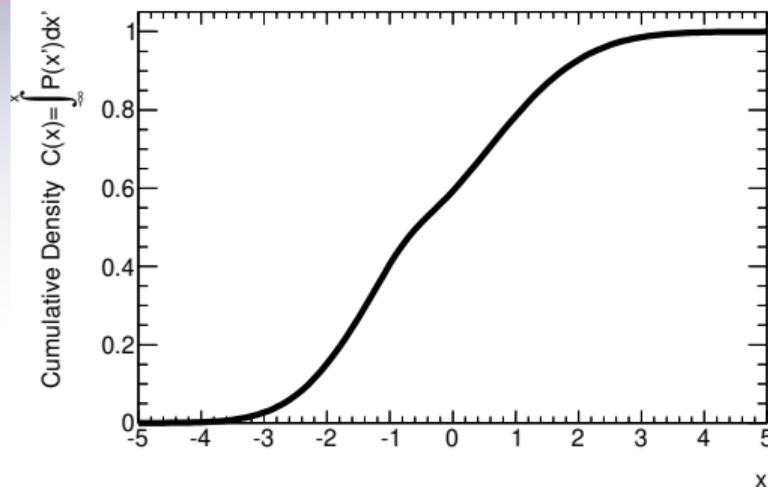
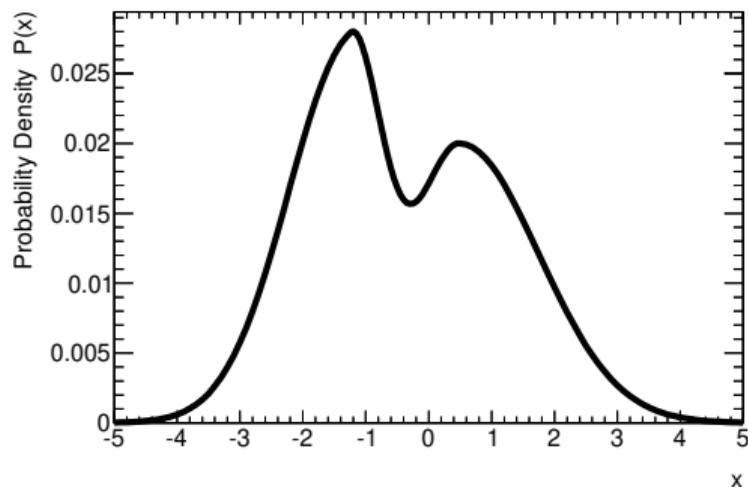
- ▶ debe satisfacer la condición de normalización previamente indicada
- ▶ debe ser definida positiva
- ▶ debe ser de soporte acotado, con valores despreciables fuera de una región finita

Fuera de esas condiciones, una PDF puede ser arbitraria, p.e. exhibir uno o varios máximos locales, tener discontinuidades...

En contraste, la CDF es una función monótonicamente creciente de x .
(ver ejemplo en la lámina siguiente)



Variables aleatorias, funciones de densidad de probabilidad (III)



Un ejemplo arbitrario de PDF con un máximo global y un segundo máximo local, y su CDF correspondiente. Fuera del intervalo en el gráfico, el valor de la PDF es totalmente despreciable.

Ejercicio : Suponer que nuestra PDF $P(x)$ es la suma de dos PDFs Gaussianas

$$P(x; \mu_1, \sigma_1, \mu_2, \sigma_2, f) = fG(x; \mu_1, \sigma_1) + (1 - f)G(x; \mu_2, \sigma_2) .$$

1. verificar que el parámetro f está limitado al intervalo $[0 : 1]$;
2. verificar que $P(x)$ está normalizada;
3. para los valores siguientes: $\mu_1 = -1$, $\sigma_1 = 1$, $\mu_2 = 1$, $\sigma_2 = 2$, $f = 0,5$. Representar gráficamente $P(x)$ y su CDF.



PDFs multidimensionales (I)

Para un evento descrito por un conjunto n -dimensional de elementos $\vec{X} = \{X_1, X_2, \dots, X_n\}$ y su correspondiente conjunto de variables aleatorias $\vec{x} = \{x_1, x_2, \dots, x_n\}$, tenemos su PDF multidimensional :

$$P(\vec{x})d\vec{x} = P(x_1, x_2, \dots, x_n)dx_1dx_2 \dots dx_n .$$

PDFs de menor dimensionalidad pueden derivarse por integración de ciertas variables. Por ejemplo, para una variable específica $x = x_j$ su densidad de probabilidad marginal unidimensional $P_X(x)$ es :

$$P_X(x)dx = dx \int_{-\infty}^{+\infty} dx_1 \dots \int_{-\infty}^{+\infty} dx_{j-1} \int_{-\infty}^{+\infty} dx_{j+1} \dots \int_{-\infty}^{+\infty} dx_n .$$

Caso bidimensional, con elementos X, Y y variables aleatorias $\vec{X} = \{x, y\}$. La probabilidad finita en un rango bidimensional rectangular es

$$P(a \leq X \leq b ; c \leq Y \leq d) = \int_a^b dx \int_c^d dy P(x, y) .$$

Para un valor fijo de Y , la función de densidad condicional de X es

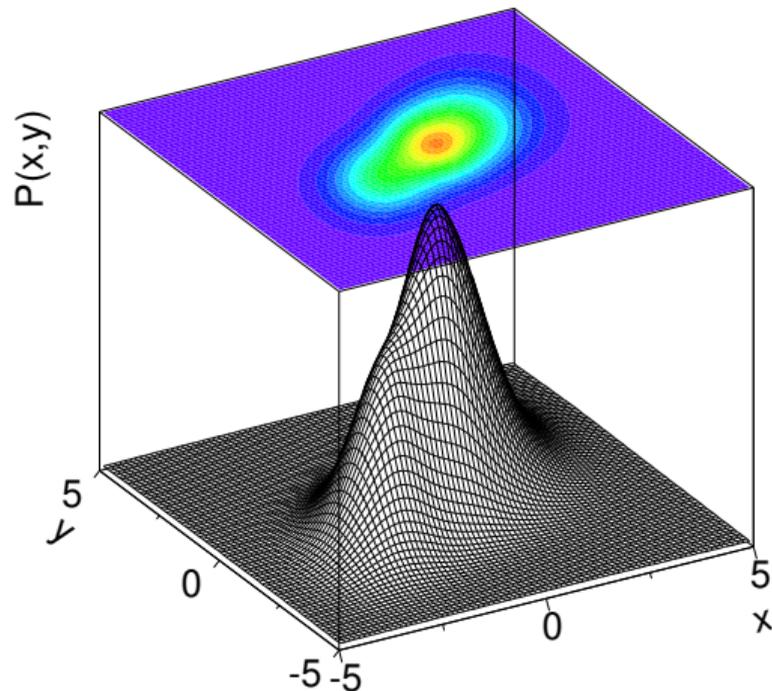
$$P(x | y) = \frac{P(x, y)}{\int dx P(x, y)} = \frac{P(x, y)}{P_Y(y)} .$$

De nuevo, la relación $P(x, y) = P_X(x) \cdot P_Y(y)$ solamente es válida para X, Y independientes.



Ejemplo de una función de densidad bidimensional con variables no-independientes,
 $P(x, y) \neq P_X(x) \cdot P_Y(y)$.

Ejercicio : Producir un gráfico que represente una PDF bidimensional razonablemente similar a la ilustrada aquí.





Modelo : descripción de un proceso aleatorio

Modelo analítico = descripción de un proceso aleatorio con funciones analíticas para las PDFs

Modelo paramétrico : sus PDFs pueden describirse completamente usando un número finito de parámetros

- ▶ este requisito no es obligatorio; las PDFs pueden también ser no-paramétricas (equivalente a suponer que se necesita un número infinito de parámetros), o pueden ser mixtas

Una implementación sencilla de una PDF paramétrica es cuando sus parámetros son argumentos analíticos de la función de densidad ; la notación

$$P(x, y, \dots ; \theta_1, \theta_2, \dots)$$

indica la dependencia funcional o *forma* de la PDF en términos de variables x_1, y_2, \dots y parámetros $\theta_1, \theta_2, \dots$

Consideremos una variable aleatoria X con PDF $P(x)$. Para una función genérica $f(x)$, su *valor de expectación* $E[f]$ es su promedio ponderado sobre el rango cubierto por x :

$$E[f] = \int dx P(x) f(x) = \frac{\int dx P(x) f(x)}{\int dx P(x)} .$$

Como describiremos más adelante, los parámetros de una PDF pueden ser estimados a partir de ciertos valores de expectación.



Valores de expectación (II)

Por ser de uso frecuente, algunos valores de expectación tienen nombre propio.

Para PDFs unidimensionales, la *media* y la *varianza* se definen así :

$$\text{Media} \quad : \quad \mu = \quad E[x] = \int dx P(x)x ,$$

$$\text{Varianza} \quad : \quad \sigma^2 = \quad V[x] = E[x^2] - \mu^2 = E[(x - \mu)^2] ;$$

y la *desviación estándar* σ es la raíz cuadrada de la varianza.

Para PDFs multidimensionales, la matriz de *covarianza* $C_{ij} = C(x_i, x_j)$ y la matriz adimensional de correlación lineal ρ_{ij} se definen así :

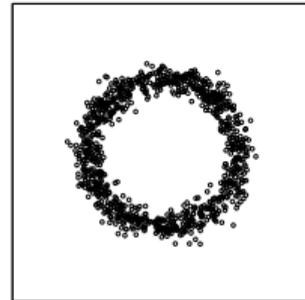
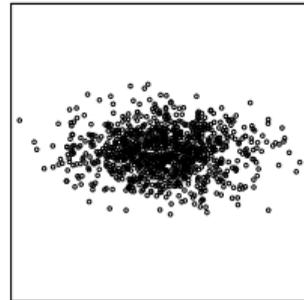
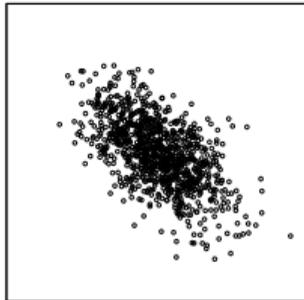
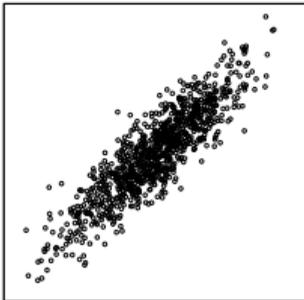
$$C_{ij} = E[x_i x_j] - \mu_i \mu_j = E[(x_i - \mu_i)(x_j - \mu_j)] , \quad \rho_{ij} = \frac{C_{ij}}{\sigma_i \sigma_j} .$$

Los coeficientes de correlación lineal tienen valores en el rango $-1 \leq \rho_{ij} \leq 1$, e indican la tendencia dominante de densidad en el patrón $(x_i; x_j)$ pattern: se habla de correlaciones positivas y negativas (o anti-correlaciones).

Para variables aleatorias X_i, X_j independientes, es decir con $P(x_i, x_j) = P_{X_i}(x_i)P_{X_j}(x_j)$, se tiene

$$E[x_i x_j] = \int \int dx_i dx_j P(x_i, x_j) x_i x_j = \mu_i \mu_j , \quad \longrightarrow \quad \rho_{ij} = 0 .$$

(pero la converso no es necesariamente cierta, p.e. ejemplo en la lámina siguiente)



De izquierda a derecha :

- ▶ $\rho = +0,9$,
- ▶ $\rho = -0,5$;
- ▶ $\rho = 0$, para variables independientes ,
- ▶ variables fuertemente correlacionadas con un patrón no lineal de correlación que “conspira” para arrojar una correlación lineal nula, $\rho = 0$.



CAPITULO III

CARACTERIZACION DE FORMAS

ESTIMACION DE PARAMETROS

PROPAGACION DE ERRORES



Caracterización de la forma de una PDF (I)

Estamos suponiendo que nuestros objetos de estudio de procesos aleatorios, que se manifiestan bajo forma de realizaciones aleatorias a partir de una PDF subyacente

- ▶ En la práctica, la verdadera dependencia funcional de una PDF es a menudo desconocida
- ▶ La información sobre su forma solamente puede extraerse a partir de una muestra de talla finita (digamos que contiene N eventos), es decir suponemos que la muestra disponible es una realización aleatoria a partir de una PDF desconocida.
- ▶ Si consideramos que esa PDF subyacente es de tipo paramétrico, la *caracterización de su forma* es un procedimiento para estimar los valores numéricos de sus parámetros, partiendo de una hipótesis “razonable” sobre la dependencia funcional sobre sus variables.
- ▶ Ahora, solamente un número finito de valores de expectación independientes pueden extraerse de una muestra de talla finita.
- ▶ No existe una receta única para la elección de los parámetros a ser estimados, con lo que el proceso es intrínsecamente incompleto.
- ▶ Se puede sin embargo confirmar en la práctica que el proceso de caracterización de forma es bastante poderoso, si los parámetros seleccionados proveen información útil y complementaria.

Ilustramos estas consideraciones con un ejemplo unidimensional para una variable aleatoria única x .



Caracterización de la forma de una PDF (II)

Ilustramos las consideraciones anteriores con un ejemplo unidimensional para una variable aleatoria única x . Consideremos el promedio empírico \bar{x} , definido como

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i .$$

Mostraremos más adelante que \bar{x} es un buen *estimador* de la media μ de la PDF subyacente $P(x)$. De modo análogo, la media cuadrática RMS (del inglés “root-mean-squared”), definida como

$$\text{RMS}^2 = \overline{x^2} - (\bar{x})^2 , \text{ con } \overline{x^2} = \frac{1}{N} \sum_{i=1}^N x_i^2 ,$$

es un estimador razonable de la varianza σ^2 (veremos más adelante una mejor definición).

En términos intuitivos, el promedio y el RMS reflejan información útil y complementaria sobre la “localización” y la “dispersión” de la región de x mayor densidad de eventos, y esta región debe corresponder de manera aproximada a los intervalos de x donde la PDF tiene valores más grandes.

Obviamente, en un caso general esos dos parámetros son insuficientes para caracterizar una PDF más genérica, que requiere un procedimiento más sistemático.



Caracterización de la forma de una PDF (III)

Para un procedimiento más sistemático, transformamos la x -dependencia de la PDF $P(x)$ en una k -dependencia de la *función característica* $C[k]$, definida como

$$C[k] = E \left[e^{ik \frac{x-\mu}{\sigma}} \right] = \sum_j \frac{(ik)^j}{j!} \mu_j .$$

Como se puede notar, la función característica es la transformación de Fourier de la PDF. Los coeficientes μ_j de la expansión *are* se llaman *momentos reducidos*; por construcción, los primeros momentos son $\mu_1 = 0$ y $\mu_2 = 1$; en términos de la variable reescalada $x' = (x - \mu)/\sigma$, la PDF fue desplazada para tener promedio nulo, y escalada para tener varianza unitaria.

En principio, mientras mayor el número de momentos μ_j sean estimados, más detallada será la caracterización de la forma de la PDF (pero una muestra finita solamente permite medir un número finito de momentos).

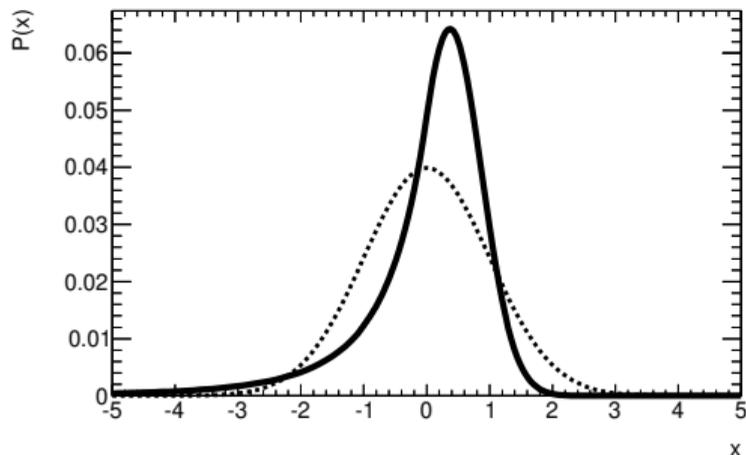
Los momentos 3 y 4 tienen nombres específicos, y sus valores son sencillos de interpretar en términos de la forma:

- ▶ el tercer momento es llamado oblicuidad (o *skewness* en inglés)
 - ▶ una distribución simétrica tiene skewness nula,
 - ▶ un valor negativo (positivo) indica una “anchura” mayor a la izquierda (derecha) de su media.
- ▶ el cuarto momento es llamado *kurtosis*
 - ▶ cantidad definida positiva, relacionada con cuán “picante” es la distribución
 - ▶ un valor pequeño indica un pico estrecho y “colas” de largo alcance: es una distribución leptokúrtica
 - ▶ un valor grande indica un pico central ancho y colas poco prominentes: es una distribución platykúrtica

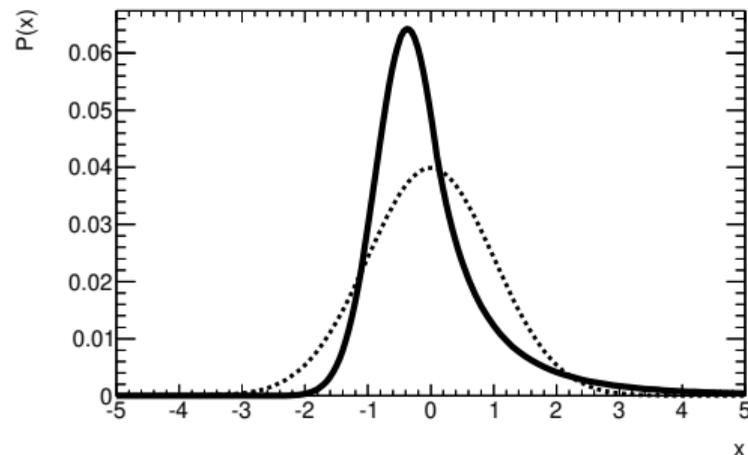


Ejemplos: skewness y kurtosis

(la curva punteada es una Gaussiana reducida, con $\mu_1 = 0$, $\mu_2 = 1$, y $\mu_i = 0 \forall i > 2$)



Cola “no-Gaussiana” a la izquierda del pico:
skewness positiva, $\mu_3 > 0$

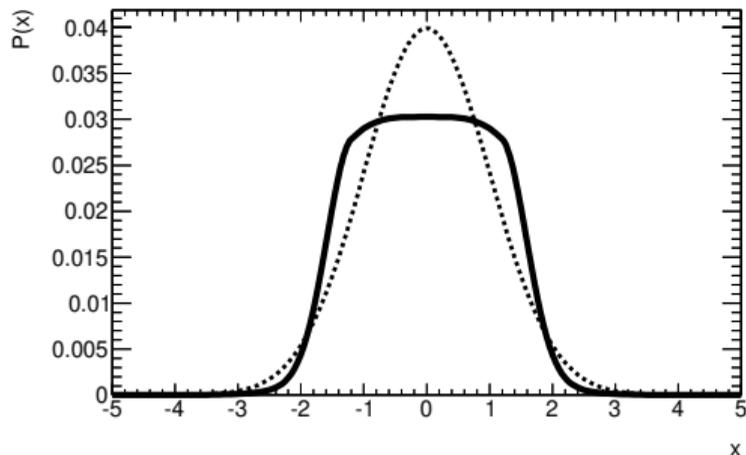


Cola “no-Gaussiana” a la derecha del pico:
skewness negativa, $\mu_3 < 0$

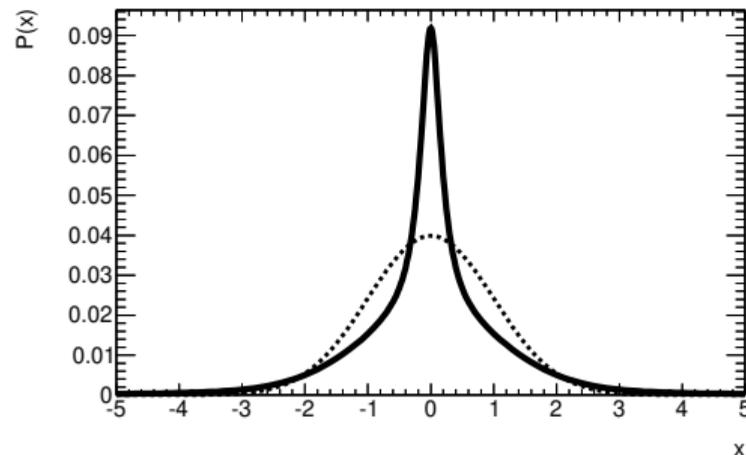


Ejemplos: skewness y kurtosis

(la curva punteada es una Gaussiana reducida, con $\mu_1 = 0$, $\mu_2 = 1$, y $\mu_i = 0 \forall i > 2$)



Pico ancho, colas de corto alcance :
kurtosis pequeña (“platykurtic”)



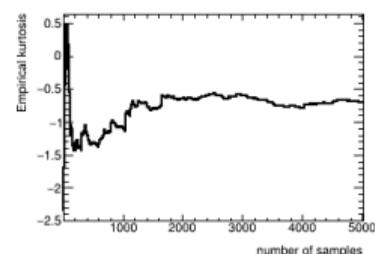
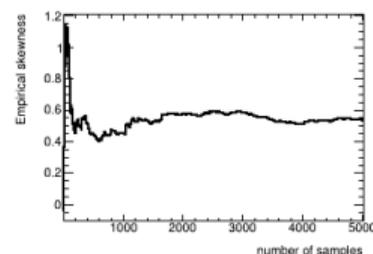
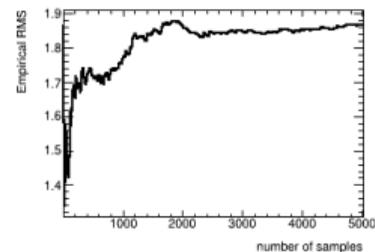
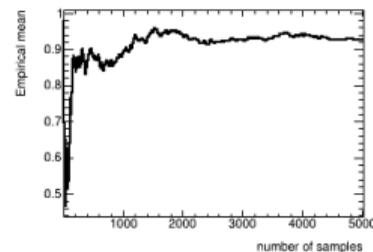
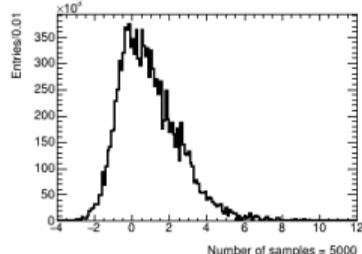
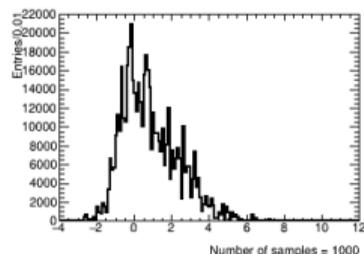
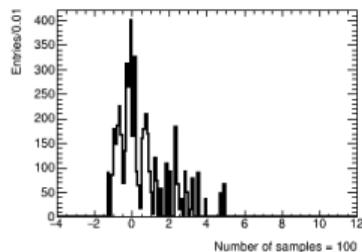
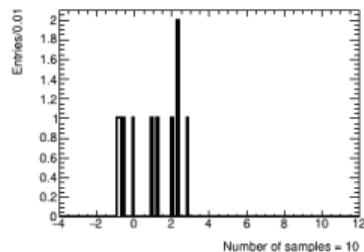
Pico estrecho, colas de largo alcance :
kurtosis grande (“leptokurtic”)



Estimación de momentos con muestras de talla finita

Un ejemplo : a partir de una misma PDF, se realizan $N = 10, 100, 1000, 5000$ realizaciones de una variable x :

- ▶ se grafican las distribuciones de frecuencia correspondientes (figuras a la izquierda),
- ▶ se estiman los valores de los 4 primeros momentos : media, RMS, skewness y kurtosis (figuras a la derecha).





La caracterización de la forma de una PDF a través de una estimación secuencial de parámetros de forma nos permitió introducir de manera cualitativa al concepto de estimación de parámetros (también llamado en inglés “point estimation”). Una receta más general sería la siguiente :

Consideremos una PDF n -dimensional, k -paramétrica,

$$P(x_1, x_2, \dots, x_n ; \theta_1, \theta_2, \dots, \theta_k) ,$$

para la cual queremos estimar los valores $\theta_1, \dots, \theta_k$ a partir de una muestra de talla finita, utilizando un conjunto de estimadores $\hat{\theta}_1, \dots, \hat{\theta}_k$.

Esos estimadores son también variables aleatorias, con sus propias medias y varianzas: sus valores diferirían al ser estimados sobre otras muestras.

Esos estimadores deben satisfacer dos propiedades clave:

- ▶ *ser consistente* : asegura que, en el límite de una muestra de talla infinita, el estimador converge al verdadero valor del parámetro;
- ▶ *ser no sesgado* : la ausencia de sesgo asegura que el valor de expectación del estimador es el verdadero valor del parámetro, para toda talla de la muestra.

Un estimador sesgado pero consistente (también llamado asintóticamente no-sesgado) es tal que el sesgo disminuye al aumentar la talla de la muestra.

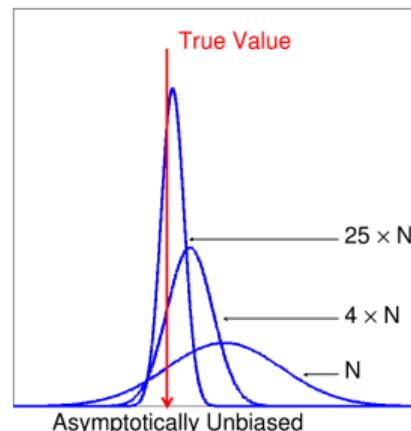
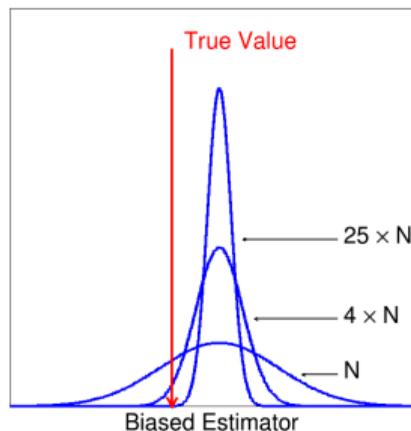
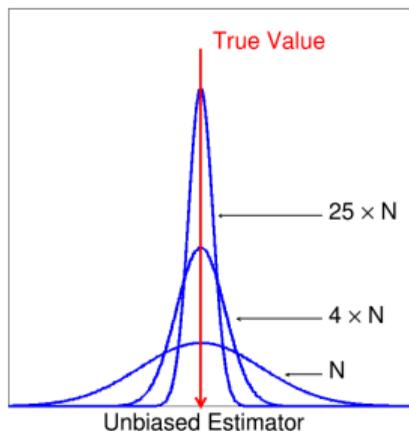


Estimación de parámetros (I bis)

Otros criterios son útiles para caracterizar la calidad de los estimadores; por ejemplo

- ▶ eficiencia: un estimador de pequeña varianza es más eficiente que uno de mayor varianza ;
- ▶ robusteza: este criterio describe la “sensibilidad” del estimador a incertidumbres en la forma de la PDF. Por ejemplo, el promedio es robusto contra incertidumbres sobre los momentos de orden par, pero es menos robusto contra incertidumbres en los momentos de orden impar.

Nota : estos criterios son en ocasiones mutuamente contradictorios; por razones prácticas, puede ser preferible tener un estimador eficiente pero sesgado, a uno no sesgado pero de pobre convergencia.





Estimación de parámetros (II)

El promedio empírico \bar{x} es un estimador convergente, no sesgado de la media μ de la PDF subyacente: $\hat{\mu} = \bar{x}$. Esto se demuestra fácilmente, evaluando el valor de expectación y la varianza de \bar{x} :

$$E[\bar{x}] = \frac{1}{N} \sum_{i=1}^N E[x] = \mu ,$$
$$V[\bar{x}] = E[(\bar{x} - \mu)^2] = \frac{\sigma^2}{N} .$$

Al contrario, el RMS empírico de una muestra es un estimador sesgado (aunque asintóticamente no-sesgado) de la varianza σ^2 . Esto se demuestra fácilmente también, reescribiendo su cuadrado en términos de la media:

$$\text{RMS}^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 = \left[\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \right] - (\bar{x} - \mu)^2 ,$$

de manera que su valor de expectación es

$$E[\text{RMS}^2] = \sigma^2 - V[\bar{x}] = \frac{N-1}{N} \sigma^2 ,$$

que si bien converge a la verdadera varianza σ^2 en el límite $N \rightarrow \infty$, subestima sistemáticamente su valor para muestras de talla finita.



Estimación de parámetros (III)

Si bien converge a la verdadera varianza σ^2 en el límite $N \rightarrow \infty$, el RMS subestima sistemáticamente su valor para muestras de talla finita. Pero es inmediato definir un estimador modificado

$$\frac{N}{N-1} \text{RMS}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 ,$$

que es, para muestras de talla finita, un estimador no sesgado de la varianza.

En resumen: para una PDF desconocida, tenemos estimadores consistentes y no sesgados de su media μ y su varianza σ^2 , que pueden ser extraídos de muestras de talla finita:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i ,$$
$$\hat{\sigma}^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \hat{\mu})^2 .$$

Nota: los factores $1/N$ y $1/(N-1)$ para $\hat{\mu}$ y $\hat{\sigma}^2$ se entienden intuitivamente: el promedio empírico puede medirse incluso en la muestra más pequeña posible de un solo evento, mientras que al menos dos eventos son necesarios para estimar la dispersión empírica de una muestra.

Ejercicio : completar los detalles relacionando la media y la varianza con sus estimadores empíricos.



El ejemplo clásico anterior trataba de una única variable aleatoria.

En presencia de múltiples variables aleatorias $\vec{x} = \{x_1, \dots, x_n\}$, la generalización del resultado anterior lleva a definir la covarianza empírica, cuyos elementos \hat{C}_{ab} son estimados en una muestra de N eventos de la manera siguiente:

$$\hat{C}_{ab} = \frac{1}{N} \sum_{i=1}^N (x_{a,i} - \hat{\mu}_a) (x_{b,i} - \hat{\mu}_b) .$$

(los índices a, b recorren la lista de variables aleatorias, $1 \leq a, b \leq n$). (Nota: para N pequeños, se debe corregir un sesgo en este estimador, c.f. el ejemplo del RMS)

Suponiendo que la verdadera covarianza es conocida, la varianza de una función arbitraria $f(\vec{x})$ de las variables aleatorias se evalúa a partir de la expansión de Taylor alrededor de las medias de sus parámetros $\hat{\vec{\mu}}$ según

$$f(\vec{x}) = f(\hat{\vec{\mu}}) + \sum_{a=1}^n \left. \frac{df}{dx_a} \right|_{\vec{x}=\hat{\vec{\mu}}} (x_a - \hat{\mu}_a) ,$$

en otras palabras, $E[f(\vec{x})] \simeq f(\hat{\vec{\mu}})$.



De manera similar,

$$E [f^2(\vec{x})] \simeq f^2(\hat{\vec{\mu}}) + \sum_{a,b=1}^n \left. \frac{df}{dx_a} \frac{df}{dx_b} \right|_{\vec{x}=\hat{\vec{\mu}}} \hat{C}_{ab} ,$$

con lo que la varianza de f se estima como

$$\hat{\sigma}_f^2 \simeq \sum_{a,b=1}^n \left. \frac{df}{dx_a} \frac{df}{dx_b} \right|_{\vec{x}=\hat{\vec{\mu}}} \hat{C}_{ab} .$$

Esta expresión, llamada *fórmula de propagación de incertidumbres*, estima la varianza de una función genérica $f(\vec{x})$ a partir de los estimadores de sus medias y covarianzas.

Ejemplos particulares de propagación de incertidumbres:

- ▶ cuando todas las variables aleatorias $\{x_a\}$ son no-correlacionadas, la matriz de covarianza es diagonal, $C_{ab} = \sigma_a^2 \delta_{ab}$ y la covarianza de $f(\vec{x})$ se reduce a

$$\hat{\sigma}_f^2 \simeq \sum_{a=1}^n \left(\left. \frac{df}{dx_a} \right|_{\vec{x}=\hat{\vec{\mu}}} \right)^2 \hat{\sigma}_a^2 .$$



Propagación de incertidumbres (III)

- ▶ para la suma de dos variables aleatorias $S = x_1 + x_2$, la varianza es

$$\sigma_S^2 = \sigma_1^2 + \sigma_2^2 + 2C_{12} = \sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2\rho_{12},$$

- ▶ la generalización a más de dos variables es:

$$\sigma_S^2 = \sum_{a,b} \sigma_a\sigma_b\rho_{ab} .$$

En ausencia de correlaciones, se dice que *los errores absolutos se suman "en cuadratura"*:

$$\sigma_S = \sigma_1 \oplus \sigma_2 = \sqrt{\sigma_1^2 + \sigma_2^2} ,$$

y si la correlación vale 1, el error en la suma es la suma directa de los errores de cada término.

Ejemplos :

- ▶ para $x_1 = (1,0 \pm 0,3)$ y $x_2 = (2,5 \pm 0,4)$ (errores similares), la suma $S = x_1 + x_2$ es $S = (3,5 \pm 0,5)$ (contribuciones similares de ambos errores) ;
- ▶ para $x_1 = (1,0 \pm 0,9)$ y $x_2 = (2,5 \pm 0,1)$ (errores muy diferentes), la suma es $S = (3,5 \pm 0,9)$ (el error total está dominado por el mayor) ;



- ▶ para el producto de dos variables aleatorias $P = x_1 x_2$, la varianza relativa es

$$\left(\frac{\sigma_P}{P}\right)^2 = \left(\frac{\sigma_1}{x_1}\right)^2 + \left(\frac{\sigma_2}{x_2}\right)^2 + 2\frac{\sigma_1}{x_1}\frac{\sigma_2}{x_2}\rho_{12} ,$$

- ▶ la generalización a más de dos variables:

$$\left(\frac{\sigma_P}{P}\right)^2 = \sum_{a,b} \frac{\sigma_a}{x_a} \frac{\sigma_b}{x_b} \rho_{ab} .$$

En ausencia de correlaciones, se dice que *los errores relativos se suman en cuadratura*:

$$\frac{\sigma_P}{P} = \frac{\sigma_1}{x_1} \oplus \frac{\sigma_2}{x_2} ,$$

y si la correlación vale 1, el error relativo sobre el producto es la suma directa de los errores relativos de cada término.



- ▶ para una función genérica en ley de potencia, $Z = x_1^{n_1} x_2^{n_2} \dots$, si todas las variables son no-correlacionadas, la varianza es

$$\frac{\sigma_Z}{Z} = n_1 \frac{\sigma_1}{x_1} \oplus n_2 \frac{\sigma_2}{x_2} \oplus \dots$$

En otras palabras: en la suma en cuadratura, los errores relativos son ponderados por los exponentes. Por tanto los términos con exponentes importantes (cuadrados, cubos, etc...) dominarán el *presupuesto de error*, mientras que términos con raíces (cuadradas, cúbicas, etc...) tendrán un impacto subdominante.