Probabilidad y estadística para la física experimental

José Ocariz

LA-CoNGA physics

ocariz@in2p3.fr

2 de marzo de 2022

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

Cofinanciado por el programa Erasmus+ de la Unión Europea

Elementos bibliográficos

- 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.rhul.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.

Algunas consideraciones iniciales

- 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**
  - 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**, `Toolkit for MultiVariate data Analysis`
  - instalé `TensorFlow` en mi móvil, pero más por curiosidad que otra cosa...
- 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 ampliamente material preparado para el curso *LA-CoNGA physics* del año pasado
  - Si bien ya cubre todos los capítulos del programa, sin duda ampliaré algunos segmentos a lo largo del curso

- Cuándo 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?
- Cuándo en otro artículo, te topas con una afirmación del estilo

$$\mathcal{BR}\left(B_s \to \mu^+\mu^-\right) < 1{,}1 \times 10^{-8}\ ,\ 95\,\%\ \text{C.L.}$$

  ¿Cómo interpretas esa afirmación?
- Cuándo 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

Probabilidad matemática (I)

- ¿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 resultances 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.

Probabilidad matemática (II)

Sea Ω el universo total de posibles realizaciones de un proceso aleatorio, y sean $X, Y \ldots$ 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 :

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

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

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

La probabilidad condicional $\mathcal{P}(X \mid Y)$ se define como la probability 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

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

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

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

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

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