Teoría de probabilidad: Una introducción para el curso de Física Moderna

Juan Gancio (jgancio@fisica.edu.uy)

Marzo 2022

Las siguientes notas resumen brevemente los conocimiento necesarios sobre teoría de probabilidad para el curso de Física Moderna. Las mismas se encuentran basadas en las notas de los cursos de "Probabilidad y Estadística" dictado por el Dr. Juan Kalamkerian en 2015, y "Análisis de datos: información y complejidad" dictado por el Dr. Nicolás Rubido en 2019. No se pretende ser exhaustivo en los temas tratados, sino que se presenta una breve introducción a ciertos conceptos necesarios para el abordaje de los temas del curso. Estos conceptos son introducidos considerando variables discretas, para luego generalizarlos para el tratamiento de variables continuas, las que serán de mayor interés para el curso.

1 Probabilidad: definiciones generales

Definición: Experimento. La noción de probabilidad esta ligada a un **experimento**. La realización de dicho experimento da lugar a un resultado. La repetición del experimento puede dar lugar a varios resultados diferentes.

Definición: Espacio de muestra. Al conjunto de todos los resultados posibles del experimento se le llama espacio de muestra o espacio muestral, comúnmente referido como Ω.

Definición: Evento. A cualquier subconjunto de resultados, $E \subset \Omega$, se le llama evento.

Definición: Variable aleatoria. Se dice variable aleatoria, X, al valor de un resultado antes de la realización del experimento. Una vez realizado el experimento, al resultado obtenido lo notamos como x.

Definición: Medida de probabilidad. Una medida de probabilidad P es una función, $P : \mathrm{X}(\Omega) \to [0, 1]$, que cumple con las siguientes propiedades:

1. $P(E) \geq 0 \; \forall \; E \subset \Omega$ (No negatividad)
2. $\sum_{x \in \Omega} P(\mathrm{X} = x) = 1$ (Normalización)
3. $P(E) = \sum_{x \in E} P(\mathrm{X} = x)$ (Adición de valores no simultáneos)

Para simplificar la notación, muchas veces notamos $P(\mathrm{X} = x) = P(x)$, es decir, la probabilidad de que la variable aleatoria tome valor x.

Ejemplo: Podemos considerar la variable aleatoria X como la tirada de un dado de 6 caras. En este caso el espacio muestral son las seis caras del dado: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si consideramos que se trata de un dado no cargado, es decir, que todas las caras tienen la misma probabilidad de ocurrencia, $P(X = 1) = P(X = 2) = P(X = 3) = P(X = 4) = P(X = 5) = P(X = 1) = C$. Como la probabilidad debe estar normalizada, tenemos que $\sum_{i=1}^{6} P(X = i) = 1$. Entonces $\sum_{i=1}^{6} C = C \sum_{i=1}^{6} 1 = 6C = 1$, lo que implica que $C = 1/6$. Si ahora consideramos el evento $E =$ números pares $= \{2, 4, 6\}$, de acuerdo con la propiedad de suma de valores no simultaneos tenemos $P(E) = P(X = 2) + P(X = 4) + P(X = 6) = 3/6 = 1/2$.

2 Probabilidades conjuntas y marginales

Dado un conjunto de N variables aleatorias, $X_1, X_2, \ldots, X_N$, se tiene la **probabilidad conjunta** de las N variables aleatorias, $P(X_1 = x_1, \ldots, X_N = x_N)$, como la probabilidad de que cada variable aleatoria tome valores $X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N$. Nuevamente, simplificamos la notación como $P(X_1 = x_1, \ldots, X_N = x_N) = P(x_1, \ldots, x_N)$.

La probabilidad conjunta es una medida de probabilidad, por lo tanto verifica las propiedades de no negatividad, normalización y de adición de valores no simultaneo. En particular, notamos que si Ω_i es el espacio muestral de la variable aleatoria X_i, entonces la propiedad de normalización se escribe como:

$$\sum_{x_1 \in \Omega_1} \cdots \sum_{x_N \in \Omega_N} P(x_1, \ldots, x_N) = 1$$

.

Definición: Decimos que las variables aleatorias son **independientes** si se cumple que

$$P(x_1, x_2 \ldots, x_N) = P(x_1)P(x_2), \ldots, P(x_N).$$

Definición: A partir de una distribución de probabilidad conjunta, definimos la **probabilidad marginal** de la variable aleatoria X_i como:

$$P(X_i = x_i) = \sum_{x_1 \in \Omega_i} \cdots \sum_{x_{i-1} \in \Omega_{i-1}} \sum_{x_{i+1} \in \Omega_{i+1}} \cdots \sum_{x_N \in \Omega_N} P(X_1 = x_1, \ldots, X_N = x_N).$$

Observese que las sumas son en todos los valores posibles de todas las variables aleatorias que intervienen en la probabilidad conjunta, menos en la que nos interesa obtener la probabilidad marginal.

Propiedad: Si la probabilidad conjunta esta normalizada, entonces todas las probabilidades marginales están también normalizadas.
Dem.: Por la definición de probabilidad marginal, tenemos
$P(X_i) = \sum_{x_1 \in \Omega_i} \cdots \sum_{x_{i-1} \in \Omega_{i-1}} \sum_{x_{i+1} \in \Omega_{i+1}} \cdots \sum_{x_N \in \Omega_N} P(x_1, \ldots, x_N)$.

Entonces la normalización de $P(x_i)$ implica

$$\sum_{x_i\in\Omega_i} P(x_i) = \sum_{x_i\in\Omega_i}\sum_{x_1\in\Omega_i}\cdots\sum_{x_{i-1}\in\Omega_{i-1}}\sum_{x_{i+1}\in\Omega_{i+1}}\cdots\sum_{x_N\in\Omega_N} P(x_1,\ldots,x_N)$$
$$= \sum_{x_1\in\Omega_i}\cdots\sum_{x_N\in\Omega_N} P(x_1,\ldots,x_N) = 1,$$

donde la última igualdad se cumple si la probabilidad conjunta está normalizada.

Ejemplo: Tabla de doble entrada. Para dos variables aleatorias discretas, X e Y, podemos representar la probabilidad conjunta, $P(x,y)$, en una tabla de doble entrada, donde las columnas representan los posibles valores que puede tomar la variable X, y las filas los posibles valores de la variable Y:

P(x,y)	$X=1$	$X=2$	$X=3$
$Y=3$	1/24	3/12	3/24
$Y=4$	1/12	0	1/24
$Y=5$	3/24	1/24	0
$Y=6.5$	2/12	1/24	1/12

Observese que sumando en todas las entradas de la tabla, verificamos la normalización de la probabilidad conjunta: $\frac{1}{24}+\frac{3}{12}+\frac{3}{24}+\frac{1}{24}+0+\frac{1}{24}+\frac{3}{24}+\frac{1}{24}+0+\frac{2}{12}+\frac{1}{24}+\frac{1}{12}=1$. Sumando en filas y en columnas obtenemos las probabilidades marginales, $P(x)$ y $P(y)$, respectivamente.

P(x,y)	$X=1$	$X=2$	$X=3$	**P(y)**
$Y=3$	1/24	3/12	3/24	5/12
$Y=4$	1/12	0	1/24	3/24
$Y=5$	3/24	1/24	0	2/12
$Y=6.5$	2/12	1/24	1/12	7/24
P(x)	5/12	1/3	3/12	

Nuevamente, sumando en la fila de $P(x)$, $\frac{5}{12}+\frac{1}{3}+\frac{3}{12}=1$, y en la columna para $P(y)$, $\frac{5}{12}+\frac{3}{24}+\frac{2}{12}+\frac{7}{24}=1$, se verifica la normalización de las probabilidades marginales.

3 Valor esperado

El valor esperado, o esperanza, de una variable aleatoria, $E(X)$, se define como

$$E(X) = \sum_{x\in\Omega} xP(X=x).$$

Obsérvese que el valor esperado es un promedio ponderado por las probabilidades de ocurrencia de los posibles resultados de un experimento.

Ejemplo: Calculemos el valor esperado para el dado de 6 caras:

$$E(X) = 1\frac{1}{6} + 2\frac{1}{6} + 3\frac{1}{6} + 4\frac{1}{6} + 5\frac{1}{6} + 6\frac{1}{6} = \frac{21}{6} = \frac{7}{2} = 3,5$$

Propiedades:

1. $E(X+Y) = E(X) + E(Y)$,
2. $E(\alpha X) = \alpha E(X)$, con $\alpha = cte.$
3. Si X e Y son variables independientes, $E(XY) = E(X)E(Y)$.

Propiedades 1 y 2 implican que el valor esperado es una función lineal. Además, para cualquier función de la variable aleatoria, $g(X)$, se calcula su valor esperado bajo la distribución de probabilidad P, como

$$E(g) = \sum_{x\in\Omega} g(x)P(X=x),$$

que también suele notarse como $\langle g(x)\rangle$ o $\bar{g}(x)$.

4 Variables aleatorias continuas

Decimos que una variables aleatoria, X, es continua si puede tomar un conjunto continuo de valores. Por ejemplo, algún intervalo continuo de los números reales ($\mathbb{R}$), o incluso todos los número reales. Más formalmente, decimos que X es una variable aleatoria continua si existe una función, $f_X : \mathbb{R} \to \mathbb{R}$, tal que $f_X(x) \geq 0$ y la probabilidad de un evento $E \subset \Omega$ está dada por

$$P(E) = \int_E f_X(x)dx.$$

La función $f_X(x)$ se conoce como la **función densidad** de X.

Observación: $P(X = a) = 0$ para cualquier a.

Observación: La función de densidad NO es una probabilidad, solo tiene sentido de probabilidad cuando es integrada. De esta manera, un infinitesimal de probabilidad puede escribirse como $dP(x) = f_X(x)dx$.

Observación: La expresión de la observación anterior NO es la probabilidad de que la variable aleatoria tome el valor x (ya que por la primera observación esta probabilidad es nula), sino que es la probabilidad de que la variable aleatoria tome un valor en el intervalo $[x, x+dx]$.

Observación: La propiedad de normalización para una variable continua implica que $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

Ejemplo: Distribución de probabilidad uniforme. Consideremos una variable aleatoria que puede tomar, con igual probabilidad, cualquier valor en el intervalo $[a, b]$. Como cualquier valor dentro del intervalo es igualmente probable, esto implica que para cualquier intervalo infinitesimal $dx \subset [a, b]$, la probabilidad de que la variable aleatoria tome un valor dentro de este intervalo es proporcional al tamaño del intervalo, es decir: $dP(x) = Cdx$, con C una constante. De acuerdo con la definición de función densidad, para este caso tenemos

$$f_X(x) = \begin{cases} C & \text{si } x \in [a,b], \\ 0 & \text{en otro caso,} \end{cases}$$

es decir, la función densidad es constante dentro del intervalo $[a, b]$, y nula fuera de este intervalo. Al igual que en el caso discreto, encontramos el valor de C a partir de la propiedad de normalización:

$$\int_{-\infty}^{+\infty} f_X(x)dx = \int_a^b Cdx = C\int_a^b dx = C(b-a) = 1,$$

por lo tanto $C = \frac{1}{b-a}$.

4.1 Probabilidades conjuntas y marginales

Por simplicidad consideraremos la probabilidad conjunta de dos variables aleatorias, $P(x, y)$, pero los conceptos pueden generalizarse sin dificultad. Para este caso, se pueden obtener las probabilidades marginales de la siguiente manera:

$$P(x) = \int_{-\infty}^{+\infty} P(x,y)dy,$$

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

La normalización de la probabilidad conjunta, $\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(x, y)dxdy = 1$, implica nuevamente que las probabilidades marginales están normalizadas:

$$\int_{-\infty}^{+\infty} P(x)dx = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(x,y)dydx = 1,$$

$$\int_{-\infty}^{+\infty} P(y)dy = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} P(x,y)dxdy = 1.$$

4.2 Valor esperado

Para el caso de una variable aleatoria continua, el valor esperado se calcula como

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx,$$

y para una función $g(X)$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x)f_X(x)dx.$$

Ejemplo: Valor esperado de una distribución uniforme. Volvamos al ejemplo dónde habíamos calculado

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a,b], \\ 0 & \text{en otro caso.} \end{cases}$$

Si ahora calculamos $E(X)$, tenemos

$$E(X) = \int_{-\infty}^{+\infty} x f_X(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \frac{(b-a)^2}{2} = \frac{b-a}{2},$$

por lo tanto el valor esperado cae en la mitad del intervalo.