

Distribuciones de Probabilidad y Distribución Gaussiana

1. Distribuciones de Probabilidad

La función que gobierna la distribución de datos en una experiencia es denominada Función de

Distribución de Probabilidad ó Densidad de Probabilidad. Veamos cómo podemos representarla. Si tomamos un número muy grande N de valores de x , veremos que cierto número dN de ellas presentan valores entre cierto x y $x + dx$. De acuerdo a la definición clásica de probabilidad, decimos que la probabilidad de obtener una medida con valor entre x y $x + dx$ está dada por la siguiente expresión

$$dP(x, x + dx) = \frac{dN}{N} \quad (1.1)$$

Cuanto mayor sea el intervalo dx (aunque siempre $dx \ll |x|$), mayor cantidad dN de medidas caerán en él.

$$\frac{dN}{N} \propto dx$$

También es obvio que el número dN de medidas comprendidas entre x y $x + dx$ no sólo depende del tamaño del intervalo dx , sino del propio valor x . La experiencia indica que para ciertos valores de x se acumulan muchas medidas dentro de un intervalo dx , en tanto que para valores de x mucho mayores o menores que el primero, hay muy pocas medidas dentro del mismo intervalo dx . De esta forma podemos escribir

$$\frac{dN}{N} = f(x)dx \quad (1.2)$$

que combinándola con la ecuación (1.1) podemos obtener

$$dP(x, x + dx) = f(x)dx \quad (1.3)$$

donde la función $f(x)$ es conocida como la Densidad de Probabilidad de la variable aleatoria x .

Puede observarse que de acuerdo a la ecuación (1.2) $f(x)$ es positiva siempre.

Luego integrando (1.2) entre $-\infty$ y $+\infty$ obtenemos

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \quad (1.4)$$

La ecuación (1.4) establece que la probabilidad que tomemos una medida cuyo valor de nuestra variable aleatoria este comprendido entre $-\infty$ y $+\infty$, es la unidad. En general, la probabilidad de obtener una medida de nuestra variable, entre dos valores cualesquiera x_1 y x_2 está dada por

$$P(x_1, x_2) = \int_{x_1}^{x_2} f(x)dx \quad (1.5)$$

1.1. Valor Esperado y Desviación Estándar

1.1.1. Valor Esperado

Si bien es cierto que un número muy grande N de medidas de la magnitud x (en rigor, $N \rightarrow \infty$) arrojará diversos valores, es necesario ponerse de acuerdo para asignarle un valor determinado $\langle x \rangle$ a la magnitud medida, en el entendido de que las condiciones del sistema no se han alterado en la sucesión de mediciones hechas. A este número $\langle x \rangle$ se le denomina Esperanza ó Valor Esperado de la magnitud x y se define como

$$\langle x \rangle = \int_{-\infty}^{+\infty} x f(x) dx \quad (1.6)$$

1.1.2. Desviación Estándar

Para tener una idea de cuánto se "desparraman" los valores medidos de x alrededor del valor esperado $\langle x \rangle$, se define la desviación estándar cuadrática de la variable aleatoria x como

$$\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle \quad (1.7)$$

o en su forma integral utilizando la ecuación (1.6)

$$\sigma^2 = \int_{-\infty}^{+\infty} (x - \langle x \rangle)^2 f(x) dx \quad (1.8)$$

Se puede probar que si la curva de $f(x)$ es simétrica alrededor de cierto valor x_0 , entonces resulta

$$\langle x \rangle = x_0$$

1.2. Promedio y Varianza

Es claro que en la práctica es imposible obtener el valor de $\langle x \rangle$ y de σ a partir de las ecs. (1.6) y (1.8) por las siguientes razones

- Se necesita conocer la densidad de probabilidad $f(x)$ de la variable aleatoria x .
- Debemos poder construir $f(x)$ en todo su extensión para $-\infty < x < +\infty$. Para ello tendríamos que tomar infinita cantidad de mediciones de la magnitud.

Experimentalmente todo lo que podemos hacer es tomar un conjunto finito o discreto de valores N valores de x : $\{x_1; x_2; \dots; x_N\}$. Este conjunto constituye una muestra de todo el universo de valores de x . Con estos datos debemos tratar de estimar los valores de la esperanza $\langle x \rangle$ y la dispersión estándar σ .

1.2.1. Promedio

Para la muestra de N valores, podemos definir el valor medio, o la media del conjunto, como el número:

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

(1.9)

La mejor estimación que tenemos del valor verdadero de la medida, es el promedio \bar{x} de la muestra.

En la práctica el promedio \bar{x} coincidiría con $\langle x \rangle$ si fuésemos capaces de tomar infinita cantidad de mediciones. Esta es la razón por la cual, en Física, se toma el promedio de la serie de medidas como el valor verdadero de la magnitud medida.

1.2.2. Varianza

Igualmente podemos definir un parámetro que nos de una medida de la dispersión de los N valores en torno a la media \bar{x} . Llamamos Varianza ó Dispersión Cuadrática de la muestra al número:

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1.10)$$

Se puede demostrar que la mejor estimación que se tiene de la varianza es

$$\langle s^2 \rangle = \frac{N-1}{N} \sigma^2 \quad (1.11)$$

1.2. Error en el Promedio

La media \bar{x} y la dispersión s de una muestra también son variables aleatorias. Pues si tomamos una nueva muestra de valores, sin duda obtendremos otro conjunto de

$\{x_1', x_2', \dots, x_N'\}$ diferente del anterior $\{x_1, x_2, \dots, x_N\}$. Esto nos dará, en general, otra media

\bar{x}' y otra dispersión s' . Este valor del promedio, que no es exactamente $\langle x \rangle$, está entonces afectado de un error que es necesario especificar. Podemos calcular entonces la desviación estándar alrededor del valor esperado $\langle x \rangle$. En lugar de σ utilizaremos E para indicar la desviación estándar del promedio, y lo llamamos *Error en el Promedio*. Se puede demostrar que

$$E^2 = \langle (\bar{x} - \langle x \rangle)^2 \rangle \quad (1.12)$$

$$E^2 = \int_{-\infty}^{+\infty} (\bar{x} - \langle x \rangle)^2 f(x) dx \quad (1.13)$$

y utilizando la ecuación (1.9) llegamos a que

$$E^2 = \langle \left(\frac{1}{N} \sum_{i=1}^N (x_i - \langle x \rangle) \right)^2 \rangle = \frac{1}{N^2} \langle \sum_{i=1}^N (x_i - \langle x \rangle)^2 \rangle$$

Operando se obtiene una relación entre el Error en el Promedio y la Desviación Estándar

$$E = \frac{\sigma}{\sqrt{N}} \quad (1.14)$$

Esta es la expresión del error o desviación estándar del promedio \bar{x} de una serie de N medidas, con respecto a su valor esperado $\langle x \rangle$. Si bien podemos calcular fácilmente \bar{x} con la serie de N medidas, no podemos aún calcular el error E , puesto que no sabemos cómo estimar la desviación estándar σ .

Utilizando la ec. (1.11), se puede demostrar que

$$\sigma^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (1.15)$$

Analicemos las ecs. (1.14) y (1.15) en detalle. En primer lugar, notemos que la definición de la desviación estándar σ es independiente del número de medidas adquiridas. Recordemos que representa la dispersión de las medidas en torno al valor más probable de la magnitud medida. En cambio, el error del promedio E depende del número de medidas. Notemos que si tomamos un número suficientemente grande de medidas podemos hacer que $E \rightarrow 0$. ¿Esto tiene sentido desde el punto de vista experimental? Evidentemente no tiene sentido de hablar de un error que tiende a cero. No olvidemos

que, más allá de la matemática, todo el formalismo desarrollado debe ser aplicado a un proceso de medición, para el cual utilizamos instrumentos que tienen una cierta apreciación. ¿Qué significa esto? Que la apreciación del instrumento utilizado es la cota inferior para el error del promedio. A partir de allí podemos estimar cual es el máximo número de medidas que tiene sentido tomar. Finalmente veamos cómo expresar el resultado de la medición realizada. Sea ϵ la mejor estimación del valor verdadero de la magnitud y s la mejor estimación de la desviación estándar. Entonces:

2. Distribución Normal o Gaussiana

La densidad de probabilidad Normal o Gaussiana es una función de variable continua. Este tipo de funciones permite modelar un sin número de sistemas físicos. En este curso aplicaremos la función Gaussiana para estudiar la distribución de una serie de datos en torno al valor esperado. Consideramos la siguiente función

$$f(x) = A e^{\left(\frac{-(x-x_0)^2}{2\sigma_0^2}\right)} \quad A > 0 \quad (2.1)$$

La definición de la ecuación (2.1) satisface las condiciones para ser considerada una distribución de probabilidad

\sum $f(x)$ es definida positiva

\sum $f(x) \rightarrow 0$ cuando $x \rightarrow \pm \infty$

\sum La integral en todo su dominio es igual a la unidad, para esto la constante A debe tener el siguiente valor.

$$1 = \int_{-\infty}^{+\infty} f(x) dx = A \int_{-\infty}^{+\infty} e^{\left(\frac{-(x-x_0)^2}{2\sigma_0^2}\right)} dx$$

donde la integral de la función Gaussiana es conocida

$$\int_{-\infty}^{+\infty} e^{\left(\frac{-(x-x_0)^2}{2\sigma_0^2}\right)} dx = \sigma_0 \sqrt{2\pi}$$

Por lo tanto se cumple que

$$A = \frac{1}{\sigma_0 \sqrt{2\pi}} \quad (2.2)$$

Entonces podemos definir la siguiente Función *de Distribución Gaussiana* ó *Normal*.

$$f(x) = \frac{1}{\sigma_0 \sqrt{2\pi}} e^{\left(\frac{-(x-x_0)^2}{2\sigma_0^2}\right)} \quad (2.3)$$

Se puede demostrar en forma muy sencilla que x_0 corresponde al máximo de la curva y σ_0 a los puntos de inflexión. La experiencia indica que cuando medimos una magnitud x con un instrumento de muy pequeña apreciación y realizamos un número N muy grande medidas todas en las mismas condiciones, la densidad de probabilidad que mejor ajusta la distribución de datos en torno del valor esperado en la función de densidad Gaussiana. En este caso x_0 representa el valor esperado x y σ_0 la desviación estándar.

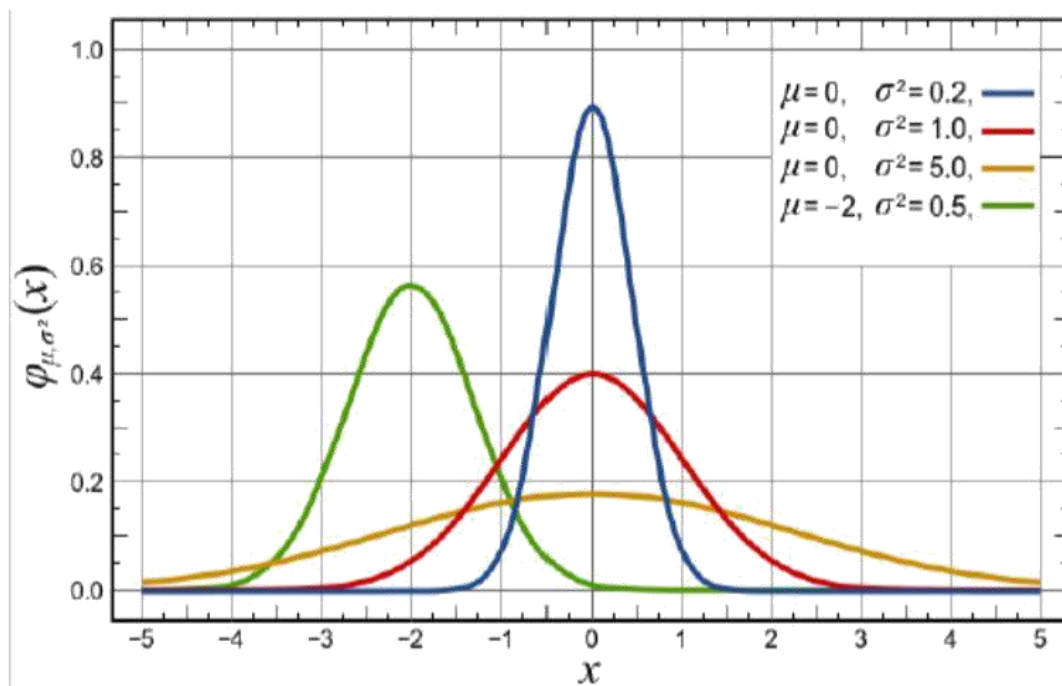


Figura 2.1: En esta figura se muestran varias funciones de distribución normales o Gaussianas ϕ , con distintos valores de σ y valores centrales μ .

Hemos determinado entonces la forma concreta de la distribución de probabilidad $f(x)$ de una variable aleatoria x . Esto significa que podemos aplicar la ec. (1.5) para calcular directamente la probabilidad de obtener algún valor de x dentro de un intervalo $[x_1, x_2]$ arbitrario. Es decir:

$$P(x_1, x_2) = \frac{1}{\sigma_o \sqrt{2\pi}} \int_{x_1}^{x_2} e^{\left(\frac{-(x-x_o)^2}{2\sigma_o^2}\right)} dx \quad (2.4)$$

El problema que presenta la ec. (2.4) es que no tiene solución analítica por no existir ninguna función elemental que sea primitiva de la ec. (2.1). Existen al menos dos soluciones posibles para este problema. La clásica que consiste en utilizar tablas de la integral de probabilidad, como la que aparece en "Manual de matemáticas" (Bronshstein). Esto es un proceso engorroso que implica realizar una serie de cambios de variable y consideraciones diversas sobre los intervalos para poder utilizar dichas tablas. La segunda alternativa que consideraremos en este curso es realizar una integración numérica

mediante el programa Matlab. Para ello utilizaremos la función `quad`, que realiza la integración numérica de funciones.

2.1. Interpretación de la Desviación Estándar

Se puede demostrar fácilmente, utilizando la ec. (2.4), que la probabilidad de que un dato de la serie de medidas pertenezca al intervalo $[\bar{x} - s, \bar{x} + s]$ es de ~ 0.6826 . Esto significa

que un dato cualquiera tiene una probabilidad de 68% de diferir en σ del valor esperado. Es claro que este hecho no es una propiedad intrínseca de σ , sino que se debe a que la distribución utilizada es Gaussiana.