CIMAT Centro de Investigación en Matemáticas, A.C.

CONACYT

Curso Propedéutico de Estadística clase 4 y 5

Temario clase 4 y 5

- Valor esperado, Varianza y mediana
- Distribución normal
- Distribuciones conjuntas
- Distribuciones marginales
- Distribuciones condicionales
  - -Independencia entre variables

Valor esperado

¿Cómo se obtiene el valor esperado en variables aleatorias discretas?

Sea X una variable aleatoria discreta con un conjunto de valores posibles D y una función de masa de probabilidad $p(x)$. El **valor esperado** o **valor medio** de X, denotado por $E(X)$ o μ_X o sólo μ, es

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

Vamos a reemplazar la **suma** por la **integración** y la **función de masa de probabilidad** por la **función de densidad de probabilidad**

El **valor esperado** o **valor medio** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ es

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx$$

Valor esperado

Ejemplo:

La función de densidad de probabilidad de una v.a. *X* es:

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es su valor esperado?

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)\, dx = \int_0^1 x \cdot \frac{3}{2}(1 - x^2)\, dx = \frac{3}{2}\int_0^1 (x - x^3)\, dx$$

$$= \frac{3}{2}\left(\frac{x^2}{2} - \frac{x^4}{4}\right)\Bigg|_{x=0}^{x=1} = \frac{3}{8}$$

Propiedades del valor esperado

Sea X una variable aleatoria con distribución de probabilidad $f(x)$. El valor esperado de la variable aleatoria $g(X)$ es

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x)f(x)$$

si X es discreta, y

$$\mu_{g(X)} = E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\,dx$$

si X es continua.

$$g(X) \Rightarrow \begin{array}{c} X^2 \\ 3X - 1 \end{array}$$

Propiedades del valor esperado

Ejemplo:
X una variable aleatoria con función de densidad dada por:

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{en otro caso.} \end{cases}$$

¿Cuál es el valor esperado de g(X) = 4X + 3

$$E(4X+3) = \int_{-1}^{2} \frac{(4x+3)x^2}{3}\, dx = \frac{1}{3}\int_{-1}^{2} (4x^3 + 3x^2)\, dx = 8.$$

Propiedades del valor esperado

Dada X v.a. otras propiedades importantes del valor esperado son

Si $g_1(X) = a, \; g_2(X) = aX + b,$ y $g_3(X) = ah_1(X) + h_2(X)$, entonces

Donde a y b son constantes, g_1, g_2, g_3 , h_1 y h_2 son funciones.

$$E(g_1(X)) = E(a) = a$$

$$E(g_2(X)) = E(aX + b) = aE(X) + b \; = a\mu + b$$

$$E(g_3(X)) = E(ah_1(X) + h_2(X)) = aE(h_1(X)\,) + E(h_2(X)\,).$$

Estas propiedades **se pueden demostrar** usando sumas (caso discreto) o integrales (caso continuo).

Varianza

¿Cómo se obtiene la varianza de un variables aleatoria discreta?

Sea $p(x)$ la función de masa de probabilidad de X y μ su valor esperado. En ese caso la varianza de X, denotada por $V(X)$ o σ_X^2, o simplemente σ^2, es

$$V(X) = \sum_D (x - \mu)^2 \cdot p(x) = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es

$$\sigma_X = \sqrt{\sigma_X^2}$$

Vamos a reemplazar la **suma** por la **integración** y la **función de masa de probabilidad** por la **función de densidad de probabilidad**

La **varianza** de una variable aleatoria continua X con función de densidad de probabilidad $f(x)$ y valor medio μ es

$$\sigma_X^2 = V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot f(x)dx = E[(X - \mu)^2]$$

La **desviación estándar** (DE) de X es $\sigma_X = \sqrt{V(X)}$.

Propiedades de la varianza

Sea X una variable aleatoria con distribución de probabilidad $f(x)$. La varianza de la variable aleatoria $g(X)$ es

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \sum_x [g(x) - \mu_{g(X)}]^2 f(x)$$

si X es discreta, y

$$\sigma^2_{g(X)} = E\{[g(X) - \mu_{g(X)}]^2\} = \int_{-\infty}^{\infty} [g(x) - \mu_{g(X)}]^2 f(x)\, dx$$

si X es continua.

Ejercicio:
Calcule la **varianza** de $g(X) = 2X + 3$, donde X es una variable aleatoria con la siguiente distribucion de probabilidad

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

Propiedades de la varianza

Ejercicio:
Calcule la **varianza** de $g(X) = 2X + 3$, donde X es una variable aleatoria con la siguiente distribucion de probabilidad

x	0	1	2	3
$f(x)$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{8}$

¿Qué hacemos primero?

Primero se calcula la media de la variable aleatoria $2X + 3$.

$$\mu_{2X+3} = E(2X+3) = \sum_{x=0}^{3}(2x+3)f(x) = 6$$

$$\sigma^2_{2X+3} = E\{[(2X+3) - \mu_{2x+3}]^2\} = E[(2X+3-6)^2]$$
$$= E(4X^2 - 12X + 9) = \sum_{x=0}^{3}(4x^2 - 12x + 9)f(x) = 4$$

Valor esperado y varianza

Observaciones

1) $E(X)$ es un **promedio ponderado** de posibles X valores, donde la **función de ponderación** es la **función de densidad de probabilidad** $f(x)$

2) La varianza y la desviación estándar son medidas cuantitativas de qué tanta **dispersión** hay en la **distribución o población de valores *x***

3) Existe una relación a partir de $E(X)$ para calcular la varianza:

$$V(X) = E(X^2) - [E(X)]^2$$

Valor esperado y varianza

Ejemplo:

La función de densidad de probabilidad de una v.a. *X* es:

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

¿Cuál es la varianza de X?

$$E(X) = \frac{3}{8}$$

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot f(x)\, dx = \int_0^1 x^2 \cdot \frac{3}{2}(1 - x^2)\, dx = \int_0^1 \frac{3}{2}(x^2 - x^4)\, dx = \frac{1}{5}$$

$$V(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2 = \frac{19}{320} = 0.059$$

$$\sigma_X = 0.244$$

Percentiles y mediana

$E(X)=\mu$ es la media de **tendencia central** que especifica la ubicación o centralización de la población (se utiliza con mucha frecuencia)

Existen otras medidas de tendencia central como la **mediana** que son bastante útiles.

Hay primero ver que son los **percentiles ¿?**.

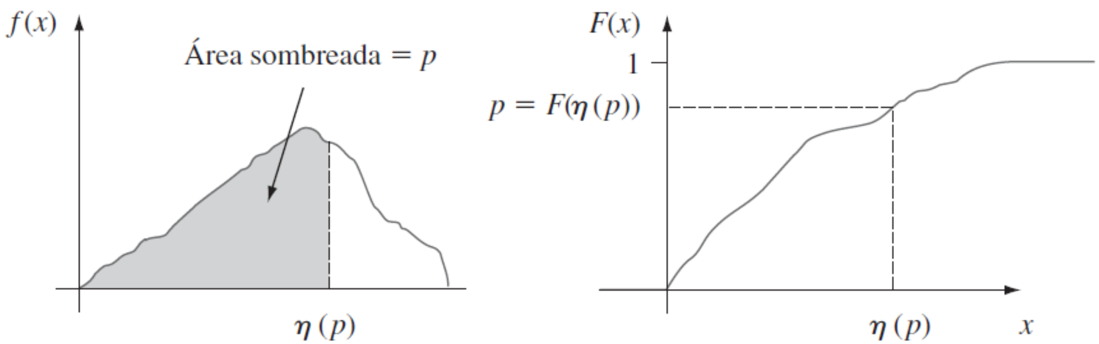
Sea p un número entre 0 y 1. El **(100*p*)° percentil** de la distribución de una variable aleatoria continua X, denotada por $\eta(p)$, se define como

$$p = F(\eta(p)) = \int_{-\infty}^{\eta(p)} f(y)\,dy$$

Sea p valor entre 0 y 1.
$\eta(p)$ es ese valor sobre el **eje de medición** (sobre los valores que toma X), de tal suerte que $100p$% del **área bajo la gráfica** de $f(x)$ queda a la **izquierda** de $\eta(p)$ y $100(1 - p)$% queda a la **derecha**

Percentiles y mediana

Sea p=0.75 valor entre 0 y 1.
$\eta(0.75)$ es ese valor sobre el **eje de medición** (sobre los valores que toma X), de tal suerte que $100(0.75)\%=75\%$ del **área bajo la gráfica** de $f(x)$ queda a la **izquierda** de $\eta(0.75)$ y $100(1 - 0.75)\%=25\%$ queda a la **derecha**



¿qué es la mediana?

La **mediana** de una distribución continua, denotada por $\widetilde{\mu}$, es el 50º percentil, así que $\widetilde{\mu}$ satisface $0.5 = F(\widetilde{\mu})$. Es decir, la mitad del área bajo la curva de densidad se encuentra a la izquierda de $\widetilde{\mu}$ y la mitad a la derecha de $\widetilde{\mu}$.

Percentiles y mediana

Una forma de **obtener los percentiles** de una distribución es resolviendo la ecuación resultante (donde es necesario conocer la expresión de F(x)

$$p = F(\eta(p))$$

Ejemplo:

La distribución de la **cantidad de grava** (en toneladas) vendida por una compañía de materiales para la construcción particular en una semana dada es una variable aleatoria continua *X* con función de densidad de probabilidad:

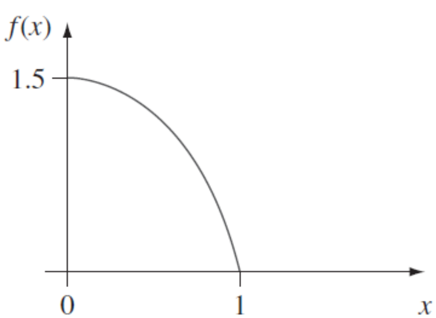
$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \leq x \leq 1 \\ 0 & \text{de lo contrario} \end{cases}$$

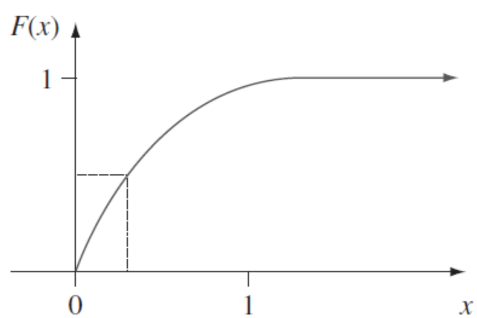
Percentiles y mediana

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{de lo contrario} \end{cases}$$

¿cómo se calcula la **F(x)?**

$$F(x) = \int_0^x \frac{3}{2}(1 - y^2)\,dy = \frac{3}{2}\left(y - \frac{y^3}{3}\right)\Bigg|_{y=0}^{y=x} = \frac{3}{2}\left(x - \frac{x^3}{3}\right)$$





¿Cuál es el percentil 50%?

$$p = F(\eta(p))$$

Percentiles y mediana

El (100*p*)º percentil de esta distribución satisface **la ecuación**

$$p = F(\eta(p)) = \frac{3}{2}\left[\eta(p) - \frac{(\eta(p))^3}{3}\right]$$

$$(\eta(p))^3 - 3\eta(p) + 2p = 0$$

$$p = 0.5 \quad \Rightarrow \quad \eta^3 - 3\eta + 1 = 0 \quad \Rightarrow \quad \eta(0.5) = 0.347$$

Si la distribución no cambia de una semana a otra, a la larga 50% de todas las semanas se realizarán ventas de menos de 0.347 ton y 50% de más de 0.347 ton

Función de distribución normal

La **distribución normal** es la **más importante** tanto en la probabilidad y en la estadística.

Muchas **poblaciones numéricas** tienen distribuciones que **pueden ser representadas muy fielmente** mediante una curva normal apropiada

Los ejemplos:

- **Estaturas**, **pesos** y otras características físicas (el famoso artículo 1903 Biometrika "On the Laws of Inheritance in Man" discutió muchos ejemplos de esta clase),
- **Errores de medición** en experimentos científicos,
- **Mediciones antropométricas** en fósiles,
- **Tiempos de reacción** en experimentos psicológicos,
- Mediciones de **inteligencia y aptitud**
- **Calificaciones** en varios exámenes
- **Indicadores económicos**

Función de distribución normal

¿Saben de otra razón por la cuál es importante la distribución normal?

Además, aun cuando las variables individuales no estén normalmente distribuidas, **las sumas y los promedios** de las variables en **condiciones adecuadas** tendrán de manera aproximada una **distribución normal**

Teorema del límite central

Teorema del límite central: Si $\bar{X}$ es la media de una muestra aleatoria de tamaño n, tomada de una población con media μ y varianza finita σ^2, entonces la forma límite de la distribución de

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}},$$

a medida que $n \to \infty$, es la distribución normal estándar $n(z; 0, 1)$.

Función de distribución normal

Se dice que una variable aleatoria continua X tiene una **distribución normal** con parámetros μ y σ (o μ y σ^2), donde $-\infty < \mu < \infty$ y $0 < \sigma$, si la función de densidad de probabilidad de X es

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} \qquad -\infty < x < \infty$$

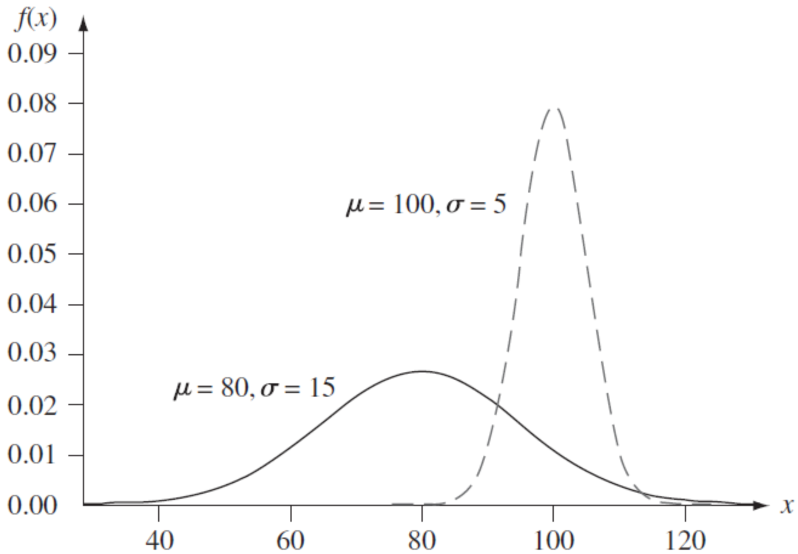
$$X \sim N(\mu,\sigma^2) \qquad f(x;\mu,\sigma) \geq 0 \qquad \int_{-\infty}^{\infty} f(x;\mu,\sigma)\,dx = 1$$

$$E(X) = \mu \qquad V(X) = \sigma^2$$

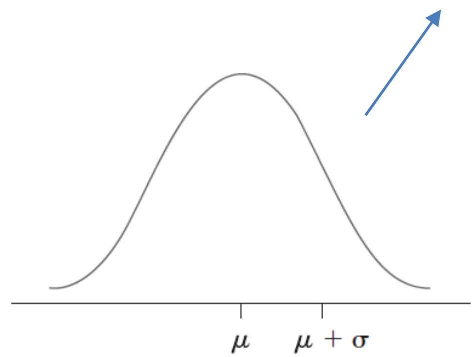
μ es un ***parámetro de ubicación***, ya que al cambiar su valor **desplaza** la curva de densidad hacia uno u otro lado

σ se conoce como un ***parámetro de escala*** porque al cambiar su valor **estira o comprime** la curva sin cambiar la forma básica.

Función de distribución normal



Punto de inflexión



Función de distribución normal

El cálculo de $P(a \leq X \leq b)$ cuando X es una **variable aleatoria normal** con parámetros μ y σ, requiere determinar la integral:

$$\int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)} dx$$

Ninguna de las **técnicas estándar** de integración puede ser utilizada para lograr esto.

Se opta por usar la **variable aleatoria normal estándar** para realizar los cálculos de manera más sencilla.

¿cómo se define la distribución normal estándar?

$$\mu = 0 \qquad \sigma = 1$$

Función de distribución normal

La distribución normal con valores de parámetro $\mu = 0$ y $\sigma = 1$ se llama **distribución normal estándar**. Una variable aleatoria que tiene una distribución normal estándar se llama **variable aleatoria normal estándar** y se denotará por Z. La función de densidad de probabilidad de Z es

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \quad -\infty < z < \infty$$

Observaciones:

La gráfica de f(z;0,1) se llama **cuva normal estándar.**

Sus **puntos de inflexión** están en -1 y 1

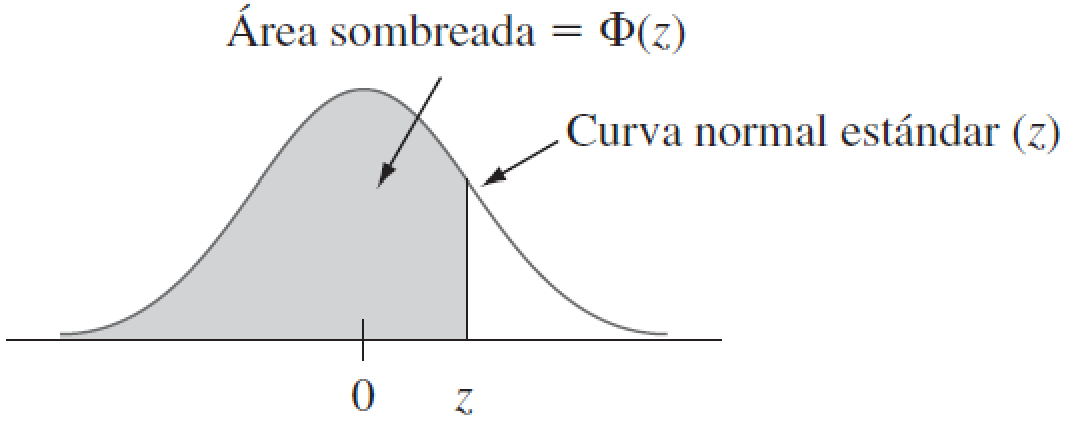
La **función de distribución acumulada** F(Z)=$\Phi(Z)$

$$P(Z \leq z) = \int_{-\infty}^{z} f(y; 0, 1)\, dy$$

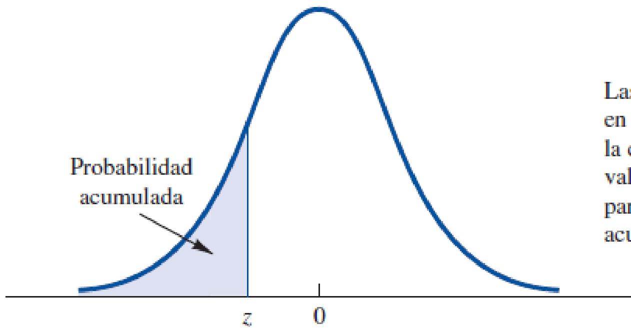
Función de distribución normal

La distribución normal estándar no siempre sirve como **modelo de una población** que surge naturalmente. En cambio, es una **distribución de referencia.**

Empleando **técnicas de integración numérica** se obtienen **tablas** para calcular el área bajo la curva para distintos valores del número real z: F(Z)=$\Phi(Z)$



Función de distribución normal

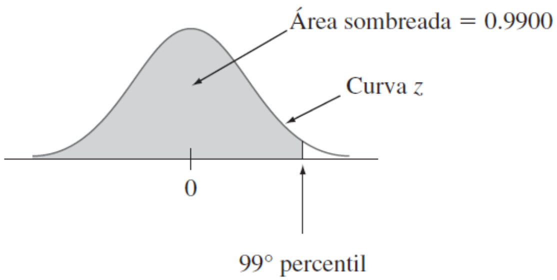


Las entradas que aparecen en la tabla dan el área bajo la curva y a la izquierda del valor de z. Por ejemplo, para $z = -0.85$, la probabilidad acumulada es 0.1977.

z	**0.00**	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559

Función de distribución normal

Se pueden usar las tablas para hallar **los percentiles** de la **distribución normal estándar**



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9913
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936

¿Cuál es el percentil 99%? Z=2.33

Si *p* no aparece a menudo se utiliza el número **más cercano** a él, aunque la **interpolación lineal** da una respuesta más precisa

¿Cuál es el percentil 65%?

Función de distribución normal

Para calcular las probabilidades de **distribuciones normales no estándar**, se puede proceder a estandarizar la variable:

$$X \sim N(\mu, \sigma^2) \Rightarrow (X-\mu)/\sigma \quad N(0,1)$$

$$Z = \frac{X-\mu}{\sigma} \Rightarrow P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

$$P(X \leq a) = \Phi\left(\frac{a-\mu}{\sigma}\right) \qquad P(X \geq b) = 1 - \Phi\left(\frac{b-\mu}{\sigma}\right)$$

Función de distribución normal

Ejemplo:
El tiempo que requiere un conductor para **reaccionar** a las luces de freno de un vehículo que está desacelerando es crítico **para evitar colisiones por alcance**.

El artículo **"Fast-Rise Brake Lamp as a Collision-Prevention Device" (*Ergonomics*, 1993: 391–395)** sugiere que el tiempo de reacción de respuesta en tráfico a una señal de luces de freno estándar puede ser modelado con una **distribución normal** que tiene un **valor medio** de **1.25 s** y **desviación estándar** de **0.46 s**.

¿Cuál es la **probabilidad de que el tiempo de reacción** sea de entre 1.00 y 1.75 segundos?:

$$1.00 \le X \le 1.75 \quad \Rightarrow \quad \frac{1.00 - 1.25}{0.46} \le \frac{X - 1.25}{0.46} \le \frac{1.75 - 1.25}{0.46}$$

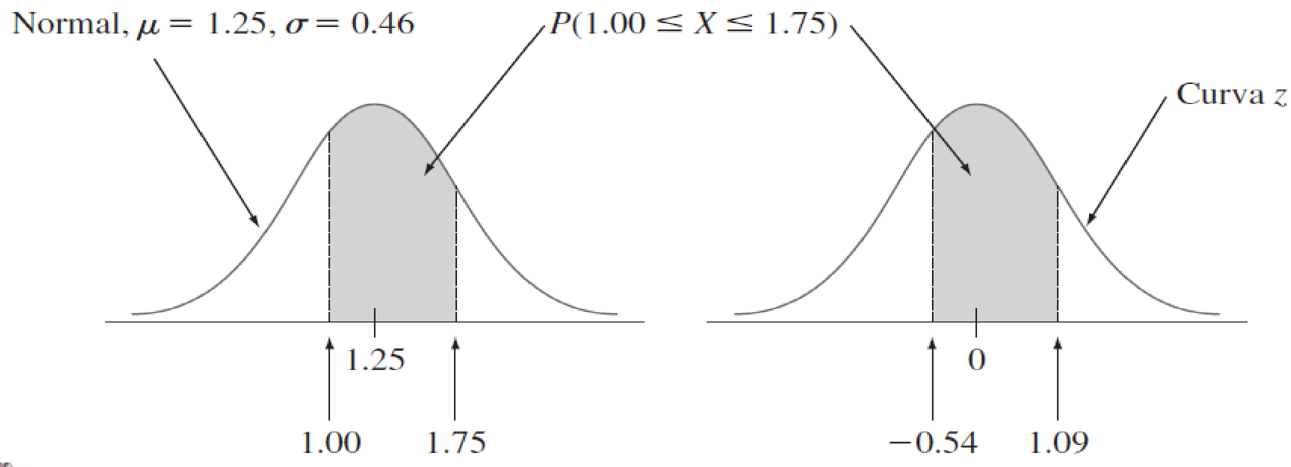
Función de distribución normal

¿Cuál es la **probabilidad de que el tiempo de reacción** sea de entre 1.00 y 1.75 segundos?:

$$P(1.00 \leq X \leq 1.75) = P\left(\frac{1.00 - 1.25}{0.46} \leq Z \leq \frac{1.75 - 1.25}{0.46}\right)$$

$$= P(-0.54 \leq Z \leq 1.09) = \Phi(1.09) - \Phi(-0.54)$$

$$= 0.8621 - 0.2946 = 0.5675$$



Función de distribución normal

Si se ven los 2 segundos como un **tiempo de reacción críticamente largo**, la probabilidad de que el **tiempo de reacción real exceda** este valor es:

$$P(X > 2) = P\left(Z > \frac{2 - 1.25}{0.46}\right) = P(Z > 1.63) = 1 - \Phi(1.63) = 0.0516$$

Cuando **estandarizamos** cantidades obtenemos una medida que nos dice que tan a la **izquierda** o **derecha** de la media μ está el un valor específico x en términos de desviaciones estándar σ, es decir:

Si μ =100 y σ =15, x=130 corresponde a $z = (130 - 100)/15 = 30/15 = 2.00.$

Lo que nos lleva a decir que x=130 está a dos desviaciones estándar a la derecha de μ =100

Función de distribución normal

Si μ =100 y σ =15, x=85 corresponde a $z = (85 - 100)/15 = -1.00$

Lo que nos lleva a decir que x=85 está a una desviaciones estándar a la **izquierda** de μ =100

La **tabla *z*** se aplica a *cualquier* distribución normal siempre que se piense en función del **número de desviaciones estándar** respecto al **valor medio**

Ejercicio:
Se sabe que el voltaje de ruptura de un diodo de un tipo particular seleccionado al azar está normalmente distribuido.

¿Cuál es la probabilidad de que el voltaje de ruptura de un diodo esté **dentro** de **1 desviación estándar** de su **valor medio**?

Función de distribución normal

Hint: Esta pregunta puede contestarse sin conocer μ o σ, en tanto se sepa que la **distribución es normal**; la respuesta es la misma para ***cualquier*** **distribución normal**:

$$P(X \text{ está dentro de 1 desviación estándar de su media})$$

$$= P(\mu - \sigma \leq X \leq \mu + \sigma) \quad = P\left(\frac{\mu - \sigma - \mu}{\sigma} \leq Z \leq \frac{\mu + \sigma - \mu}{\sigma}\right)$$

$$= P(-1.00 \leq Z \leq 1.00) \quad = \Phi(1.00) - \Phi(-1.00) = 0.6826$$

Si la distribución de la población de una variable es (aproximadamente) normal, entonces

1. Alrededor de 68% de los valores está dentro de 1 DE de la media.
2. Alrededor de 95% de los valores está dentro de 2 DE de la media.
3. Alrededor de 99.7% de los valores está dentro de 3 DE de la media

Función de distribución normal

El $(100p)^{\circ}$ percentil de una **distribución normal** con media μ o y desviación estándar σ es fácil de relacionar con el $(100p)^{\circ}$ percentil de la **distribución normal estándar.**

$$\begin{array}{c}(100p)^{\circ}\text{ percentil}\\ \text{para }(\mu,\sigma)\text{ normal)}\end{array} = \mu + \left[\begin{array}{c}(100p)^{\circ}\text{ para}\\ \text{normal estándar)}\end{array}\right]\cdot\sigma$$

Ejercicio:

Los autores de **"Assessment of Lifetime of Railway Axle" (*Intl. J. of Fatigue*, 2013: 40–46)** utilizan los datos recolectados de un experimento con una **longitud de grieta inicial específica** y un número de ciclos para proponer una **distribución normal** con **valor promedio** de carga 5.496 mm y **desviación estándar** 0.067 mm para la variable aleatoria X **profundidad final de la grieta**.

Función de distribución normal

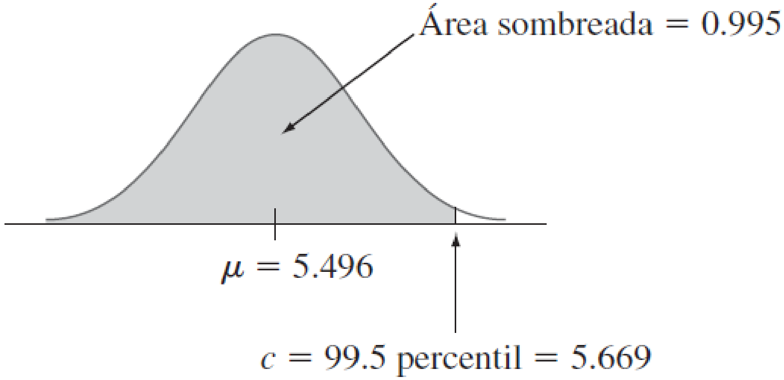
Ejercicio:
¿qué **valor de profundidad final de la grieta** podría ser superado por sólo 0.5% de todas las grietas bajo estas circunstancias?

$$P(X > c) = 0.005 \qquad P(X \leq c) = 0.995$$

Por tanto c es el percentil 99.5 $\qquad \mu = 5.496$ y $\sigma = 0.067$

El percentil 99.5 de la distribución normal estándar es 2.58, por lo que:

$$c = \eta(0.995) = 5.496 + (2.58)(0.067) = 5.669 \text{ mm}$$



Distribuciones de probabilidad conjunta

Habrá situaciones en las que se busque registrar los **resultados simultáneos** de **diversas variables aleatorias**

Ejemplo 1: en un **experimento químico controlado** podríamos medir la cantidad del **precipitado *P*** y la del **volumen *V*** de gas liberado, lo que daría lugar a un **espacio muestral bidimensional** que consta de los resultados (p, v)

Ejemplo 2: podríamos interesarnos en la **dureza *d*** y en la resistencia a la **tensión *T*** de cobre estirado en frío que produciría los resultados (d, t)

Ejemplo 3: En un estudio realizado con estudiantes universitarios para determinar la probabilidad de que tengan éxito en la universidad, basado en los datos previos, se podría utilizar un **espacio muestral tridimensional**: la **calificación** que obtuvieron una prueba de aptitudes, el **lugar** que ocuparon en la preparatoria y la **calificación promedio** que obtuvieron al final de su primer año en la universidad

Distribuciones de probabilidad conjunta

Si X y Y son **dos variables aleatorias discretas**, la distribución de probabilidad para sus **ocurrencias simultáneas** se representa mediante una función con valores $f(x, y)$, para cualquier par de valores (x, y) dentro del **rango** de las variables aleatorias X y Y

$$f(x, y) = P(X = x, Y = y)$$

Dan la probabilidad de que los resultados x y y **ocurran al mismo tiempo**

Por ejemplo, si se le va a dar servicio a los neumáticos de un **camión de transporte pesado**, y X representa el número de KM que éstos han recorrido y Y el número de neumáticos que deben ser reemplazados, entonces $f(30{,}000, 5)$ es la **probabilidad** de que los neumáticos hayan recorrido 30,000 KM **y** que el camión necesite 5 neumáticos nuevos.

Distribuciones de probabilidad conjunta

La función $f(x,y)$ es una **distribución de probabilidad conjunta** o **función de masa de probabilidad conjunta** de las variables aleatorias discretas X y Y, si

1. $f(x, y) \geq 0$ para toda (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

Para cualquier región A en el plano xy, $P[(X, Y) \in A] = \sum\sum_A f(x,y)$.

Distribuciones de probabilidad conjunta

Ejemplo:
Se seleccionan al azar 2 repuestos para bolígrafo de una caja que contiene 3 repuestos **azules**, 2 **rojos** y 3 **verdes**. Si *X* es el número de **repuestos azules** y *Y* es el número de **repuestos rojos** seleccionados, calcule:

1) La **función de probabilidad conjunta** $f(x, y)$,

¿Cuáles son los **posibles valores** que puede tomar el par (x, y)?

$$(0, 0), (0, 1), (1, 0), (1, 1), (0, 2) \text{ y } (2, 0)$$

Ahora bien, $f(0, 1)$, por ejemplo, representa la probabilidad de seleccionar un repuesto **verde** y uno **rojo**.

¿**Cuáles es la probabilidad** de que esto ocurra?

Distribuciones de probabilidad conjunta

Ejemplo:
Se seleccionan al azar 2 repuestos para bolígrafo de una caja que contiene 3 repuestos **azules**, 2 **rojos** y 3 **verdes**. Si *X* es el número de **repuestos azules** y *Y* es el número de **repuestos rojos** seleccionados, calcule:

Primero:
¿Cuáles son las combinaciones de obtener 2 repuestos de 8 posibles?

$$\binom{8}{2} = 28$$

Segundo:
¿Cuáles son las posibilidades de obtener 1 repuesto verde de 3 **y** un repuesto rojo de 2?

$$\binom{2}{1}\binom{3}{1} = 6$$

Tercero:
$f(0, 1)$ = es el cociente 6/28 = 3/14

Distribuciones de probabilidad conjunta

Ejemplo:
Se seleccionan al azar 2 repuestos para bolígrafo de una caja que contiene 3 repuestos **azules**, 2 **rojos** y 3 **verdes**. Si *X* es el número de **repuestos azules** y *Y* es el número de **repuestos rojos** seleccionados, calcule:

Ahora bien, *f* (2, 0), por ejemplo, representa la probabilidad de seleccionar dos repuestos **azules**.

¿**Cuáles es la probabilidad** de que esto ocurra? $\frac{3}{28}$

¿Cómo se podría establecer de forma funcional la **función de masa de probabilidad conjunta**?

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

Distribuciones de probabilidad conjunta

	$f(x, y)$	$x = 0$	$x = 1$	$x = 2$	Totales por renglón
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Totales por columna		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

2) $P[(X, Y) \in A]$, donde A es la región $\{(x, y) \mid x + y \leq 1\}$.

¿Cuáles son los puntos del soporte del vector aleatorio (x,y) que cumplen esta condición $\{(x, y) \mid x + y \leq 1\}$.?

$$P[(X, Y) \in A] = P(X + Y \leq 1) = f(0,0) + f(0,1) + f(1,0)$$

$$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}$$

Distribuciones de probabilidad conjunta

La función *f* (*x*,*y*) es una **función de densidad conjunta** de las variables aleatorias continuas *X* y *Y* si

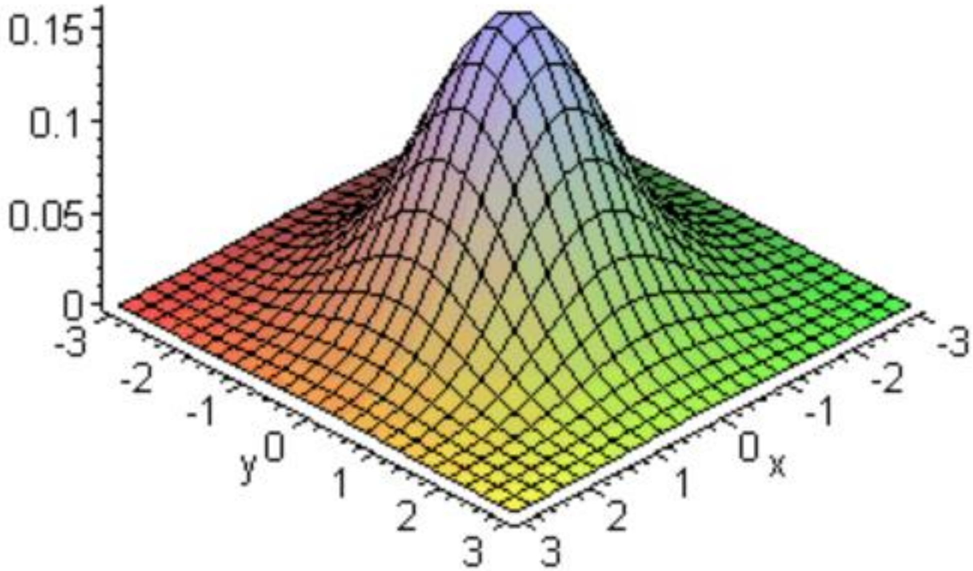
1. $f(x, y) \geq 0$, para toda (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,

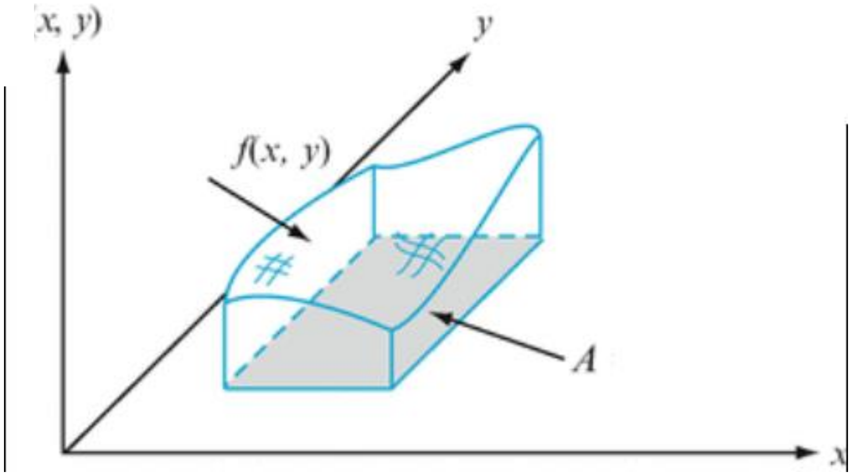
$$P[(X, Y) \in A] = \int \int_A f(x, y)\, dx\, dy, \text{ para cualquier región } A \text{ en el plano } xy.$$

*La **función de densidad conjunta** *f* (*x*,*y*) es una **superficie** que yace sobre el plano *xy*

P*[(*X*,*Y*) ∈ *A*], donde *A* es cualquier región en el plano *xy*, que es igual al **volumen del cilindro recto limitado por la base *A* y la superficie

Distribuciones de probabilidad conjunta





Distribuciones de probabilidad conjunta

Ejemplo:
Una empresa privada opera un local que da servicio a clientes que llegan en **automóvil** y otro que da servicio a clientes que llegan **caminando**.

En un día elegido al azar, sean *X* y *Y*, respectivamente, **las proporciones de tiempo que ambos locales están en servicio**, con función de densidad conjunta de estas variables aleatorias dada por:

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \le x \le 1, 0 \le y \le 1 \\ 0, & \text{en otro caso.} \end{cases}$$

1) Verifique la condición que el **volumen bajo la superficie** es igual a 1

2) Calcule $P[(X, Y) \in A]$, donde $A = \{(x, y) \mid 0 < x < 1/2, 1/4 < y < 1/2\}$

Distribuciones de probabilidad conjunta

La integración de $f(x,y)$ sobre la **totalidad de la región** es:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dx\ dy = \int_{0}^{1}\int_{0}^{1} \frac{2}{5}(2x + 3y)\, dx\ dy$$

$$= \int_{0}^{1} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1} dy = \int_{0}^{1} \left(\frac{2}{5} + \frac{6y}{5}\right) dy$$

$$= \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Bigg|_{0}^{1} = \frac{2}{5} + \frac{3}{5} = 1.$$

2) Calcule $P[(X, Y) \in A]$, donde $A = \{(x, y) \mid 0 < x < 1/2,\ 1/4 < y < 1/2\}$

Distribuciones de probabilidad conjunta

Para calcular la probabilidad utilizamos:

$$P[(X,\ Y) \in A] = P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right)$$

$$= \int_{1/4}^{1/2} \int_{0}^{1/2} \frac{2}{5}(2x + 3y)\ dx\ dy\ = \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1/2} dy$$

$$= \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5}\right) dy\ = \left(\frac{y}{10} + \frac{3y^2}{10}\right)\Bigg|_{1/4}^{1/2} = \frac{1}{10}\left[\left(\frac{1}{2} + \frac{3}{4}\right) - \left(\frac{1}{4} + \frac{3}{16}\right)\right]$$

$$= \frac{13}{160}$$

Distribuciones marginales

Dada la **distribución de probabilidad conjunta** $f(x, y)$ de las variables aleatorias **discretas** X y Y:

- La distribución de probabilidad $g(x)$ **solo de** X se obtiene sumando $f(x, y)$ **sobre los valores de** Y

- La distribución de probabilidad $h(y)$ **solo de** Y se obtiene sumando $f(x, y)$ sobre los valores de Y

Definimos $g(x)$ y $h(y)$ como **distribuciones marginales** de X y Y

Cuando X y Y son **variables aleatorias continuas**, las **sumatorias** se reemplazan por **integrales**

Distribuciones marginales

Las **distribuciones marginales** sólo de X y sólo de Y son

$$g(x) = \sum_y f(x, y) \quad \text{y} \quad h(y) = \sum_x f(x, y)$$

para el caso discreto, y

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{y} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

para el caso continuo.

Observación:
El término ***marginal*** se utiliza aquí porque, en el caso discreto, los valores de $g(x)$ y $h(y)$ son precisamente los **totales marginales** de las columnas y los renglones respectivos, cuando los valores de $f(x, y)$ se muestran en una tabla rectangular.

Distribuciones marginales

Ejercicio:

Muestre que los totales de columnas y renglones de la siguiente tabla dan las distribuciones marginales de **sólo *X*** y **sólo *Y***

	$f(x, y)$	$x = 0$	$x = 1$	$x = 2$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

Para X:

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28}$$

Distribuciones marginales

		x		
	$f(x, y)$	0	1	2
	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
y	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28}$$

De manera similar podemos mostrar que los valores de *h*(*y*) están dados por los **totales de los renglones**

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

¿Son g(x) y h(y) funciones de masa de probabilidad **legítimas**?

Distribuciones marginales

Ejercicio:

Calcule *g*(*x*) y *h*(*y*) para la siguiente **función de densidad conjunta** de las variables X y Y:

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{en otro caso.} \end{cases}$$

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 \frac{2}{5}(2x + 3y)\, dy = \left(\frac{4xy}{5} + \frac{6y^2}{10} \right) \Bigg|_{y=0}^{y=1} = \frac{4x + 3}{5}$$

para $0 \leq x \leq 1$, y $g(x) = 0$ en otro caso.

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 \frac{2}{5}(2x + 3y)\, dx = \frac{2(1 + 3y)}{5}$$

para $0 \leq y \leq 1$, y $h(y) = 0$ en otro caso

Distribuciones marginales

$$g(x) = \frac{4x+3}{5} \quad \text{para } 0 \le x \le 1, \text{ y } g(x) = 0 \text{ en otro caso.}$$

$$h(y) = \frac{2(1+3y}{5} \quad \text{para } 0 \le y \le 1, \text{ y } h(y) = 0 \text{ en otro caso}$$

¿Son g(x) y h(y) **funciones de densidad legítimas**?

Se puede verificar mostrando que se satisfacen las **condiciones**

$$\int_{-\infty}^{\infty} g(x)\, dx = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1.$$

$$P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$$

$$= \int_{a}^{b} \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = \int_{a}^{b} g(x)\, dx$$

Distribuciones condicionales

Si utilizamos la definición de **probabilidad condicional**

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ siempre que } P(A) > 0,$$

Sa ahora A y *B* son los **eventos definidos** por $X = x$ y $Y = y$, respectivamente, entonces

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{f(x, y)}{g(x)}, \text{ siempre que } g(x) > 0$$

donde *X* y *Y* son **variables aleatorias** **discretas**

No es difícil mostrar que la función $f(x, y)/g(x)$, que es estrictamente **una función de *y*** con ***x* fija**, satisface todas las **condiciones** de una distribución de probabilidad

Distribuciones condicionales

Lo anterior es también cierto cuando $f(x, y)$ y $g(x)$ son la densidad conjunta y la distribución marginal, respectivamente, de **variables aleatorias** **continuas**

Este tipo de distribución se llama **distribución de probabilidad condicional** y se define formalmente como sigue

Sean X y Y dos variables aleatorias, discretas o continuas. La **distribución condicional** de la variable aleatoria Y, dado que $X = x$, es

$$f(y|x) = \frac{f(x, y)}{g(x)}, \text{ siempre que } g(x) > 0.$$

De manera similar, la distribución condicional de la variable aleatoria X, dado que $Y = y$, es

$$f(x|y) = \frac{f(x, y)}{h(y)}, \text{ siempre que } h(y) > 0.$$

Distribuciones condicionales

Si deseamos encontrar la **probabilidad** de que la variable aleatoria discreta *X* caiga entre *a* y *b* cuando sabemos que la **variable discreta** *Y* = *y*

$$P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x|y)$$

Cuando *X* y *Y* son **variables aleatorias continuas**, evaluamos

$$P(a < X < b \mid Y = y) = \int_a^b f(x|y)\, dx$$

Distribuciones condicionales

Ejemplo 1:
Remitámonos a ejemplo sobre la distribución conjunta que resulta de obtener 2 repuestos de bolígrafos donde *X* es el número de **repuestos azules** y *Y* es el número de **repuestos rojos** seleccionados al azar de una caja que contiene 3 repuestos **azules**, 2 **rojos** y 3 **verdes**.

$$f(x, y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

	$f(x, y)$	$x = 0$	$x = 1$	$x = 2$
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0
	2	$\frac{1}{28}$	0	0

Calcule la distribución condicional de *X*, dado que *Y* = 1, y utilice el resultado para determinar $P(X = 0 \mid Y = 1)$

¿Qué calculamos primero?

Distribuciones condicionales

$$f(x|y) = \frac{f(x, y)}{h(y)}$$

Necesitamos f(x|y=1) por lo que primero calculamos h(1)

$$h(1) = \sum_{x=0}^{2} f(x, 1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

$$f(x|1) = \frac{f(x, 1)}{h(1)} = \left(\frac{7}{3}\right) f(x, 1) \qquad x = 0, 1, 2.$$

Calculemos f(0|1), f(1|1), y f(2|1)

Distribuciones condicionales

$$f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$$

$$f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right)(0) = 0.$$

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

$$P(X = 0 \mid Y = 1) = f(0|1) = \frac{1}{2}$$

Por lo tanto, si se sabe que 1 de los 2 repuestos seleccionados es **rojo**, tenemos una probabilidad igual a 1/2 de que el otro repuesto **no sea azul**

Distribuciones condicionales

Ejemplo 2:
La **densidad conjunta** para las variables aleatorias (*X*,*Y*), donde ***X* es el cambio unitario de temperatura** y ***Y* es la proporción de desplazamiento espectral** que produce cierta **partícula atómica** es:

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{en otro caso.} \end{cases}$$

a) Calcule las **densidades marginales** *g*(*x*), *h*(*y*) y la **densidad condicional** *f* (*y* | *x*)

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_x^1 10xy^2\, dy = \frac{10}{3} xy^3 \Big|_{y=x}^{y=1} =$$

$$\frac{10}{3} x(1 - x^3),\ 0 < x < 1,$$

Distribuciones condicionales

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{0}^{y} 10xy^2\, dx = 5x^2y^2\Big|_{x=0}^{x=y}$$

$$= 5y^4,\ 0 < y < 1$$

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1 - x^3)} = \frac{3y^2}{1 - x^3}\ ,\ 0 < x < y < 1.$$

$$P\left(Y > \frac{1}{2}\ \middle|\ X = 0.25\right) = \int_{1/2}^{1} f(y \mid x = 0.25)\, dy = \int_{1/2}^{1} \frac{3y^2}{1 - 0.25^3}\, dy$$

$$= \frac{8}{9}$$

Independencia estadística

Si $f\,(x \mid y)$ no depende de y ⇒ entonces $f\,(x \mid y) = g(x)$ ⇒ $f\,(x, y) = g(x)h(y)$

$$f\,(x|y) = \frac{f\,(x, y)}{h(y)}$$

Debería tener sentido que si $f\,(x \mid y)$ **no depende de *y***, entonces, por supuesto, **el resultado de la variable aleatoria *Y*** **no repercute** en el **resultado de la variable aleatoria *X***

En otras palabras, decimos que X y Y son **variables aleatorias independientes**.

Independencia estadística

Sean X y Y dos variables aleatorias, discretas o continuas, con distribución de probabilidad conjunta $f(x,y)$ y distribuciones marginales $g(x)$ y $h(y)$, respectivamente. Se dice que las variables aleatorias X y Y son **estadísticamente independientes** si y sólo si

$$f(x, y) = g(x)h(y)$$

para toda (x,y) dentro de sus rangos.

Dada la función **de densidad conjunta del vector aleatorio (X, Y):**

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1 \\ 0, & \text{en otro caso,} \end{cases}$$

¿Son independientes X y Y?

Independencia estadística

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1 \\ 0, & \text{en otro caso,} \end{cases}$$

para $0 < y < 1$ $\quad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^2 \frac{x(1+3y^2)}{4} dx$

$$= \left(\frac{x^2}{8} + \frac{3x^2y^2}{8}\right)\Bigg|_{x=0}^{x=2} = \frac{1+3y^2}{2}$$

Por lo tanto, usando la definición de la densidad condicional para $0 < x < 2$

$$f(x|y) = \frac{f(x,y)}{h(y)} = \frac{x(1+3y^2)/4}{(1+3y^2)/2} = \frac{x}{2}$$

f(x|y) no depende de y, por lo que efectivamente son independientes.

Independencia estadística

Observación:

La comprobación de la independencia estadística de **variables aleatorias discretas** requiere una **investigación más profunda**, ya que es posible que el producto de las distribuciones marginales sea igual a la distribución de probabilidad conjunta para algunas, aunque no para todas**, las combinaciones de (*x*,*y*)**

Regresando al ejemplo de distribución conjunta que resulta de obtener 2 repuestos de bolígrafos donde X es el número de **repuestos azules** y Y es el número de **repuestos rojos**

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}}$$

		x			Totales
	$f(x,y)$	0	1	2	por renglón
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Totales por columna		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Independencia estadística

Consideremos el punto (0,1) ⇨ $f(0,1)$, $g(0)$ y $h(1)$ son

$$f(0,1) = \frac{3}{14}$$

$$g(0) = \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14}$$

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}$$

$$f(0,1) \neq g(0)h(1)$$

Distribución conjunta de n v.a.

Todas las definiciones anteriores respecto a dos variables aleatorias se pueden **generalizar** al caso de ***n* variables aleatorias**

Sea $f(x_1, x_2, \ldots, x_n)$ la **función de probabilidad conjunta** de las n variables aleatorias $X_1, X_2, \ldots, X_n$

Caso discreto:

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

Caso continuo:

$$g(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_n$$

Ahora podemos obtener **distribuciones marginales conjuntas** como $g(x_1, x_2)$,

Distribución conjunta de n v.a.

$$g(x_1, x_2) = \begin{cases} \sum\limits_{x_3} \cdots \sum\limits_{x_n} f(x_1, x_2, \ldots, x_n) & \text{(caso discreto)}, \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_3\, dx_4 \cdots dx_n & \text{(caso continuo)}. \end{cases}$$

Podríamos considerar numerosas distribuciones condicionales. Por ejemplo, la **distribución condicional conjunta** de X_1, X_2 y X_3, dado que $X_4 = x_4$, $X_5 = x_5$, . . . , $X_n = x_n$

$$f(x_1, x_2, x_3 \mid x_4, x_5, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n)}{g(x_4, x_5, \ldots, x_n)}$$

donde $g(x_4, x_5, \ldots, x_n)$ es la **distribución marginal conjunta** de las variables aleatorias X_4, X_5, . . . , X_n

Distribución conjunta de n v.a.

Se tiene la siguiente definición para la **independencia mutua** de las variables $X_1, X_2, \ldots, X_n$

Sean $X_1, X_2, \ldots, X_n$, n variables aleatorias, discretas o continuas, con distribución de probabilidad conjunta $f(x_1, x_2, \ldots, x_n)$ y distribuciones marginales $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectivamente. Se dice que las variables aleatorias $X_1, X_2, \ldots, X_n$ son recíproca y **estadísticamente independientes** si y sólo si

$$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \cdots f_n(x_n)$$

para toda $(x_1, x_2, \ldots, x_n)$ dentro de sus rangos.

Ejemplo: Suponga que el **tiempo de vida** en anaquel de cierto producto comestible perecedero empacado en cajas de cartón, en años, es una variable aleatoria cuya función de densidad de probabilidad está dada por:

Distribución conjunta de n v.a.

$$f(x) = \begin{cases} e^{-x} & x > 0, \\ 0, & \text{en otro caso.} \end{cases}$$

Represente la función de distribución conjunta de 3 variables aleatorias $X1$, $X2$ y $X3$

$$f(x_1, x_2, x_3) = f(x_1)f(x_2)f(x_3) = e^{-x_1}e^{-x_2}e^{-x_3}$$

$$= e^{-x_1 - x_2 - x_3}$$

CIMAT

Centro de Investigación en Matemáticas, A.C.

Unidad Monterrey
CIDICS, Campus de la Salud, UANL Ave. Carlos Canseco s/n
Col Mitras Centro, Monterrey N.L., México, C.P.66460 México
Tel +52 (81) 8329 4000 Ext 1778
cimat@cimat.mx / www.cimat.mx

CIMAT
Unidad Monterrey
Centro de Investigación en Matemáticas, A.C.