Probabilidad

1. **Conceptos Básicos de Probabilidad**
2. **Espacio Muestral**
3. **Tipos de Eventos**
4. **Eventos Dependientes e Independientes**
5. **Probabilidad Condicional**
6. **Reglas de la Probabilidad**
7. **Combinatoria Básica**
8. **Distribuciones de Probabilidad**
9. **Aplicación de la Probabilidad en la Ciencia Política**

Introducción a la probabilidad

Distintos tipos de probabilidad

- PROBABILIDAD CLÁSICA (equiprobable) →

  P(A) = casos favorables a A / casos totales

- PROBABILIDAD COMO FRECUENCIA RELATIVA →

  P(A) = cant de veces que ocurre A / nro total de experimentos

- PROBABILIDAD SUBJETIVA (BAYESIANA) →

  la probabilidad se actualiza a medida que se obtiene nueva información

- DEFINICIÓN AXIOMÁTICA (MATEMÁTICA) DE PROBABILIDAD →

  reglas de la probabilidad

¿Qué es la probabilidad?

¿Qué es la probabilidad?

¿Qué la diferencia de la inferencia?

- Teoría de probabilidad: provee modelos para entender (ordenar) los experimentos o situaciones inciertas o azarosas, sobre las cuales no conocemos el resultado hasta que ocurren

- Inferencia estadística: provee herramientas para, a partir de una muestra, afirmar con fundamentos respecto de algún elemento o característica de la población. Posibilidad de tomar decisiones.

- Probabilidad asociada al azar ⇒ Probabilidad frecuentista
- Probabilidad asociada a la incertidumbre/ignorancia/desconocimiento ⇒ Probabilidad bayesiana

Ejemplo: Tirar una moneda 5 veces

Resultado = {Cara, Cara, Cruz, Cara, Cruz}

¿Cuál es la probabilidad de que salga cara?

Prob (cara) = ?

¿ 3 / 5 ?

~~¿ 3 / 5 ?~~

~~¿ 3 / 5 ?~~

Frecuencia relativa (de cara)

3 / 5 $\Rightarrow$ Dato que viene de la observación

$\neq$

Probabilidad

3 / 5 $\Rightarrow$ Dato que viene de la observación

$\neq$

Probabilidad

Teoría frecuentista: relación entre teoría y observación:

Cuando Lim tiros $\rightarrow \infty$

Freq. relativa (cara) $\approx$ Prob(cara)

Definiciones de probabilidad

- **EXPERIMENTO ALEATORIO: proceso con resultado incierto. Puede repetirse en iguales condiciones para la obtención de “datos”**

Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de “datos”
- **RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado**

Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de “datos”
- RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado
- **ESPACIO MUESTRAL: conjunto de todos los posibles resultados de un experimento aleatorio**

Definiciones de probabilidad

- EXPERIMENTO ALEATORIO: proceso con resultado azaroso. Puede repetirse en iguales condiciones para la obtención de “datos”
- RESULTADOS: realizaciones del experimento aleatorio. No conocemos con seguridad el resultado, pero sí podemos conocer la probabilidad de ocurrencia de cada resultado
- ESPACIO MUESTRAL: conjunto de todos los posibles resultados de un experimento aleatorio
- **EVENTO: algún subconjunto del espacio muestral**

Experimento aleatorio

Por ejemplo: tirar un dado, una moneda, sacar una carta de un mazo.

Todos estos experimentos pueden repetirse bajo las mismas condiciones (diferencia con los experimentos o eventos sociales).

Si bien conocemos cuáles son sus posibles salidas o resultados (sacar alguna de las cartas del mazo, sacar números del 1-6 en el dado, etc), no es posible predecir con exactitud cuál será el resultado al ejecutarlo.

Experimento aleatorio $\neq$ experimento social

Evaluar la probabilidad de ocurrencia de cierto evento social no es independiente del contexto y de lo social. Por eso no es un experimento que pueda repetirse siempre bajo las mismas condiciones. Probabilidad **condicional**.

Por ejemplo: ¿qué probabilidad existe de que durante este mes se produzca una huelga laboral en Argentina?

Resultados

Lo que se obtiene al “tirar” la moneda o el dado o sacar la carta es un resultado.

Resultados posibles de tirar un dado = $\{1, 2, 3, 4, 5, 6\}$

Resultados posibles de tirar una moneda = {cara, cruz}

Existe una probabilidad asociada a cada resultado:

Prob(cara)

Prob(cruz)

Espacio muestral Ω

Conjunto de todas las posibles salidas, resultados, eventos que pueden obtenerse

Ejemplo:

$\Omega = \{\varnothing, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{2, 4, 6\}, \{1, 3, 5\} \ldots \}$

Eventos

A partir de la combinación de resultados obtenemos eventos posibles.

$$\boldsymbol{\Omega} = \{\varnothing, \{1, 2, 3, 4, 5, 6\}, \{1, 2\}, \{2, 4, 6\}, \{1, 3, 5\}$$

En algunos casos puede interesarnos estimar la probabilidad de un evento más que la de un resultado.

Por ejemplo: una apuesta a partir de si sale un dado par o impar

Definición de la probabilidad: teoría de conjuntos

A partir de dos eventos del espacio muestral $\boldsymbol{\Omega}$ (A y B):

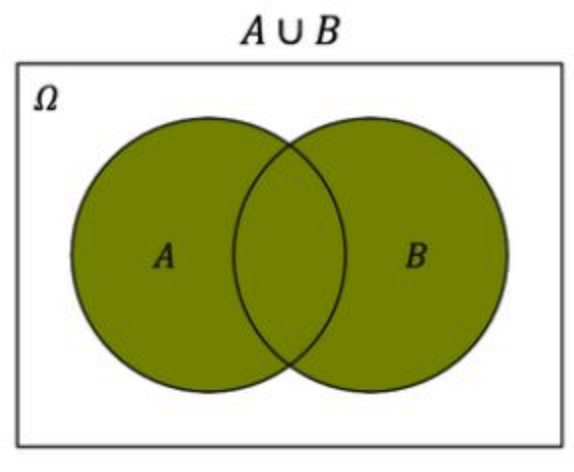
Evento A = {1, 3, 5} Evento B = {menor a 3}

Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

A U B (unión de A y B): significa A o B o ambos

A U B = {1, 2, 3, 5}

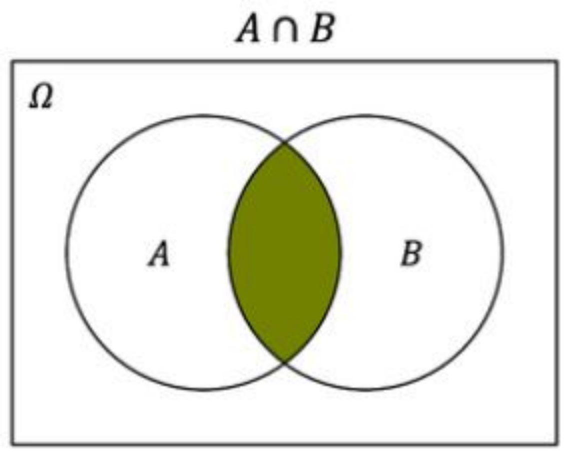


Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

A ∩ B (intersección de A y B): sólo los resultados que ocurren en ambos eventos

A ∩ B = {1}

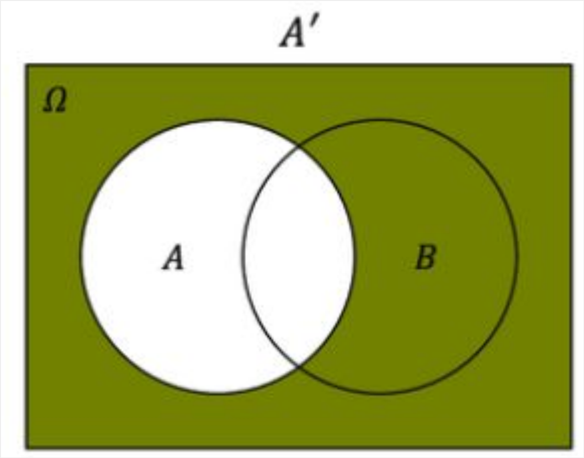


Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

A^c (Complemento de A): todos los resultados que no están en A

A^c = {2, 4, 6}



Definición de la probabilidad: teoría de conjuntos

Evento A = {1, 3, 5} Evento B = {menor a 3}

*No son mutuamente excluyentes en este ejemplo pero…

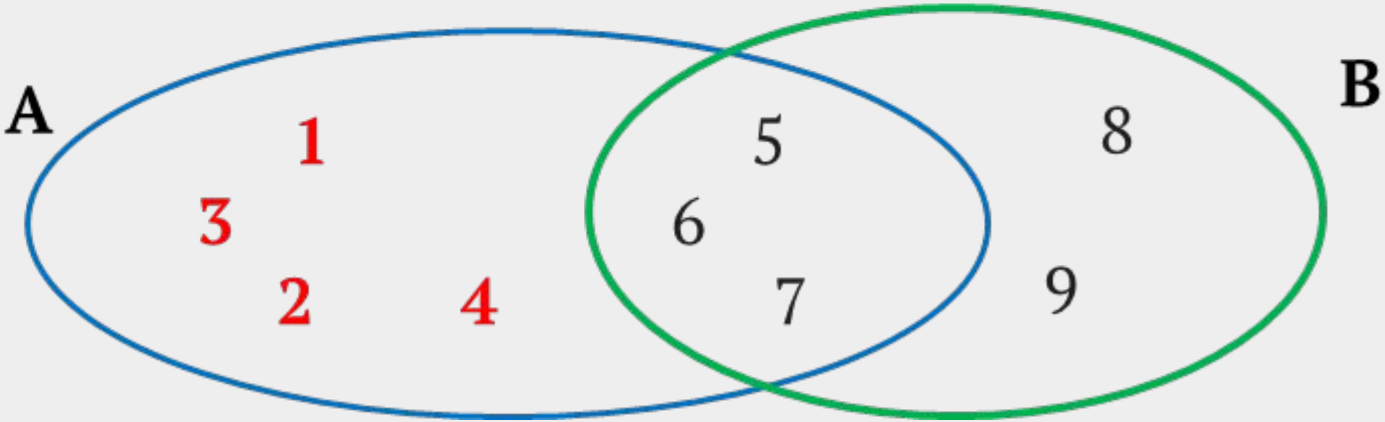
Si $A \cap B = \varnothing$ $\longrightarrow$ A y B son mutuamente excluyentes

Ejemplo: ¿puede un dado salir par e impar a la vez?

Definición de la probabilidad: teoría de conjuntos

Diferencia de conjuntos

- Ejemplo:
- A={1, 2, 3, 4, 5, 6, 7} y B={5, 6, 7, 8, 9}
- A-B= {1, 2, 3, 4}



Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 bolas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos AUB, A∩B y B'∩A' (complemento).

Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 bolas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos AUB, A∩B y B'∩A' (complemento).

E={ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } (espacio muestral)

A={7, 8, 9}; B={0, 1, 2, 3, 4, 5, 7, 8, 9}; C={0, 1, 2, 3, 4, 5}

Ejemplo de unión e intersección de conjuntos

Una bolsa tiene 10 bolas numeradas del 0 al 9. Describir los sucesos: A="Mayor que 6“; B="No obtener 6“; C="Menor que 6“, definiendo por extensión; hallar los sucesos A∪B, A∩B y B'∩A' (complemento).

E={ 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 } (espacio muestral)

A={7, 8, 9}; B={0, 1, 2, 3, 4, 5, 7, 8, 9}; C={0, 1, 2, 3, 4, 5}

A U B={0, 1, 2, 3, 4, 5, 7, 8, 9}

A∩B={7, 8, 9} que es =A;   B’∩A’={6} que es =B’

Definición axiomática de la probabilidad: algunas reglas

- No negatividad: $P(A) \geq 0, \forall A \subseteq \Omega$
- Las probabilidades suman 1: $P(\Omega) = 1$
- Las probabilidades están entre 0 y 1: $0 \leq P(A) \leq 1$
- Probabilidad de un subconjunto: Siendo A y B eventos en Ω $(A, B \subseteq \Omega)$, si $B \subseteq A \rightarrow P(B) \leq P(A)$
- Regla de la suma para eventos mutuamente excluyentes: Si $P(A) = 0.20$ y $P(B) = 0.35$

  $P(A \text{ o } B) = 0.55$

(reiteramos)

- Regla de la suma para eventos mutuamente excluyentes:

  Si P(A) = 0.20 y P(B) = 0.35 P(A o B) = 0.55

- Si los sucesos no son mutuamente excluyentes, la regla de la adición se formula del siguiente modo:

$$P\,(A \text{ o } B) = P\,(A) + P(B) - P\,(A \text{ y } B)$$

Donde P (A y B) denota la probabilidad de que A y B ocurran al mismo tiempo. -INTERSECCIÓN

Regla de Suma para eventos no mutuamente excluyentes

Si esto lo mostramos en un diagrama de Venn, y tomamos la notación de conjuntos:

$$P\,(A \cup B) = P\,(A) + P\,(B) - P\,(A \cap B)$$

Eventos equiprobables (probabilidad clásica)

$$1 = P(\Omega) = P(\text{cara U cruz})$$ -eventos mutuamente exluyentes-

$$= P(\text{cara}) + P(\text{cruz}) = P(\text{cara}) + P(\text{cara})$$

$$= 2^* P(\text{cara})$$

$$= P(\text{cara}) = P(\text{cruz}) = 1/2$$

Si hay 2 resultados posibles → c/u ocurre con probabilidad 1/2

Si hay 6 resultados posibles (ejemplo dado) → c/u ocurre con probabilidad 1/6

Pasemos a un ejemplo...

¿Qué tan probable es que dos personas, elegidas al azar, compartan el día de cumpleaños?

¿Qué les parece?

Pasemos a un ejemplo…

Punto clave: personas elegidas al azar supone considerar todas las combinaciones posibles (de personas que existan)

Supongamos equiprobabilidad e invirtamos el problema:

$P(P2 \neq P1) = 364/365$

¿Y qué ocurre si quiero pensar el problema para estimar la probabilidad de que 3 personas compartan cumpleaños?

$$P\left(P2 \neq P1 \cap P3 \neq P1 \text{ y } P2\right) =$$

$$= 364/365 * 363/365 \ldots.\ 365 - (n-1)/365$$

Si nos preguntamos por la probabilidad de que en un grupo de 30, al menos 2 de ellas compartan cumpleaños...

P(30 personas cumplan en días distintos) =

= 364/365 * 363/365.... 365 - 29/365

= 0.3

P(30 personas cumplan en días distintos)^c = P(al menos dos personas entre las 30 compartan cumpleaños)

O

1 - P(30 personas cumplan en días distintos) = P(al menos dos personas entre las 30 compartan cumpleaños)

1 - 0.3 = 0.7

La probabilidad es del 70%

Probabilidad condicional

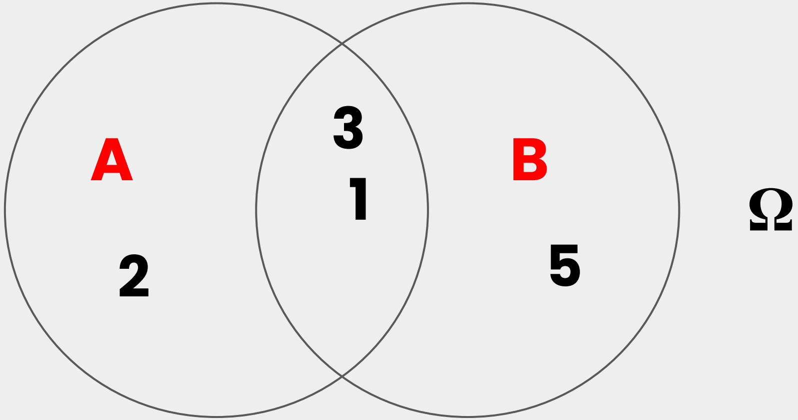
Supongamos que dado Ω, conocemos que ocurrió el evento B

$B = \{1, 3, 5\}$

$A = \{1, 2, 3\}$

¿Cuál es la probabilidad de ocurrencia del evento A?

Probabilidad condicional

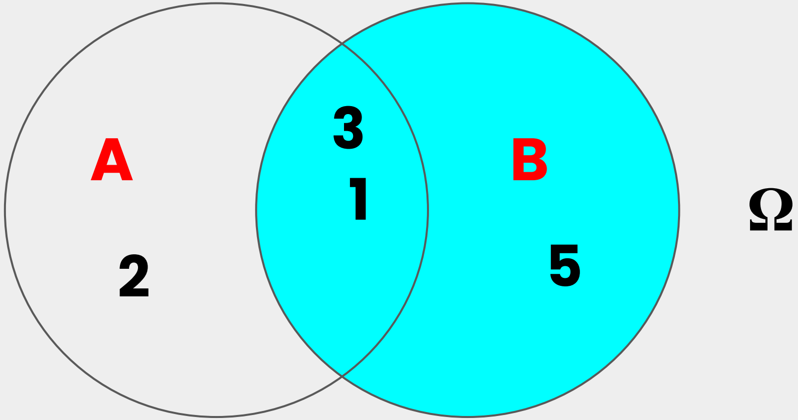


Inicialmente:

P(A) = 3/6 (o 1/2)

P(B) = 3/6 (o 1/2)

Probabilidad condicional



Sabiendo que ocurrió B:

$P(A|B) = ?$

$P(B) = 1/2$

$P(A|B) = P(A \cap B) | B$

$= P(A \cap B) / P(B) = (2/6) / (3/6) = 2/3$

Probabilidad condicional: ley de la multiplicación

$$P(A|B) = P(A \cap B) \; / \; P(B)$$

$$\Downarrow$$

$$P(A \cap B) = P(A|B) * P(B)$$

Probabilidad condicional: ley de la probabilidad total

- Permite determinar la probabilidad incondicional de un evento a partir de probabilidades condicionales:

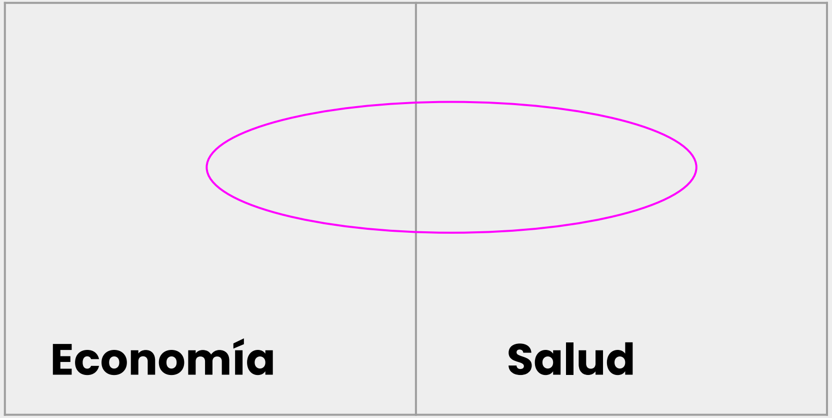
Dado un evento A y una partición del espacio muestral en distintas partes: B1, B2,… Bn, la probabilidad total de A es:

$$P(A) = P(A \mid B1) * P(B1) + P(A \mid B2) * P(B2) + P(A \mid Bn) * P(Bn)$$

Notemos que los Bi, 1=1…, n conforman la totalidad del espacio muestral y son mutuamente excluyentes (no presentan elementos en común)

Probabilidad total: ejemplo

Probabilidad de ser mujer en distintas áreas de la administración pública

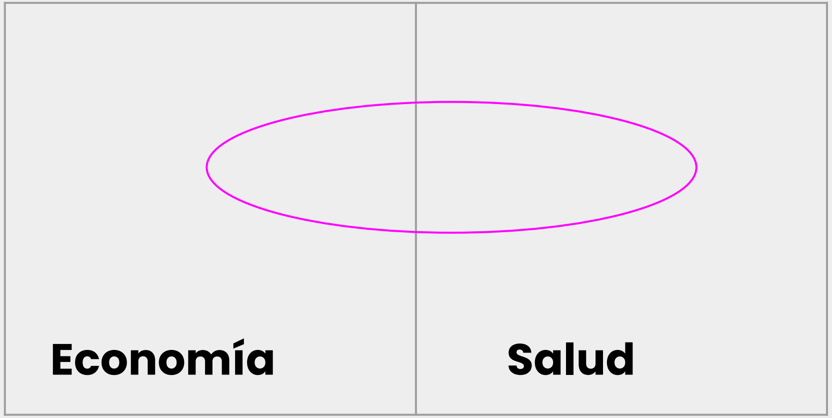


A = ser mujer

Nos preguntamos por la probabilidad del evento A dado que estoy en un “mundo” o el otro

Probabilidad total: ejemplo

Probabilidad de ser mujer en distintas áreas de la administración pública



A = ser mujer

$P(A) = P(A|E) * P(E) + P(A|S) * P(S)$ – **en cada caso se pondera por la probabilidad de que estemos en uno u otro grupo**

Y esto es igual a:

$P(A) = P(A \cap E) + P(A \cap S)$ - porque ambos son eventos mutuamente excluyentes

Un ejemplo interesante en ciencias sociales:
Aplicación de la ley de probabilidad total para Encuestas sensibles

¿Cómo medir la incidencia de eventos no deseados socialmente?:

Por ejemplo:

-Consumo de drogas

-ETS

-Religión

-Algunas posturas políticas

-Sueldo

Un ejemplo interesante en ciencias sociales: Aplicación de la ley de probabilidad total para Encuestas sensibles

Para realizar una encuesta se le pide a los participantes que saquen al azar (y observen en privado) una carta de un mazo.

40% pregunta inocua

60% pregunta de interés

Pregunta inocua: ¿Usted ha nacido en año par?

Pregunta de interés: ¿Alguna vez usted ha consumido drogas ilegales?

Cada encuestado responde **"sí"** o **"no"** pero no revela a qué pregunta contestó ➡ **¿Podemos conocer qué porcentaje de la muestra ha consumido drogas ilegales?**

Un ejemplo interesante en ciencias sociales:
Aplicación de la ley de probabilidad total para Encuestas sensibles

	P. inocua	P. interés
Sí		
No		

Ω

$P(\text{sí}) = P(\text{sí}|PI) * P(PI) + P(\text{sí}|Pint) * P(Pint)$
$P(\text{sí}) = \mathbf{P(\text{sí}|PI)} * 0.4 + P(\text{sí}|Pint) * 0.6$

Podemos aproximar la proporción de personas que nacen en año par e impar, y conocer **P(sí|PI) Supongamos que P(sí|PI) = 0.5**

$P(\text{sí}) = 0.5 * 0.4 + P(\text{sí}|Pint) * 0.6$
P(sí) = total de participantes que responden sí / sobre total de participantes
$\mathbf{P(\text{sí}|Pint)} * 0.6 = P(\text{sí}) - 0.5 * 0.4$

Eventos independientes

Vimos un ejemplo en el que conocer la probabilidad de ocurrencia de un evento B nos brinda información acerca de la probabilidad de un evento A. Esto supone que A y B no son eventos independientes.

En contraste, 2 eventos son independientes entre sí cuando conocer información sobre la probabilidad de uno no es relevante para estimar la probabilidad del otro. Esto supone que:

$$\mathbf{P(A|B) = P(A)}$$

¿Qué ejemplos de eventos independientes se pueden pensar?

Eventos independientes

Vimos un ejemplo en el que conocer la probabilidad de ocurrencia de un evento B nos brinda información acerca de la probabilidad de un evento A. Esto supone que A y B no son eventos independientes.

En contraste, 2 eventos son independientes entre sí cuando conocer información sobre la probabilidad de uno no es relevante para estimar la probabilidad del otro. Esto supone que:

$$\mathbf{P\,(A|B) = P(A)}$$

$$\mathbf{P(A|Bc) = P(A)}$$

Ejemplo:

A = Mañana va a aumentar el subte

B = Esta clase terminará a las 22 hs

Podemos suponer que ambos eventos son independientes, en tanto que un aumento en el subte no aumenta ni disminuye las probabilidades de que la clase termine a las 22 hs y viceversa

Eventos independientes

Bajo eventos dependientes, partíamos de la siguiente ecuación:

$$P(A|B) = P(A \cap B) \; / \, P\,(B)$$

Y al pasar **P(B)** multiplicando se obtenía la ley de la multiplicación

Con eventos independientes, siendo que **$P\,(A|B) = P(A)$,** hacemos un pasaje de términos análogo y obtenemos:

$$P(A) = P(A \cap B) \; / \, P\,(B)$$

$$P(A \cap B) = P(A) * P(B)$$

Si se cumple esta última igualdad entonces se trata de eventos independientes, de lo contrario, se trata de eventos dependientes.

Eventos independientes

¿Qué sucede con la independencia y los eventos mutuamente excluyentes?

Eventos (in)dependientes

Veamos un ejemplo en el que conocer la ocurrencia de un evento A nos da información sobre la probabilidad de ocurrencia de un evento B y cómo no se cumple dicha igualdad

Evento A = Una persona tiene un título universitario

Evento B = La persona gana un salario alto

Suponemos las siguientes probabilidades:

$P(A) = 0.3$

$P(B|A) = 0.6$

$P(B|Ac) = 0.2$

$P(B) = 0.4$

Y queremos ver si se cumple la igualdad de eventos independientes:

$$\mathbf{P(A \cap B) = P(A) * P(B)}$$

$P(A \cap B)=P(B|A)*P(A)$

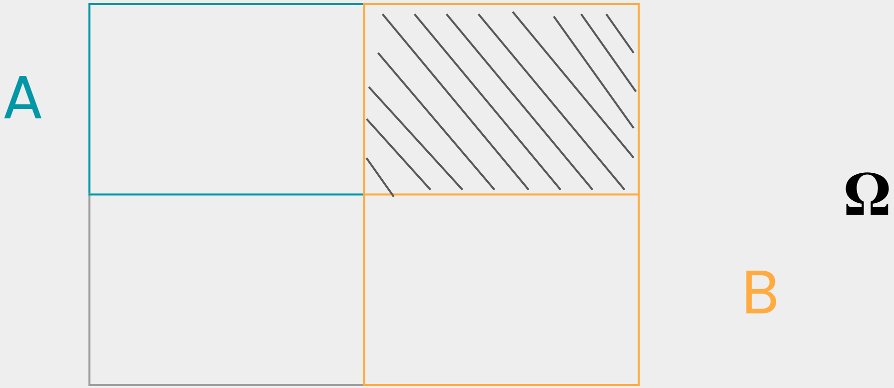
$\mathbf{P(A \cap B)=0.6*0.3 = 0.18}$

$\mathbf{P(A)*P(B) = 0.3*0.4 = 0.12}$

$P(A \cap B) \neq P(A) * P(B)$ □ A y B **no** son eventos independientes

Eventos independientes

Ahora verifiquemos la independencia entre dos eventos gráficamente:



$$P(A|B) = P(A \cap B) \; / \; P(B)$$

Y bajo eventos independientes:

$$\mathbf{P(A) = P(A \cap B) \; / \; P(B)}$$

Lo que equivale a decir que $P(A)$ tiene que ser a Ω como $P(A \cap B)$ a $P(B)$:

$$P(A)/\Omega = ½ / 1 = ½$$

$$P(A \cap B) / P(B) = ¼ / ½ = ½$$

Teorema de Bayes

Probabilidad subjetiva:

Ya no hablamos de equiprobabilidad o de probabilidad frecuentista, sino que la probabilidad de un evento se relaciona con las creencias del observador. La probabilidad se actualiza a partir de la obtención de nueva información.

Aplicaciones:

-Medicina

-Juicios, criminología

-Economía y finanzas

-Marketing, diseño web (AB testing)

Teorema de Bayes

Frecuentemente sucede que se desea calcular **$P(A|B)$**, cuando en realidad lo que se conoce es:

$P(B|A)$, o

$P(A)$

$P(Ac)$ …

Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

Dada la existencia de tests médicos poco específicos (alto % de falsos positivos), no se suele someter a la población a un testeo frecuente de todas las enfermedades. Más bien se utilizan los tests como confirmación a partir de que un paciente presenta síntomas.

FALSOS POSITIVOS: ¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X?

$P(A|B)$

$P(A)$ = Sano, $P(B)$ = +

Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

Supuestos:

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05 (0.99 de probabilidad de estar sano)

¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X? - $P(A|B)$ -

$P(A)$ = la persona está sana

$P(B)$ = la persona dió +

→ **No conocemos $P(A|B)$** → **¿Y conocemos $P(B|A)$?**

Teorema de Bayes: motivación

Ejemplo del uso de tests médicos para detectar una enfermedad rara

Supuestos:

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05
  (0.99 de probabilidad de estar sano)

¿Cuál es la probabilidad de que una persona esté sana siendo que obtuvo un test positivo de una enfermedad X? - P(A|B) -

P(A) = la persona está sana

P(B) = la persona dió +

➡ **No conocemos P(A|B)** ➡ **¿Y conocemos P(B|A)?**

El teorema de bayes es una igualdad matemática resultante de aplicar:

-Ley de la multiplicación
-Probabilidad conjunta
-Ley de probabilidad total

Repasemos un poco lo que estuvimos viendo para así llegar al teorema de Bayes y poder hallar **$P(A|B)$**...

Probabilidad condicional

Ley de la multiplicación

$$P(A|B) = P(A \cap B) \ / \ P(B)$$

$$\Downarrow$$

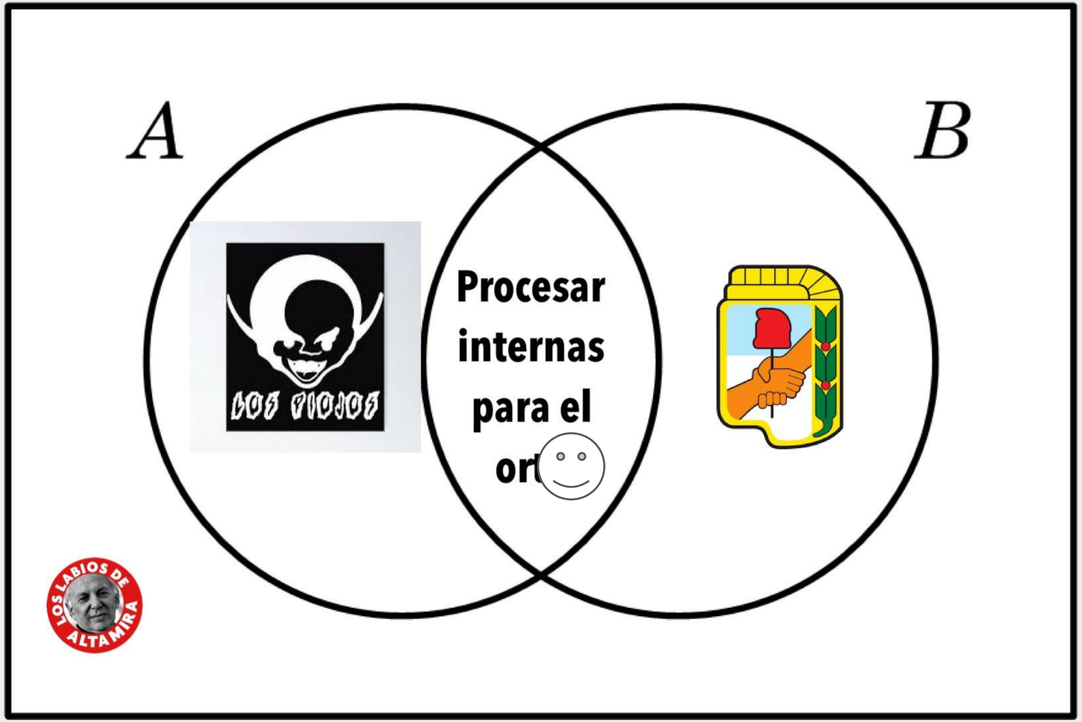
$$P(A \cap B) = P(A|B) * P(B)$$

Simetría de la probabilidad conjunta

$$P(A \cap B) = P(B \cap A)$$

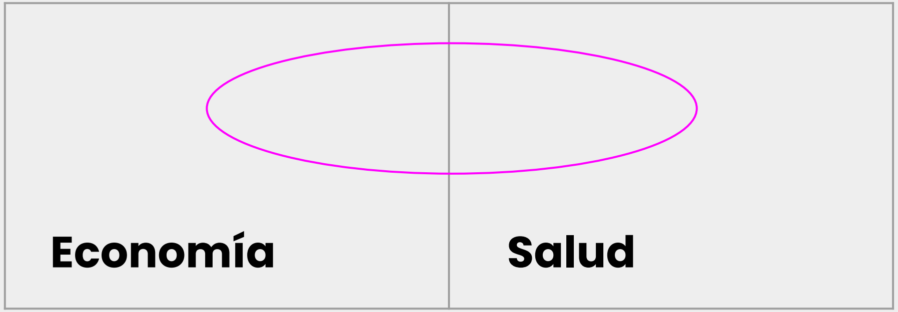
¿Recuerdan lo que era la intersección entre 2 eventos?

¿Recuerdan lo que era la intersección entre 2 eventos?



Probabilidad condicional

Ley de la probabilidad total



- Permite determinar la probabilidad incondicional de un evento a partir de probabilidades condicionales:

Dado un evento A y una partición del espacio muestral en distintas partes: B1, B2,… Bn, la probabilidad total de A es:

$$P(A) = P(A \mid B1) * P(B1) + P(A \mid B2) * P(B2) + P(A \mid Bn) * P(Bn)$$

Notemos que los Bi, 1=1…, n conforman la totalidad del espacio muestral y son mutuamente excluyentes (no presentan elementos en común)

TEOREMA DE BAYES

Invertimos las probabilidades conjuntas (intersección) y usamos ley de la multiplicación en el numerador

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) * P(A)}{P(B)}$$

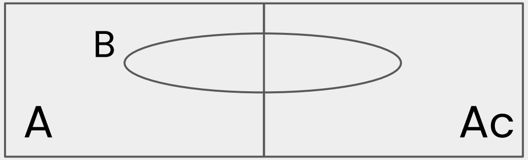
TEOREMA DE BAYES

Invertimos las probabilidades conjuntas (intersección) y usamos ley de la multiplicación en el numerador

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A) * P(A)}{P(B)}$$

Ley de probabilidad total en el denominador

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A)*P(A) + P(B|Ac) * P(Ac)}$$



��

TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A)*P(A) + P(B|Ac) * P(Ac)}$$

Datos:

- Una enfermedad rara afecta sólo al 1% de la población
- Si una persona está enferma, el test dará positivo con probabilidad 0.98
- Si la persona está sana, el test dará positivo con probabilidad 0.05 (0.99 de probabilidad de estar sano)

Definición de los eventos:

$P(Sana) = 0.99 \rightarrow P(A) = 0.99$

$P(Sanac) = 0.01 \rightarrow P(Ac) = 0.01$

$P(+|Sana) = 0.05 \rightarrow P(B|A) = 0.05$

$P(+|Sanac) = 0.98 \rightarrow P(B|Ac) = 0.98$

TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A)*P(A) + P(B|Ac) * P(Ac)}$$

Definición de los eventos:

$P(Sana) = 0.99 \longrightarrow P(A) = 0.99$

$P(Sanac) = 0.01 \longrightarrow P(Ac) = 0.01$

$P(+|Sana) = 0.05 \longrightarrow P(B|A) = 0.05$

$P(+|Sanac) = 0.98 \longrightarrow P(B|Ac) = 0.98$

$$P(sana|+) = \frac{0.05 * 0.99}{0.05 * 0.99 + 0.98 * 0.01} = 0.83$$

Cuando un test da positivo, lo más probable es que la persona esté sana

TEOREMA DE BAYES: aplicación al estudio de los falsos positivos

¿Qué sucede si no se trata de una enfermedad rara?
Si se tratara de una enfermedad que afecta al 50% de la población (una gripe por ejemplo)...

$$P(\text{sana}|+) = \frac{0.05 * 0.5}{0.05 * 0.5 + 0.98 * 0.5}$$

$$P(\text{sana}|+) = \frac{0.05 * \cancel{0.5}}{\cancel{0.5} * (0.05 + 0.98)} = 0.048$$

En este caso cuando el test da positivo lo más probable es que la persona esté enferma

TEOREMA DE BAYES: aplicación en juicios

¿De qué forma se podría aplicar el teorema de Bayes a la hora de resolver un caso judicial?

TEOREMA DE BAYES: aplicación en juicios

- El uso de este teorema se relaciona con el estudio de la probabilidad de observar ciertas conductas o comportamientos sospechosos, dada la inocencia de una persona en un caso judicial

Estar cerca de la escena del crimen, portar ropa de la víctima, haber estado en contacto con ella poco antes (algunos ejemplos) $\longrightarrow$ **A** = tener comportamientos "sospechosos", tomar decisiones "dudosas"

Presunción de inocencia hasta que se demuestre lo contrario $\longrightarrow$ **B** = ser inocente

Al observar **$P(A|B)$** nos estamos preguntando por **¿qué tan probable es observar conductas sospechosas de una persona inocente?** Y el objetivo sería rechazar la H0 de que una persona es inocente únicamente si existe suficiente evidencia en su contra, es decir se busca que:
$P(A|B) \approx 0$ al condenar a una persona. (De lo contrario $\longrightarrow$ error de tipo I)

El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

- En el año 1999 Sally Clarke, una abogada de Cheshire, Inglaterra fue acusada de haber asesinado a sus dos hijos. El primero de ellos murió a las 8 semanas de edad (1996), y el segundo a las 11 semanas (1998).

- Tres años después fue puesta en libertad. La sentencia de 1999 se considera uno de los mayores errores judiciales de la historia de Inglaterra.

El error se debió a lo que se conoce como prosecutor fallacy (el **error probabilístico** de hallar a alguien culpable a raíz de un falso razonamiento científico).
Veamos este error más de cerca...

El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

El principal problema del razonamiento que dio lugar a la sentencia tiene que ver con haber considerado a los dos eventos como **independientes**. Esto llevó al siguiente cálculo por parte de un pediatra inglés:

$$P(A \cap B) = P(A) * P(B)$$
$$P(A \cap B) = 1/8543 * 1/8543 \approx 1 / 73.000.000$$

Por aquel entonces, nacían alrededor de 700.000 bebés por año, por lo que se trataría de un suceso con ocurrencia cada 100 años.

El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + P(B|Ac) * P(Ac)}$$

$P(H)$ = los dos hijos murieron de muerte súbita
$P(H) = 1/1300 * 1/100 = 0.0000077$
$P(Hc) = 1 - P(H) = 0.9999923$. Prob de que mueran de otra causa distinta
$P(D)$ = ambos hijos murieron de forma repentina y sorpresiva
$P(D|H) = 1$
$P(D|Hc) = 30/650.000*1/10 = 0.0000046 \longrightarrow$ la prosecución equipara esto a un crimen

Queremos estimar $P(H|D)$ = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

$$P(H|D) = \frac{P(D|H) * P(H)}{P(D|H) * P(H) + P(D|Hc) * P(Hc)}$$

$P(H)$ = los dos hijos murieron de muerte súbita
$P(H) = 1/1300 * 1/100 = 0.0000077$
$P(Hc)$ = mueren de otra causa distinta
$P(D)$ = ambos hijos murieron de forma repentina y sorpresiva
$P(D|H) = 1$
$P(D|Hc) = 30/650.000*1/10 = 0.0000046$ → la prosecución equipara esto a un crimen

Queremos estimar $P(H|D)$ = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

El uso de la probabilidad y estadística en juicios: el caso de Sally Clarke

En este caso se debería haber utilizado el Teorema de Bayes, que da lugar a la ocurrencia de un evento altamente improbable al compararlo con otros también improbables eventos

$$P(H|D) = \frac{1*0.0000077}{1*0.0000077 + 0.0000046 * 0.999992} \approx 0.6$$

$P(H)$ = los dos hijos murieron de muerte súbita
$P(H) = 1/1300 * 1/100 = 0.0000077$
$P(Hc)$ = mueren de otra causa distinta
$P(D)$ = ambos hijos murieron de forma repentina y sorpresiva
$P(D|H) = 1$
$P(D|Hc) = 30/650.000*1/10 = 0.0000046$ → la prosecución equipara esto a un crimen

Queremos estimar $P(H|D)$ = probabilidad de que la causa haya sido muerte súbita dado lo inesperado y repentino de la misma

Variables Aleatorias - Elementos

1. ~~**Definición de Variable Aleatoria**~~
2. ~~**Variable Aleatoria Discreta (VAD)**~~
3. ~~**Distribución de Probabilidad de las VAD**~~
4. ~~**Función de Distribución Acumulada de las VAD**~~
5. ~~**Media y Varianza**~~
6. ~~**Variable Aleatoria Continua (VAC) (función de densidad, acumulado)**~~
7. ~~**Media y Varianza**~~
8. ~~**Transformación de una Variable Aleatoria: la estandarización**~~

Variables Aleatorias - Definición y sus elementos

Una **variable aleatoria** es una **función** que asigna un **valor numérico a cada resultado** de un experimento aleatorio. Esto permite modelar la incertidumbre al mapear cada resultado a un número real definido por una función X cuyo valor dependerá de los resultados del experimento

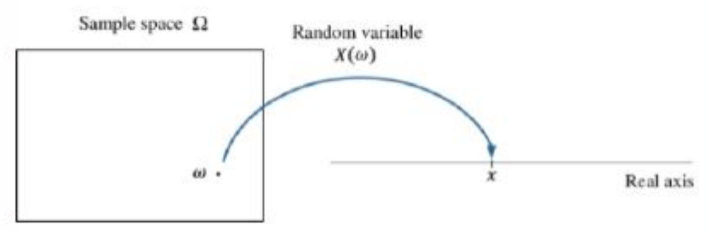


Figura 13.3 From the sample space Ω to the real axis.

El resultado de un experimento aleatorio no puede predecirse con certeza antes de realizarlo (a priori). Por tanto, no podemos predecir con certeza el valor de la variable aleatoria. Sin embargo, podemos calcular la probabilidad de observar sus posibles valores, pero ¿cómo?

Variables Aleatorias - Discretas y su f(x) de probabilidad

Las variables aleatorias, como en la estadística descriptiva, se dividen entre discretas y continuas.

Las variables aleatorias **discretas** son aquellas a las que se les asigna un número específico al resultado del experimento. Por ejemplo, si en una encuesta determinada se le pregunta a los votantes si apoyan o no a un determinado candidato, la variable aleatoria puede tomar valores como "1" si el votante apoya al candidato y "0" si no lo apoya. Esta sería una variable aleatoria discreta. Se entiende que **los valores que toma la variable son finitos o contables.**

A cada valor x de X se le asigna una probabilidad que es igual a la probabilidad de la unión de los resultados elementales para los que X = x. La probabilidad del evento X=x está dado por P (X=x), de aquí es posible hacer corresponder los valores de la variable aleatoria X con una **distribución de probabilidad** determinada. De ella surge la **función de probabilidad f(x) = P(X=x)**. Por lo tanto, f(x) debe ser mayor a 0 y la suma de f(x) igual a 1.

Variables Aleatorias - Discretas y f(x) de probabilidad

1. **Variable aleatoria X:** Imaginen que están analizando cuántos votantes apoyan a un candidato específico en una encuesta. La variable X podría representar **el número de personas que apoyan al candidato**, y x sería un **número específico** (por ejemplo, que X=5 significa que 5 personas apoyan al candidato).
2. **Evento asociado Ax:** El evento Ax representa el conjunto de condiciones o situaciones que hacen que la variable X tome el valor específico x. En este caso, Ax podría ser el conjunto de personas encuestadas que respondieron "sí" al apoyo del candidato, y el número de apoyos fue igual a 5.
3. **Probabilidad:** Entonces, la función te dice que la probabilidad de observar exactamente 5 apoyos (es decir, X=5) es igual a la probabilidad de que se cumpla el evento A5, es decir, que cinco personas respondan "sí".
4. La función simplemente indica que la probabilidad de que la variable aleatoria X tome un valor específico x es igual a la probabilidad del evento Ax

Variables Aleatorias - Discretas y su f(x) de probabilidad

Ejemplo

Supongamos que encuestamos a 10 personas y queremos saber cuál es la probabilidad de que exactamente 5 de ellas apoyen a un candidato. La probabilidad P(X=5) sería la probabilidad de que ocurran las condiciones A_5 (que consisten en obtener 5 respuestas "sí" y 5 respuestas "no" en la encuesta). Esta probabilidad se puede calcular usando distribuciones de probabilidad, como la **distribución binomial**, pero más adelante entramos en las funciones de distribución propiamente dichas.

Variables Aleatorias - Discretas y su f(x) de probabilidad

A la función de distribución de probabilidad, también se le vincula **la función de distribución acumulada,** la cual nos dice la probabilidad de que una variable aleatoria X tome un valor **menor o igual** a x. En otras palabras, es la suma de todas **las probabilidades de que X sea hasta x.** La función se representa de la siguiente manera:

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t)$$

Esta es la suma de las probabilidades individuales para todos los valores posibles de X hasta x, lo que da lugar a la probabilidad acumulada. La función f(t) es la **función de densidad** o la **función de probabilidad** (dependiendo si X es continua o discreta), que nos da la probabilidad de que X tome el valor exacto t.

Supongamos entonces el siguiente ejemplo:

<u>Tiramos una moneda 3 veces (experimento aleatorio)</u>

- Definimos una variable aleatoria como X = número de caras,
- X ∈ {0, 1, 2, 3} estos son los valores que puede asumir, es decir, x. Donde 0 es que nunca salga cara, 1 que sólo una vez salga, 2 que salga dos veces y 3 que las tres veces sea cara.

La **distribución de frecuencia acumulada** nos dice la probabilidad de obtener hasta un número determinado de caras (es decir, la probabilidad acumulada de obtener 0, 1, 2 o 3 caras).

Variables Aleatorias - Discretas y su f(x) de probabilidad

Siguiendo la función, entonces, tenemos el siguiente procedimiento:

$$F(x) = P(X \leq x)$$

X = número de caras, $X \in \{0, 1, 2, 3\}$

- $F(0) = Pr(X \leq 0) = Pr(X = 0) = \frac{1}{8}$

La probabilidad de obtener 0 caras (es decir, ninguna cara) da $\frac{1}{8}$ ya que hay 2 posibles resultados por cada lanzamiento (cara o cruz) que se repiten 3 veces ($2^3 = 8$) y sólo es posible un resultado {cruz,cruz,cruz}

- $F(1) = Pr(X \leq 1) = Pr(X = 0) + Pr(X = 1) = 4/8 \rightarrow F(1)$: La probabilidad de obtener 1 o menos caras

Hay 3 maneras de obtener 1 cara: {cara, cruz, cruz}, {cruz, cara, cruz} y {cruz, cruz, cara} cada una con probabilidad $\frac{1}{8}$, por lo que se suman la probabilidad de F(0) y F(1) cuyo total es 4/8

- $F(2) = Pr(X \leq 2) = Pr(X = 0) + Pr(X = 1) + Pr(X = 2) = \frac{7}{8} \rightarrow F(2)F(2)F(2)$: La probabilidad de obtener 2 o menos caras

Para las al menos dos caras, hay 3 combinaciones posibles {cara, cara, cruz}, {cara, cruz, cara} y {cruz, cara, cara}. Por lo que la suma acumulada resulta $4/8 + \frac{3}{8} = \frac{7}{8}$

- $F(3) = Pr(X \leq 3) = 1 \rightarrow F(3)$: La probabilidad de obtener 3 o menos caras (es decir, cualquier resultado)

Se trata de la probabilidad de obtener **cualquier resultado posible** en los tres lanzamientos, lo que es **100%** o 1, porque todas las combinaciones posibles ya han sido cubiertas.

Variables Aleatorias - Discretas y la esperanza y varianza

La media de una variable aleatoria discreta X se obtiene de la suma de los valores multiplicados por su probabilidad. En probabilidad se llama media o esperanza o valor esperado de una variable aleatoria X denotada por μ or E(X) y se calcula de la siguiente manera:

$$\mu = \sum_{x \in S_x} x\, f(x)$$

La esperanza o media es el valor promedio que esperas de un fenómeno aleatorio si repites el experimento muchas veces. Nos da una idea de la escala o tamaño promedio del fenómeno que estamos observando y tiene la misma unidad que la variable aleatoria X. Ayuda a entender el comportamiento "normal" o "esperado" del fenómeno que estás analizando

Variables Aleatorias - Discretas y la esperanza y varianza

La Esperanza tiene una propiedad importante con respecto a su transformación lineal:

Dada una V.A. con media μ_X, **la V.A. Y,** que se obtiene a través de una transformación lineal **Y = a +bX** , con a y b como constantes,

tiene media $\mu_Y = a + b\mu_X$

Esto significa que cuando aplicamos una transformación lineal a una variable aleatoria, la media de la nueva variable Y se calcula **ajustando la media de X** de acuerdo con los valores de a y b, siendo **a** una constante que **desplaza** los valores de X, mientras que **b** es una constante que **escala** los valores de X.

Por ejemplo:

Piensen que estamos estudiando el **nivel de apoyo a un candidato** en distintas encuestas. El número de personas que apoyan al candidato en cada encuesta es una variable aleatoria X con una **media μ_X= 60** (es decir, en promedio, 60 personas apoyan al candidato en cada encuesta).

Variables Aleatorias - Discretas y la esperanza y varianza

Ahora, supongamos que queremos crear una nueva variable Y que refleje un **ajuste en los resultados** por algún factor externo, como una campaña publicitaria que se espera que aumente el apoyo en **10 personas** en cada encuesta que además **duplica** el nivel de apoyo existente. En este caso:

- $a=10$ (el aumento constante de apoyo debido a la campaña),
- $b=2$ (el apoyo se duplica debido a la campaña).

La transformación sería:

$Y=10+2X$

Para calcular la **media de Y**, usamos la propiedad:

$$\mu_y = a + b\mu_X$$

$\mu_Y=10+2\mu_X = 10+2(60) = 10+120 =130$

Entonces, la media de la nueva variable Y, que incluye el aumento y la duplicación del apoyo, sería 130. Esto significa que, en promedio, después de la campaña, esperarías ver **130 personas apoyando al candidato** en cada encuesta, en lugar de las 60 iniciales.

Variables Aleatorias - Discretas y la esperanza y varianza

Igual que en las distribuciones de frecuencia, la varianza de una V.A. es muy importante en tanto que es la suma de las desviaciones de la media al cuadrado multiplicado por sus respectivas probabilidades.

Por lo tanto, siendo X una V.A. discreta con media μ, la varianza de X, denotada por σ^2 o Var(X) se calcula de la siguiente manera:

$$\sigma^2 = \sum_{x \in S_X} (x - \mu)^2 f(x)$$

La varianza de la V.A. Y = a + bX está dada por

$$\sigma_Y^2 = b^2 \sigma_X^2$$

Variables Aleatorias - Continua y su f(x) de densidad

Una V.A. es continua cuando no podemos calcular la probabilidad de que la variable tome un valor exacto (porque hay infinitos valores posibles). Sin embargo, podemos calcular la **probabilidad de que la variable esté dentro de un rango** de valores. Esto es lo que hace la **función de densidad de probabilidad**.

X es una V.A. continua si existe una función **f(x)** tal que la función de distribución acumulada F(x) = P(X≤x) está dada por el área debajo de f(x) hacia la izquierda de x. La **función de densidad de probabilidad** f(x) nos da una idea de **qué tan probable es que la variable esté alrededor de un valor x**, pero no exactamente en x. Para encontrar la probabilidad de que X esté en un cierto rango, calculamos el área bajo la curva de f(x) en ese rango.

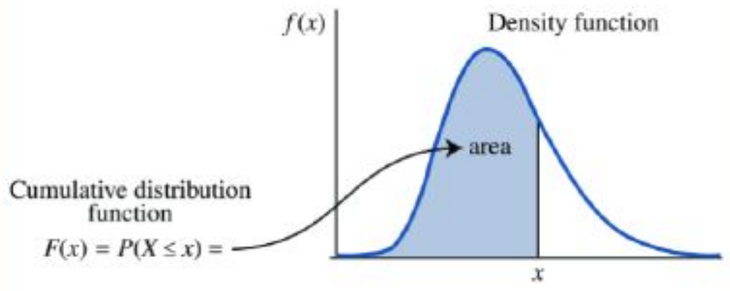


Figura 13.6 Relationship between cumulative distribution function and density function for a continuous random variable.

Variables Aleatorias - Continua y su f(x) de densidad

La construcción de la función de densidad de una VA X resulta de tomar valores entre l_1 y l_2 y dividir el intervalo en N subintervalos con tamaño **$\delta = (l_2 - l_1) / N$**

dados por $(x_1-\delta/2, x_1-\delta/2)$, $(x_2-\delta/2, x_2+\delta/2)$,...,$(x_N-\delta/2, x_N+\delta/2)$, donde x_1, x_2, …, x_N son los puntos medios de los subintervalos. De ello, diremos que p_1, p_2, …p_N son las probabilidades de que X tome valores entre los subintervalos.

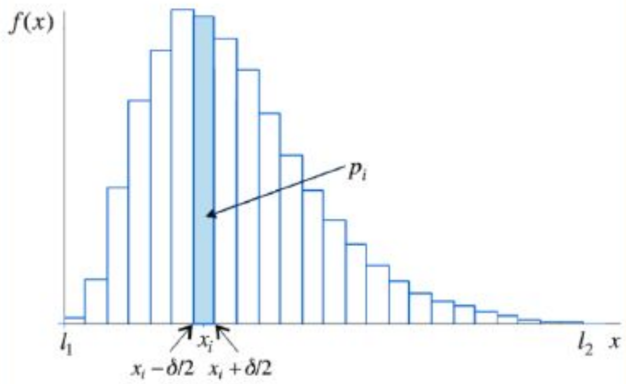


Figura 13.7 Probability histogram of a continuous random variable.

Variables Aleatorias - Continua y su f(x) de densidad

Un ejemplo:

Supongamos que estamos estudiando el porcentaje de votantes que apoyan a un candidato en una elección. El porcentaje puede ser cualquier valor entre 0% y 100%, por lo cual, tenemos una variable aleatoria continua. Ahora digamos que nos interesa ver el rango de porcentaje de apoyo que va desde 40% a 60%.

Nuestro intervalo de interés: $(I_1,I_2) = (40\%,60\%)$

Definimos la longitud de nuestros intervalos en igual tamaño en 5 subintervalos (N=5)

$\delta = (I_2-I_1) / N = (60\%-40\%)/5 = 4\%$, por lo tanto cada intervalo tendrá una longitud de 4%

Lo siguiente es establecer los puntos medios tomando el valor inicial de cada subintervalo sumado por la mitad del subintervalo:

$x_1 = 40\% + 2/\delta = 40\% + 2\% = 42\% \rightarrow (x_1-\delta/2, x_1+\delta/2) = (42\% - 2\% , 42\% + 2\%) = (40\%,44\%) = p_1$

$x_2 = 44\% + 2/\delta = 44\% + 2\% = 46\% \rightarrow (x_2-\delta/2, x_2+\delta/2) = (46\% - 2\% , 46\% + 2\%) = (44\%,48\%) = p_2$

$x_3 = 48\% + 2/\delta = 48\% + 2\% = 50\% \rightarrow (x_3-\delta/2, x_3+\delta/2) = (50\% - 2\% , 50\% + 2\%) = (48\%,52\%) = p_3$

$x_4 = 52\% + 2/\delta = 52\% + 2\% = 54\% \rightarrow (x_4-\delta/2, x_4+\delta/2) = (54\% - 2\% , 54\% + 2\%) = (52\%,56\%) = p_4$

$x_5 = 56\% + 2/\delta = 56\% + 2\% = 58\% \rightarrow (x_5-\delta/2, x_5+\delta/2) = (58\% - 2\% , 58\% + 2\%) = (56\%,60\%) = p_5$

Variables Aleatorias - Continua y su f(x) de densidad

Volviendo al histograma presentado anteriormente, tenemos que la **altura** del rectángulo se calcula como la **densidad de probabilidad por unidad de longitud (**p_i/δ), mientras que **el área está representado por** p_i

Ahora bien, el área total de los rectángulos del histograma es igual a 1 porque es la probabilidad de que X asuma cualquier valor dentro del intervalo (l_1, l_2). Si dividimos el intervalo en **más subintervalos** (es decir, aumentamos N y reducimos δ, el ancho de cada subintervalo), a medida que los subintervalos se hacen más pequeños, los rectángulos del histograma se vuelven más estrechos y la representación del histograma comienza a parecerse a una **curva suave**. Esta curva es la **función de densidad de probabilidad**.

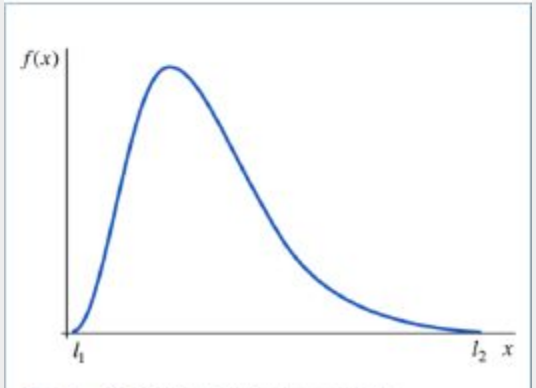


Figura 13.8 Density function of a continuous random variable.

Variables Aleatorias - Continua y su f(x) de densidad

La propiedades de la función de densidad son las siguientes:

- es **no negativa** en todas partes, lo que significa que siempre tiene valores iguales o mayores a 0 (porque las probabilidades no pueden ser negativas).
- el **área bajo la curva** de la función de densidad es **igual a 1**, lo que asegura que la probabilidad total de que la variable aleatoria tome un valor en el intervalo completo es 100%.

Variables Aleatorias - Continua y su f(x) de densidad

Por otro lado, la probabilidad de que una VA X tome valor entre *a* y *b* está representado por el área bajo la curva f(x) sobre el intervalo (*a,b*). Esto es posible de obtener como la diferencia entre los valores de la función de distribución acumulada en *b* y *a* donde

$$P(a < X < b) = P(X < b) - P(X < a) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

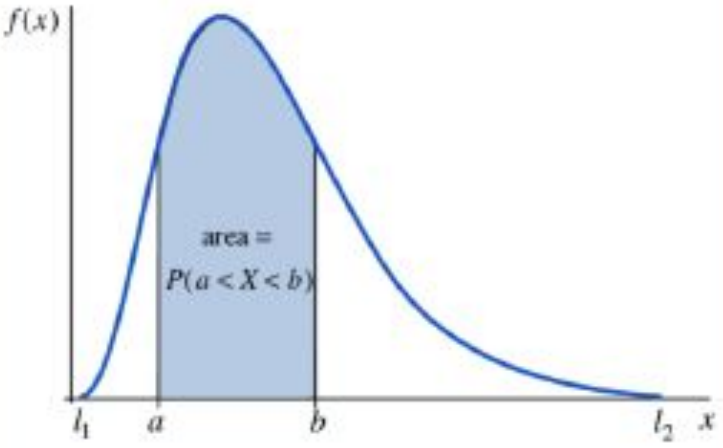


Figura 13.9 Probability of an interval for a continuous random variable.

Variables Aleatorias - Continua y la esperanza y varianza

La media y la varianza de una VA continua se expresan a través de integrales

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx, \quad Var(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

De manera similar a una VA discreta, una VA continua Y se obtiene a partir de una transformación lineal

Y = .$a + b_X$ con $\mu_Y = a + b\mu_X$ y $\sigma^2_Y = b^2\sigma^2_Y$

Variables Aleatorias - Continua y la estandarización

La **estandarización** es un tipo de **transformación lineal** que se aplica a una variable aleatoria X para convertirla en una nueva variable Y, de modo que tenga una **media (esperanza) de 0** y una **varianza de 1**.

Esto nos permite comparar variables que tienen diferentes escalas o unidades de medida de una manera más clara y objetiva. En el contexto de Ciencia Política, imaginemos que tenemos diferentes **encuestas** de apoyo a candidatos y queremos comparar cómo varían esos apoyos entre sí, aunque tengan **promedios y dispersiones diferentes**. Estandarizar los datos hace que las comparaciones sean más fáciles.

De aquí resulta la fórmula

$$Y = \frac{X - \mu}{\sigma}$$

Variables Aleatorias - Continua y la estandarización

De la función anterior desagregamos lo siguiente:

X es la variable original, μ es la **media** de X y σ es la **desviación estándar** de X.

Restar la media μ centra los datos alrededor de 0. Es decir, hace que la nueva variable Y tenga un promedio de 0. Si X es el porcentaje de apoyo a un candidato, al restar la media estamos midiendo **cuánto se desvía** cada encuesta del promedio de apoyo.

Dividir por la desviación estándar σ normaliza las diferencias respecto a la media en la nueva variable Y escalando los valores para que la dispersión sea más fácil de comparar entre distintas variables.

Por lo tanto $E(Y) = 0$, $Var(Y) = 1$

Variables Aleatorias - Estandarización vs transformación lineal

Una última aclaración para diferencias la estandarización y la transformación lineal

Estandarización es específica para resolver el problema de **comparar** variables o conjuntos de datos que tienen diferentes escalas, **normalizando** los valores.

Transformación lineal es más general y se utiliza cuando simplemente se quiere **reajustar** los valores de una variable a una escala distinta, sin necesariamente normalizarla.

Variables Aleatorias - Elementos

~~9. Definición de Distribuciones Probabilísticas Paramétricas~~
~~10. Distribución DIscreta Uniforme~~
~~11. Distribución de Bernoulli~~
~~12. Distribución Binomial~~

~~14. Distribución Continua Uniforme~~
~~16. Distribución Normal~~
~~17. Propiedades de la Distribución Normal~~
~~18. Regla Empírica de la Distribución Normal~~
19. Función de la Distribución Acumulativa de la Normal
20. Distribución Normal Estándar
21. Tablas (z y t)

Distribuciones Paramétricas de Probabilidad

Una **distribución paramétrica de probabilidad** es una herramienta que nos permite hacer predicciones sobre un conjunto de datos, basándonos en ciertos parámetros que describen cómo se comportan esos datos. Estos parámetros son **constantes numéricas** que afectan la forma en que se distribuyen los datos y que determinan las **probabilidades** de que ocurran ciertos valores.

En definitiva, las distribuciones paramétricas nos permiten modelar muchas situaciones diferentes con **una fórmula general** que cambia según los valores de sus parámetros.

Veamos algunas…

Distribución Uniforme Discreta - V. A. Discretas

La **distribución uniforme discreta** es útil cuando analizamos situaciones donde **todos los resultados tienen la misma probabilidad**. Podemos aplicarla en situaciones como la elección aleatoria de candidatos, sorteos, o cualquier proceso donde haya un número igual de oportunidades para cada opción.

Una VA X tiene una distribución uniforme discreta en sus primeros n números naturales si la función de probabilidad de X es

$$f(x) = \frac{1}{n}, x = 1, 2, \ldots, n$$

Con $E(X) = \frac{n+1}{2}, \ Var(X) = \frac{n^2-1}{12}.$

Esto significa que **cada número** del 1 al n tiene una probabilidad igual de 1/n, por lo tanto, el valor promedio estará aproximadamente en el medio y, a su vez, si el valor de n aumenta, la varianza también crece, lo que significa que los valores estarán más dispersos.

Distribución Uniforme Discreta - V. A. Discretas

Veamos un ejemplo…

Supongamos un bolillero con 8 bolitas numeradas del 1 al 8. El experimento consiste en sacar una bolilla al azar, lo que significa que cada bolita tiene la misma probabilidad de ser seleccionada. Esta situación es un ejemplo de una **distribución uniforme discreta**, ya que todas las bolitas tienen la misma probabilidad de ser extraídas.

$$f(x) = \frac{1}{n}, x = 1, 2, \ldots, n$$

$$E(X) = \frac{n+1}{2},\ Var(X) = \frac{n^2-1}{12}.$$

La variable X representa el número de bolilla que se extrae, por lo tanto X puede asumir valores 1, 2, 3, 4, 5, 6, 7, 8. Así definimos entonces que la función de probabilidad es la siguiente:

$f(x) = 1/n$, $X = 1, 2, .., n$ $f(x) = \frac{1}{8} = 0{,}125$, $X=1, 2, 3, 4, 5, 6, 7, 8$

$E(X) = n + 1/ 2 = 8+1/2 = 4{,}5$ y $Var(X) = n^2-1/12 = 8^2-1/12 = 5{,}25$

Según la media, es probable que el resultado de la bolita sea 4 o 5, pero la alta varianza, indica que el resultado pueda distar ampliamente, lo que indica equiprobabilidad (cada evento es igualmente probable)

Distribución de Bernoulli - V.A. Discretas

Consideremos un experimento aleatorio cuyos resultados son dos eventos mutuamente excluyentes A y A′, convencionalmente llamados **éxito** y **fracaso**, codificados como 1 y 0, respectivamente. Denotemos por p la probabilidad de que ocurra A (éxito) y por 1−p la probabilidad de que ocurra A′ (fracaso). La ejecución única de este experimento se llama un **ensayo de Bernoulli**, denominado así en honor al matemático suizo Jacob Bernoulli.

Una **variable aleatoria discreta** X tiene una **distribución de Bernoulli** si toma los valores 0 y 1 con probabilidades p y 1−p, respectivamente, donde $0<p<1$. La **función de probabilidad** de esta variable aleatoria es:

$$f(x) = p^x(1-p)^{1-x}, x = 0, 1.$$

- La **media** es $E(X) = p$
- La **varianza** es $Var(X) = p(1-p)$

Ensayo de Bernoulli y la Distribución Binomial

Veamos un ejemplo

Consideren un proceso industrial donde se producen bienes en el sector de electrónica. La experiencia sugiere que el 1% de los productos son defectuosos. La selección aleatoria de una pieza puede ser pensada como un ensayo de Bernoulli, donde el evento “éxito” es identificado con una pieza defectuosa. La probabilidad de éxito es de 0,01 (1%) con reposición. La reposición indica que la probabilidad no cambie (que se mantenga el n)

El ensayo de Bernoulli entonces tiene las siguientes características

Éxito: Obtener una pieza defectuosa.

Fracaso: Obtener una pieza no defectuosa.

Probabilidad de éxito: $p = 0.01$ (1%).

Probabilidad de fracaso: $1-p = 0.99$

Ensayo de Bernoulli y la Distribución Binomial

Ahora supongamos que de un lote, 20 piezas son seleccionadas con reposición. El **número de piezas defectuosas** en las 20 extracciones es una **variable aleatoria**, cuyos posibles valores son los números enteros del 0 al 20.

Dado que estamos haciendo 20 ensayos de Bernoulli, esta situación sigue una **distribución binomial**, que es utilizada para contar cuántos éxitos (piezas defectuosas) ocurren en un número fijo de ensayos independientes (20 extracciones).

Las características principales son:

- Un ensayo de Bernoulli se repite un cierto número de veces.
- La probabilidad de éxito no cambia de un ensayo a otro.
- Los ensayos son independientes.

La probabilidad de éxito se mantiene sin cambios y los ensayos son independientes debido al hecho de que las selecciones se realizan con reemplazo.

Distribución Binomial - V.A. Discretas

La distribución binomial es una distribución de probabilidad discreta que es, en definitiva, una generalización de la distribución de Bernoulli. Cuenta el número de éxitos en *n* intentos de Bernoulli.

Una variable aleatoria discreta X tiene distribución binomial si su función de probabilidad es

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$$

donde p ∈ (0, 1) ;

n es un número entero positivo y $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ es el coeficiente binomial.

Con E(X) = np , Var(X) = np(1-p)

La distribución binomial depende de los parámetros n y p que pueden cambiar la forma de su distribución de probabilidad de manera tal que

- symmetric if $p = 0.5$
- positively skewed if $p < 0.5$
- negatively skewed if $p > 0.5$

Distribución Binomial - V.A. Discreta

Repongamos el ejemplo de las piezas defectuosas. Dijimos que se seleccionaron 20 piezas al azar con reposición y establecimos que la Variable Aleatoria X es la probabilidad de sacar una pieza defectuosa. Según el ejemplo, la probabilidad de encontrar una pieza defectuosa es el 1%. Ya tenemos algunos parámetros establecidos.

p = 0,01

n = 20

por lo tanto, f (x) = $\binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n.$ con $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ como coeficiente,

si reemplazamos entonces tenemos la siguiente función de distribución

f(x) = (20/x)0,01^{x}(1-0,01), x= 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20

Con la función, lo que resta es establecer la probabilidad de encontrar x cantidad de piezas defectuosas.

Distribución Binomial - V.A. Discreta

Reemplacemos entonces:

Ejemplo 1: ¿Cuál es la probabilidad de que no haya piezas defectuosas? Acá estamos buscando P(X=0)

$$P(X=0) = \binom{20}{0}(0.01)^0(0.99)^{20} = (0.99)^{20} \longrightarrow P(X=0) \approx 0.817$$

Ejemplo 2: ¿Cuál es la probabilidad de que una pieza sea defectuosa? Acá estamos buscando P(X=1)

$$P(X=1) = \binom{20}{1}(0.01)^1(0.99)^{19} \longrightarrow P(X=1) = 20 \cdot (0.01) \cdot (0.99)^{19} \approx 0.165$$

Con E(X) = np = 20.0,01 = 0,2 y Var(X) = np(1-p) = 20.0,01(1-0,01) = 0,198

$$E(X) = np, \; Var(X) = np(1-p)$$

Distribución Binomial - V.A. Discreta

La distribución binomial depende de los parámetros n y p que pueden cambiar la forma de su distribución de probabilidad de manera tal que

- es **simétrica** si p = 0,5
- tiene **sesgo positivo** si p < 0,5
- tiene **sesgo negativo** si p > 0,5

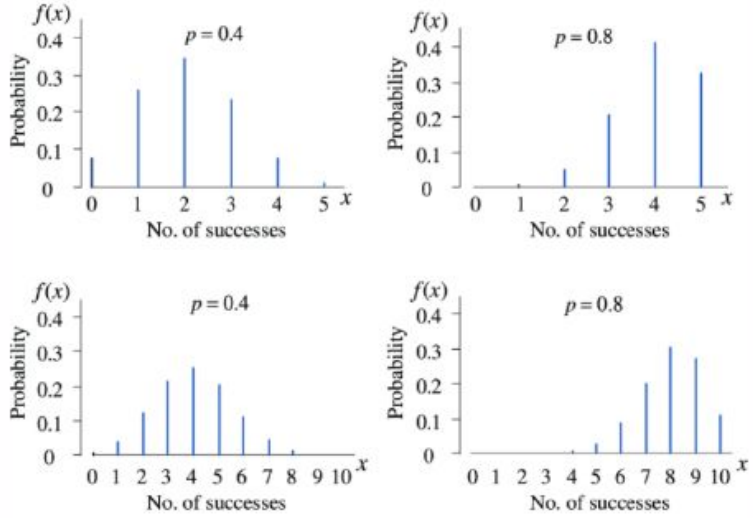


Figura 14.2 Binomial distribution for $p = 0.4, 0.8$ and $n = 5, 10$.

Distribución Uniforme Continua - V.A. Continua

Nuevamente, estamos frente a una distribución donde todos los valores en un intervalo tienen la misma probabilidad de ocurrir. Una VA X tiene una distribución uniforme continua cuando su función de densidad es la siguiente

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{otherwise} \end{cases}$$

Esto significa que la probabilidad de encontrar un valor en cualquier subintervalo del intervalo [a,b] es proporcional al tamaño de ese subintervalo.

A su vez, la media y la varianza se calculan como $E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(a-b)^2}{12}$

Distribución Uniforme Continua

Distribución Normal o de Gauss - V.A. Continua

La distribución normal o de Gauss es una distribución de probabilidad continua que es simétrica y tiene forma de campana. Es una de las distribuciones más importantes en estadística debido a su presencia en muchos fenómenos naturales y sociales. Una VA continua X tiene distribución normal si su función de densidad está dada por

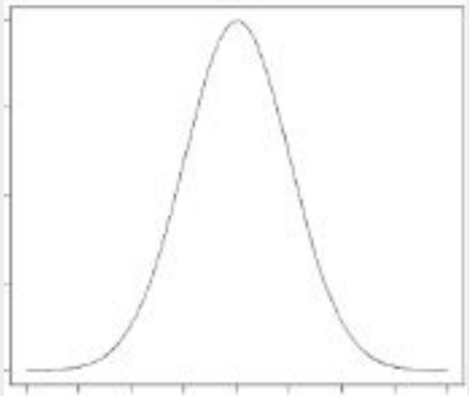
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$

Donde la media y la varianza están dadas por $E(X) = \mu$ and $Var(X) = \sigma^2$

Distribución Normal - V.A. Continua

La función de densidad f(x) cumple con las siguientes propiedades:

- Es **simétrica** alrededor de μ ya que $f(\mu-\delta) = f(\mu+\delta)$ para cualquier $\delta > 0$.
- Es **creciente** en el intervalo $(-\infty,\mu)$ y **decreciente** en el intervalo (μ,∞); alcanzando su **máximo**, $(\sigma\sqrt{2\pi})^{-1}$, en $x=\mu$.
- Tiene el **eje X** como asíntota horizontal, es decir, $f(x)\to 0$ cuando $x\to-\infty$ o $x\to\infty$
- Tiene dos **puntos de inflexión** en $x=\mu-\sigma$ y $x=\mu+\sigma$; es **cóncava** (hacia abajo) en el intervalo $(\mu-\sigma,\mu+\sigma)$ y **convexa** en otros lugares.

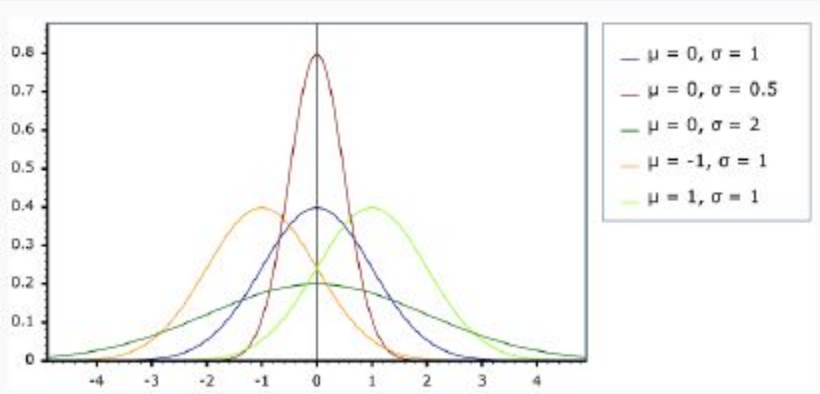


Distribución Normal - V.A. Continua

Para indicar que una VAC X tiene distribución normal con media μ y varianza σ^2 se expresa de la siguiente manera:

$$X \sim N(\mu, \sigma^2)$$

- La **media** determina la **ubicación** de la función de densidad en el eje x: es el punto donde la curva normal alcanza su pico.
- La **varianza** determina la **altura** y el **ancho** de la curva normal. Una varianza grande produce una curva plana y ancha, mientras que una varianza pequeña produce una curva pronunciada y estrecha.



Distribución Normal - La Regla Empírica

En una distribución normal, aproximadamente el 68% de los valores caen dentro de una desviación estándar de la media, alrededor del 95% dentro de dos desviaciones estándar, y el 99.7% dentro de tres desviaciones estándar.

Esta es la llamada Regla Empírica según la cual las áreas bajo la curva normal en los intervalos $(\mu - \sigma, \mu + \sigma)$,$(\mu - 2\sigma, \mu + 2\sigma)$ y $(\mu - 3\sigma, \mu + 3\sigma)$ son aproximadamente iguales a 0.68, 0.95 y 0.997, respectivamente.

Esto significa que la probabilidad de observar un valor de X dentro de:

- (a) una desviación estándar por encima o por debajo de la media es 0.68
- (b) dos desviaciones estándar por encima o por debajo de la media es 0.95
- (c) tres desviaciones estándar por encima o por debajo de la media es 0.997

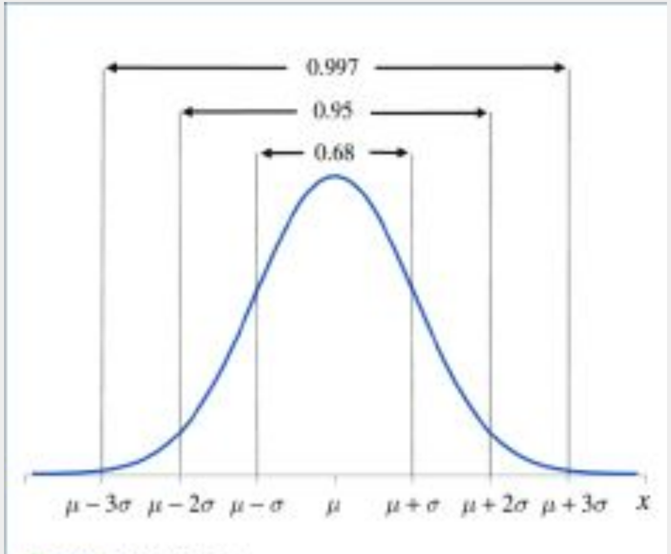
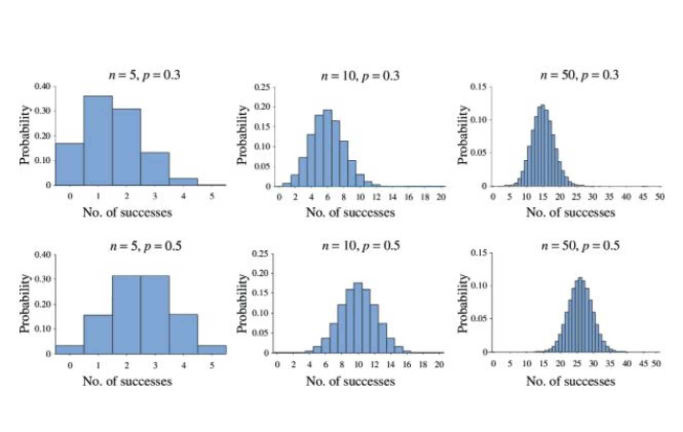


Figura 14.11 Empirical rule.

Distribución Normal - Teorema del Límite Central

Este teorema es uno de los pilares de la estadística porque permite utilizar la distribución normal para hacer inferencias sobre poblaciones, incluso si la población no sigue una distribución normal.

El Teorema del Límite Central establece que, dado un número suficientemente grande de muestras independientes extraídas de una población, la distribución de la suma (o la media) de estas muestras tiende a ser normal (o gaussiana), independientemente de la forma de la distribución original de la población.



Distribución Normal - Teorema del Límite Central

Acá lo que podemos observar, es la

aproximación normal de una serie

de distribuciones binomiales

donde se va cambiando el n en cada una

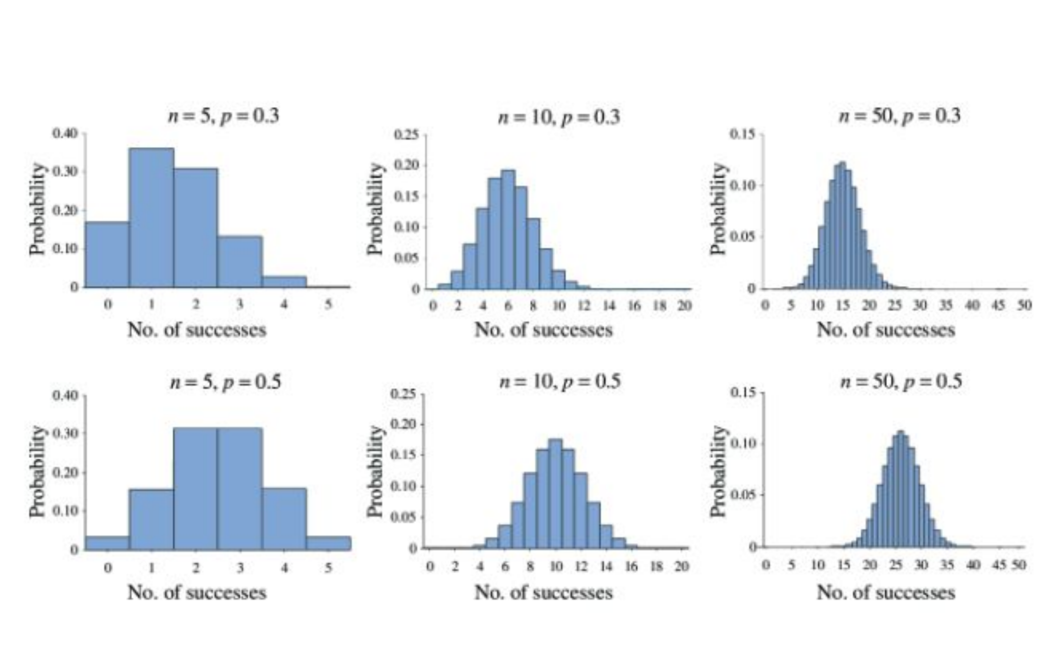
conjunto con su probabilidad p.

Como se puede ver, cuando p es 0,5 y

aumenta el n, la curva tiene una

distribución asimétrica alrededor

del la media.



Distribución Normal o Gaussiana - Estandarización

La distribución normal estandarizada es la distribución normal con media 0 y varianza 1 denotada por N(0,1). Su función de densidad es

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

Z deriva de una transformación estandarizada de X llamada Z = (X − μ)/σ. Esto significa que la función de distribución acumulada de X puede ser obtenida de

$$F(x) = P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

Donde Φ(·) indica la función de distribución acumulada de la distribución normal estandarizada.

Los valores de Φ(z), para muchos valores de z están tabulados en las tablas de estandarización normal

$$\Phi(0) = 0.5 \quad \Phi(0.1) = 0.5398$$

$$\Phi(1) = 0.8413 \quad \Phi(3) = 0.9986$$

Distribución Normal o Gaussiana

La Ley Empírica

Las áreas bajo la curva normal en los intervalos $(\mu - \sigma, \mu + \sigma)$,$(\mu - 2\sigma, \mu + 2\sigma)$ y $(\mu - 3\sigma, \mu + 3\sigma)$ son aproximadamente iguales a 0.68, 0.95 y 0.997, respectivamente. Esto significa que la probabilidad de observar un valor de X dentro de:

- (a) una desviación estándar por encima o por debajo de la media es 0.68
- (b) dos desviaciones estándar por encima o por debajo de la media es 0.95
- (c) tres desviaciones estándar por encima o por debajo de la media es 0.997

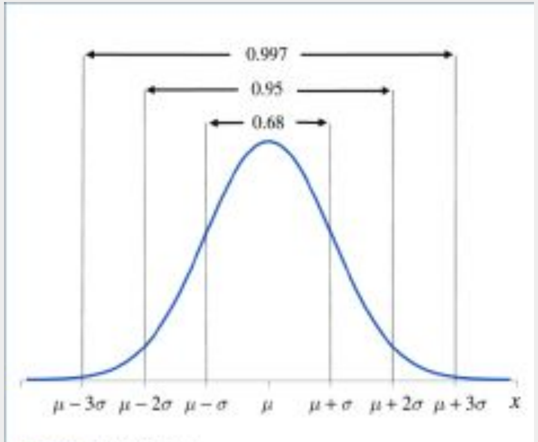


Figura 14.11 Empirical rule.