Probabilidad y Variables Aleatorias

(abreviaremos VVAA)

Prof. Magdiel ABLAN BORTONE

Prof. Paolo RAMONI PERAZZI

Facultad de Ingeniería

Universidad de Los Andes

Aleatoriedad

Quizás, a lo mejor, puede que, porsiacaso, posiblemente, tal vez, acaso…

Aleatoriedad: concepto un poco difícil de definir…

… pero al cual dejamos con frecuencia las decisiones complejas…

Aleatoriedad

… a todo nivel

TRIPLE GORDO millonarios cada semana

Seguros Caracas de Liberty Mutual

Pirámide SEGUROS

Mercantil Seguros

Seguros Constitución

Probabilidad...

Postgrado en Modelado y Simulación

- La teoría de probabilidades nos ayuda a modelar los fenómenos que percibimos como aleatorios e impredecibles viéndolos como resultado de un experimento aleatorio
- Al conjunto de todos los resultados posibles de un experimento se le conoce como espacio muestral (denotado como S, Ω, entre otros)

- $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$
- $\Omega = \{$  $\}$

$\Omega = \{$  $\}$

Evento...

- Un *evento* o *suceso* consiste en uno o más de los resultados posibles de un experimento, es decir, un subconjunto del Ω
  - **Elemental:** por ejemplo, lanzo el dado y obtengo {5}
  - **Compuesto:** por ejemplo, obtengo pares {2, 4, 6}

Teoría de conjuntos

- Los posibles subconjuntos que integran Ω pueden ser relacionados mediante uniones, intersecciones o complementos
- Similar a sumar y restar, pero como se trata de conjuntos, se usa una notación diferente

- **Uniones**

Por ejemplo, A es el evento que se obtiene un número impar y B es el evento que se obtiene un número mayor que 2

- $A = \{1, 3, 5\}$ y $B = \{3, 4, 5, 6\}$
- $A \cup B = \{1, 3, 4, 5, 6\}$
- Es decir, reunimos los elementos de ambos conjuntos en uno nuevo
- Note que 3 y 5 no se repiten
- ¿Qué resulta de $A \cup C$, si $C = \{2, 4, 6\}$?

Teoría de conjuntos

- **Intersecciones**
- Recuperan únicamente aquellos elementos comunes a los conjuntos involucrados

$A = \{1, 3, 5\}$ y $B = \{3, 4, 5, 6\}$

- $A \cap B = \{3, 5\}$
- Recordemos que $C = \{2, 4, 6\}$
- $A \cap C = \varnothing$

- **Complementos**
- El complemento de A es el conjunto de elementos presentes en el Ω, pero que están ausentes en A

Si $\Omega = \{1, 2, 3, 4, 5, 6\}$ y $A = \{1, 3, 5\}$

- $A^c = \{2, 4, 6\}$

Probabilidades de eventos A y/o B

- Los tipos de probabilidades a encontrar varían según la pregunta planteada pero, en general, pertenecen a alguna de las siguientes cinco:
    - Probabilidad marginal
    - Probabilidad de unión
    - Probabilidad de intersección
    - Probabilidad condicionada, y
    - Probabilidad complementaria

Ejemplo, suponga que lanza un dado…

Probabilidades de eventos A y/o B

…puede querer saber la probabilidad de obtener un número…

… par (*p* marginal)

… par **o** menor a 4 (*p* unión)

… par **y** menor que 4 (*p* intersección)

… 5 si sabe que es impar (*p* condicionada)

Probabilidad marginal

- Si únicamente estamos interesados en la probabilidad de una única característica
- Si estamos buscando únicamente la probabilidad de un conjunto A, estamos encontrando la *probabilidad marginal* de A

Ejemplo, la probabilidad de obtener par

- El evento aquí es $A = \{2, 4, 6\}$
- Así, la probabilidad par es $P(A) = 3/6 = 1/2$

Probabilidad unión

- La probabilidad de la unión de dos eventos se denota $P(A \cup B)$
- Asociado al conector "**o**", es decir A o B
- Naturalmente, también en ambos
- Ejemplo $A = \{2, 4, 6\}$ y $B = \{1, 2, 3\}$
- $A \cup B = \{1, 2, 3, 4, 6\}$
- $P(A \cup B) = 5/6$

Probabilidad intersección

- La probabilidad intersección de dos eventos se denota $P(A \cap B)$

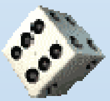
- Asociado al conector "**y**", es decir A y B

Ejemplo $A = \{2, 4, 6\}$ y $B = \{1, 2, 3\}$

- $A \cap B = \{2\}$
- $P(A \cap B) = 1/6$

Probabilidad complementaria

- A^c es cada ítem en Ω ausente en A

Ejemplo, si $A = \{2,4\}$ entonces $A^c = \{1, 3, 5, 6\}$

- Así, $P(A^c) = 4/6 = 2/3$

Probabilidad condicional

- En ocasiones tener un conocimiento previo del resultado cambia la probabilidad del mismo
- Como cuando se subdividen las salidas en subconjuntos (i.e. pares, impares)
- Las probabilidades condicionales tratan con el cambio que provoca el uso de información previa
- La probabilidad que ocurra un evento, dado que otro ya ocurrió

Probabilidad condicional

- Estas probabilidades pueden resolverse sin formula
- La notación de la probabilidad del evento A dado B, el cual ya ha ocurrido, es $P(A \mid B)$
- $P(A \mid B) \neq P(B \mid A)$
- Ejemplo, $P(5 \mid \text{impar})$
- $P(5 \mid \text{impar}) = 1/3 = 0{,}33$
- ¿$P(\text{impar}|5)$?

Probabilidad condicional

- Estas probabilidades también pueden resolverse con formula
- $P(A \mid B) = \frac{P(A \cap B)}{P(B)}$
- Note que no puede hallar P(A | B) si P(B) = 0
- En el caso del dado: B = {1, 3, 5} y A = {5}
- P(A ∩ B) = 1/6, porque A ∩ B = {5}
- P(B) = 3/6
- $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/6}{3/6} = \frac{1}{6} * \frac{6}{3} = \frac{1}{3} = 0{,}33$

Aplicando reglas de probabilidad

Postgrado en Modelado y Simulación

- Se puede encontrar las probabilidades de los eventos, o sus combinaciones, sumando restando, multiplicando y dividiendo las probabilidades de los eventos originales

- Naturalmente, hay que seguir ciertas reglas y fórmulas

- Mientras mejor las entienda, mejor

Aplicando reglas de probabilidad

- Cualquier probabilidad debe seguir tres propiedades básicas:
  - ✓ Debe estar en $0 \leq p \leq 1$
  - ✓ Encontrar la probabilidad de un conjunto de resultados individuales de Ω, implica sumar sus probabilidades (esto no necesariamente es así cuando los resultados se combinan)
  - ✓ La suma de todas las salidas en Ω, deben sumar uno
- Esto último es la base de la *regla del complemento*

$$P(A) + P(A^c) = 1$$

$$P(A) = 1 - P(A^c)$$

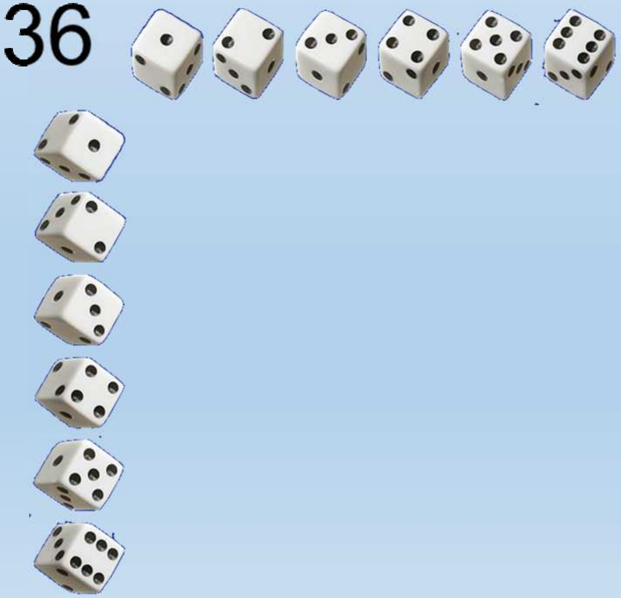
Aplicando reglas de probabilidad

- A veces los eventos de interés son tan complicados y difíciles que es más sencillo enfocarse en el complemento, es decir, en las salidas que no nos interesan

- Ilustremos con el siguiente ejemplo, el lanzamiento de dos dados puede dar 36 resultados diferentes

Aplicando reglas de probabilidad

- Suponga A → al menos uno de los dados > 1
- Para encontrar esta P(A) se debe sumar MUCHAS probabilidades individuales
- De hecho, todas excepto (1, 1). Suma = $\frac{35}{36}$
- A^c = (1,1) → $P(A^c)$ = 1/36, por lo tanto
- $P(A) = 1 - P(A^c) = 1 – 1/36 = 35/36$



Aplicando reglas de probabilidad

- La *regla de multiplicación* se usa para encontrar la probabilidad de una intersección de dos eventos A y B. Observe que:

- $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} \rightarrow P(A \cap B) = P(B)^{*}\, P(A \mid B)$

- La regla de la multiplicación divide la probabilidad conjunta en dos pasos: primero ocurre B y luego A dado B

Aplicando reglas de probabilidad

Postgrado en Modelado y Simulación

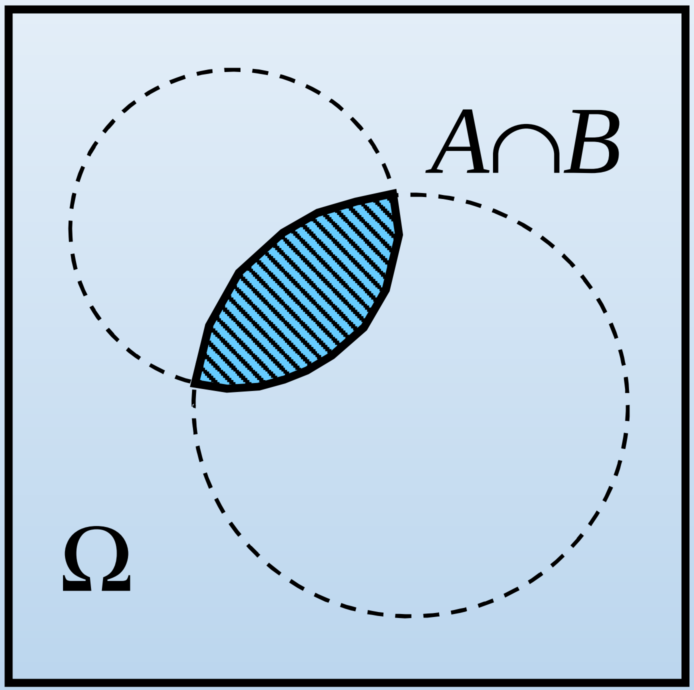
- Ejemplo, 60 % de un salón son mujeres (M), 40% de las cuales están casadas (C) ¿Cuál es la probabilidad que una muestra aleatoria sea mujer y casada? $\rightarrow$ ¿$P(M \cap C)$?
- $P(M) = 0{,}6$ y $P(C \mid M) = 0{,}4$
- $P(M \cap C) = P(M) * P(C \mid M) = 0{,}6 * 0{,}4 = 0{,}24$
- 24% del salón son mujeres casadas

Aplicando reglas de probabilidad

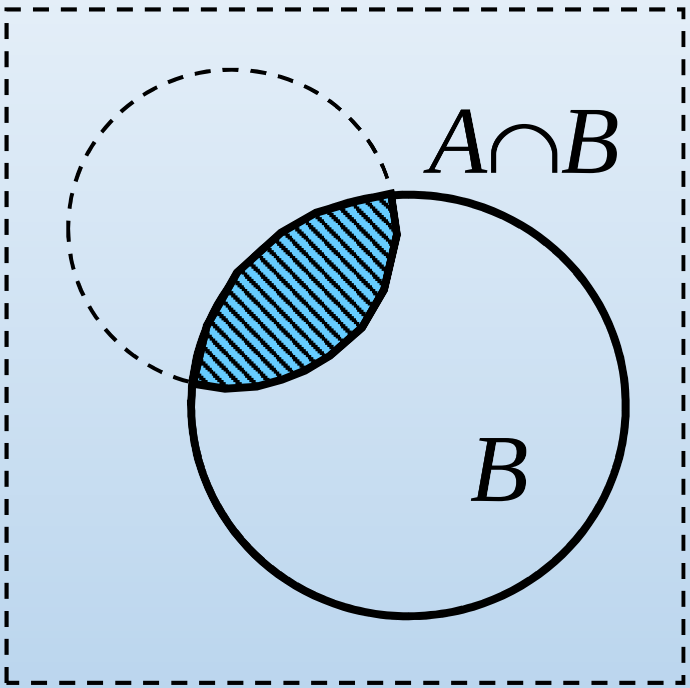
- Diferencia entre *probabilidad conjunta* y *condicional*: la primera selecciona una muestra que tiene dos características, la segunda busca la probabilidad de una muestra que ya cumple con una característica. Por ejemplo:
  - A llueve un día dado y B ese día cae en mayo
  - La probabilidad conjunta, $P(A \cap B)$, es el cociente del número de días que llueve en mayo entre el total de días del año.
  - La probabilidad condicionada, $P(A|B)$, es el cociente del número de días lluviosos del mes de mayo entre el número de días que tiene el propio mes de mayo.

Aplicando reglas de probabilidad

- Graficamente:



Probabilidad conjunta



Probabilidad condicionada

Aplicando reglas de probabilidad

- Para encontrar la probabilidad de la unión de A y B se aplica la *regla de la adición*

- No es sólo sumar, pues se duplicarían los elementos $A \cap B$, incrementando artificialmente la probabilidad

- Por eso se resta $A \cap B$ una vez

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

Identificando independencia

- Independencia: una de las asunciones más importantes de los modelos básicos de probabilidad
- Significa que la ocurrencia de un evento no afecta la ocurrencia de los otros
- Puede evaluarse de dos maneras:
  - ✓ Use la definición de independencia; revise si $P(A|B) = P(A)$ o si $P(B|A) = P(B)$
  - ✓ Revise la regla de la multiplicación para independencia; vea si $P(A \cap B) = P(A) * P(B)$, si la igualdad se cumple A y B son independientes

Identificando independencia

Ejemplo, ¿son A = {impar} y B = {1} eventos independientes?

- Contesto primero: Si sé que es impar ¿Cuál es la probabilidad que sea 1? R.- 1/3
- Ahora contesto: ¿Cuál es la probabilidad que sea 1 sin saber si es impar? R.- 1/6
- $P(A|B) \neq P(A)$, así que los eventos A y B no son independientes: el conocimiento de un evento afecta la probabilidad del segundo

Identificando independencia

Ahora suponga que C = {1 o 2} ¿Son A y C independientes?

- Revisamos ¿$P(C \mid A) = P(C)$?, $P(C) = 2/6 = 1/3$

- $C \cap A = \{1\} \rightarrow P(C \cap A) = 1/6$
- $P(C \mid A) = \frac{1/6}{3/6} = \frac{1}{6} * \frac{6}{3} = \frac{1}{3}$

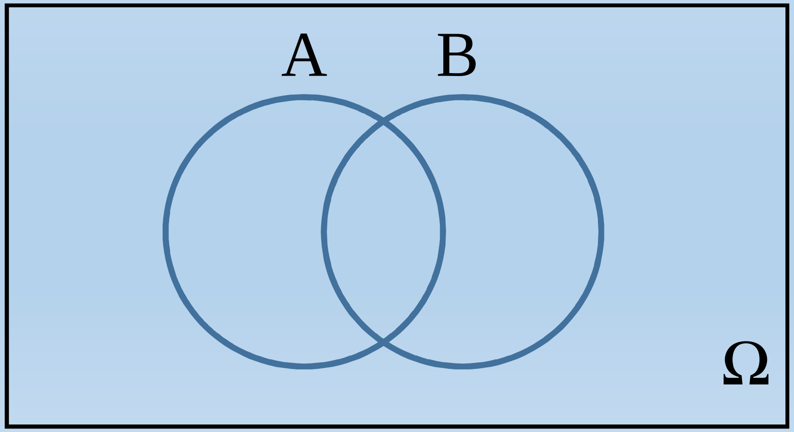
- Ambos eventos son independientes

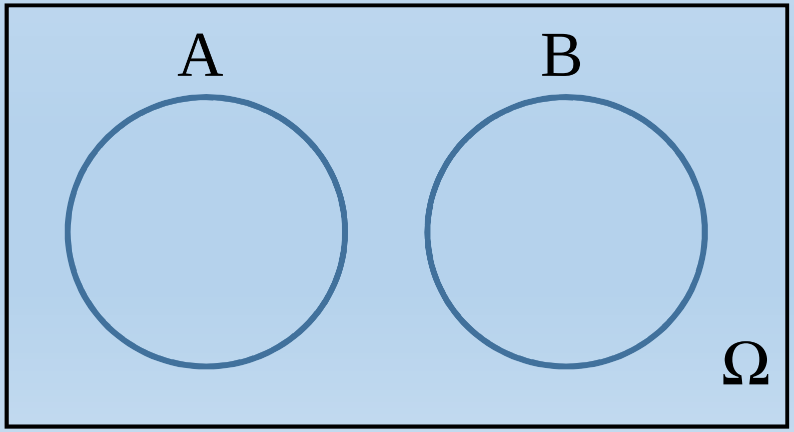
Representando probabilidades

- Hemos visto ejemplos sencillos, pero usualmente las situaciones son complicadas
- Por ejemplo, dado P(A | B), P(B) y P(B^c), calcule P(A)
- Por ejemplo, dado P(B | A) y sus complementos, encuentre P(A | B)
- Sí, Usted puede calcular esto, sólo necesita una buena representación gráfica y una buena fórmula
- Veamos cómo graficar

Diagramas de Venn

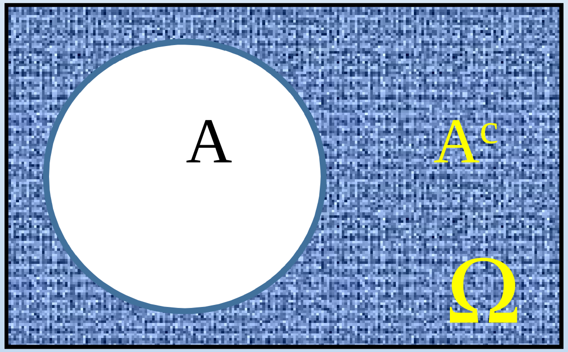
- Un modo de organizar información representando el Ω como una caja, y círculos para los distintos eventos en el problema
- Si los eventos intersectan, los círculos se superponen
- Si los eventos son mutuamente excluyentes, los círculos están separados



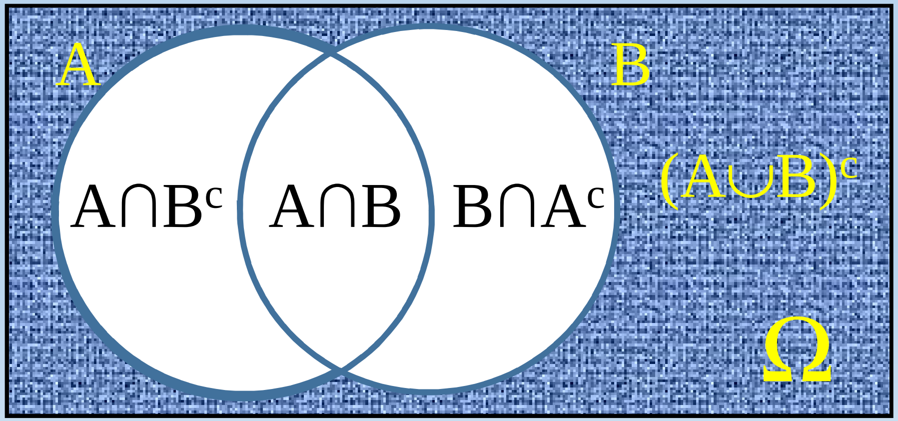


Diagramas de Venn

- Permite organizar y contar todos los conjuntos posibles en el escenario de probabilidad
- Cada parte del diagrama tiene un significado y una probabilidad



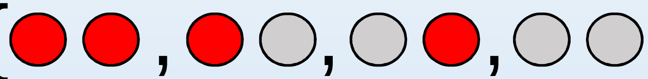
- Permite visualizar relaciones importantes entre eventos

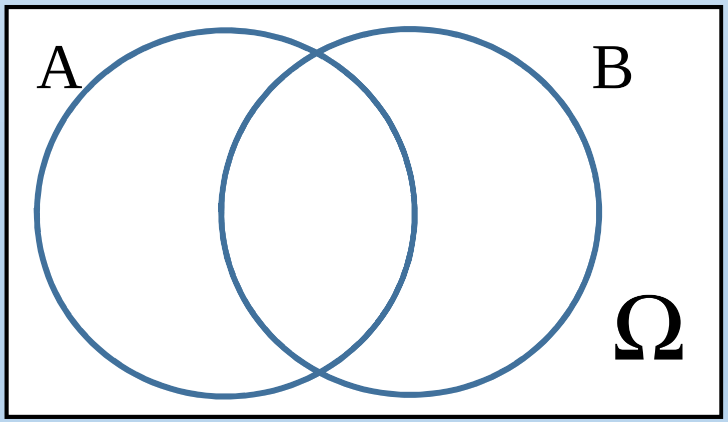


Diagramas de Venn

- Ejemplo, suponga una calle con dos semáforos. La probabilidad que el 1ero esté en rojo es 0,4 y que el 2do esté en rojo es 0,3. La probabilidad que ambos estén en rojo es de 0,1
- ¿Cuál es la probabilidad que ninguna luz esté en rojo?
- ¿Cuál es la probabilidad que exactamente una de las luces esté en rojo?

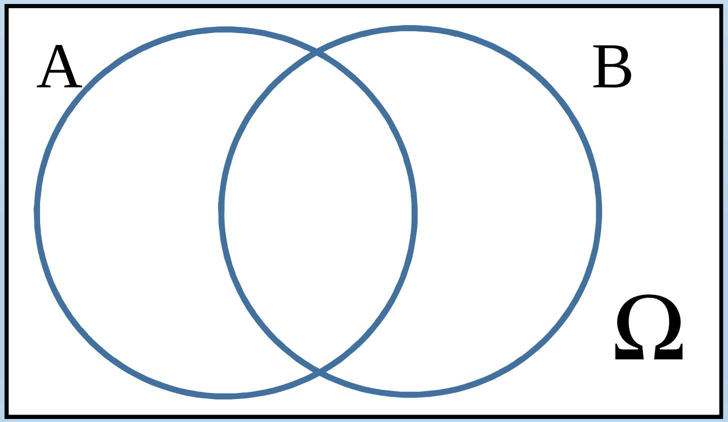
Diagramas de Venn

- En este caso (excluimos amarillo y verde):
- Ω ={}
- A = {primer semáforo en rojo} = 0,4
- B = {segundo semáforo en rojo} = 0,3



Diagramas de Venn

- $P(A) = P(A \cap B) + P(A \cap B^c) = 0{,}4$
- Si $P(A \cap B) = 0{,}1 \rightarrow P(A \cap B^c) = ?$
- $P(B) = P(B \cap A) + P(B \cap A^c) = 0{,}3$
- Si $P(A \cap B) = 0{,}1 \rightarrow P(B \cap A^c) = ?$
- ¿Terminamos?



Diagramas de árboles

- Algunos problemas de probabilidades involucran una secuencia de eventos
- En esos casos, necesitamos métodos que muestren cada etapa del mismo Ω
- El diagrama de árboles usa conjuntos de ramas para mostrar cada etapa y cada resultado posible dentro de cada etapa es representado por más ramas
- Al final se pueden encontrar las rutas que siguen las ramas a todos los elementos en Ω

Diagramas de árboles

1er 2do 3er

Ω

$C \cap C \cap C = CCC$

$C \cap C \cap S = CCS$

$C \cap S \cap C = CSC$

$C \cap S \cap S = CSS$

$S \cap C \cap C = SCC$

$S \cap C \cap S = SCS$

$S \cap S \cap C = SSC$

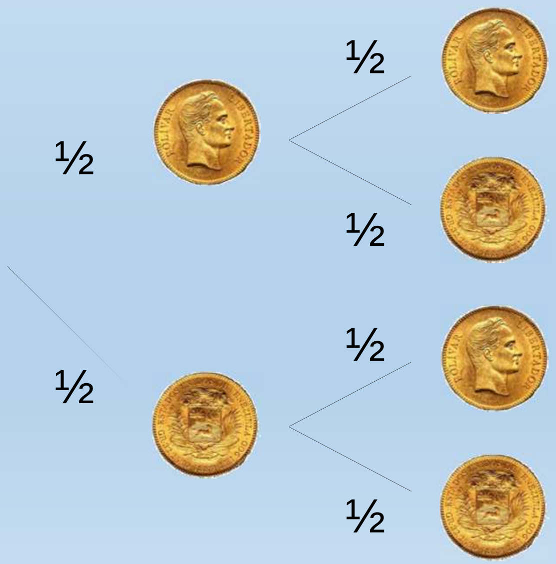
$S \cap S \cap S = SSS$

Diagramas de árboles

- En una pizzería, los clientes pueden ordenar…
  - o Uno de tres tamaños: Pequeña (P), Mediana (M) o Familiar (F)
  - o Masa delgada (D) o masa gruesa (G)
  - o Dos sabores: Salame (S) y/o Champiñones (C)
- El Ω es el resultado de cuatro etapas:
  - ✓Tamaño de la pizza {P, M, F}
  - ✓Grosor de la masa {D, G}
  - ✓Salame {S o S^c}
  - ✓Champiñon {C o C^c}
- ¿Cómo sería este diagrama de árbol?

Diagramas de árboles

- Una vez construido el diagrama, el paso siguiente consiste en incluir las probabilidades de cada rama individual
- Una secuencia que comienza con probabilidades marginales (primer evento)
- Ejemplo de eventos independientes



$P(C \cap C) = ½ * ½ = ¼$

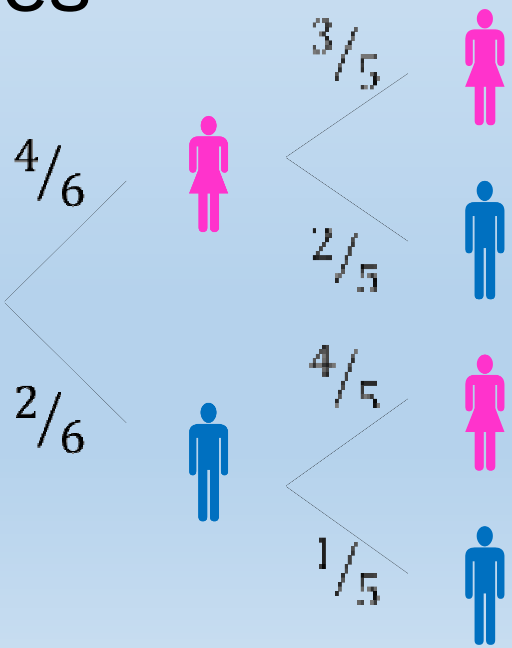
$P(C \cap S) = ¼$

$P(S \cap C) = ¼$

$P(S \cap S) = ¼$

Diagramas de árboles

- ¿Si los eventos son dependientes?
- De un grupo de cuatro damas (D) y dos caballeros (C) y debe escoger dos para que integren un comité. Lo escoge al azar
- No puede escoger la misma persona dos veces



$$P(D \cap D) = \tfrac{4}{6} * \tfrac{3}{5} = \tfrac{12}{30}$$

$$P(D \cap C) = \tfrac{8}{30}$$

$$P(C \cap D) = \tfrac{8}{30}$$

$$P(C \cap C) = \tfrac{2}{30}$$

Ley de Probabilidad Total

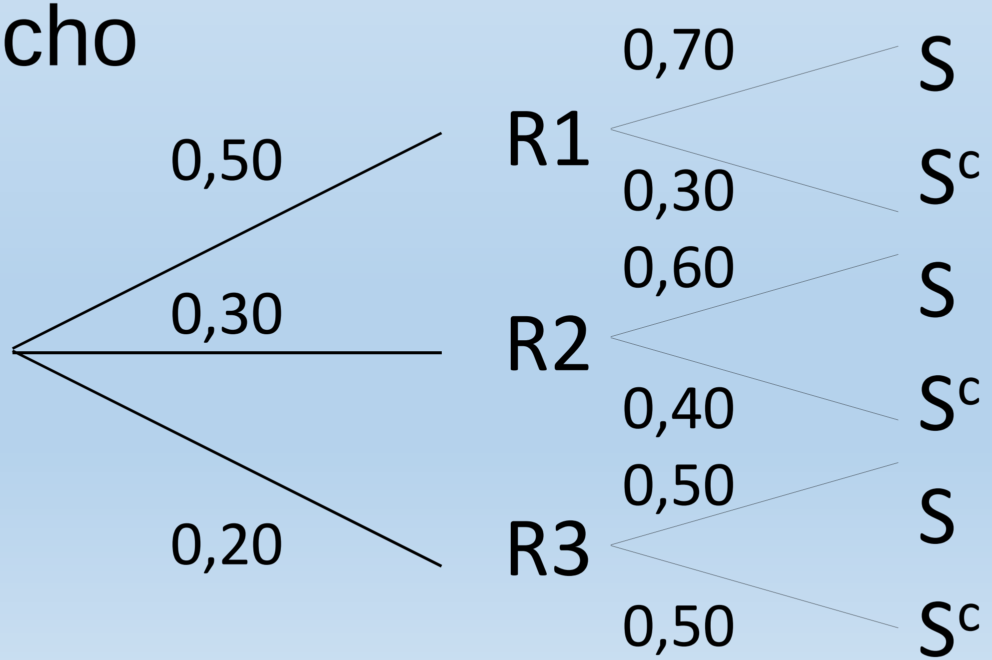
- A veces un problema ofrece varias probabilidades condicionales diferentes que involucran al evento A, pero no P(A)
- Si los eventos $A_1,...A_k$ particionan el Ω en eventos mutuamente exhaustivos y exclusivos, la *Ley de la Probabilidad Total* establece que la probabilidad de un evento B esta dada por:

$$P(B) = P(B|A_1)\,P(A_1) + P(B|A_2)\,P(A_2) + \ldots + P(B|A_k)\,P(A_k)$$

$$P(B) = \sum_{i=1}^{k} P(B|A_i)\,P(A_i)$$

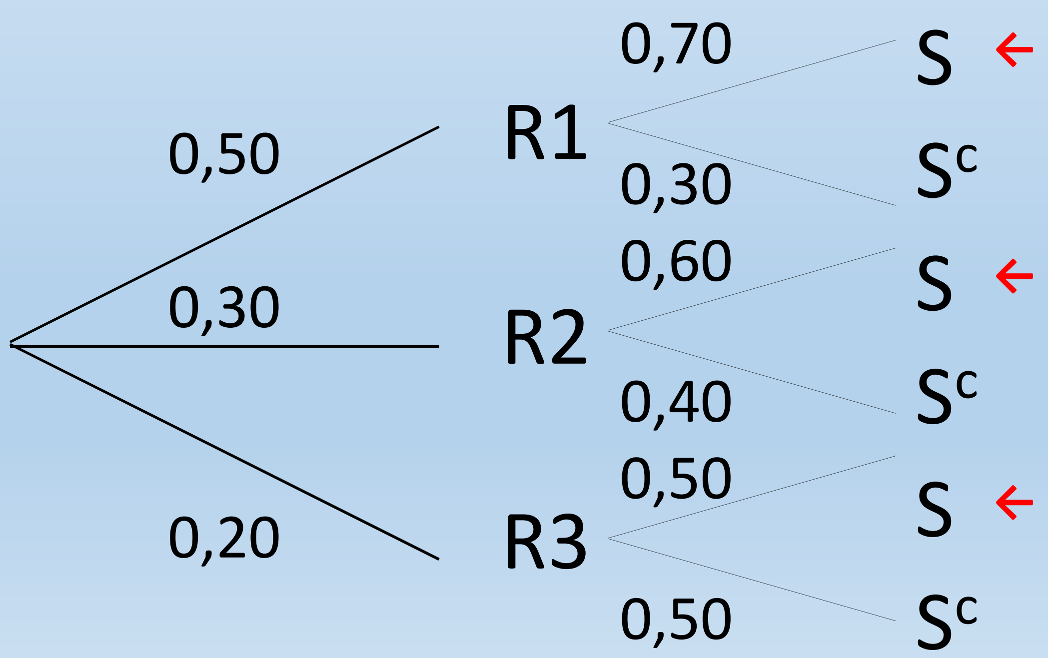
Ley de Probabilidad Total

- Queremos saber la probabilidad que un cliente que vaya a un restaurant de la localidad quede satisfecho
- Etapas: (1) un cliente va a uno de tres restaurantes, (2) el cliente queda satisfecho o insatisfecho



Ley de Probabilidad Total

- La satisfacción puede alcanzarse a través de tres vías diferentes
- 0,50 * 0,70 = 0,35; 0,30 * 0,60 = 0,18; 0,20 * 0,50 = 0,10
- P(S) = 0,35 + 0,18 + 0,10 = 0,63



Teorema de Bayes

Postgrado en Modelado y Simulación

- Suponga que un doctor le cuesta diagnosticar una enfermedad sin un test sanguíneo
- El paciente tiene o no la enfermedad
- El doctor espera que la probabilidad de acertar sea alta y fallar sea baja
- Los que hacen el test quieren saber la probabilidad de acertar dado que el paciente tiene la enfermedad

Teorema de Bayes

- El doctor quiere saber algo diferente: la probabilidad que el paciente tenga la enfermedad dado que diagnosticó positivamente al paciente
- Una *probabilidad posterior*, encontrada **después** del hecho
- La P(A | B), o la probabilidad que A ocurrió en la etapa uno dado que observamos B en la etapa dos… Lo contrario al diagrama de árbol
- Se resuelve mediante *Teorema de Bayes*

Teorema de Bayes

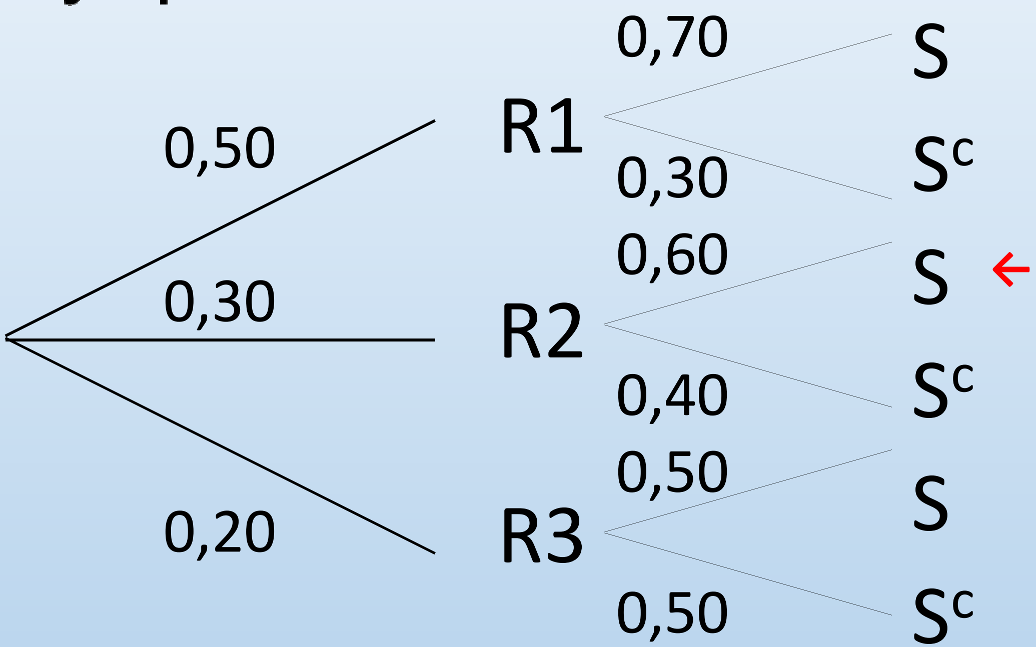
- Volviendo al ejemplo de los restaurantes, ¿Dado un cliente que quedó satisfecho, cuál es la probabilidad que comió en el R2?
- Es decir, ¿Cuál es $P(R2 \mid S)$
- Usando la probabilidad condicional, se tiene:

$$P(R2 \mid S) = \frac{P(R2 \cap S)}{P(S)}$$

- Note que $P(S)$ ya lo calculamos mediante el *Teorema de Probabilidad Total*

Teorema de Bayes

- $P(R2 \cap S)$ es la probabilidad que el cliente fue a R2 y quedó satisfecho



- $$P(R2 \mid S) = \frac{P(R2 \cap S)}{P(S)} = \frac{0{,}18}{0{,}63} = 0{,}286$$

VVAA

Postgrado en Modelado y Simulación

- Una variable aleatoria es *variable* pues toma distintos valores y *aleatoria* pues el valor observado no puede ser predicho antes de la realización
- En ocasiones se sabe cuáles son sus posibles valores
- Define una cantidad que representa:
  - Algo medido de forma incompleta,
  - Una muestra extraída al azar de una población usando un mecanismo aleatorio

¿Aleatorio?

"Nueve máquinas diferentes para lanzar monedas, diseñadas y construidas por Nitipak Samsen del Programa de Design Interactions del Royal College of Art. Su objetivo es construir una máquina que pueda lanzar una moneda y que esta caiga cara o sello de manera predecible. No es fácil controlar el resultado de una moneda lanzada. Samsen ha identificado 31 factores que influencian el resultado y cree que es posible controlarlos"

http://makezine.com/2009/08/14/nitipak-samsens-coin-flipping-machi/

¿Aleatorio?

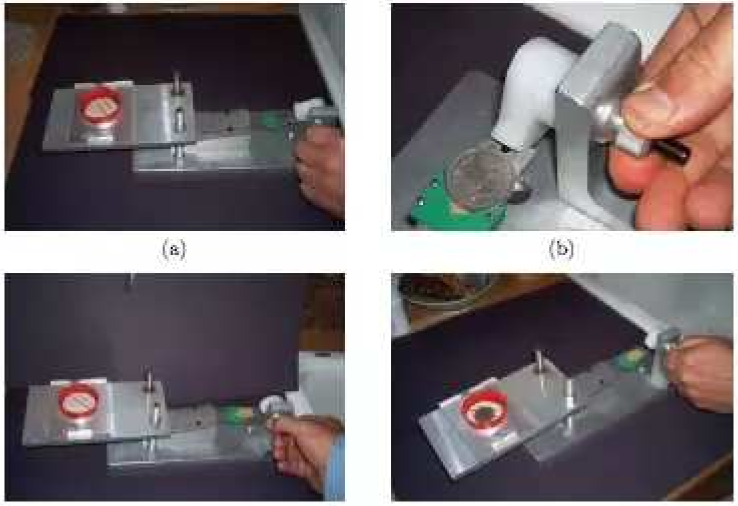


Fig. 1

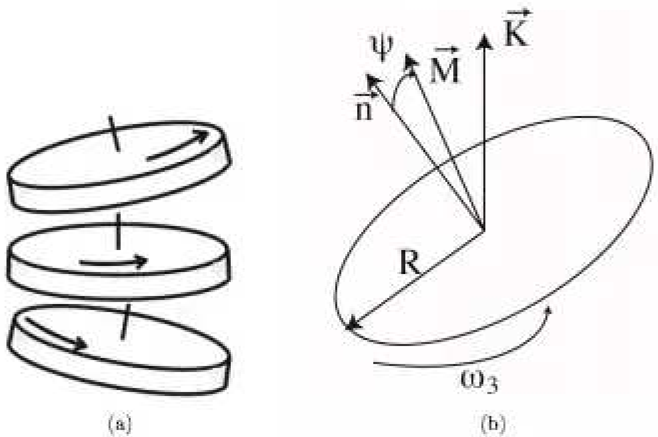


Fig. 2 *(a) Diagram of a precessing coin. (b) Coordinates of precessing coin: $\vec{K}$ is the upward direction, $\vec{n}$ is the normal to the coin, $\vec{M}$ is the angular momentum vector, and ω_3 is the rate of rotation around the normal $\vec{n}$.*

SIAM REVIEW
Vol. 49, No. 2, pp. 211–235

Dynamical Bias in the Coin Toss*

Persi Diaconis[†]
Susan Holmes[‡]
Richard Montgomery[§]

Abstract. We analyze the natural process of flipping a coin which is caught in the hand. We show that vigorously flipped coins tend to come up the same way they started. The limiting chance of coming up this way depends on a single parameter, the angle between the normal to the coin and the angular momentum vector. Measurements of this parameter based on high-speed photography are reported. For natural flips, the chance of coming up as started is about .51.

Key words. Berry phase, randomness, precession, image analysis

AMS subject classifications. 62A01, 70B10, 60A99

DOI. 10.1137/S0036144504446436

VVAA

Postgrado en Modelado y Simulación

- Cuando se refiere a un valor específico de la variable, cuando éste valor se ha observado, deja de ser aleatorio

- En este caso se escribe $X = x$ ó $X = 1$ para indicar que se ha observado un valor específico de X

VVAA

- Formalmente, una VA es una función de un Ω, el cual incluye a todas las probabilidades posibles, el intervalo [0,1]

- Se denotan con letras mayúsculas (X, Y, etc.)

- Una vez que ha sido determinado el Ω de un experimento este nunca cambia, pero se pueden tener muchas VVAA asociadas a él

VVAA

Ejemplo (A), se arroja dos veces un dado equilibrado. El Ω asociado es:

$$\Omega = \{(x_1, x_2)\ /\ x_i \in \{1, 2, 3, 4, 5, 6\}\}$$

- Posibles VVAA asociadas a este experimento:
  - o X: ”número de caras pares”
  - o Y: “máximo puntaje”
  - o Z: “suma de puntos”
- Algunos resultados:

$$X(3,5) = 0 \quad Y(3,5) = 5 \quad Z(3,5) = 8$$

$$X(2,1) = 1 \quad Y(2,1) = 2 \quad Z(2,1) = 3$$

$$X(4,6) = 2 \quad X(4,6) = 6 \quad Z(4,6) = 10$$

- Otros ejemplos:

B. Arrojamos una moneda equilibrada tres veces

$$X = \begin{cases} 1 \text{ si el número de caras es impar} \\ 0 \text{ en caso contrario} \end{cases}$$

C. Arrojamos una moneda hasta que se obtiene la primera cara

X: "número de tiros necesarios"

D. Tiempo que tarda en concretarse una conexión a un servidor

X: "tiempo requerido para la conexión"

- En los ejemplos (A), (B) y (C) las VVAA toman un número finito o infinito, pero numerable, de valores.

- En el ejemplo (D) la VA toma valores en un conjunto infinito no numerable del intervalo [0, ∞] ó [0, *M*] (si existe un tiempo máximo *time out*).

VVAA

- En los ejemplos (A), (B) y (C) las VVAA toman un número finito o infinito, pero numerable, de valores. En este caso hablamos de VVAA **DISCRETAS**
- En el ejemplo (D) la VA toma valores en un conjunto infinito no numerable del intervalo [0, ∞] ó [0, *M*] (si existe un tiempo máximo *time out*). En este caso hablamos de VVAA **CONTÍNUAS**

Modelo probabilístico

- Por supuesto, en este curso nos interesan situaciones en las que las probabilidades siguen un cierto patrón predecible, útil para establecer y describir un modelo de probabilidad
- Formalmente, un modelo probabilístico provee formulas para calcular probabilidades, determinar salidas promedio de múltiples experimentos y estimar la variabilidad esperada de un experimento aleatorio al siguiente

Modelo probabilístico

- En otras palabras, un modelo probabilístico es un modelo matemático que permite ajustar un proceso aleatorio

Modelo probabilístico

Postgrado en Modelado y Simulación

- Un modelo probabilístico debe cumplir con ciertas asunciones, a fin de garantizar un correcto ajuste
- Su desempeño suele evaluarse usando datos reales
- Las partes fundamentales de un modelo probabilístico son la(s) variable(s) aleatoria(s) y su(s) respectiva(s) distribución(es) de probabilidad

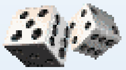
Función de masa de probabilidad

- La función que asigna probabilidades para una VA discreta se llama fmp
- Muestra cuánta probabilidad, o *masa* (en el sentido de *peso*), se le otorga a un valor dado de las VVAA
- Para una VA continua no asigna *masa* sino *densidad*, o qué tan densa es la probabilidad alrededor de un valor x dado (no para valores particulares de X sino intervalos de X)

Distribución de probabilidad

- *Distribución de probabilidad* para X: lista de todos los valores posibles de X acompañada de sus respectivas probabilidades
- La distribución de probabilidad de una VA discreta X tiene las siguientes propiedades:
  - $\mathrm{P(x)}$ está en [0,1] para cualquier valor de X
  - Valores individuales de X son mutuamente excluyentes, por lo tanto $\mathrm{P(a) + P(b)}$
  - Dado que a cada elemento en Ω se le asigna un valor de X, la suma de todas las probabilidades en la distribución de probabilidad de X debe ser uno

Función de masa de probabilidad

 Retomando el ejemplo de los dos dados…

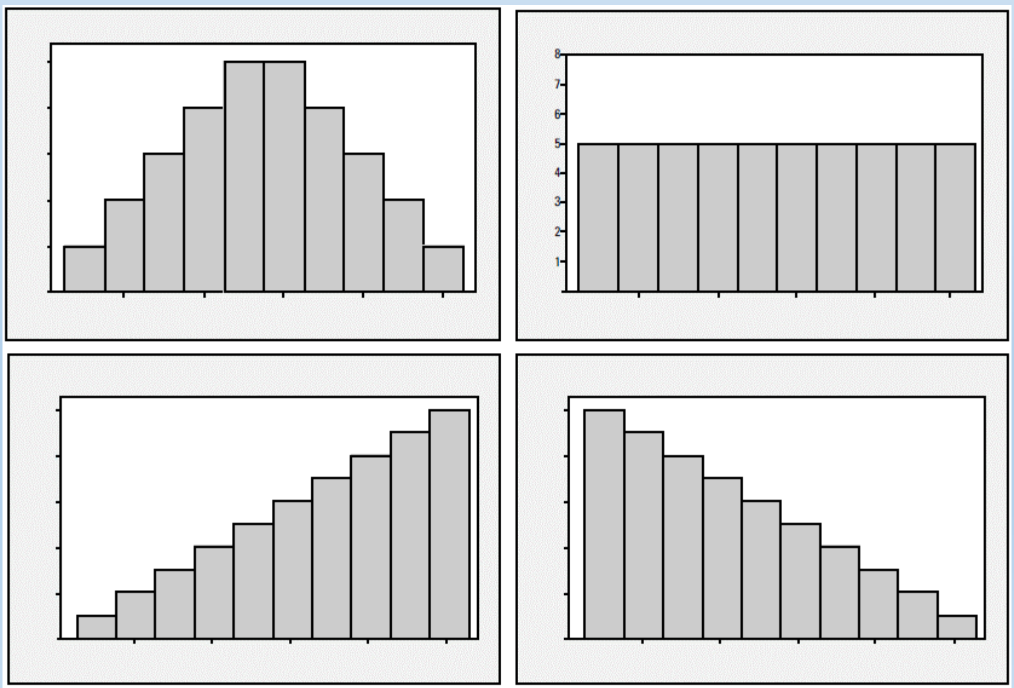
Ω

Resultados (ω)						X(ω)	Frecuencias	P(x)
(1,1)						2	1	1/36
(1,2)	(2,1)					3	2	2/36
(1,3)	(2,2)	(3,1)				4	3	3/36
(1,4)	(2,3)	(3,2)	(4,1)			5	4	4/36
(1,5)	(2,4)	(3,3)	(4,2)	(5,1)		6	5	5/36
(1,6)	(2,5)	(3,4)	(4,3)	(5,2)	(6,1)	7	6	6/36
(2,6)	(3,5)	(4,4)	(5,3)	(6,2)		8	5	5/36
(3,6)	(4,5)	(5,4)	(6,3)			9	4	4/36
(4,6)	(5,5)	(6,4)				10	3	3/36
(5,6)	(6,5)					11	2	2/36
(6,6)						12	1	1/36
						Total	36	1

Función de masa de probabilidad

- Por supuesto, en las VVAA discretas la distribución de probabilidad puede graficarse como histograma
- Uno de los factores que permiten la identificación de la distribución de probabilidad es la forma que esta toma cuando se la grafica…

P(x)



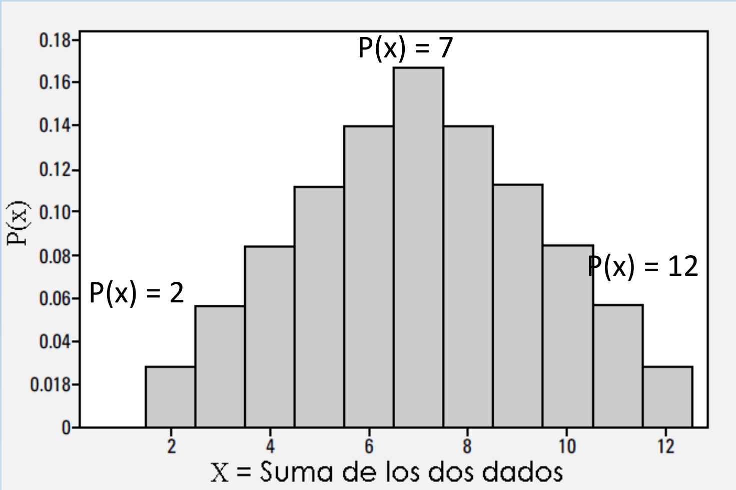
Valores de X

Función de masa de probabilidad

- En el ejemplo de la suma de dos dados, la distribución de probabilidad de X tiene forma de campana y es simétrica

- La suma igual a siete es la más probable (por eso su importancia en algunos juegos de azar)

- La probabilidad disminuye al alejarnos del centro, hasta llegar a las más bajas



P por arriba, por debajo, etc.

- Una vez establecida la distribución de probabilidad, usualmente se requiere usarla para calcular varias probabilidades
- Por ejemplo, en el caso de la suma de los dados, puede interesar que la suma sea:
    - Al menos 7
    - Menor que 7
    - A lo sumo 10
    - Más de 10
- Estas situaciones también representan eventos

Probabilidad...

- El evento “al menos 7” significa que siete es el menor valor posible
- Incrementa de ahí en adelante hasta el máximo valor posible [7,12] adicionando probabilidades

$$P(7 \leq X \leq 12) = P(X=7) + P(X=8) + P(X=9) + P(X=10) + P(X=11) + P(X=12)$$

- O, usando la tabla

$$P(7 \leq X \leq 12) = 6/36+5/36+4/36+3/36+2/36+1/36$$

$$P(7 \leq X \leq 12) = 21/36 = 0,58 = 58\%$$

Probabilidad...

- Note que los eventos “menos que 7” y “al menos 7” son complementarios
- Dividen Ω en dos conjuntos separados, los cuales no contienen elementos en común
- La suma de las probabilidades de ambos eventos suman uno.
- Por ejemplo, si hemos calculado $P(7 \leq X \leq 12)$, el cálculo del complemento se simplifica a

$$P(X < 7) = 1 - P(7 \leq X \leq 12)$$

Función de distribución acumulada

- Cuando lidiamos con estos "menor que", "al menos", etc. Lo mejor es acudir a una función de distribución acumulada
- Una función que representa la probabilidad que X es menor o igual a cualquier valor x, siendo igual a la suma de todas las probabilidades que X sea menor o igual a x
- La denotaremos como $F(x) = P(X \leq x)$

Función de distribución acumulada

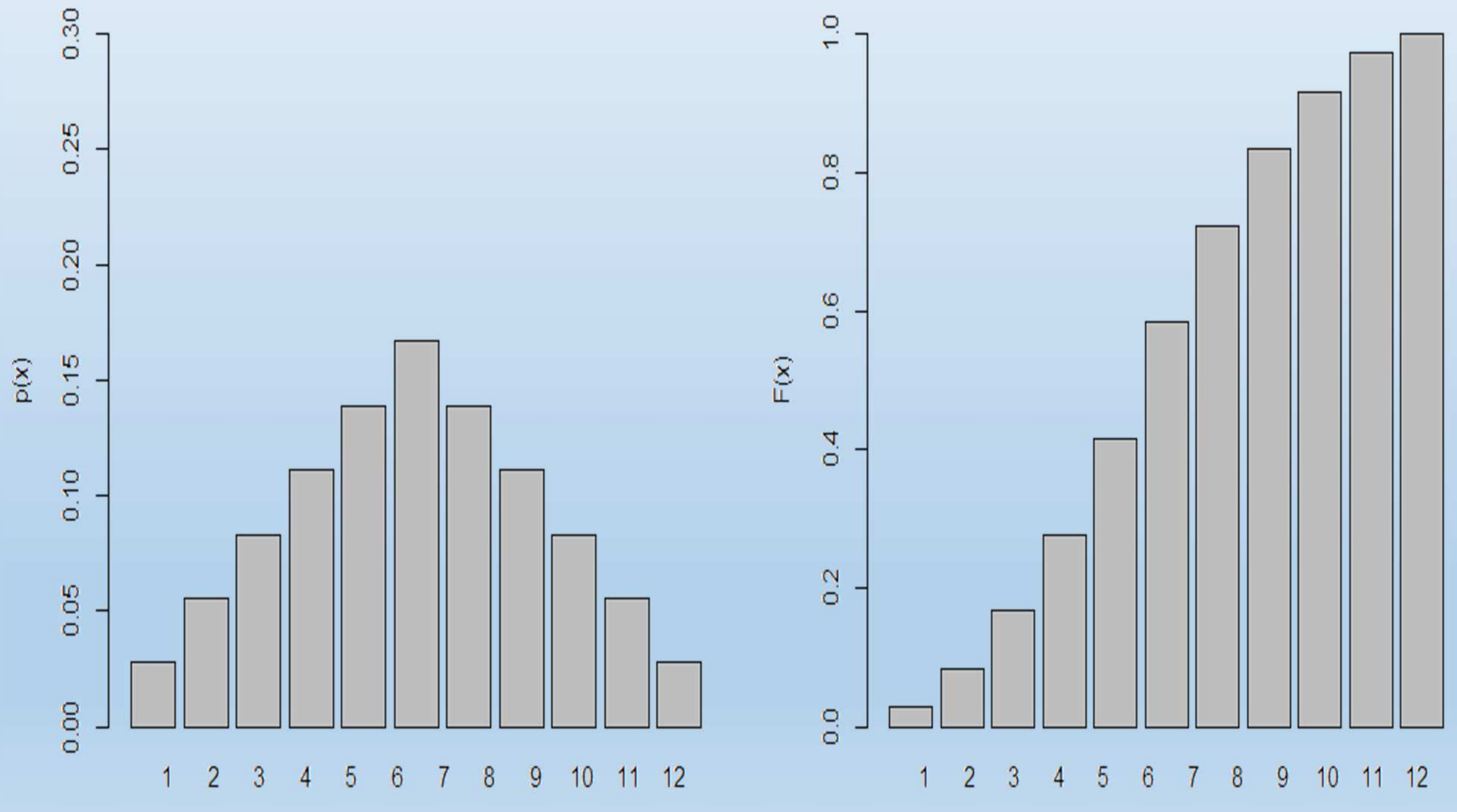
En nuestro ejemplo de los dos dados

Ω

Resultados (ω)						$X(\omega)$	Frecuencias	$P(x)$	$F(x)$
(1,2)						2	1	1/36	1/36
(1,2)	(2,1)					3	2	2/36	3/36
(1,3)	(2,2)	(3,1)				4	3	3/36	6/36
(1,4)	(2,3)	(3,2)	(4,1)			5	4	4/36	10/36
(1,5)	(2,4)	(3,3)	(4,2)	(5,1)		6	5	5/36	15/36
(1,6)	(2,5)	(3,4)	(4,3)	(5,2)	(6,1)	7	6	6/36	21/36
(2,6)	(3,5)	(4,4)	(5,3)	(6,2)		8	5	5/36	26/36
(3,6)	(4,5)	(5,4)	(6,3)			9	4	4/36	30/36
(4,6)	(5,5)	(6,4)				10	3	3/36	33/36
(5,6)	(6,5)					11	2	2/36	35/36
(6,6)						12	1	1/36	36/36
						Total	36	1	

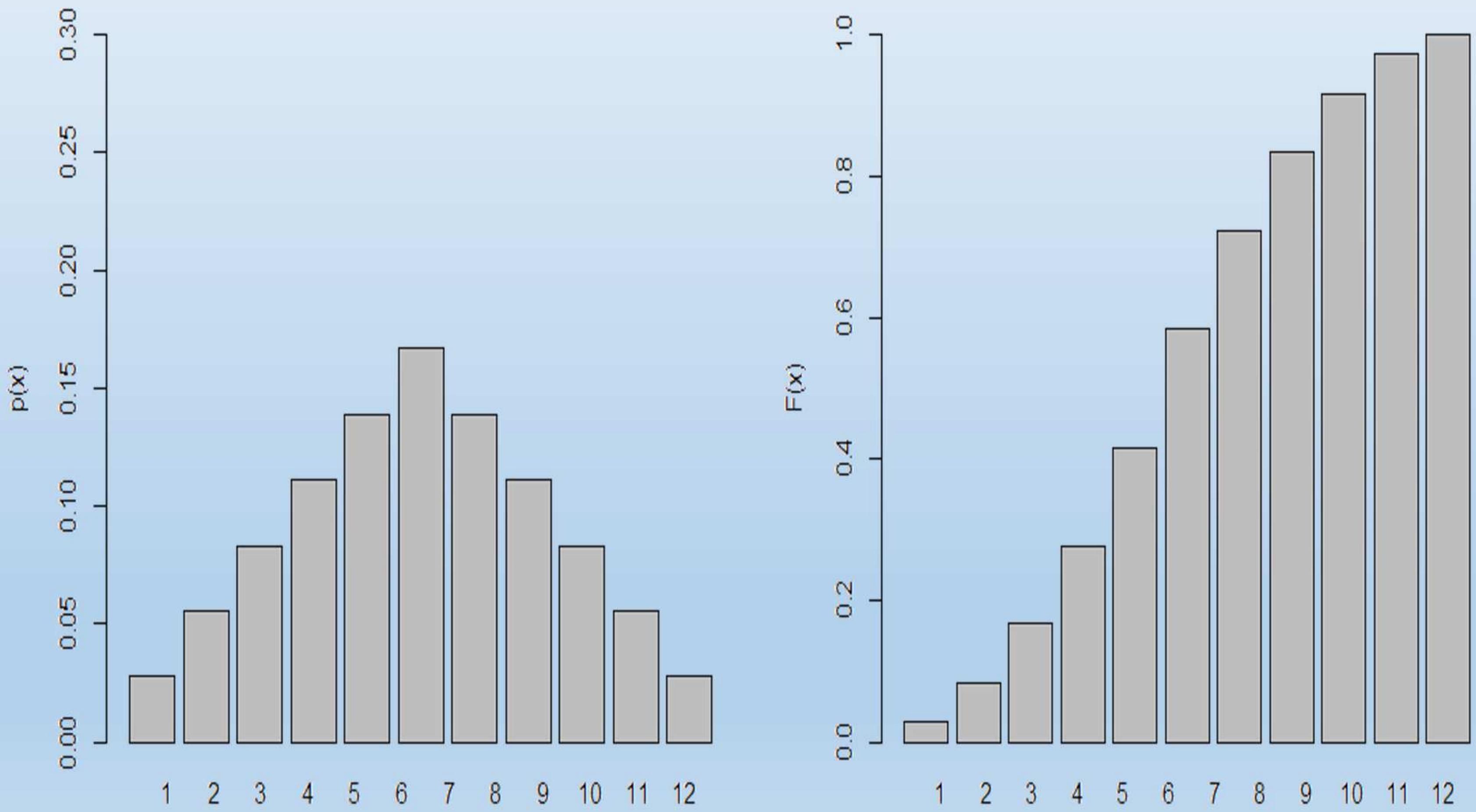
Función de distribución acumulada

Gráficamente…



Función de distribución acumulada

Gráficamente… ¿Qué pasa con una VA continua?



Esperanza

- Ejemplo: una empresa de televisión por cable tiene 20.000 clientes y ofrece de uno a cinco paquetes.
- La distribución del número de paquetes x contratados por los 20.000 clientes es

x	1	2	3	4	5
Número de clientes	8000	6000	4000	2000	1000
Proporción	0,38	0,28	0,19	0,10	0,05

- Si nos interesa conocer el número promedio de paquetes contratados, o sea el valor promedio de X, entonces calculamos:

$$\frac{1*8000+2*6000+3*4000+4*2000+5*1000}{21000} =$$

$$\frac{45000}{21000} = 2.14$$

Esperanza

- Es decir, la esperanza de una VA es su centro de masa o centro de probabilidad:

$$E[X] = \begin{cases} \sum_{i=1} x_i P(X = x_i) \\ \int_{-\infty}^{\infty} x f(x) dx \end{cases}$$

Dispersión

- Formalmente, varianza es la distancia promedio de los valores de los datos respecto al promedio

$$\mathrm{Var}[X] = \mathrm{E}[X - \mathrm{E}[X]]^2 = \sigma_x^2$$

$$\mathrm{Var}[X] = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \mathrm{E}[X^2] - \mu_X^2$$

- Las unidades están al cuadrado, la raíz cuadrada mantiene las unidades iguales, así tenemos la desviación estándar: $\sigma_x = \sqrt{\sigma_x^2}$

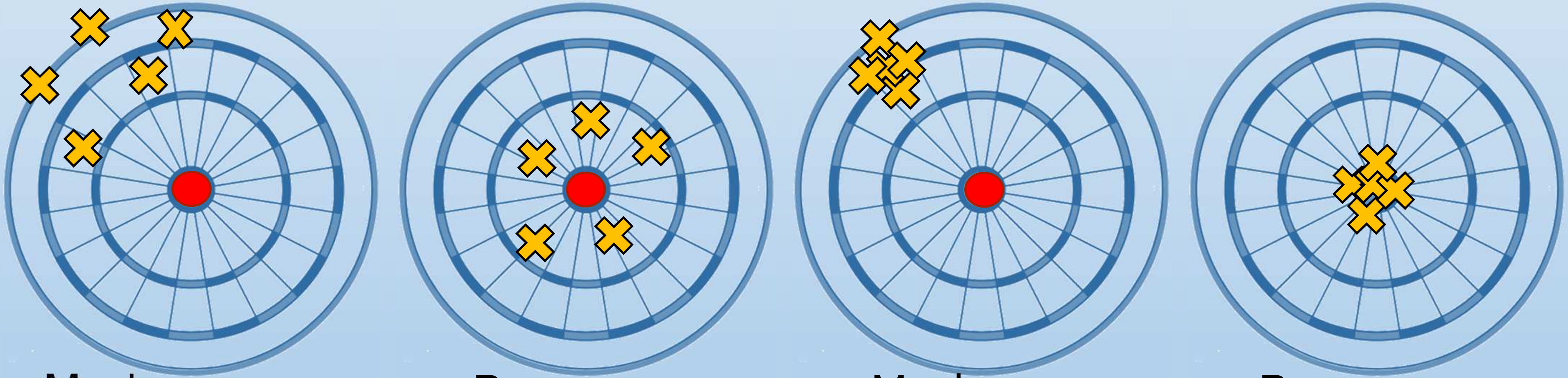
- Algunas propiedades

$$\mathrm{Var}[X] \geq 0$$

$$\mathrm{Var}[X] = 0 \; sii \; X = \text{constante}$$

Dispersión y sesgo

En la imagen siguiente, el punto rojo es el objetivo. Cualquier lanzamiento cercano a él se considera poco sesgado. Si los lanzamientos posteriores son cercanos al primero, entonces se considera que la varianza es baja.



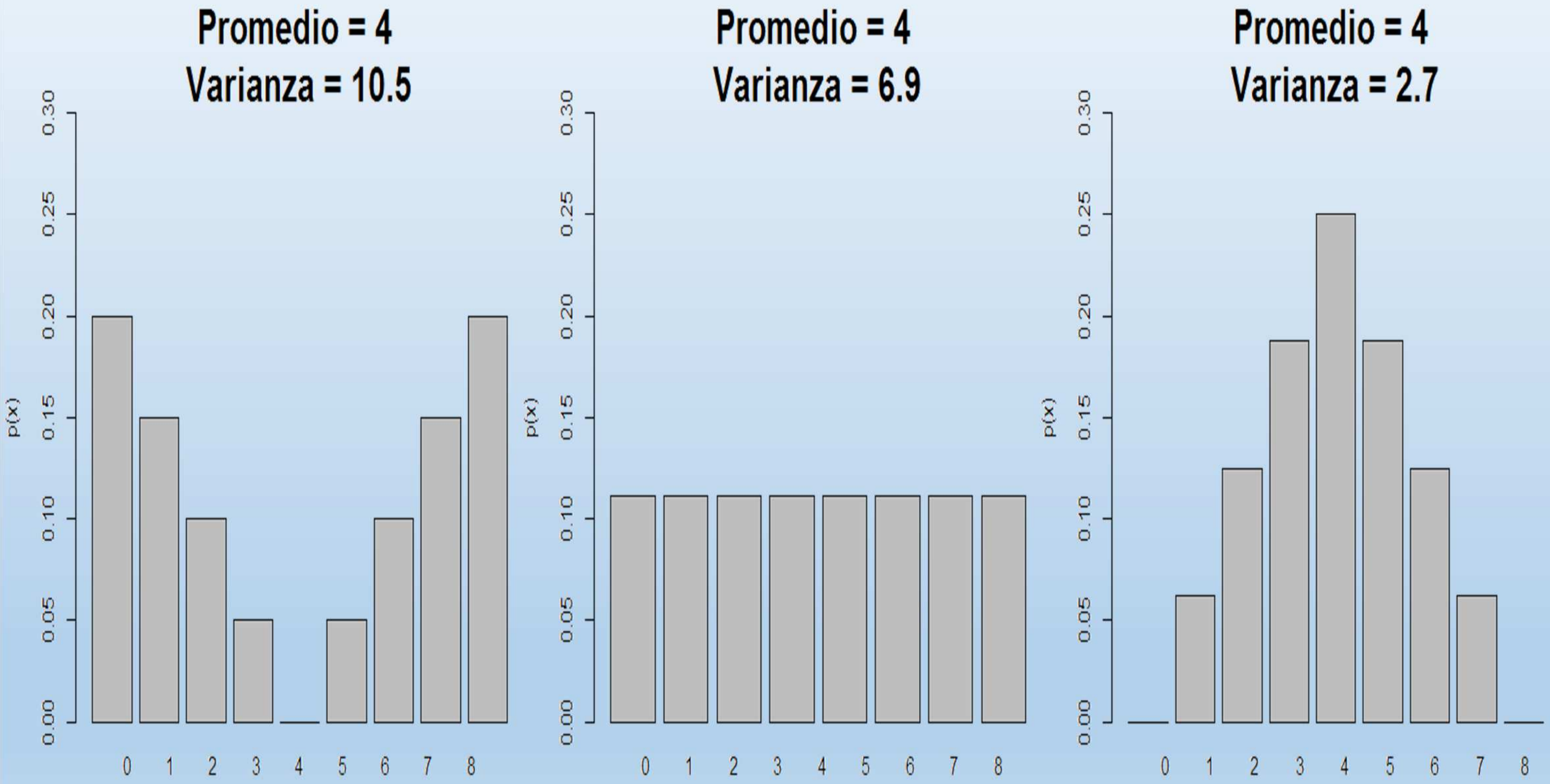
Mucho sesgo
Varianza elevada

Poco sesgo
Varianza elevada

Mucho sesgo
Varianza baja

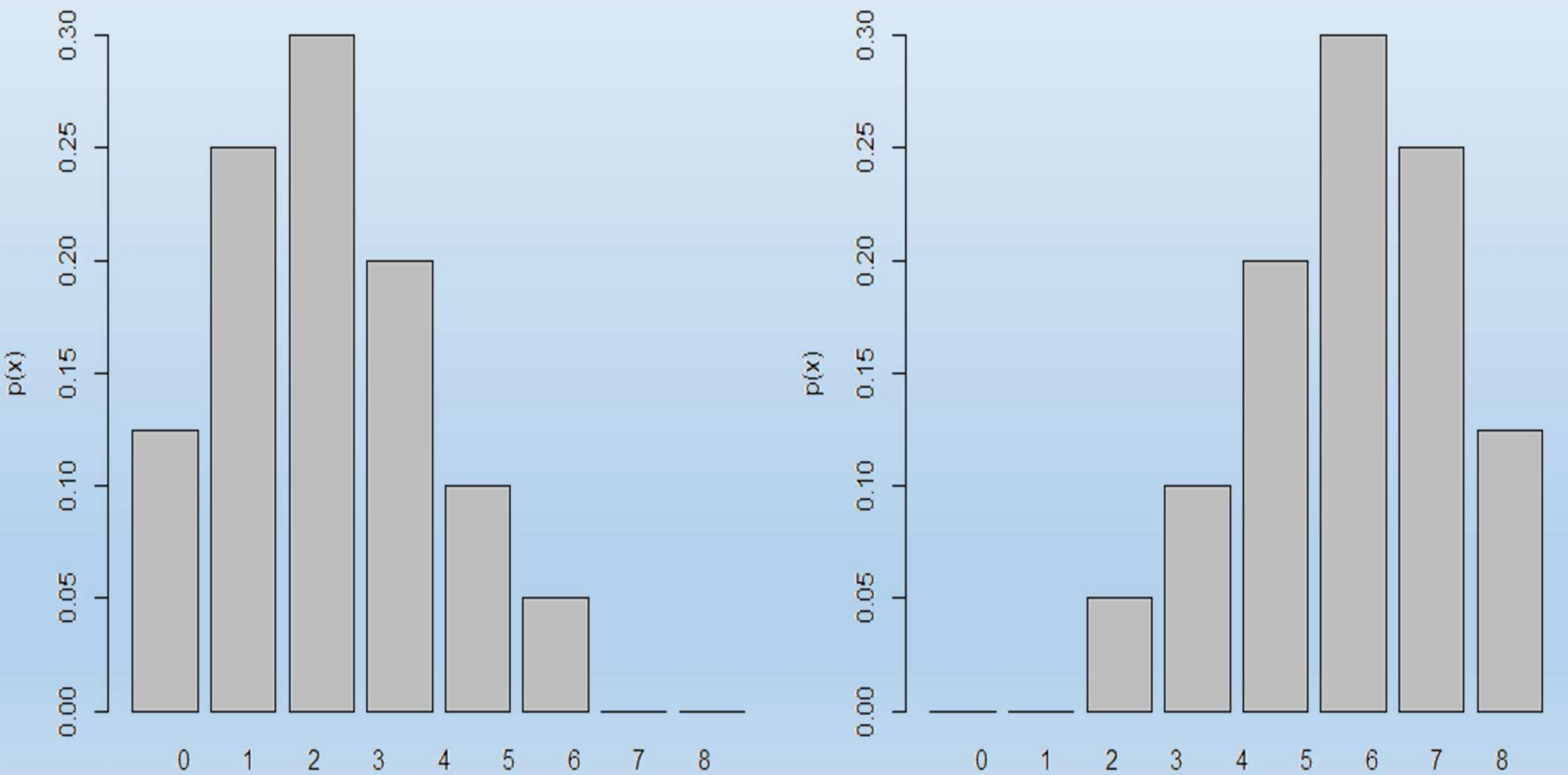
Poco sesgo
Varianza baja

Representación gráfica



Representación gráfica

- Por supuesto hay distribuciones sesgadas…



Momentos

Postgrado en Modelado y Simulación

- Para cada VA existe un conjunto de medidas descriptivas que pueden emplearse para caracterizar su distribución de probabilidad
- Pueden ser especificarlas si son conocidas
- Los momentos de una VA X son los valores esperados de ciertas funciones de X
- Sea X una VA y r un número natural, se llama momento de orden r al valor, cuando existe (es decir: la suma que lo define o la integral converge), de $m_r = E[X^r]$

Momentos

- El primer momento o momento de orden uno es la esperanza o valor esperado de la variable aleatoria y se denota por (μ_X)

- La cual, como ya vimos se considera como una cantidad numérica alrededor de la cual los valores de la VA tienden a agruparse

- Por lo tanto, la media es una medida de tendencia central

Momentos

- El momento de orden dos es la varianza, definida como:

$$Var[X] = E[X - E[X]]^2$$

- Como ya vimos, la varianza de una variable aleatoria es una medida de la dispersión de la distribución de probabilidad de esta.

Momentos

- El tercer momento está relacionado con la asimetría de la distribución de probabilidad de X
- También llamado sesgo (skewness)

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{E[(X-\mu)^2]^{3/2}}$$

Momentos

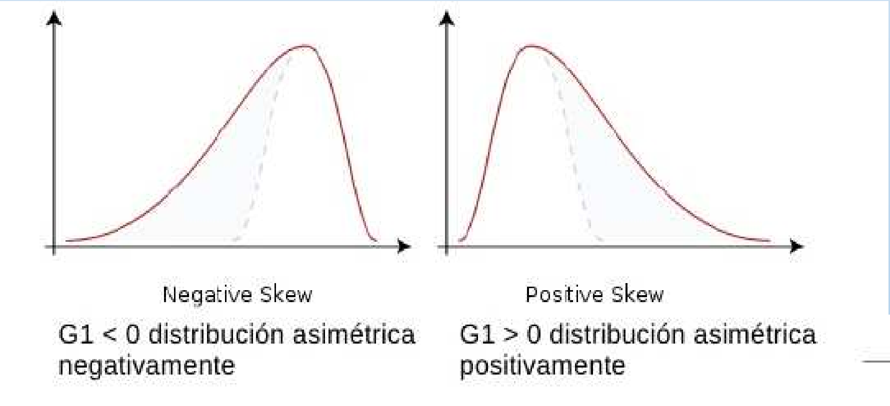
Postgrado en Modelado y Simulación

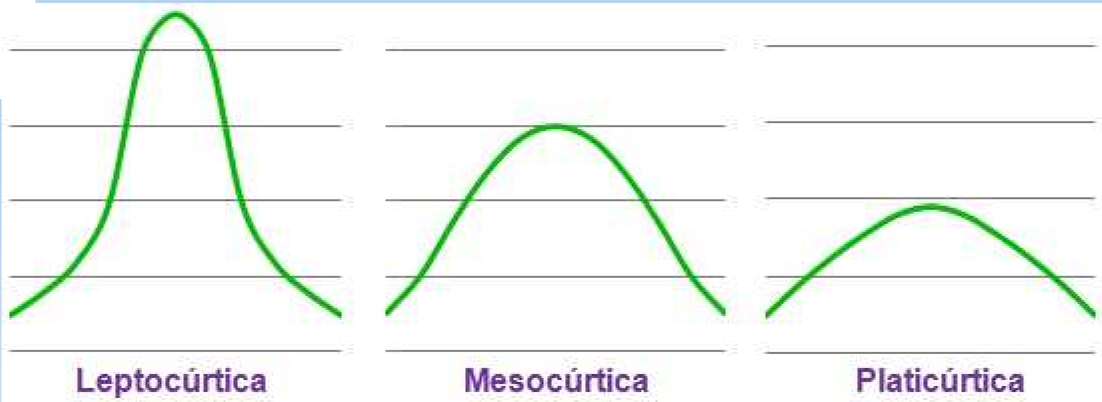
- El cuarto momento central es una medida de qué tan “puntiaguda” es la distribución de probabilidad y recibe el nombre de kurtosis

$$\gamma_2 = \frac{\sum \frac{(X - \bar{X})^4 n(n+1)}{n-1} - 3\left[\sum (X - \bar{X})^2\right]^2}{(n-2)(n-3)}$$

Momentos

- Los momentos tercero y cuarto también se conocen, respectivamente, como los factores de forma primero y segundo de la distribución de probabilidad pues, en gran medida, determinan la forma de la distribución de probabilidad.





Densidad conjunta y marginal

- La función de distribución conjunta de dos variables aleatorias cualesquiera X y Y se define como:

$$F(a,b) = P\{X \leq a, Y \leq b\}, -\infty < a,b < \infty$$

- La función de probabilidad de X puede obtenerse a partir de la conjunta:

$$\begin{aligned} F_X(a) &= P\{X \leq a\} \\ &= P\{X \leq a, Y < \infty\} \\ &= F(a, \infty) \end{aligned}$$

Densidad conjunta: Caso continúo

- Cuando X y Y son continuas, se define la probabilidad conjunta como:

$$P\{X \in A, Y \in B\} = \int_B \int_A f(x,y)dx\, dy$$

Integrales dobles, ejemplo

Postgrado en Modelado y Simulación

$$\int_0^1 \int_{\sqrt{x}}^{1+x} 2xy\, dy\, dx$$

$$\int_0^1 \left\{ \int_{\sqrt{x}}^{1+x} 2xy\, dy \right\} dx$$

$$\int_0^1 \left\{ 2x \left. \frac{y^2}{2} \right]_{\sqrt{x}}^{1+x} \right\} dx$$

$$\int_0^1 \left\{ \left. xy^2 \right]_{\sqrt{x}}^{1+x} \right\} dx$$

$$\int_0^1 \{x(1+x)^2 - x(\sqrt{x})^2\}\, dx$$

$$\int_0^1 (x + x^2 + x^3) dx$$

$$\left. \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} \right]_0^1$$

$$\left(\frac{1}{2} + \frac{1}{3} + \frac{1}{4}\right) - (0)$$

$$= \frac{13}{12}$$

Densidad conjunta: Caso continúo

- Continuamos, cuando X y Y son continuas, se define la probabilidad conjunta como:

$$P\{X \in A,\ Y \in B\} = \int_B \int_A f(x,y)dx\,dy$$

- La función de probabilidad o densidad de X puede obtenerse a partir de la conjunta:

$$P\{X \in A\} = P\{X \in A,\ Y \in (-\infty,\ \infty\}$$

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

$$= \int_A f(x)dx$$

- Donde $\quad f_x(x) = \int_{-\infty}^{\infty} f(x,y)dy$

Densidad conjunta: Caso discreto

Postgrado en Modelado y Simulación

- Cuando X y Y son discretas se define la probabilidad conjunta como:

$$P(x,y) = P\{X = x,\ Y = y\}$$

- La función de probabilidad de masa de X puede obtenerse a partir de la conjunta

$$p_X(x) = \sum_{y:p(x,y)>0} p(x,y)$$

Independencia de VVAA

- Intuitivamente, dos VVAA son independientes si los valores que toma una de ellas no afectan a los de la otra ni a sus probabilidades.
- La independencia puede ser evidente a partir del experimento, por ejemplo:  

 También la dependencia puede ser clara, por ejemplo:

  - X = resultado de un lanzamiento
  - Y = indicadora de puntuación par
  - Si $Y = 1 \rightarrow X = \{2, 4, 6\}$; si $X = 3 \rightarrow Y = 0$

Independencia de VVAA

- En ocasiones podemos suponer la existencia de alguna relación entre variables, aunque difícil de entrever

- Por ejemplo, si a diferentes elevaciones medimos la temperatura media mensual, vemos que hay una relación: los valores de la una influirán en los de la otra.

- Explorar la naturaleza exacta de la relación entre ambas es lo que en estadística conocemos como un problema de regresión

Independencia de VVAA

- Dos VA X y Y son independientes si para cualesquiera *a* y *b*:

$$P(X \leq a \cap Y \leq b) = P(X \leq a) \cdot P(Y \leq b)$$

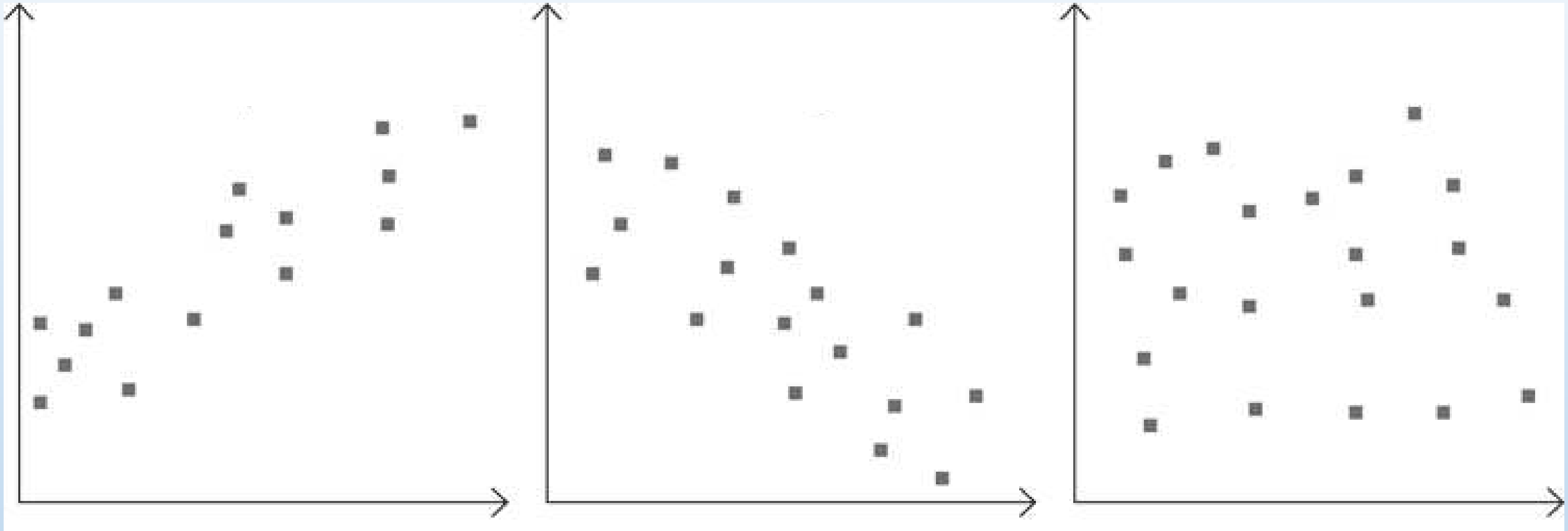
- En el caso discreto

$$p(x,y) = p_X(x)p_Y(y)$$

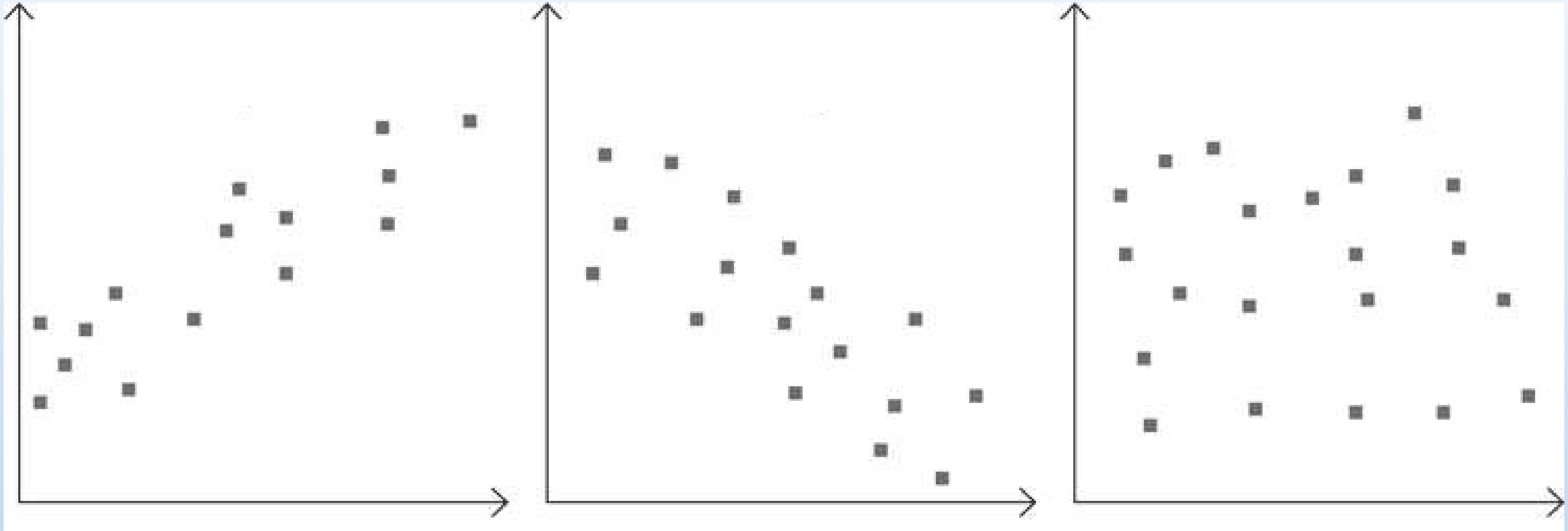
- En el caso contínuo

$$f(x,y) = f_X(x)f_Y(y)$$

Gráficamente…



Covarianza y correlación



- Ambos conceptos describen las relaciones y medidas de dependencia entre dos VVAA
- Similares pero diferentes…

Covarianza y correlación

- Covarianza: mide de cuánto varían dos VVAA en tándem
- Correlación: representa qué tanto están relacionadas dos VVAA
- La covarianza es una medida de correlación, la correlación es una forma escalar de la covarianza

Covarianza y correlación

- La correlación toma valores entre -1 y +1, la covarianza entre -∞ y +∞
- La covarianza es afectada por el cambio de escala, la correlación no
- La correlación no tiene dimensión (no tiene unidades), en la covarianza tiene las unidades producto de las variables

Covarianza y correlación

- Formalmente, la covarianza de dos variables aleatorias X y Y se define como:

$$\mathrm{Cov}(X,Y) = E[(X - E[X])(Y - E[Y])]$$

$$= E[XY - XE[Y] - YE[X] + E[X]E[Y]]$$

$$= E[XY] - E[Y]E[X] - E[X]E[Y] + E[X]E[Y]$$

$$= E[XY] - E[X]E[Y]$$

- Si X y Y son independientes $\mathrm{Cov}(X,Y) = 0$

Covarianza y correlación

- Dadas las VVAA X, Y y Z y c constante, algunas propiedades son:
    1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
    2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$
    3. $\mathrm{Cov}(\mathrm{c}X,Y) = c\mathrm{Cov}(X,Y)$
    4. $\mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$

Correlación

- La correlación de dos variables aleatorias X y Y se define como:

$$\rho = \frac{E[(X-\mu_x)(Y-\mu_x)]}{\sigma_x\sigma_y} = \frac{Cov(X,Y)}{\sigma_x\sigma_y}$$

- Si X y Y son independientes su correlación es 0

- Las funciones de R para calcular covarianza son `cov` y para la correlación es `cor`

Referencias

Postgrado en Modelado y Simulación

- Jones et al. 2009. Capítulos 13, 14 y 15
- Rumsey, D. 2006. Probability for dummies. Wiley Publishing, Inc. Indianapolis
- Rizzo, 2008. Capítulo 2,
- Jeff Leek. Coursera Data Analysis Course
- Interesante: http://www.openintro.org/
- Ejercicios y material adicional: libros de Sheldon Ross (Introduction to Probability Models y Simulation)