Primera sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

26 de marzo de 2023

Experimentos aleatorios

La ciencia de la probabilidad se remonta al siglo XVII cuando surgió del estudio de los juegos de azar. En la actualidad se extiende a una amplia gama de problemas en ingeniería, administración, economía, etc.

Definición 1 (Experimento aleatorio, ε)

Es una actividad o procedimiento que produce posibilidades distintas y bien definidas llamadas resultados o sucesos. Es aleatorio si:

(i) Puede repetirse tantas veces como desee, y
(ii) produce resultados impredecibles.

Tipos de experimentos:

1. **Experimento simple.-** es aquel producido por una sola entidad, por ejemplo, un dado. Se refiere a una sola experiencia, por ejemplo, el dado es lanzado, o el dado es dejado caer, etc. En general es aquel experimento que no puede dividirse en otros experimentos.

Ejemplo 1

ε_1 : Designar un delegado de un grupo de 20.
ε_2 : Elegir un punto de $[0, 1]$.
ε_3 : Verificar el estado de un transmisor (0 ó 1).

2. **Experimento compuesto.-** está formado por experimentos simples que tienen lugar simultánea o consecutivamente.

Ejemplo 2

ε_4 : Se lanza una moneda o un dado.
ε_5 : Lanzar dos monedas (distintas) simultáneamente.
ε_6 : Lanzar dos monedas indénticas.

Observación 1

En los experimentos compuestos es conveniente usar un diagrama de árbol para determinar sus sucesos.

Espacio muestral y eventos

Definición 2 (Espacio muestral, Ω)

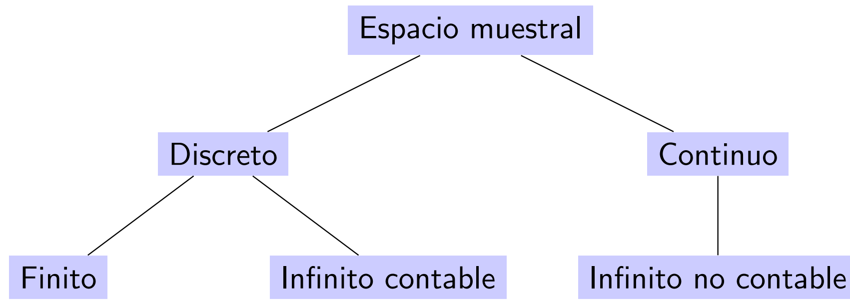
Es el conjunto de todos los resultados posibles de un experimento aleatorio.

$$\Omega := \{\omega \,:\, \omega \,\text{es un punto muestral}\}.$$

Un punto muestral o "suceso" es una observación, resultado o salida del experimento.

Definición 3 (Eventos)

Se dice que un conjunto, E, es un evento del espacio muestral Ω, si es un subconjunto de él, es decir $E \subseteq \Omega$.



Ejemplo 3

Halle los espacios muestrales de :

1. ε_1 : Lanzar una moneda al aire

$$\Omega_1 := \{C, S\}$$

2. ε_2 : Lanzar un dado al aire

$$\Omega_2 := \{1, 2, 3, 4, 5, 6\}$$

3. ε_3 : Elegir una carta de un mazo (4 palos: Espadas (E), Corazones (C), Diamantes (D) y Tréboles (T))

$$\Omega_3 := \Big\{1E, \cdots, 13E, 1C, \cdots, 13C, \\ 1D \cdots, 13D, 1T, \cdots, 13T\Big\}$$

Algebra de Eventos

El álgebra de eventos es crucial en la construcción de espacios de probabilidad.

Tipos de eventos

- Evento imposible: $E = \emptyset$.
- Evento seguro: $E = \Omega$.
- Evento unitario: $E = \{\omega\}$.

Ocurrencia de eventos

- E ocurre si el resultado o suceso es un elemento de E.
- E no ocurre si no se observa ninguno de sus sucesos.

Es decir, si $\omega \in \Omega$ es el resultado del experimento. Entonces

- $\omega \in E \Rightarrow E$ ocurre.
- $\omega \notin E \Rightarrow E$ no ocurre.

Sean E y F dos eventos de Ω.

- F es el evento de que E no ocurre:

$$F = E^c := \Big\{\omega \subset \Omega : \omega \notin E\Big\}.$$

- F ocurre siempre que E ocurra:

$$E \subset F \equiv \Big\{\omega \in E \Rightarrow \omega \in F\Big\}.$$

- Ocurre E, ó F, ó ambos:

$$E \cup F := \Big\{\omega \in \Omega : \omega \in E \vee \omega \in F\Big\}.$$

- Ocurren ambos eventos:

$$E \cap F := \Big\{\omega \in \Omega : \omega \in E \wedge \omega \in F\Big\}.$$

- F ocurre pero no E:

$$F - E = F \cap E^c := \Big\{\omega \in \Omega : \omega \in F \wedge \omega \notin E\Big\}.$$

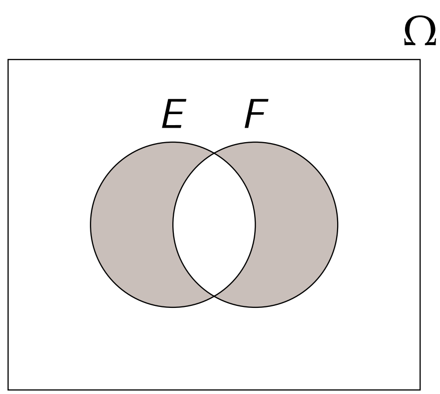
- Ocurre exactamente uno de los dos eventos:

$$E \triangle F := \Big\{\omega \in \Omega : \omega \in E \triangle \omega \in F\Big\}.$$

Disyunción excluyente

Diferencia simétrica ($\triangle$)

$$E \triangle F$$



(Diagramas de Venn - Euler)

Sean $E_1, E_2, \cdots, E_n$, n eventos de un espacio muestral Ω.

- Halle el evento de que ocurra al menos uno:

$$E = \bigcup_{i=1}^{n} E_i := \left\{\omega \in \Omega \,:\, \omega \in E_i \text{ para algún } i \in \{1, \cdots, n\}\right\}.$$

- Halle el evento en que todos ocurran:

$$E = \bigcap_{i=1}^{n} E_i := \left\{\omega \in \Omega \,:\, \omega \in E_i \text{ para todo } i \in \{1, \cdots, n\}\right\}.$$

- Halle el evento de que ninguno ocurra:

$$E = \bigcap_{i=1}^{n} E_i^c.$$

Definición 4

Sean $E_1, E_2, \cdots, E_n$, una familia de n eventos de un espacio muestral Ω.

- Se dice que los eventos son mutuamente excluyentes si

$$E_i \cap E_j = \emptyset, \quad i \neq j \text{ (Disjuntos dos a dos).}$$

  Es decir cada suceso pertenece a lo más a un evento.

- Se dice que son colectivamente exhaustivos si

$$E_1 \cup E_2 \cdots \cup E_n = \Omega.$$

  Es decir cada suceso pertenece al menos a un evento.

Definición 5

Se dice que una familia $E_1, E_2, \cdots, E_n$ de n eventos es una partición del espacio muestral si

$$\Omega = \bigcup_{i=1}^{n} E_i, \quad E_i \cap E_j = \emptyset, \quad i \neq j.$$

Es decir cada suceso pertenece a uno y sólo un evento de la partición.

Observación 2

Una partición de Ω es una familia colectivamente exhaustiva y mutuamente excluyente.

Definición 6 (Conjunto potencia, 2^Ω)

Dado un espacio muestral Ω, su conjunto potencia, 2^Ω, es la familia de subconjuntos de Ω más grande, es decir::

$$2^\Omega := \{E \,:\, E \subset \Omega\}.$$

Ejemplo 12

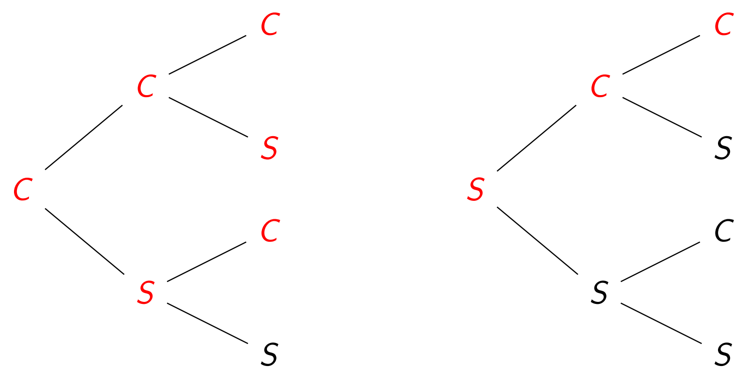
Se lanzan 3 monedas (distintas) al mismo tiempo. Sea A el evento de obtener al menos 2 caras, B el evento de no obtener cara y C el evento de obtener caras en la segunda moneda. ¿Cuál de éstos son mutuamente excluyentes?

Espacio muestral:

$$\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}.$$

Evento:

$A = \{CCC, CCS, CSC, SCC\}$, $B = \{SSS\}$, $C = \{CCC, CCS, SCC, SCS\}$



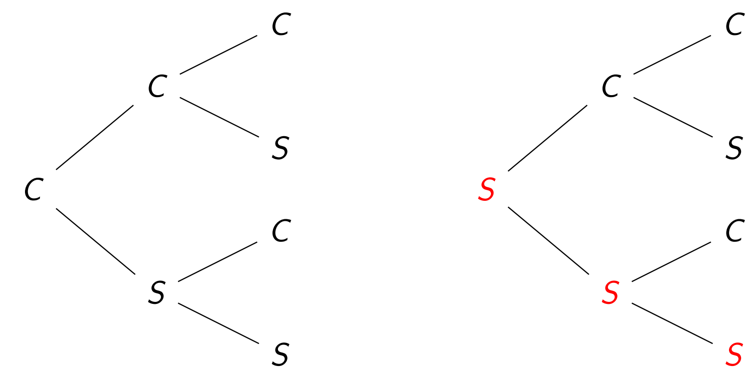
Conceptos básicos | Algebra de Eventos

Espacio muestral:

$$\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}.$$

Evento:

$A = \{CCC, CCS, CSC, SCC\}$, $B = \{SSS\}$, $C = \{CCC, CCS, SCC, SCS\}$



Los profesores (FC-UNI) | Primera sesión | 26 de marzo de 2023 | 17 / 39

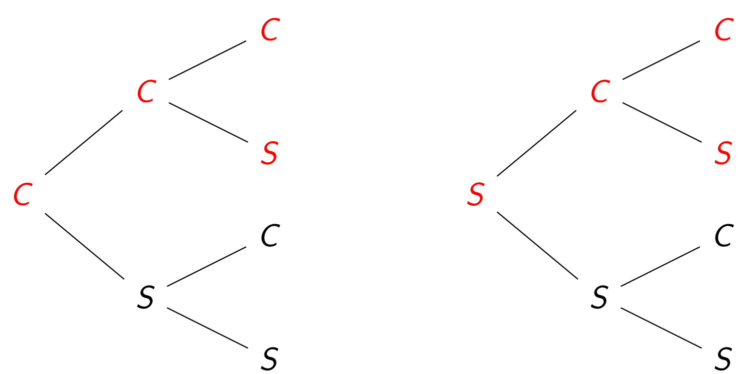
Conceptos básicos | Algebra de Eventos

Espacio muestral:

$$\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}.$$

Evento:

$A = \{CCC, CCS, CSC, SCC\}$, $B = \{SSS\}$, $C = \{CCC, CCS, SCC, SCS\}$



Los profesores (FC-UNI) | Primera sesión | 26 de marzo de 2023 | 17 / 39

Conceptos básicos | Algebra de Eventos

Ejemplo 12

Se lanzan 3 monedas (distintas) al mismo tiempo. Sea A el evento de obtener al menos 2 caras, B el evento de no obtener caras y C el evento de obtener cara en la segunda moneda. ¿Cuál de éstos son mutuamente excluyentes? ¿Son colectivamente exhaustivos?. Sino lo son, ¿cómo lo soluciona?

Solución

- A y B son mutuamente excluyentes.
- B y C son mutuamente excluyentes.
- No son colectivamente exhaustivos.
- Agregamos el evento:

$$D = \{CSS, SSC\},$$

luego A, B, C y D son colectivamente exhaustivos.

Propiedades de Eventos

- Leyes conmutativas:

$$E \cup F = F \cup E \quad , \quad E \cap F = F \cap E.$$

- Leyes asociativas:

$$(E \cup F) \cup G = E \cup (F \cup G) \quad , \quad (E \cap F) \cap G = E \cap (F \cap G).$$

- Leyes distributivas:

$$E \cap \left(\cup_{i=1}^{\infty} F_i \right) = \cup_{i=1}^{\infty} (E \cap F_i),$$

$$E \cup \left(\cap_{i=1}^{\infty} F_i \right) = \cap_{i=1}^{\infty} (E \cup F_i).$$

Propiedades de Eventos

- Leyes De Morgan:

$$\left(\bigcup_{i=1}^{\infty} E_i \right)^c = \bigcap_{i=1}^{\infty} E_i^c \quad , \quad \left(\bigcap_{i=1}^{\infty} E_i \right)^c = \bigcup_{i=1}^{\infty} E_i^c .$$

- Diferencia simétrica:

$$E \triangle F = (E - F) \cup (F - E) = (E \cup F) - (E \cap F).$$

- Conjunto potencia:

$$A \subset B \iff 2^A \subset 2^B.$$

Técnicas de conteo

Son métodos sofisticados para determinar, sin enumeración directa, el número de resultados posibles de un evento particular. Comunmente se denomina análisis combinatorio.

Principios básicos de conteo

Sean $E_1, E_2, \cdots, E_n$, eventos de experimentos aleatorios posiblemente diferentes. Se tiene los sgtes (principios):

1. **Multiplicación:** El número de sucesos de su producto cartesiano es

$$|E_1 \times \cdots \times E_n| = |E_1| \times \cdots \times |E_n|.$$

2. **Adición:** Si los eventos son mutuamente excluyentes, se tiene

$$|E_1 \cup \cdots \cup E_n| = |E_1| + \cdots + |E_n|.$$

Interpretación de estos principios

Sin pérdida de generalidad consideremos sólo dos eventos:

1. **Multiplicación:** Si un evento E_1 puede ocurrir de m formas y por cada una de estas, un segundo evento E_2 puede ocurrir de n formas, entonces la combinación de estos eventos ocurre de $m \times n$ formas.
2. **Adición:** Suponga que algún evento E_1 puede ocurrir en m formas y que un segundo evento E_2 puede ocurrir en n formas, pero ambos eventos no pueden ser simultáneos. Entonces $E_1 \cup E_2$ puede ocurrir de $m + n$ formas.

Los procedimientos utilizados en el análisis combinatorio para la elección y ordenación de los elementos de un conjunto son

- Permutaciones.
- Variaciones.
- Combinaciones.

Observación 3

El término "permutación" fue introducido por Bernoulli. Mientras que, combinación fue usado por Pascal y Leibniz. Y el último término en aparecer en la literatura fue "variación". Aquí diferenciaremos entre los términos permutación y variación, pero en general son usados indistintamente.

Permutaciones

Definición 7 (Permutación)

Una permutación es un arreglo ordenado (con o sin repetición) de los n elementos de un conjunto.

Teorema 1

El número de permutaciones de n objetos, sin repetición (distinguibles), es

$$n! := n(n-1)\cdots 2 \times 1.$$

Demostración

i) Consideramos una $n-$tupla ordenada cualquiera que contenga cada elemento del conjunto exactamente una vez.

ii) Para la 1ra posición se tiene n posibilidades,

Permutaciones

Demostración (cont...)

iii) para la 2da, $n-1$, debido a que se fijó un elemento en la 1ra.

iv) Así sucesivamente, se tendrá en la última posición un sólo elemento por elegir.

v) Luego por el principio de multiplicación se tiene $n!$ permutaciones.

Ejemplo 4

¿De cuántas formas pueden colocarse los 11 jugadores de un equipo de fútbol teniendo en cuenta que el portero no puede ocupar otra posición distinta que la portería?

Permutaciones

Teorema 2

El número de permutaciones de n objetos, donde se repiten (o son indistingibles), n_1, n_2, $\cdots$, n_r de los objetos, es

$$\frac{n!}{n_1!n_2!\cdots n_r!}$$

Demostración

i) Primero, observemos que permutar m ($m \leq n$) objetos indistinguibles produce un único arreglo.

ii) Segundo, recordamos que permutar n objetos distintos produce $n!$ arreglos.

Permutaciones

Demostración (cont...)

iii) Las permutaciones que contienen a n_i objetos indistinguibles, $i = 1, \cdots, r$, se cuentan como 1.

iv) Lo que se traduce en dividir nuestro número original de permutaciones, $n!$, por los $n_i!$ permutaciones que ahora se repiten.

Ejemplo 5

En una urna, hay 5 bolas del mismo tamaño y peso, de los cuales, 3 son rojas y 2 son azules. ¿De cuántas maneras se pueden extraer una a una las bolas de la urna?

Permutaciones

Teorema 3

El número de permutaciones (circulares) de n objetos distintos, donde no hay ni primer, ni último objeto, es

$$(n-1)!\ .$$

Demostración

i) Primero, fijamos un elemento inicial.

ii) Luego, permutamos los $n-1$ restantes, obteniendo $(n-1)!$ permutaciones.

iii) El elemento fijado solo se contará una vez en el arreglo, debido a que cualquiera de las posiciones que se le asigne será equivalente por rotaciones.

iv) Así, por principio de multiplicación se obtiene $(n-1)!$ permutaciones.

Combinaciones

Definición 8 (Combinación)

Las combinaciones son selecciones de objetos, con o sin repetición, donde el orden no importa.

Teorema 4

El número de combinaciones de k objetos de un grupo de n objetos, sin repetición (distinguibles), es el sgte coeficiente binomial:

$$\binom{n}{k} := \frac{n!}{k!(n-k)!}.$$

Combinaciones

Demostración

i) Consideramos un ordenamiento cualquiera de los n objetos y elegimos los k primeros.

ii) Si consideramos solo las permutaciones de los subconjuntos formados por los k primeros y los restantes $n-k$ obtenemos $k! \times (n-k)!$ permutaciones.

iii) Tenemos en total $n!$ permutaciones de las cuales podemos agrupar $k! \times (n-k)!$ permutaciones que representarán una elección de dos subconjuntos de k y $n-k$ elementos.

iv) Por tanto el número de subconjuntos se obtendrán dividiendo las $n!$ permutaciones por $k! \times (n-k)!$.

Combinaciones

Ejemplo 6

Halle las combinaciones de dos elementos del conjunto $A = \{a, b, c\}$.

Solución

Hallamos las permutaciones posibles y elegimos las dos primeras letras:

$$\begin{array}{ccc} a\,b \mid c & a\,c \mid b & b\,c \mid a \\ b\,a \mid c & c\,a \mid b & c\,b \mid a \end{array}$$

Entonces, las combinaciones seria el cociente

$$\frac{3!}{2!1!} = \frac{6}{2} = 3.$$

Combinaciones

Teorema 5

El número de combinaciones con repetición de n objetos, tomados de un conjunto de r elementos distintos es

$$\binom{n+r-1}{n} = \frac{(n+r-1)!}{r!(n-1)!}.$$

Es decir la elección de n objetos de r conjuntos o clases.

Demostración

i) Consideremos que deseamos etiquetar n objetos en r clases.

ii) Ordenamos linealmente los n objetos y como deseamos separarlos en r clases necesitaremos $r-1$ separadores.

iii) Tenemos $n+1$ posiciones (geométricas) para colocar los separadores de los objetos.

Cont...

iv) Como las elecciones de algún grupo puede ser vacía, existe la posibilidad de que haya separadores juntos. Por tanto, agregamos virtualmente $r-2$ posiciones más, lo que nos permitirá tener hasta $r-1$ separadores juntos. Obteniendo en total $n+r-1 = (n+1)+(r-2)$ posiciones.

v) Como no importa el orden como coloquemos los separadores en las posiciones se tiene: $\binom{n+r-1}{r-1}$ formas de colocar los separadores.

vi) Esto es equivalente a elegir n objetos de las r clases.

Combinaciones

Ejemplo 7

En una pastelería hay 5 tipos distintos de pasteles. ¿De cuántas formas se pueden elegir 3 pasteles?.

Solución

- Si nos gusta mucho un pastel lo podemos pedir hasta 3 veces.
- No nos importa el orden y podemos repetir.
- Entonces, estamos en el caso de combinaciones con repetición.
- Por tanto, el número de formas de elegir será

$$\binom{5+3-1}{3} = \frac{7!}{3!4!} = 35.$$

Variaciones

Definición 9 (Variación)

Las variaciones son arreglos de selecciones de objetos (con o sin repetición), donde el orden de los objetos seleccionados importa.

Teorema 6

El número de variaciones de k elementos tomados de n elementos, sin repetición, es

$$\frac{n!}{(n-k)!}.$$

Demostración

i) Primero hallamos las combinaciones de k en n elementos:

$$\binom{n}{k}.$$

Variaciones

Demostración (cont...)

ii) Luego, permutamos los elementos de cada combinación: $k!$

iii) Por último, aplicamos el principio de multiplicación, obteniendo

$$\underbrace{\binom{n}{k}}_{\text{combinaciones}} \times \underbrace{k!}_{\text{permutaciones}} = \underbrace{\frac{n!}{(n-k)!}}_{\text{variaciones}}.$$

Ejemplo 8

¿Cuántos números de tres cifras (todas distintas) se pueden formar con los números 1, 2, 3, 4, 5?

Variaciones

Teorema 7

El número de variaciones de k elementos tomados de n elementos distintos, con repetición, en el sentido de que los elementos elegidos pueden aparecer hasta k veces, es

$$n^k \ .$$

Demostración

i) Sobre cada arreglo de k elementos, consideramos que cada entrada tienes n posible elecciones. Luego, aplicamos principio de multiplicación para concluir.

Ejemplo 9

La población de una ciudad es de 20,000 habitantes. Si cada residente tiene tres iniciales, ¿es cierto que al menos dos personas tienen las mismas iniciales?

Solución

Hay $26^3 = 17576$ posibilidades de acuerdo con el alfabeto inglés, lo que significa que habrá al menos dos habitantes con las mismas iniciales ya que $17576 < 20{,}000$.

FIN

Segunda sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

3 de abril de 2023

Los modelos de equiprobabilidad (idea intuitiva) y frecuencia relativa motivan el siguiente modelo axiomático:

Definición 1 (Función de probabilidad)

Sea Ω un espacio muestral y $\mathcal{F}$ un espacio de eventos. Se dice que, $\mathbb{P} : \mathcal{F} \to [0, 1]$ es una función de probabilidad, si cumple:

Normalización: $\mathbb{P}(\Omega) = 1.$

No negatividad:

$$\mathbb{P}(A) \geq 0 \quad \forall A \in \mathcal{F}.$$

Aditividad: Para cada sucesión, $\{A_i\}_{i\in\mathbb{N}}$, de eventos mutuamente excluyentes de $\mathcal{F}$, se tiene que

$$\mathbb{P}\left(\biguplus_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

Observación 1

1. Si un evento tiene probabilidad cero, ese evento no puede ocurrir;
2. si tiene una probabilidad 1, entonces es seguro que ocurrirá.

Observación 2

Se puede obtener funciones de probabilidad al normalizar alguna medida de una magnitud, ejemplos:

1. Si el espacio muestral (Ω) es finito. Se puede medir sus eventos a través de la medida de conteo, $|\cdot|$. Luego su probabilidad correspondiente será:

$$\mathbb{P}(\cdot) := \frac{|\cdot|}{|\Omega|}.$$

Observación (cont...)

2. En el caso de eventos de $\mathbb{R}$, se puede usar una medida de longitud, *long*$(\cdot)$. Entonces, la función de probabilidad será:

$$\mathbb{P}(\cdot) := \frac{long(\cdot)}{long(\Omega)}.$$

3. En caso de eventos de $\mathbb{R}^2$, se puede usar una medida de area, *area*$(\cdot)$. Luego, se tiene que

$$\mathbb{P}(\cdot) := \frac{area(\cdot)}{area(\Omega)}.$$

4. Finalmente, para eventos de $\mathbb{R}^3$, se puede usar una medida de volumen, *vol*$(\cdot)$. Obteniendo,

$$\mathbb{P}(\cdot) := \frac{vol(\cdot)}{vol(\Omega)}.$$

Teorema 1

Dados $E, F \in \mathcal{F}$, se cumple las sgtes propiedades sobre $\mathbb{P}$:

(i) $\mathbb{P}(\varnothing) = 0$.

(ii) $E \cap F = \varnothing \Rightarrow \mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F)$.

(iii) $\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$.

(iv) $E \subset F \Rightarrow \mathbb{P}(E) \leq \mathbb{P}(F)$.

(v) $0 \leq \mathbb{P}(E) \leq 1$.

Demostración

(i) Aplicando el axioma de aditividad a la sucesión de eventos $A_1 = \Omega$, $A_i = \varnothing$, para $i > 1$, se tiene

$$\mathbb{P}(\Omega) = \mathbb{P}(\Omega) + \sum_{i>1} \mathbb{P}(\varnothing) \Rightarrow \sum_{i>1} \mathbb{P}(\varnothing) = 0.$$

Luego, necesariamente $\mathbb{P}(\varnothing) = 0$.

Demostración (cont...)

(ii) Dada la sucesión de eventos $A_1 = E$, $A_2 = F$, $A_i = \varnothing$, para $i > 2$, aplicando nuevamente el axioma de aditividad,

$$\begin{aligned}\mathbb{P}(E \cup F) &= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \underbrace{\sum_{i>2} \mathbb{P}(\varnothing)}_{0} \\ &= \mathbb{P}(E) + \mathbb{P}(F),\end{aligned}$$

donde hemos concluido, usando (i).

(iii) Del hecho que $\Omega = E \cup E^c$, usando el axioma de normalización y (ii), obtenemos que

$$1 = \mathbb{P}(E) + \mathbb{P}(E^c),$$

de lo cual se concluye inmediatamente.

Demostración (cont...)

(iv) De los axiomas de aditividad y no negatividad, y del hecho $F = E \cup (F - E)$, se tiene que

$$\mathbb{P}(F) = \mathbb{P}(E) + \underbrace{\mathbb{P}(F - E)}_{\geq 0} \geq \mathbb{P}(E).$$

(iv) Se sigue del hecho que $E \subset \Omega$, de (iv), y del axioma de no negatividad.

Propiedades importantes de la probabilidad

Proposición 1 (Aditividad finita)

Para cada secuencia $A_1, \ldots, A_n$ de eventos de $\mathcal{F}$, mutuamente excluyentes. Se cumple que

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = \mathbb{P}(A_1) + \cdots \mathbb{P}(A_n).$$

Proposición 2 (Principio de inclusión-exclusión)

Dados dos eventos cualesquiera A y B de $\mathcal{F}$, se tiene que

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

Demostración

Dado que $A \cup B = A \cup (B - A)$ y $B = (A \cap B) \cup (B - A)$, se tiene por el axioma de aditividad:

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B - A), \tag{1}$$

$$\mathbb{P}(B) = \mathbb{P}(A \cap B) + \mathbb{P}(B - A). \tag{2}$$

Luego, se concluye de (1) y (2).

Proposición 3

El principio de inclusión-exclusión se puede generalizar como

$$\mathbb{P}\left(\bigcup_{l=1}^{n} A_l\right) = \sum_{l=1}^{n} (-1)^{l+1} \alpha_l,$$

Proposición 3 (Cont...)

donde los α_l son las sumas de las probabilidades de todas las posibles intersecciones de l eventos tomados de los n.

$$\begin{aligned}
\alpha_1 &= \sum_{i=1}^{n} \mathbb{P}(A_i), \\
\alpha_2 &= \sum_{1 \leq i < j \leq n} \mathbb{P}(A_i \cap A_j), \\
\alpha_3 &= \sum_{1 \leq i < j < k \leq n} \mathbb{P}(A_i \cap A_j \cap A_k), \\
&\vdots \\
\alpha_n &= \mathbb{P}(A_1 \cap \cdots \cap A_n).
\end{aligned}$$

Proposición 4 (Continuidad)

Si $\{A_n\}_{n \in \mathbb{N}}$ es una sucesión monótona de eventos de $\mathcal{F}$, es decir:

i) *Si $A_1 \subset A_2 \subset \cdots$, entonces $\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \text{lím}_{n\to\infty} \mathbb{P}(A_n)$.*

ii) *Si $A_1 \supset A_2 \supset \cdots$, entonces $\mathbb{P}(\bigcap_{n=1}^{\infty} A_n) = \text{lím}_{n\to\infty} \mathbb{P}(A_n)$.*

Demostración

Verificaremos solo (i), para lo cual consideremos la nueva sucesión de eventos mutuamente excluyentes: $B_1 = A_1$, $B_j = A_j - A_{j-1}$ si $j > 1$.

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{j=1}^{\infty} \mathbb{P}(B_j) = \lim_{n\to\infty} \sum_{j=1}^{n} \mathbb{P}(B_j) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{j=1}^{n} B_j\right) = \lim_{n\to\infty} \mathbb{P}(A_n),
\end{aligned}$$

donde hemos aplicado el axioma de aditividad.

Ejemplo 1

Sea el experimento aleatorio de lanzar una moneda. Sean los eventos:

1. H_1 : sale cara en el primer lanzamiento,
2. H_2 : sale cara en el segundo lanzamiento.

Halle $\mathbb{P}(H_1 \cup H_2)$.

Solución

Si suponemos que son igualmente probable los sucesos, entonces

$$\mathbb{P}(H_1 \cup H_2) = \mathbb{P}(H_1) + \mathbb{P}(H_2) - \mathbb{P}(H_1 \cap H_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}.$$

Definición 2 (Probabilidad condicional)

Dado un evento $B \subset \Omega$ con $P(B) > 0$. Se define para todo evento A, la probabilidad condicional de A, dado que B ha ocurrido, como

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

.

Propiedades

Si $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$, tenemos

- Si $B \subset A$ entonces $\mathbb{P}(A|B) = 1$.
- Si $A \cap B = \emptyset$ entonces $\mathbb{P}(A|B) = 0$.
- $A \subset B$ entonces $\mathbb{P}(A|B) = \mathbb{P}(A)/\mathbb{P}(B)$.

Propiedades de la probabilidad condicional

Observación 3

Dado $B \subset \Omega$ fijo arbitrario tal que $P(B) > 0$. La probabilidad condicional $\mathbb{P}(\cdot|B)$, puede ser vista como una función de probabilidad:

$$\mathbb{P}_B(E) = \mathbb{P}(E|B) \quad \forall B \in \mathcal{F}.$$

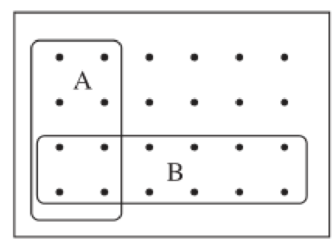
Todas las propiedades e intuiciones que se aplican a la función probabilidad se aplican a $\mathbb{P}_B$ también, es decir

- $\mathbb{P}(\Omega|B) = 1$.
- $\mathbb{P}(A|B) \geq 0$,
- Si $\{A_i\}_{i=1}^{\infty}$ son mutuamente excluyentes, entonces $\mathbb{P}(\bigcup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i|B)$.

Ejemplo explicativo

Ejemplo 2

De la figura se observa que si B ocurre entonces solo ocurren 4 puntos muestrales del evento A. Por tanto, la intuición nos dice que $\mathbb{P}(A|B) = \dfrac{4}{12}$, donde hemos consideramos a B como un nuevo espacio muestral. (Suponemos equiprobabilidad)



Ley de multiplicación

Si $\mathbb{P}(A) > 0$ y $\mathbb{P}(B) > 0$, tenemos que

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B|A) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Podemos generalizar este resultado de la probabilidad condicional a múltiples eventos. Sean $A_1, A_2, \cdots, A_k$ eventos para los cuales $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_k) > 0$. Entonces

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots A_k) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}).$$

Ejemplo contraintuitivo Feller 1968

Sean dos familias con dos hijos, donde la probabilidad de género de cada niño es $1/2$. Si M representa masculino y F femenino y ordenamos los niños de mayor a menor. Se tiene que el espacio muestral de seleccionar al azar una familia es $\Omega = \{MM, MF, FM, FF\}$. Sean además los eventos:

- A : "Ambos hijos en la familia son niños".
- B : "La familia tiene un niño".
- C : "El primer hijo es un niño".

Halle (suponga equiprobabilidad)

1) La probabilidad de que ambos hijos sean niños sabiendo que el primero fue niño.
2) La probabilidad de que ambos hijos sean niños suponiendo que hay un niño en la familia.

Solución

Los eventos son $A = \{MM\}$, $B = \{MF, FM, MM\}$ y $C = \{MF, MM\}$.

1) Dado que el primer hijo es un niño, la probabilidad de que ambos hijos sean niños es $\mathbb{P}(A|C) = \mathbb{P}(A \cap C)/\mathbb{P}(C) = \mathbb{P}(A)/\mathbb{P}(C) = (1/4)/(1/2) = 1/2$.
2) Dado que la familia tiene un niño, la probabilidad de que ambos hijos sean niños es

$$\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B) = \mathbb{P}(A)/\mathbb{P}(B) = (1/4)/(3/4) = 1/3.$$

Observamos que 1) coincide con la intuición de que obtener 2 hijos del mismo género es $1/2$, sin embargo 2) parece contradecir la intuición arrojando un valor menor.

Independencia

Definición 3

Dados dos eventos A y B con probabilidades no nulas. Se dice que A y B son independientes si cumplen alguna de las sgtes afirmaciones equivalentes:

i) $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.
ii) $\mathbb{P}(A|B) = \mathbb{P}(A)$.
iii) $\mathbb{P}(B|A) = \mathbb{P}(B)$.

Observación 4

Si la probabilidad de un evento es nula, entonces este es independiente de cualquier otro evento.

Definición 4

Un número finito de eventos $A_1, A_2, \ldots, A_n$ son independientes si

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdots \mathbb{P}(A_n)$$

y son independientes por pares si cada par $A_i, A_j, i \neq j$ son independientes.

Ejemplo 3

Una moneda es lanzada 4 veces. Consideremos los eventos: A donde el primer lanzamiento muestra cara; B el tercer lanzamiento muestra sello; y C hay un número igual de caras y sellos. ¿Son estos eventos independientes dos a dos?. ¿Son $A, B,$ y C eventos independientes ?.

Ley de la probabilidad total

Sea A un evento, entonces se sabe que la intersección de A y el evento universal Ω es A. Se sabe además que B y su complemento B^c constituye una partición. Así

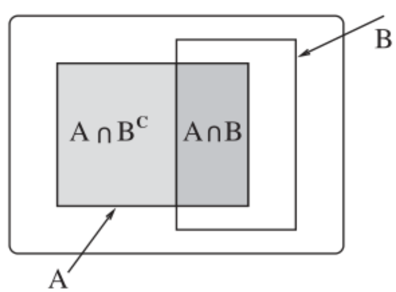
$$A = A \cap \Omega \quad \text{y} \quad B \cup B^c = \Omega$$

Sustituyendo el segundo resultado en el primero en la ecuación anterior y aplicando la Ley de Morgan, tenemos

$$A = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c).$$

Ley de la probabilidad total(1)

Los eventos $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y de acuerdo al siguiente gráfico,



Se muestra que B y B^c no pueden tener salidas en común, la intersección de A y B no puede tener salidas en común con la intersección de A y B^c,

Ley de la probabilidad total(2)

Usando el hecho que $(A \cap B)$ y $(A \cap B^c)$ son mutualmente exclusivos y por propiedad de la función probabilidad, se tiene

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c).$$

Esto significa que para evaluar la probabilidad de un evento A, es suficiente encontrar las probabilidades de la intersección de A y B y A y B^c y sumarlos. En general sean n eventos $B_i, i = 1, 2, \ldots, n$, una partición del espacio muestral Ω. Entonces para algún evento A, podemos escribir

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i), \quad n \geq 1$$

Esta es la <u>ley de la probabilidad total</u>.

Ley de la probabilidad total(3)

En efecto, los conjuntos $A \cap B_i, i = 1, 2, \ldots, n$ son mutualmente exclusivos (desde que los B_i lo son) y el hecho que los $B_i, i = 1, 2, \ldots$ forman una partición de Ω implica que

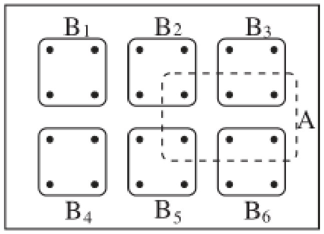
$$A = \bigcup_{i=1}^{n} A \cap B_i, \ \ n \geq 1,$$

y por el axioma de aditividad, se tiene

$$\mathbb{P}(A) = \mathbb{P}\left\{\bigcup_{i=1}^{n} A \cap B_i\right\} = \sum_{i=1}^{n} \mathbb{P}(A \cap B_i).$$

Ejemplo explicativo

La siguiente figura muestra una partición de un espacio muestral, que contine 24 puntos muestrales equiprobables, en seis eventos $B_1, \cdots, B_6$. Halle la probabilidad de A usando la información parcial de los B_i's.



Demostración

Debido a que los eventos B_i constituyen una partición, cada punto de A está en uno y sólo uno de los eventos B_i y la probabilidad del evento A se pueden encontrar sumando las probabilidades de los eventos $A \cap B_i$ para $i = 1, 2, \ldots, 6$, es decir

$$\mathbb{P}(A \cap B_1) = \mathbb{P}(A \cap B_4) = 0,$$
$$\mathbb{P}(A \cap B_2) = \mathbb{P}(A \cap B_5) = \frac{1}{24},$$
$$\mathbb{P}(A \cap B_3) = \mathbb{P}(A \cap B_6) = \frac{2}{24}.$$

Por el Teorema de probabilidad total, se tiene

$$\mathbb{P}(A) = \sum_{i=1}^{6} \mathbb{P}(A \cap B_i) = \frac{1}{24} \times 2 + \frac{2}{24} \times 2 = \frac{1}{4}.$$

Uso de árboles para organizar el cálculo

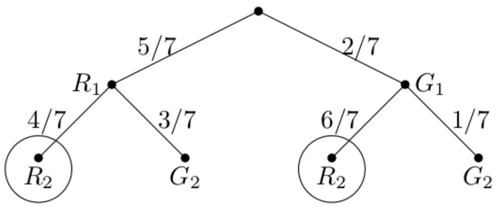
Los árboles son útiles para organizar el proceso en una secuencia de acciones donde se involucra cálculos con probabilidad condicional, la regla del producto y la ley de probabilidad total.

Ejemplo 4

Una urna contiene 5 pelotas rojas y 2 pelotas grises. Se saca una pelota, si es gris, se agrega una pelota roja a la urna y si es roja, se agrega una pelota gris a la urna. (La pelota original no se devuelve a la urna.) Luego se saca una segunda pelota. ¿Cuál es la probabilidad de que la segunda pelota sea roja?.

Solución

La secuencia de acciones es: primero se saca la pelota 1 (y añade la bola apropiada a la urna) y luego se saca la pelota 2, según se muestra en el siguiente diagrama de árbol:



i) En las ramas del primer nivel se colocan las probabilidades. Por ejemplo, la probabilidad de sacar R_1 (rojo en la primera extracción) que es $5/7$, se coloca sobre la rama que va del nodo raiz a R_1.

Solución (cont...)

ii) En el siguiente nivel ponemos probabilidades condicionales. Por ejemplo, la probabilidad a lo largo de la rama de R_1 a R_2 es $\mathbb{P}(R_2|R_1) = 4/7$ que representa la probabilidad de ir al nodo R_2 dado que ya se está en R_1.

iii) La regla de multiplicación nos dice que la probabilidad de llegar a cualquier nodo es sólo el producto de las probabilidades a lo largo su ruta. Por ejemplo, el nodo etiquetado por R_2, representa realmente el evento $R_1 \cap R_2$ porque proviene del nodo R_1, es decir

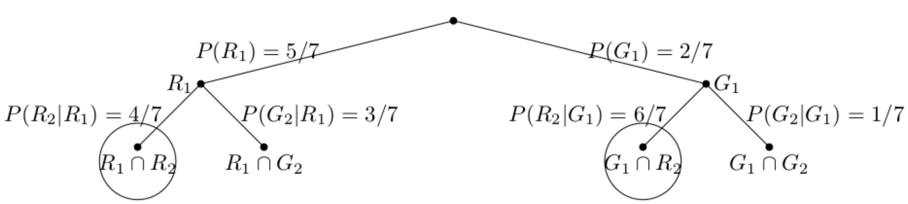
$$\mathbb{P}(R_1 \cap R_2) = \mathbb{P}(R_1)\mathbb{P}(R_2|R_1) = \frac{5}{7} \cdot \frac{4}{7}.$$

Solución (cont...)

iv) La ley de probabilidad total es simplemente la afirmación de que $\mathbb{P}(R_2)$ es la suma de las probabilidades de todos las rutas que conducen a R_2 (los dos nodos en círculo). Esto es,

$$\mathbb{P}(R_2) = \frac{5}{7} \times \frac{4}{7} + \frac{2}{7} \times \frac{6}{7} = \frac{32}{49}.$$

Aquí está el mismo árbol etiquetado con precisión:



El Teorema de Bayes

El teorema de Bayes es un pilar de la probabilidad y estadística y es central para el resto de este curso y otros cursos más avanzados.

Lema 1

Dados dos eventos A y B con probabilidades no nulas, se tiene que

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)\mathbb{P}(A|B)}{\mathbb{P}(A)}.$$

Demostración

Basta utilizar la regla de multiplicación

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Teorema 2 (Regla de probabilidad de Bayes)

Sean $B_1, B_2, \cdots, B_k$ una partición del espacio muestral Ω, entonces para todo $A \subset \Omega$ con $\mathbb{P}(A) > 0$, se tiene

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(B_j)\mathbb{P}(A|B_j)}{\mathbb{P}(A)} \quad \forall j = 1, \cdots, k,$$

con $\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(B_i)\mathbb{P}(A|B_i)$.

- La regla de Bayes nos dice cómo invertir las probabilidades condicionales, es decir, encontrar $\mathbb{P}(B_j|A)$ de $\mathbb{P}(A|B_j)$.
- A menudo $\mathbb{P}(A)$ se calcula utilizando la ley de probabilidad total.

A las probabilidades $\mathbb{P}(B_j)$ se les llama probabilidades a priori y luego de conocer la ocurrencia de la información A, es decir $\mathbb{P}(A)$, se obtiene las probabilidades condicionales $\mathbb{P}(B_j|A)$ que se les llama probabilidades a posteriori o actualizadas.

Ejemplo 5

Considere una prueba de rutina de detección de una enfermedad. Supongamos que la frecuencia de la enfermedad en la población es del $0,5\,\%$ (tasa de base). La prueba es muy precisa con un $5\,\%$ de tasa de falsos positivos y un $10\,\%$ de tasa de falsos negativos. Si se toma la prueba y sale positivo. ¿Cuál es la probabilidad de que se tenga la enfermedad?

Solución

Sean los eventos:

- H = 'tiene la enfermedad'.
- $-H$ = 'no tiene la enfermedad'.
- E = 'prueba positiva'
- $-E$ = 'prueba negativa'.

Solución (cont...)

Se nos da $\mathbb{P}(H) = 0{,}5\,\%$ y por lo tanto $\mathbb{P}(-H) = 99{,}5\,\%$. Las tasas falsas positivas y falsas negativas son probabilidades condicionales:

$$\mathbb{P}(\text{falso positivo}) = \mathbb{P}(E|-H) = 5\,\%$$
$$\mathbb{P}(\text{falso negativo}) = \mathbb{P}(-E|H) = 10\,\%$$

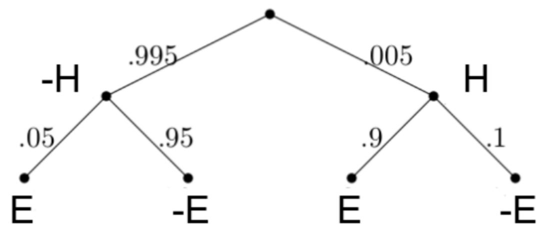
Las probabilidades complementarias se conocen como las tasas negativas y positivas verdaderas:

$$\mathbb{P}(-E|-H) = 1 - \mathbb{P}(E|-H) = 95\,\%$$
$$\mathbb{P}(E|H) = 1 - \mathbb{P}(-E|H) = 90\,\%$$

Nos piden la probabilidad de que una persona tenga la enfermedad dado que dió positivo, es decir, $\mathbb{P}(H|E) = ?$.

Solución (cont...)

Árboles: Las probabilidades se pueden mostrar en un diagrama de árbol:



Por el teorema de Bayes, se tiene

$$\mathbb{P}(H|E) = \frac{\mathbb{P}(E|H) \cdot \mathbb{P}(H)}{\mathbb{P}(E)}.$$

Sólo nos falta hallar $\mathbb{P}(E)$.

Solución (cont...)

Luego, por la regla de probabilidad total y del diagrama de árbol sólo bastaría sumar las probabilidades para cada uno de los nodos marcados con E, es decir

$$\mathbb{P}(E) = ,995 \times ,05 + ,005 \times ,9 = ,05425$$

Así,

$$\mathbb{P}(H|E) = \frac{,9 \times ,005}{0,5425} = 0,082949 \approx 8,3\,\%.$$

Esto se llama la falacia de la tasa de base porque la tasa de base de la enfermedad en la población es tan baja que la gran mayoría de las personas que se toman la prueba son saludables, e incluso con una prueba precisa la mayoría de los positivos serán personas sanas.

FIN

Tercera sesión

Cálculo de Probabilidades - CM1H2

Jonathan Munguia[1] José Zamudio[2]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

13 de Abril de 2023

Outline

1. Variables aleatorias
   - Definición de Variable aleatoria
   - Variable aleatoria discreta
   - Principales variables aleatorias

En esta sección nos enfocamos en cuantificar ciertas características de los sucesos de los espacios muestrales, las cuales serán consideradas como variables del experimento.

Definición 1

Dado un experimento aleatorio ε y asociado al mismo, un espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$. Se define una variable aleatoria (v.a.), como una función real $X : \Omega \to \mathbb{R}$ t.q.

$$[X \leq x] := \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad \forall x \in \mathbb{R}.$$

Observación 1

Se puede definir a partir de una v.a X una función de probabilidad. Para lo cual se necesita:

- Espacio muestral: $\Omega_X = R_X$.
- Espacio de eventos: $\mathcal{G}$ generado por los intervalos $(-\infty, x]\, \forall x \in \mathbb{R}$. (llamado $\sigma-$álgebra de Borel)

Luego, se define la función de probabilidad generada por X,
$\mathbb{P}_X : \mathcal{G} \to [0, 1]$ t.q.

$$\mathbb{P}_X(A) := \mathbb{P}[X \in A] \quad \forall A \in \mathcal{G}.$$

Definición 2 (Función de Distribución)

Se define la Función de Distribución (acumulada), denotada por F_X, de una v.a. X, como, $F_X : \mathbb{R} \to [0, +\infty)$ t.q.

$$F_X(x) := \mathbb{P}[X \leq x] \quad \forall x \in \mathbb{R}.$$

Outline

1. Variables aleatorias
   - Definición de Variable aleatoria
   - Variable aleatoria discreta
   - Principales variables aleatorias

Definición 3 (V.a Discreta)

Se dice que X es discreta si su rango, R_X, es finito o enumerable.

Definición 4

Se define la función de masa (cuantía o de frencuencia) respecto a una v.a Discreta X, como $p : R_X \to [0, 1]$ t.q.

$$p(x) := \mathbb{P}[X = x] \quad \forall x \in R_X.$$

Observación 2

Se expresa la función de probabilida generada por X, como

$$\mathbb{P}_X(A) = \sum_{x \in A} p(x) \quad \forall A \in R_X$$

Ejemplo 1

Dado el experimento ε : "Lanzar 3 monedas distintas al mismo tiempo". Halle la función de masa respecto a la variable aleatoria;

$$X : \text{"Número de caras obtenidas al lanzar las 3 monedas"}$$

Solución

1) Espacio muestral:

$$\Omega = \{CCC, CCS, CSC, CSS, SCC, SCS, SSC, SSS\}.$$

2) Espacio de probabilidad: $(\Omega, 2^\Omega, \mathbb{P})$, donde $\mathbb{P}$ es obtenida de la medida de conteo.

Solución

3) Calculamos las imágenes:

$$\begin{aligned} &X(CCC) = 3, \\ &X(CCS) = X(CSC) = X(SCC) = 2, \\ &X(CSS) = X(SCS) = X(SSC) = 1, \\ &X(SSS) = 0\,. \end{aligned}$$

4) A través de X se define el nuevo espacio muestral:

$$\Omega_X = \{0, 1, 2, 3\},$$

con frecuencias absolutas:

$$\begin{array}{lcccc} x: & 0 & 1 & 2 & 3 \\ p_x: & 1 & 3 & 3 & 1. \end{array}$$

Solución

5) Por último, la función de masa se define como la frecuencia relativa:

$$p(x) := \frac{p_x}{|\Omega|}, \quad x \in \Omega_X.$$

Observación 3

La función de masa generaliza la idea de probabilidad por frecuencia utilizada para variables aleatorias de rango finito.

Observación 4

De la definición de probabilidad, se deduce que

$$\sum_{x \in R_X} p(x) = 1$$

Ejemplo 2

Lanzamos una moneda dos veces y sea X el número de caras obtenidas. Hallar la función de distribución (acumulada).

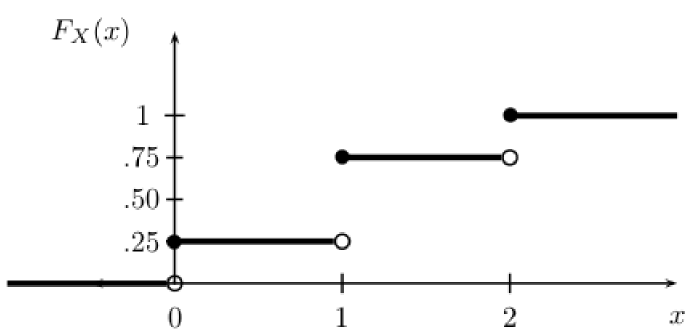
Solución

Como $\mathbb{P}(X=0) = \mathbb{P}(X=2) = 1/4$ y $\mathbb{P}(X=1) = 1/2$. La función distribución acumulada es

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1/4 & 0 \leq x < 1 \\ 3/4 & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$

Solución (cont...)

Se nota que F_X es continua por la derecha, no decreciente y es definida para todo $x \in \mathbb{R}$.



Propiedades de F_X

1. $0 \leq F_X(x) \leq 1$ para todo $x \in \mathbb{R}$.
2. Se cumple que $\lim_{x \to -\infty} F_X(x) = 0$ y $\lim_{x \to +\infty} F_X(x) = 1$.
3. Si la variable aleatoria X tiene una imagen finita, entonces
   - $F_X(x) = 0$ para todo x suficientemente pequeño.
   - $F_X(x) = 1$ para todo x suficientemente grande.
4. F_X es no decreciente:

$$F_X(x_2) \leq F_X(x_1) \quad \text{para} \quad x_1 \leq x_2.$$

5. F_X es `continua por la derecha`.

Introducimos a continuación la idea intuitiva del valor medio esperado.

Ejemplo 3

Sean

- Experimento ε : "Lanzar 20 veces 3 monedas".
- Variable aleatoria X : "Números de caras obtenidas en un lanzamiento".
- Rango: $R_X := \{0, 1, 2, 3\}$.
- Frecuencias absolutas para cada $x \in R_X$:

x :	0	1	2	3
p_x :	4	6	7	3.

Calcule el valor promedio de caras obtenidas por lanzamiento.

solución

- El valor promedio es la media ponderada:

$$\bar{x} := \frac{0 \times 4 + 1 \times 6 + 2 \times 7 + 3 \times 3}{20} = 1{,}45$$

- Obtenemos las frecuencias relativas: $p(x) = \frac{p_x}{20}$. Así valor medio se escribe como

$$\bar{x} = \sum_{x \in R_X} x\, p(x)\,.$$

- De la probabilidad por frecuencias, se deduce que cuando un experimento se realiza muchas veces, se obtiene aproximadamente para cada suceso la probabilidad teórica (a continuación).

solución

- La probabilidad teórica coincide con la función de masa:

$$p(0) = \frac{1}{8} = p(3)\ ,\ p(1) = \frac{3}{8} = p(2).$$

- Luego, el valor promedio esperado (a la larga) es

$$\mu = 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = 1{,}5.$$

- El valor medio experimental se acerca al valor medio esperado.

Definición 5 (Esperanza)

Se define la esperanza de una v.a X, como el valor medio esperado (teórico):

$$\mu = \mathbb{E}(X) := \sum_{x \in R_X} x\, p(x)\,.$$

Definición 6 (Varianza)

La varianza de una v.a X, se define como:

$$\sigma^2 = Var(X) := \sum_{x \in R_X} (x - \mu)^2\, p(x)\,.$$

Observación 5

La varianza representa la media de las desviaciones cuadráticas con respecto a la media esperada.

Proposición 1

Dado un espacio de probabilidad $(\Omega, 2^{\Omega}, \mathbb{P})$ y X una v.a discreta. La esperanza de X se puede hallar directamente sobre los puntos muestrales:

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega)\, \mathbb{P}(\{\omega\}).$$

Teorema 1

Sea X una v.a discreta con función de masa de probabilidad p_X y $g : D_g \subset \mathbb{R} \to \mathbb{R}$ tal que $R_X \subset D_g$. La función de masa de probabilidad de la v.a $Y = g \circ X$ se obtiene como

$$p_Y(y) = \sum_{\{x \in R_X : g(x) = y\}} p_X(x) \quad \forall y \in R_Y.$$

Además, la esperanza de Y se puede hallar como

$$\mathbb{E}(Y) = \mathbb{E}[g(X)] = \sum_{x \in R_X} g(x)\, p_X(x).$$

Demostración

Ver Proposición 4.1 de [1].

cont...

Sea $Y = g(X)$. Se tiene que

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_y y\mathbb{P}[Y = y] = \sum_y y\mathbb{P}[g(X) = y] \\
&= \sum_y y \sum_{\substack{x \\ g(x)=y}} \mathbb{P}[X = x] \\
&= \sum_y \sum_{\substack{x \\ g(x)=y}} y\mathbb{P}[X = x] \\
&= \sum_x \sum_{\substack{y \\ g(x)=y}} y\mathbb{P}[X = x],
\end{aligned}$$

donde hemos cambiado el orden de las sumatorias y debido a que hay un solo y, obtenemos

$$\mathbb{E}(Y) = \sum_x g(x)\mathbb{P}[X = x].$$

Ejemplo 4

Dada la v.a X, la función real g y la v.a inducida $Y := g \circ X$, verifique los resultados anteriores, según la sgte. tabla.

Ω	X	g
w_1	x_1	y_1
w_2		
w_3	x_2	
w_4	x_3	y_2
w_5		
w_6	x_4	

Solución

a) La esperanza expresada en función de sus puntos muestrales:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=1}^{4} x_i\, \mathbb{P}[X = x_i] \\
&= x_1\mathbb{P}(w_1, w_2) + x_2\mathbb{P}(w_3) + x_3\mathbb{P}(w_4, w_5) + x_4\mathbb{P}(w_6) \\
&= x_1\mathbb{P}(w_1) + x_1\mathbb{P}(w_2) + x_2\mathbb{P}(w_3) \\
&+ x_3\mathbb{P}(w_4) + x_3\mathbb{P}(w_5) + x_4\mathbb{P}(w_6) \\
&= \sum_{i=1}^{6} X(w_i)\mathbb{P}(w_i).
\end{aligned}$$

Solución (cont...)

a) La esperanza de la composición $Y = g \circ X$:

$$\begin{aligned}
\mathbb{E}[Y] &= y_1\mathbb{P}[Y = y_1] + y_2\mathbb{P}[Y = y_2] \\
&= y_1\mathbb{P}[\{w_1, w_2, w_3\}] + y_2\mathbb{P}[\{w_4, w_5, w_6\}] \\
&= y_1\Big(\mathbb{P}[\{w_1, w_2\}] + \mathbb{P}[\{w_3\}]\Big) + y_2\Big(\mathbb{P}[\{w_4, w_5\}] + \mathbb{P}[\{w_6\}]\Big) \\
&= y_1\mathbb{P}[X = x_1] + y_1\mathbb{P}[X = x_2] + y_2\mathbb{P}[X = x_3] + y_2\mathbb{P}[X = x_4] \\
&= \sum_{i=1}^{4} g(x_i)\mathbb{P}[X = x_i].
\end{aligned}$$

Outline

Distribución de masa puntual δ

Se dice que una v.a X tiene distribución de masa puntual en a, $X \sim \delta_a$, si es una variable aleatoria constante. Se tiene función de masa igual a

$$p_X(x) = \begin{cases} 1 & x = a \\ 0 & x \neq a. \end{cases}$$

Y función de distribución

$$F_X(x) = \begin{cases} 1 & x \geq a \\ 0 & x < a. \end{cases}$$

Distribución uniforme discreta

Una v.a X tiene distribución uniforme discreta, si dado $R_X := \{x_1, \cdots, x_n\}$, con $n \in \mathbb{N}$. Su función de masa es de la forma:

$$p_X(x_i) = \begin{cases} \frac{1}{n} & \text{para } i = 1, \ldots, n\,, \\ 0 & \text{en otros casos.} \end{cases}$$

Definición 7

Se dice que un experimento aleatorio es de Bernoulli si su espacio muestral es

$$\Omega := \{E, F\}.$$

Es decir sólo hay dos sucesos posibles, que denotaremos por:

$$E : \text{exito} \quad , \quad F : \text{fracaso}.$$

Distribución de Bernoulli

Se dice que una v.a X tiene distribución de Bernoulli, $X \sim Be(p)$, con probabilidad de éxito p. Si

$$X(E) = 1 \quad , \quad X(F) = 0.$$

Y

$$\mathbb{P}(X = 1) = p \quad \wedge \quad \mathbb{P}(X = 0) = 1 - p,$$

para algún $p \in (0, 1)$

Distribución de Bernoulli (cont...)

i) Función de masa:

$$p_X(x) = p^x(1-p)^{1-x}\,, \quad x \in \{0, 1\}.$$

ii) La esperanza:

$$\mu = \mathbb{E}(X) = 0 \times (1-p) + 1 \times p = p.$$

iii) La varianza:

$$\begin{aligned} \sigma^2 = Var\,(X) &= (0-p)^2(1-p) + (1-p)^2 \times p \\ &= p(1-p)[p + 1 - p] = p(1-p). \end{aligned}$$

Ejemplo 5

Determine la función de masa del experimento de obtener 6 al lanzar un dado.

Solución

E : "obtener un 6",
F : "obtener un número distinto de 6".
La función de masa es:

$$p_X(x) = \left(\frac{1}{6}\right)^x \left(\frac{5}{6}\right)^{1-x}.$$

Distribución binomial

Una v.a X tiene distribución Binomial, $X \sim \text{Bi}(n,p)$, si se verifica que

i) Consiste de n experimentos de Bernoulli $\varepsilon_1, \cdots, \varepsilon_n$.
ii) Los experimentos son independientes.
iii) La probabilidad de éxito, p, es constante en cada ensayo.
iv) Espacio muestral:

$$\Omega = \{\omega = (w_1, \cdots, w_n) \, : \, w_i = E \vee F\}$$

v) Variable aleatoria:

$$\begin{aligned} X : \quad & \Omega \to \mathbb{R} \\ & \omega \to X(\omega) = N^\circ \, \text{de exitos} \end{aligned}$$

Distribución binomial (cont...)

vi) Dado $w \in [X = x]$ cualesquiere

$$\mathbb{P}(\{w\}) = \prod_{i=1}^{n} \mathbb{P}(\{w_i\}) = p^x (1-p)^{n-x}$$

vii) El cardinal de $[X = x]$ se basa en dividir el conjunto de n elementos:
- x repeticiones de E,
- $n - x$ repeticiones de F,

es decir $\binom{n}{x}$.

viii) Luego,

$$\begin{aligned} \mathbb{P}[X = x] &= \sum_{w} \mathbb{P}(\{w\}) \\ &= |[X = x]| \cdot \mathbb{P}(\{w\}) \\ &= \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned}$$

Distribución binomial (cont...)

ix) La función de masa es

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{para } x = 0, \ldots, n \\ 0 & \text{en otros casos.} \end{cases}$$

Ejemplo 6

Si se lanza un dado 10 veces. Calcular la probabilidad de obtener 3 veces 6.

Solución

Considerando los lanzamientos independientes, se tiene que $X \sim Bi(10, 1/6)$, luego la función de masa en 3 es

$$\mathbb{P}[X = 3] = \binom{10}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^7.$$

Definición 8 (Geométrica)

- Experimento geométrico.
  ε : Repetir un experimento de Bernoulli hasta que ocurra un primer éxito.
- Espacio muestral.
  $\Omega = \{E, FE, FFE, \cdots\}$.
- Variable aleatoria.
  X : N° de ensayos de Bernoulli hasta que resulte el 1er éxito.
- Rango: $R_X = \{1, 2, \cdots\}$.
- $X \sim Ge(p)$.

- Función de masa, usando la independencia.

$$p_X(x) = (1-p)^{x-1}p \quad \forall x \in R_X.$$

- Momentos:

$$\mathbb{E}(X) = \frac{1}{p} \quad \wedge \quad Var(X) = \frac{1-p}{p^2}.$$

Ejemplo 7

Se dispone de una selladora de botellas de vidrio. Esta se utiliza hasta que se detecta la primera botella mal sellada. Se sabe que la probabilidad de que una botella sea mal sellada es de $1/9$ y la probabilidad de que la botella sea sellada correctamente es $8/9$. Sea X la v.a. que nos da el número de botellas selladas hasta que aparezca el primer sello defectuoso. Determine la distribución de probabilidad de X.

Solución

X :"# de botellas selladas hasta aparecer el primer sello defectuoso"
El rango $R_X = \{1, 2, 3, \cdots\}$, la probabilidad de fracaso $\mathbb{P}(F) = \frac{8}{9}$.
Luego, la distribución es geométrica con función de masa:

$$p(x) = \mathbb{P}[X = x] = \left(\frac{8}{9}\right)^{x-1}\left(\frac{1}{9}\right) \quad \forall x \in R_X.$$

Referencias bibliográficas

1. Sheldon Ross. A First Course in Probability, Pearson, 9th ed. 2014.

FIN

Cuarta sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

4 de septiembre de 2021

Momentos

Para una variable aleatoria X (discreta o continua), la esperanza $\mathbb{E}(X)$ indica el `centro` de la distribución de X. Este es el primer elemento de una secuencia de números conteniendo información acerca de la distribución de X. Esta secuencia de números $\mathbb{E}(X), \mathbb{E}(X^2), \mathbb{E}(X^3), \ldots$ es llamada `momentos` de X.

Definición 1

Sea $k \geq 1$. El `k-ésimo momento` de la variable aleatoria X es el valor $\mathbb{E}(X^k)$ y lo denotaremos por μ'_k.

Ejemplo 1

El momento más representativo es la esparanza de una v.a. X, la cual usualmente se denota por $\mu := \mathbb{E}(X)$ y se denota por μ_k.

Proposición 1

Dados $X, Y : \Omega \to \mathbb{R}$ variables aleatorias y $c \in \mathbb{R}$. Se tiene que $\mathbb{E}(X + cY) = \mathbb{E}(X) + c\mathbb{E}(Y)$, siempre que las esperanzas sean finitas.

Demostración

Por la Proposición 1 de la 5ta sesión, se tiene

$$\begin{aligned}
\mathbb{E}(X + cY) &= \sum_{w \in \Omega} (X(\omega) + cY(\omega))\mathbb{P}(\omega) \\
&= \sum_{w \in \Omega} X(\omega)\mathbb{P}(\omega) + c \sum_{w \in \Omega} Y(\omega)\mathbb{P}(\omega) \\
&= \mathbb{E}(X) + c\mathbb{E}(Y).
\end{aligned}$$

Definición 2

El momento central de orden k de una variable aleatoria X se define como $\mathbb{E}[(X - \mu)^k]$ y denota por μ_k.

Ejemplo 2

Uno de los momentos centrales más importante es la varianza de X, la cual gracias al Teorema 1 de la 5ta sesión, puede expresarse como

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

Luego, por la Proposición 1, se tiene

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}(X^2 - 2X\mathbb{E}(X) + \mathbb{E}(X)^2) \\
&= \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + \mathbb{E}(X)^2 \\
&= \mathbb{E}(X^2) - \mathbb{E}(X)^2.
\end{aligned}$$

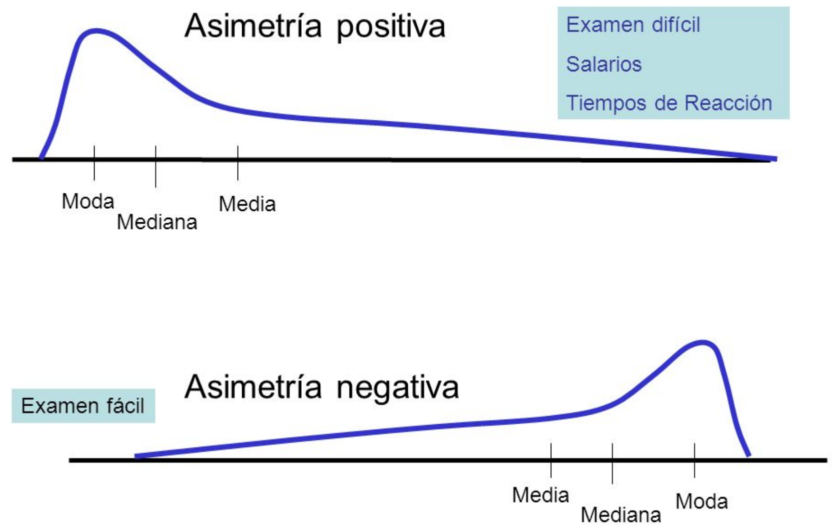
Ejemplo 3

Otras caracterísitica importantes de una variable aleatoria relacionadas con momentos centrales son:

- Coeficiente de asimetría:

$$\alpha = \frac{\mu_3}{\sigma^3}$$

  Se dice que la distribución es sesgada a la derecha si $\alpha > 0$ y sesgada a la izquierda si $\alpha < 0$.



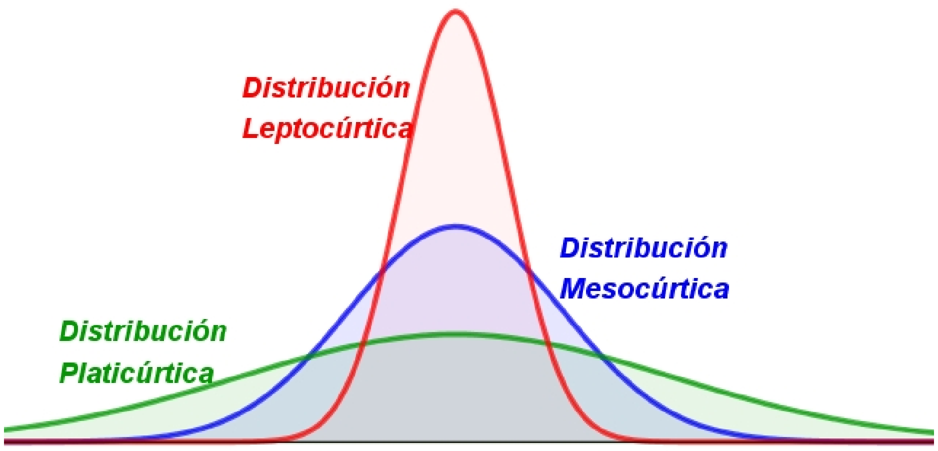
Ejemplo 4 (cont...)

- Coeficiente de curtosis:

$$k = \frac{\mu_4}{\sigma^4}$$

  A menudo en estadística se redefine este coeficiente como $\hat{k} = k - c$, donde c es la curtosis de la normal estándar (distribución de tipo continuo, se verá en próximas sesiones)

  a) Leptocúrtica: $\hat{k} > 0$
  b) Mesocúrtica: $\hat{k} = 0$
  c) Platicúrtica: $\hat{k} < 0$



Definición 3 (Función generadora de Probabilidad)

Dada una v.a discreta no negativa $X \geq 0$, se define la Función generadora de probabilidad de X, denotada por G_X, como

$$G_X(s) := \mathbb{E}[s^X] = \sum_{x \in R_X} s^x p_X(x),$$

donde s^X es una v.a. tal que $s^X(\omega) := s^{X(\omega)}$ para todo $\omega \in \Omega$.

Teorema 1

Sea X una v.a discreta no negativa con una función generadora de probabilidad G_X. Luego, se tiene que las derivadas de orden r de G_X en $s = 1$, se hallan como

$$G_X^{(r)}(1) = \mathbb{E}(X[X-1]\cdots[X-r+1]).$$

Ejemplo 4

De la proposición anterior, se puede calcular los sgtes momentos de X, como

1. $\mathbb{E}(X) = G_X'(1)$.
2. $\mathbb{E}(X^2) = \mathbb{E}(X[X-1]+X) = \mathbb{E}(X[X-1]) + \mathbb{E}(X) = G_X''(1) + G_X'(1)$.
3. $\mathrm{Var}(X) = G_X''(1) + G_X'(1) - G_X'(1)^2$.

Ejercicio 1

Probar que, $p(x) = \binom{n}{x} p^x (1-x)^{n-x}$ es una función de masa. Hallar la media y la varianza de $X \sim Bi(n,p)$ usando la función generadora de probabilidad.

De manera intuitiva dos variables aleatorias son independientes si los valores que toma una no afecta los valores de la otra.

Definición 4 (Independencia)

Dadas $X, Y : \Omega \subset \mathbb{R} \to \mathbb{R}$ variables aleatorias discretas. Se dice que son independientes si los eventos inducidos por ellas son independientes, es decir $\forall (x,y) \in R_X \times R_Y$:

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

Observación 1

Cuando se trate de vectores aleatorios, se demostrará: Si X, Y son independientes, entonces

$$\mathbb{E}[XY] = \mathbb{E}(X)\mathbb{E}(Y).$$

Proposición 2

Dados $X, Y : \Omega \to \mathbb{R}$ v.a discretas independientes y $c \in \mathbb{R}$. Se tiene que $Var(X + cY) = Var(X) + c^2 Var(Y)$.

Demostración

De la linealidad de la esperanza, la Observación 1 y reordenando, se tiene

$$\begin{aligned} Var(X + cY) &= \mathbb{E}[(X + cY)^2] - \mathbb{E}[X + cY]^2 \\ &= Var(X) + c^2 Var(Y). \end{aligned}$$

Ejemplo 5

Si $X \sim Bi(n,p)$, entonces existen $X_1, \cdots X_n : \Omega \to \mathbb{R}$ v.a de Bernoulli independientes con $X_i \sim Be(p)$ para $i = 1, \cdots, n$, tal que $X = X_1 + \cdots + X_n$. Luego, se tiene que
$\mathbb{E}(X) = \mathbb{E}(X_1) + \cdots + \mathbb{E}(X_n) = np$,
$Var(X) = Var(X_1) + \cdots + Var(X_n) = np(1-p)$.

Teorema 2

La función de masa de una distribución discreta no negativa X se puede determinar de manera única de su función generadora de probabilidad (si existe):

$$p_X(n) := \mathbb{P}[X = n] = \frac{1}{n!} G_X^{(n)}(0) \quad \textit{para } n = 0, 1, 2, \cdots$$

Ejemplo 6

Sea $X \geq 0$ v.a. discreta con $G_X(s) = \frac{s}{5}(2 + 3s^2)$. Encontrar la distribución de X.

Muchos de los problemas en Probabilidad, están relacionados con la suma de variables aleatorias. En este aspecto las funciones generadoras de momentos son un herramienta importante.

Proposición 3

Si $X, Y : \Omega \subset \mathbb{R} \to \mathbb{R}$ son v.a. discretas independientes, entonces la función generadora de probabilidad de la suma es:

$$G_{X+Y}(s) = G_X(s) G_Y(s)$$

Ejercicio 2

a) Calcula la función generadora de probabilidad de una variable aleatoria con distribución de Poisson de parámetro λ:

$$p(k) = e^{-\lambda} \frac{\lambda^k}{k!} \quad , \; k \geq 0.$$

b) Utiliza la biyección entre funciones generadoras de probabilidad y y ditribuciones discretas no negativas para probar que la suma de dos variables aleatorias independientes con distribución de Poisson es a su vez una variable con distribución de Poisson de parámetro igual a la suma de los parámetros de las dos variables.

Definición 5 (Hipergeométrica)

- ε : Extraer una muestra aleatoria de tamaño n sin reposición de una población de N elementos.
- El conjunto de N elementos se divide en dos subconjuntos de r y $N - r$ elementos. La extracción de un elemento se clasifica como
  - Grupo de r : ÉXITO,
  - Grupo de $N - r$: FRACASO.
- X : N° de éxitos en la muestra de tamaño n.
- Rango: $R_X = \{0, 1, \cdots, \min(n, r)\}$.
- $X \sim H(N, n, r)$.

- Función de masa.

$$p(x) := \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}, \quad x \in R_X := \{0, 1, 2, \cdots, n\}.$$

- Momentos:

$$\mathbb{E}(X) = np \quad \wedge \quad Var(X) = np(1-p)\frac{N-n}{N-1},$$

donde p representa una probabilidad de éxito: $p = r/N$.

Aproximación de la Hipergeométrica a la Binomial

a) Cuando se considera N muy grande en comparación a n, las extracciones se pueden ver como extracciones con reposición, debido a que en cada extracción se puede considerar N casi invariante.

b) Así la probabilidad de obtener un elemento del conjunto de r es constante e igual a r/N.

c) Luego, las extracciones de n elementos se podrán hacer de manera independiente y nuestro experimento Hipergeométrico de reducirá a un experimento Binomial con probabilidad de éxito $p = r/N$ en una muestra de tamaño n.

Aproximando la función de masa

Dado $x \in R_X$ y N muy grande. Se puede asumir $N \gg n$, y aproximar

$$\frac{N!}{(N-n)!} = N(N-1)(N-2)\cdots(N-n+1) \approx N^n.$$

Lo que implica que, $\binom{N}{n} \approx \frac{N^n}{n!}$.

De la identidad $p = r/N$, se deduce que $r = pN$ y $N - r = (1-p)N$, luego r y $N - r$ también son muy grandes. Así, del mismo modo se deduce que

$$\binom{N-r}{n-x} \approx \frac{(N-r)^{n-x}}{(n-x)!} \quad \text{y} \quad \binom{r}{x} \approx \frac{r^x}{x!}.$$

Luego, reemplazando estas aproximaciones, se obtiene la siguiente aproximación binomial:

$$\begin{aligned} p(x) &\approx \frac{\frac{r^x}{x!} \times \frac{(N-r)^{n-x}}{(n-x)!}}{\frac{N^n}{n!}} = \frac{n!}{(n-x)!\,x!} \times \frac{r^x(N-r)^{n-x}}{N^n} \\ &= \binom{n}{x} \times \left(\frac{r}{N}\right)^x \times \left(1 - \frac{r}{N}\right)^{n-x} = \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned}$$

Definición 6 (Binomial negativa)

Replicar un experimento de `Bernoulli(p)` de forma independiente hasta que ocurre el r-ésimo éxito $(r \geq 1)$ Sea X el número de pruebas en el cual se produce el r-ésimo éxito. Entonces se dice que X tiene una distribución `binomial negativa(r,p)`. Denotada por $X \sim BN(r,p)$.

Observación 2

Existe otra definición de distribución binomial negativa, a saber, la distribución del número de fallos antes del r-ésimo éxito. (Esta es la definición empleada por el lenguaje de programación estadístico **R**, en los cálculos).
La relación entre las dos definiciones es bastante simple. Si X es binomial negativo en el sentido usual y F es binomial negativo en el sentido de **R**, entonces:

$$F = X - r$$

¿Cuál es la probabilidad de que el r-ésimo éxito ocurre en la prueba t, para $t \geq r$?. Para que esto suceda, debe haber $t-r$ fracasos y $r-1$ éxitos, en las primeras $t-1$ pruebas, con un éxito en la prueba t.
Por independencia, esto sucede con la probabilidad binomial para $r-1$ éxitos en las primeras $t-1$ pruebas, por la probabilidad p de éxito en la pruena t:

$$\mathbb{P}(X = t) = \binom{t-1}{r-1} p^r (1-p)^{t-r} \quad (t \geq r).$$

de esta manera la probabilidad es cero, cuando $t < r$. El caso especial $r = 1$ es llamada `distribución geométrica`. Es decir, X tiene una distribución geométrica con parámetro $p \in (0,1)$, denotada como $X \sim \text{Geometrica}(p)$, si

$$\mathbb{P}(X = k) = p(1-p)^{k-1}, \quad k \geq 1.$$

Tenemos a partir de este resultado que

$$\sum_{k=1}^{\infty} \mathbb{P}(X = k) = p \sum_{k=1}^{\infty} (1-p)^k = \frac{p}{1-(1-p)} = 1.$$

X es el número de lanzamientos necesarios hasta que la primera cara salga, cuando una moneda es lanzada. La media de una distribución geométrica, tiene el valor de:

$$\mathbb{E}(X) = \sum_{t=1}^{\infty} tp(1-p)^{t-1} = \frac{1}{p}.$$

Además que se tiene:

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{1-p}{p^2}.$$

La distribución `binomial negativa(r,p)` es la suma de r de variables aleatorias geométricas independientes `Geometrica(p)`. Como la esperanza es un operador lineal positivo, se concluye que la media de `binomial negativa(r,p)` es r veces la media que `Geometrica(p)` y la varianza de una suma independiente es la suma de varianza, así: Si $X \sim$ `binomial negativa(r,p)`, entonces:

$$\mathbb{E}(X) = \frac{r}{p} \quad \text{y} \quad \text{Var}(X) = \frac{r(1-p)}{p^2}.$$

Definición 7

Distribución de Rademacher Si contamos el resultado k como `éxito`, entonces es obvio que cada X_k es simplemente una variable aleatoria *binomial*(n, p_k). Pero los componentes no son independientes, ya que suman a n.

La distribución `Rademacher(p)` es una variación de la distribución de Bernoulli, 1 todavía indica éxito, pero el fracaso se codifica como -1. Si Y es una variable aleatoria `Bernoulli(p)`, entonces $X = 2Y - 1$ es una variable aleatoria `Rademacher(p)`. La función de masa de probabilidad es:

$$\mathbb{P}(X = x) = \begin{cases} p & x = 1 \\ 1 - p & x = -1 \end{cases}$$

Una variable aleatoria X `Rademacher(p)` tiene una media:

$$\mathbb{E}(X) = \sum_{x=-1,1} xp(X = x) = -1(1 - p) + 1p = 2p - 1.$$

Además $X^2 = 1$, así:

$$\mathbb{E}(X^2) = 1,$$

por tanto la varianza es:

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = 1 - (2p - 1)^2 = 4p(1 - p).$$

Definición 8 (Camino aleatorio)

Una secuencia de sucesivas sumas de variables aleatoria independientes de `Rademacher(p)` es llamado **camino aleatorio**.

Esto es, si X_i son idénticamente distribuidas a variables aleatorias `Rademacher(1/2)`, la secuencia $S_1, S_2, \ldots$ es un camino aleatorio, donde:

$$S_n = X_1 + \cdots X_n.$$

Desde que el operador esperanza es un operador lineal:

$$\mathbb{E}(S_n) = 0, \quad \text{para todo} \quad n,$$

y desde que la varianza de la suma de variables aleatorias independientes es la suma de varianzas:

$$\mathrm{Var}(S_n) = n, \quad \text{para todo} \quad n.$$

FIN

Quinta sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

4 de septiembre de 2021

Outline

"Una ley de probabilidad que es capaz de explicar que la ocurrencia de los accidentes no es tan accidental como parece"

Origen de e

El origen de e se remonta a un curioso problema financiero:

- Invertimos $S/.1$ en un banco durante un año. Recibiendo $S/.1$ como interés al final del año. Es decir, a una tasa de interés del 100 %.
- En vez de cobrar el interés al año, lo cobramos en dos partes cada 6 meses, a una tasa de 50 %, obteniendo un monto de

$$\left(1+\frac{1}{2}\right)^2 = 2{,}25.$$

- Si repetimos el proceso de reinvertir n veces, a intervalos regulares, con la tasa proporcional, obtenemos la sucesión creciente y acotada:

$$x_n = \left(1+\frac{1}{n}\right)^n.$$

Origen de e

- Si reinvertimos los interes recibidos de forma continua (interés compuesto), recibiremos una cantidad finita, denotada por e,

$$e := \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n .$$

Proposición 1

Para todo $x \in \mathbb{R}$, se tiene

$$e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n . \tag{1}$$

Solución

Dado $x \in \mathbb{R}$, tomando $h = x/n$, y la continuidad de la función exponencial, se tiene

$$\begin{aligned}
\lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n &= \lim_{n\to\infty}\left(\exp\left(n\ln\left(1+\frac{x}{n}\right)\right)\right)\\
&= \lim_{n\to\infty}\left(\exp\left(x\frac{\ln\left(1+\frac{x}{n}\right)}{\frac{x}{n}}\right)\right)\\
&= \lim_{h\to0}\exp\left(x\frac{\ln(1+h)}{h}\right)\\
&= \exp\left(x\lim_{h\to0}\left(\frac{\ln(1+h)-\ln(1)}{h}\right)\right),
\end{aligned}$$

se concluye, del hecho que $\frac{d}{dx}(\ln x) = \lim_{h\to0}\left(\frac{\ln(x+h)-\ln(x)}{h}\right)$.

Definición 1 (Distribución de Poisson)

- ε : Consiste en observar las veces que ocurre un evento en particular en un determinado periodo de tiempo, $I := [a, b]$.
- Caracterísiticas del evento:
  - i) ocurre de manera aleatoria,
  - ii) tiene una duración relativamente breve (puntual) comparada con el período de tiempo de observación, I (por ejemplo, un año).
  - iii) las repeticiones de este evento son independientes y suceden con una frecuencia promedio uniforme pequeña, denota por λ.

  Cuando el evento cumpla con (i), (ii), y (iii), lo llamaremos evento de Poisson (evento "raro").

Definición 1 (cont...)

- N : número de eventos independiente de Poisson que ocurren en un intervalo de tiempo I.
- $N \sim \mathcal{P}(\lambda)$.
- Rango: $R_N := \{0, 1, 2, \cdots\}$.
- Función de masa:

$$p_N(k) := \mathbb{P}(N = k) = e^{-\lambda}\frac{\lambda^k}{k!} \quad k = 0, 1, \ldots .$$

- Momentos:

$$\mathbb{E}(N) = \lambda \;\text{ y }\; \mathrm{Var}(N) = \lambda.$$

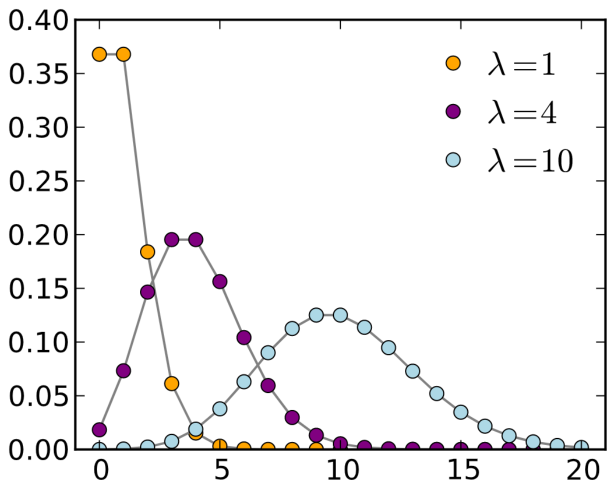


Figura: Función de masa

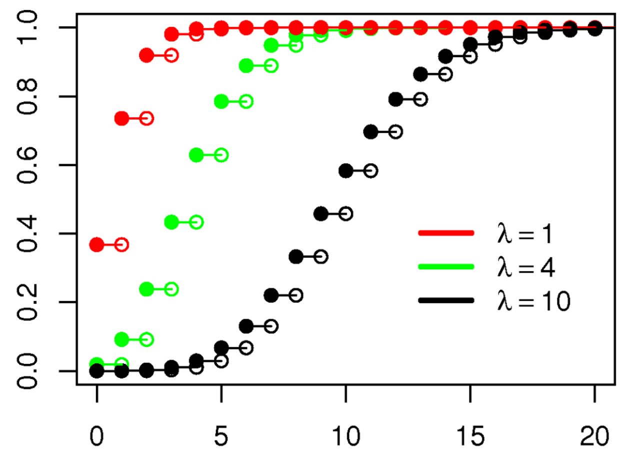


Figura: Función de distribución

Observación 1

- Abraham De Moivre (1718): introdujo por 1ra vez la distribución de Poisson como el límite de una distribución binomial, pero sin dar un prueba contundente.
- Simeon Denis Poisson (1837): derivó de manera formal dicha distribución directamente como una aproximación a la distribución binomial negativa.
- Ladislaus von Bortkiewicz (1898): fue el primero en descubrir la aplicación de la distribución de Poisson. La enunció como la Ley de los "Pequeños Números". Esta se aplica a sucesos con una baja frecuencia en una población grande. Ver [1].

Observación 2

- El parámetro λ nos indica el número de eventos de Poisson que se espera observar en una unidad de tiempo.
- En general, se puede buscar eventos de Poisson en una unidad de medida de cualquier magnitud física (no necesariamente tiempo). Por ejemplo, longitud, superficie, volumen, temperatura, etc.
- Si buscamos eventos de Poisson en t unidades de medida, entonces el parámetro promedio de ocurrencias será λt. Luego, la función de masa es

$$p_N(k) := \mathbb{P}(N = k) = e^{-\lambda t}\frac{(\lambda t)^k}{k!} \quad k = 0, 1, \ldots .$$

Ejemplo 1

La distribución de Poisson es usada en diferentes modelos, de los cuales se dese contabilizar ciertos eventos raros, por ejemplo

- Número de accidentes automovilísticos en una carrera de fórmula 1.
- Número de desintegraciones radiactivas que se observan durante 1h.
- Número de manchas en un metro cuadrado.
- Número de vehículos que llegan a una estación durante una hora.
- Número de llamadas telefónicas que llegan a un teleoperador en un turno de $4h$.
- Número de bacterias en $1cm^3$.

Ejemplo 2

Suponga que llegan en forma aleatoria una serie de llamadas a una central telefónica con un promedio de tres llamadas por minuto. Calcular la probabilidad de que en el periodo de un minuto:

a) No ocurra llamada alguna.

b) Ocurran al menos cuatro llamadas.

Solución

Sea X el número de llamadas que ocurren en el periodo de un minuto. $X \sim \mathcal{P}(\lambda)$, donde $\lambda = 3$ es el promedio de llamadas por minuto.

a) La probabilidad de que ocurran k llamadas en el periodo de un minuto:

$$\mathbb{P}[X = k] = \frac{e^{-\lambda}(\lambda)^k}{k!} = \frac{e^{-3}3^k}{k!} \ , \ k = 0, 1, 2, \ldots$$

Reemplazando para $k = 0$, se obtiene: $\mathbb{P}[X = 0] = 0{,}0498$

b) $\mathbb{P}[X \geq 4] = 1 - \mathbb{P}[X \leq 3] = 1 - \sum_{k=0}^{3} \mathbb{P}[X = k] = 0{,}35277$

Aproximación de Poisson a la Binomial

Una v.a. de Poisson $N \sim \mathcal{P}(\lambda)$ puede aproximar el número de éxitos de una v.a. Binomial $X \sim Bi(n, p)$, cuando la probabilidad de éxitos p es pequeña, y el número de pruebas (independientes) n es grande, tal que su esperanza o tasa de éxito promedio $np = \lambda$ sea pequeña. Véase la proposición siguiente:

Proposición 2

Dada una familia de v.a. $X_n \sim Bi(n, p_n)$ tal que $np_n = \lambda \in \mathbb{R}$, donde λ es una constante suficientemente pequeña. Entonces, existe una v.a $N \sim \mathcal{P}(\lambda)$, que cumple:

$$\forall k \in \{0, 1, 2 \cdots, [\![\lambda]\!] + 1\} : \quad \lim_{n \to \infty} \mathbb{P}[X_n = k] = \mathbb{P}[N = k].$$

Solución

- Como λ es pequeño, entonces solo tiene sentido estudiar valores de k cercanos a λ. Luego, se tiene las aproximaciones

$$\lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} = 1. \tag{2}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^k = 1. \tag{3}$$

$$\lim_{n\to\infty} \left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}. \tag{4}$$

- Escribimos la función de masa Binomial como

$$\mathbb{P}[X_n = k] = \frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

Solución (...)

$$\begin{aligned}\mathbb{P}[X_n = k] &= \frac{n!}{(n-k)!k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}\\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \times \frac{\lambda^k}{k!} \times \frac{(1-\lambda/n)^n}{(1-\lambda/n)^k},\end{aligned}$$

luego de (2), (3), y (4), y tomando límite, se concluye

$$\lim_{n\to\infty} \mathbb{P}[X_n = k] = e^{-\lambda}\frac{\lambda^k}{k!}.$$

Observación 3

En la práctica se considera una buena aproximación de Poisson a la Binomial si

- $p < 0{,}1$ y $np \leq 5$.
- Ó $p \leq 0{,}05$ y $n \geq 20$.

Ejemplo 3

Una fábrica pone galletas en cajas de 100. La probabilidad de que una galleta se rompa es 0,03. Encuentre la probabilidad de que una caja contenga 2 galletas rotas.

Solución

- Variable aleatoria:
  X : número de galletas rotas.
- X es una distribución binomial con $n = 100$ y $p = 0{,}03$, es decir $X \sim Bi(100, 0{,}03)$.
- Estos valores están fuera del rango de las tablas estadísticas e implican largos cálculos.

Solución (cont ...)

- Usamos la aproximación de Poisson ya que en el problema, se tiene que $p = 0{,}03 < 0{,}1$ y $np = 100x0{,}03 = 3 < 5$, lo cuál verifica la Observación 3.
- Aproximación: $X \sim \mathcal{P}(\lambda)$, donde $\lambda = np = 3$.
- La probabilidad de que una caja contenga dos galletas rotas es

$$\mathbb{P}[X = 2] = e^{-3}\frac{3^2}{2!} = 0{,}224.$$

Proposición 3 (Definición por series)

Para todo $x \in \mathbb{R}$, se tiene que

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

- Por la Proposición 3, se verifica que p_N es una función de masa:

$$\sum_{k\geq 0} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k\geq 0}\frac{\lambda^k}{k!} = e^{-\lambda} \times e^{\lambda} = 1.$$

- Hallamos la esperanza:

$$\mathbb{E}(N) = \sum_{k\geq 0} k e^{-\lambda}\frac{\lambda^k}{k!} = \lambda \sum_{k\geq 1} e^{-\lambda}\frac{\lambda^{k-1}}{(k-1)!},$$

haciendo $j = k - 1$, y usando la propiedad de la función de masa, p_N, se tiene

$$\mathbb{E}(N) = \lambda \sum_{j\geq 0} e^{-\lambda}\frac{\lambda^j}{(j)!} = \lambda \times 1 = \lambda.$$

Ejercicio 1

Halle la función generadora de $X \sim \mathcal{P}(\lambda)$, deduzca su esperanza y varianza.

Solución

Hallamos la función generadora de probabilidad:

$$\begin{aligned} G_X(s) = \mathbb{E}[s^X] &= \sum_{k=0}^{\infty} s^k \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(s\lambda)^k}{k!} = e^{-\lambda}e^{\lambda s} \\ &= e^{-\lambda(1-s)}. \end{aligned}$$

Luego, $G'(s) = \lambda e^{-\lambda(1-s)}$, $G''(s) = \lambda^2 e^{-\lambda(1-s)}$, entonces se tiene $G'(1) = \lambda$ y $G''(1) = \lambda^2$.

Solución (cont...)

Obtenemos finalmente $\mathbb{E}(X) = \lambda$ y como $G_X''(1) = \mathbb{E}(X(X-1)) = \mathbb{E}(X^2) - \mathbb{E}(X)$, se obtiene

$$\begin{aligned} Var(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2 \\ &= G''(1) + \mathbb{E}(X) - \mathbb{E}(X)^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

Referencias bibliográficas

1. `https://www.antirion.cl/entry-11/`

FIN

Sexta sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

14 de octubre de 2021

Outline

Existen v.a. cuyo conjunto de valores es incontable, por ejemplo:

- El tiempo en que un tren llega a una parada.
- La vida útil de un transmisor, etc.

Definición 1 (Función de densidad)

Dada X una v.a con valores incontables, diremos que es CONTINUA si existe una función no negativa $f_X : \mathbb{R} \to \mathbb{R}$, a la cual llamaremos función de densidad, tal que:

$$\forall\, B \subset \mathbb{R}, \quad \mathbb{P}[X \in B] = \int_B f_X(t)\, dt.$$

Observación 1

Para calificar a f_X como una función de densidad debe verificar:

- No negatividad: $f_X \geq 0$,
- Normalización: $\int_{\mathbb{R}} f_X(t)\, dt = 1$.

Observación 2

La probabilidad de que X tome valores en un intervalo, $I = [a, b]$, es

$$\mathbb{P}[a \leq X \leq b] = \int_a^b f_X(t)\, dt.$$

Podemos verla como el área bajo la gráfica f_X. Ver la Figura siguiente.

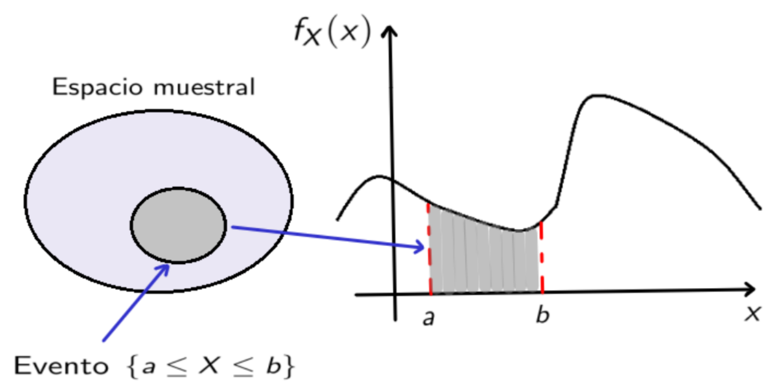


Figura: La probabilidad de que X tome valores en $[a, b]$ es la región sombreada bajo la gráfica de f_X.

Observación 3

- La probabilidad en algún valor puntual es nulo: $\mathbb{P}[X = a] = \int_a^a f_X(t)\,dt = 0$.
- La inclusión o no de los extremos de un intervalo no afecta la probabilidad:
$$\mathbb{P}[a \leq X \leq b] = \mathbb{P}[a < X < b].$$
- La función de densidad en un punto, $f_X(x)$, puede interpretarse como una función de masa por unidad de longitud cerca de x. Es decir, dado un diferencial $dx > 0$ pequeño, se tiene
$$\mathbb{P}[x \leq X \leq x + dx] = \int_x^{x+dx} f_X(t)\,dt \approx f_X(x)\,dx,$$
donde hemos aproximado el valor promedio de f_X sobre $[x, x + dx]$ por $f_X(x)$.

Outline

Definición 2

Dada X v.a continua, la función de distribución puede escribirse como
$$F_X(x) = \mathbb{P}[X \leq x] = \int_{-\infty}^{x} f_X(t)\,dt.$$

Proposición 1

La función de densidad puede obtenerse de la función de distribución por diferenciación:
$$\forall\, x \in \mathbb{R}, \quad f_X(x) = F'_X(x).$$

Demostración

Sigue resultados de "Teoría de la medida", pero en el caso de que f_X sea continua se reduce a aplicar el Teorema Fundamental de Cálculo (TFC).

Proposición 2

F_X es una función de distribución porque verifica (recordando) las propiedades:

i) *F_X es no decreciente.*

ii) *F_X es continua por la derecha.*

iii) *F_X es acotada:*
$$\lim_{x \to -\infty} F_X(x) = 0 \quad \wedge \quad \lim_{x \to +\infty} F_X(x) = 1.$$

Ejemplo 1

Sea X es una v.a. continua con función de distribución F_X y función de densidad f_X. Encuentre la función de densidad de $Y = 2X$.

Solución

- Hallamos primero la función de distribución:

$$\begin{aligned} F_Y(a) &= \mathbb{P}[Y \leq a] \\ &= \mathbb{P}[2X \leq a] \\ &= \mathbb{P}[X \leq a/2] \\ &= F_X(a/2). \end{aligned} \tag{1}$$

- Hallamos la densidad de Y derivando (1):

$$f_Y(a) = \frac{1}{2} f_X(a/2).$$

Fórmula de cambio de variables en 1D

Si deseamos hallar la función de densidad de una variable aleatoria $Y = g(X)$. Un enfoque es empezar con la función de distribución acumulativa de $Y = g(X)$ y trasladar el evento $\{g(X) \leq y\}$ en un evento equivalente que implique X. En general, no hay una fórma fácil de hacer esto. En el caso que g sea diferenciable y estrictamente creciente, podemos encontrar una expresión conveniente, ya que el evento $\{g(X) \leq y\}$ es igual al evento $\{X \leq g^{-1}(y)\}$. Así obtenemos que:

$$F_Y(y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}y).$$

Entonces podemos diferenciar con respecto a y para obtener la función de densidad de $Y = g(X)$.

Teorema 1 (Cambio de variable en una dimensión)

Sea X una v.a. continua con función de densidad f_X y sea $Y = g(X)$, donde g es diferenciable y estrictamente creciente (o estrictamente decreciente). Entonces la función de densidad de Y es dado por:

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|,$$

donde $x = g^{-1}(y)$.

Demostración

Sea g una función estrictamente creciente. La función de distribución (acumulada) de Y es:

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y),$$

Demostración (cont...)

Además

$$F_Y(y) = \mathbb{P}(X \leq g^{-1}(y)) = F_X(g^{-1}(y)) = F_X(x).$$

Así por la regla de la cadena, la función de densidad de Y es:
$f_Y(y) = f_X(x) \frac{dx}{dy}$.
La prueba para g estrictamente decreciente es análogo. En ese caso, se tiene que $-f_X(x) \frac{dx}{dy}$, que no es negativa, desde que $\frac{dx}{dy} < 0$ si g es estrictamente decreciente. Usando $\left| \frac{dx}{dy} \right|$ como en la declaración del teorema, se abarca ambos casos.

Definición 3 (Soporte)

El soporte de una variable aleatoria continua X y su distribución, es el conjunto de todos los puntos x donde $f_X(x) > 0$.

Ejemplo 2

La función de distribución acumulativa de una variable aleatoria exponencial X con parámetro λ es dado por:

$$\mathbb{P}(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0 & \text{en otro caso.} \end{cases}$$

Calculemos la función de distribución acumulativa y la función de densidad de una variable aleatoria Y, definida como $Y = X^2$.

Solución

Nosotros tenemos:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(X^2 \leq y) = \mathbb{P}(X \leq \sqrt{y}) = 1 - e^{-\lambda\sqrt{y}} \quad \text{para } y \geq 0.$$

Podemos calcular la función densidad de Y, por diferenciación. Así tenemos:

$$f_Y(y) = \frac{d}{dy}\mathbb{P}(Y \leq y) = \lambda e^{-\lambda y^{1/2}} \times \frac{1}{2}y^{-1/2} = \frac{\lambda}{2\sqrt{y}}e^{-\lambda\sqrt{y}} \quad \text{para } y > 0.$$

Observación 4

En experimentos de simulación es frecuente que se tenga acceso a secuencias de números aleatorios que se distribuyen uniformemente en el intervalo $(0, 1)$, pero lo que realmente se requiere son números aleatorios de una distribución no uniforme. En particular, en teoría de colas, a menudo se necesitan números aleatorios que se distribuyan de acuerdo con una distribución exponencial (negativa). Este último ejemplo muestra cómo se pueden obtener tales números.

Ejemplo 3

Sea X una v.a. con función de densidad f_X, y de acumulación Φ y sea $Y = e^X$. Encontrar la función de densidad de Y por el teorema de cambio de variable y también diferenciando Φ.

Solución

Desde que $g(x) = e^X$ es estrictamente creciente, se tiene que $x = \log y$ y $dy/dx = e^x$. Luego,

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(x)\frac{1}{e^x} = f_X(\log y)\frac{1}{y}, \ y > 0.$$

Ten en cuenta que después de aplicar la fórmula de cambio de variables, escribimos todo a la derecha en términos de y y especificamos el soporte de la distribución. Para determinar el soporte, sólo observamos que cuando x varía desde $-\infty$ a ∞, e^X varía desde 0 a ∞.

Solución (cont...)

Podemos obtener el mismo resultado trabajando a partir de la definición de la función de distribución acumulativa, transladando el evento $Y \leq y$ en un evento equivalente involucrando X. Para $y > 0$,

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log y) = \Phi(\log y),$$

y así el PDF, otra vez es:

$$f_Y(y) = \frac{d}{dy}\Phi(\log Y) = f_X(\log y)\frac{1}{y}, \;\; y > 0.$$

Ejercicio 1

Sea la función densidad de probabilidad de una variable aleatoria X, es dado por:

$$f_X(x) = \begin{cases} \alpha/x^5, & 1 \leq x < \infty \\ 0 & \text{en otros casos.} \end{cases}$$

Sea una nueva variable aleatoria definida como $Y = 1/X$. Encontrar la función densidad de Y.

Ejercicio 2

Sea X una variable aleatoria continua con PDF, f_X y sea $Y = a + bX$, con $b \neq 0$. Sea $y = a + bx$, el espejo de la relación entre Y y X. Entonces $\frac{dy}{dx} = b$ y así el PDF de Y es:

$$f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X\left(\frac{y-a}{b}\right)\frac{1}{|b|}.$$

Independencia de variables aleatorias

Intuitivamente, si dos variables aleatorias X e Y son independientes, entonces conocer el valor de X no da ninguna información sobre el valor de Y y viceversa. La siguiente definición, formaliza esta idea.

Definición 4

La variables aleatorias X e Y se dice que son independientes si:

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

para todo $y \in \mathbb{R}$. En el caso discreto, este caso es equivalente a la condición:

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

para todo x, y con x en el soporte de X y y en el soporte de Y.

La definición para más variables aleatorias es análogo:

Definición 5

La variables aleatorias $X_1, \ldots X_n$ son independientes si:

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_1 \leq x_1)\ldots\mathbb{P}(X_n \leq x_n)$$

para todo $x_1, \ldots, x_n \in \mathbb{R}$. Para infinitas variables aleatorias, decimos que son independientes si cada subconjunto finito de las variables aleatorias es independiente.

Desigualdad de Chebyshev

Se utiliza para proporcional cotas superiores para la probabilidad que toma una variable aleatoria en una región de $\mathbb{R}$, donde solo se conoce su media y varianza.

Teorema 2

Sea X una v.a. con media μ y varianza σ^2 finita. Para $k > 1$ se cumple

$$\mathbb{P}[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}.$$

Función de densidad condicional

Para una V.A continua X y un evento B, podemos escribir:

$$f_{X|B}(x)dx = \mathbb{P}(x < X \leq x+dx|B) = \frac{\mathbb{P}([x < X \leq x+dx] \cap B)}{\mathbb{P}(B)}.$$

Además, se tiene

$$f_{X|B}(x) = \begin{cases} f_X(x)/\mathbb{P}(B) & \text{si } [x < X \leq x+dx] \subset B \\ 0 & \text{si } [x < X \leq x+dx] \cap B = \emptyset \end{cases}$$

Ejercicio 3

Sea X una variable aleatoria con función densidad dada por:

$$f_X(x) = \begin{cases} (1/2)e^{-x/2} & \text{si } x \geq 0 \\ 0 & \text{en otros casos.} \end{cases}$$

Sea B el evento que $X < 1$ y deseamos encontrar $f_{X|X<1}(x)$. Para ello, calculemos $\mathbb{P}(X < 1)$ como:

$$\mathbb{P}(X < 1) = \int_0^1 (1/2)e^{-x/2}dx = -e^{-x/2}\Big|_0^1 = 1 - e^{-1/2}.$$

La función densidad de probabilidad condicional de X es entonces dado por:

$$f_{X|X<1}(x) = \begin{cases} f_X(x)/\mathbb{P}(X<1) & \text{para } 0 \leq x < 1, \\ 0 & \text{en otro caso,} \end{cases}$$
$$= \begin{cases} (1/2)e^{-x/2}/(1-e^{-1/2}) & \text{para } 0 \leq x < 1 \\ 0 & \text{en otro caso,} \end{cases}$$

La función de distribución acumulativa condicional, $F_{X|B}(x|B)$, de una variable aleatoria X, dado que B ha ocurrido, es definida como:

$$F_{X|B}(x|B) = \frac{\mathbb{P}(X \leq x, B)}{\mathbb{P}(B)}$$

donde $(X \leq x, B)$ es la intersección de los eventos $[X \leq x]$ y B. Además se tiene que:

$$F_{X|B}(-\infty|B) = 0 \text{ y } F_{X|B}(\infty|B) = 1$$

También:

$$\mathbb{P}(x_1 < X \leq x_2|B) = F_{X|B}(x_2|B) - F_{X|B}(x_1|B) = \frac{\mathbb{P}([x_1 < X \leq x_2], B)}{\mathbb{P}(B)}.$$

La función densidad condicional, puede ser obtenida como la derivada:

$$f_{X|B}(x|B) = \frac{d}{dx}F_{X|B}(x|B)$$

que debe ser no negativa y debe tener un área igual a 1.

FIN

Sétima sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

4 de septiembre de 2021

Distribución uniforme

Una variable aleatoria continua $X \sim U(\alpha, \beta)$ se dice uniformemente distribuida sobre el intervalo $[\alpha, \beta]$, con $\alpha < \beta$ si la función de densidad de probabilidad es dada por

$$f(x) = \begin{cases} \frac{1}{\beta-\alpha} & x \in [\alpha, \beta] \\ 0 & \text{en otros casos.} \end{cases}$$

Su función de distribución acumulada

$$F(x) = \begin{cases} 0 & x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & \alpha \leq x < \beta \\ 1 & x \geq \beta. \end{cases}$$

Sus momentos: $\mathbb{E}[X] = \frac{\alpha+\beta}{2}$, $Var[X] = \frac{(\beta-\alpha)^2}{12}$

Ejercicio 1

Se supone que un autobús llega a una parada determinada a las 10 : 00 a.m., pero la hora real de llegada es un VA. X que se distribuye uniformemente en el intervalo de 16 minutos de 9 : 52 a 10 : 08. Si un pasajero llega a la parada del autobús exactamente a las 9 : 50, ¿cuál es la probabilidad de que el pasajero suba al autobús a más tardar 10 minutos después de su llegada?

Solución

El pdf de X es $f(x) = \frac{1}{16}$ para x entre 9 : 52 a 10 : 08, y 0 en otro caso. El pasajero abordará el autobús a más tardar 10 minutos desde el momento de su llegada a la parada de autobús si el autobús llega a la parada entre las 9 : 52 y las 10 : 00 (ya que el pasajero necesariamente tendrá que esperar 2 minutos, entre las 9 : 50 y las 9 : 52). La probabilidad de que el autobús llegue entre las 9 : 52 y las 10 : 00 es $8/16 = 0{,}5$. Esta es la probabilidad requerida.

Ejercicio 2

Un número real X es elegido aleatoriamente en $[0, 1]$, además este número es un cuadrado de U, el cual está distribuido uniformente en $[0, 1]$. Determine su función de distribución acumulada y su función de densidad.

Solución

Se tiene que $X = U^2$ y si $0 \leq x \leq 1$, obtenemos

$$F_X(x) = P(X \leq x) = P(U^2 \leq x) = P(U \leq \sqrt{x}) = \sqrt{x}.$$

Además es claro que X toma valores entre 0 y 1, asi que su función de distribución acumulativa de X es dado por

$$F_X(x) = \begin{cases} 0 & , x \leq 0 \\ \sqrt{x} & , 0 \leq x \leq 1 \\ 1 & , x \geq 1. \end{cases}$$

Derivando F, se obtiene la función de densidad:

$$f_X(x) = \begin{cases} 0 & , x \leq 0 \\ 1/(2\sqrt{x}) & , 0 \leq x \leq 1 \\ 0 & , x > 1. \end{cases}$$

Notamos que F_X es continuo pero f_X no.

Distribución exponencial

Definición 1

- X :"N° de eventos independientes que ocurren en $[0, 1]$".
- λ representa la frecuencia promedio con que ocurren los eventos de Poisson.
- $X \sim \mathcal{P}(\lambda)$
- T :"representa el tiempo transcurrido hasta ocurrir el 1er evento de Poisson".
- $\beta := \frac{1}{\lambda}$ representa el periodo promedio para la ocurrencia de un evento de Poisson.

Función de distribución

Hallamos la función de ditribución usando su relación con la de Poisson:

$$\begin{aligned} F_T(t) &= \mathbb{P}[T \leq t] \\ &= 1 - \underbrace{\mathbb{P}[T > t]}_{\text{1er evento después de } t} \\ &= 1 - \underbrace{\mathbb{P}[X = 0]}_{\text{ocurre 0 eventos en } [0,t]} \\ &= 1 - f_X(0) \\ &= 1 - e^{-\lambda t}, \end{aligned}$$

donde λt representa el valor esperado de ocurrencias en el intervalo $[0, t]$.

- Función de Distribución: Dado $t \in \mathbb{R}$,

$$F(t) = \begin{cases} 1 - e^{-t/\beta} & , \ 0 \leq t, \\ 0 & , \ \text{en otro caso.} \end{cases}$$

- Función de densidad: Dado $t \in \mathbb{R}$,

$$f(t) = \begin{cases} \frac{1}{\beta} e^{-t/\beta} & , \ 0 \leq t, \\ 0 & , \ \text{en otro caso.} \end{cases}$$

- Momentos:

$$\mathbb{E}(T) = \beta \quad \wedge \quad Var(T) = \beta^2.$$

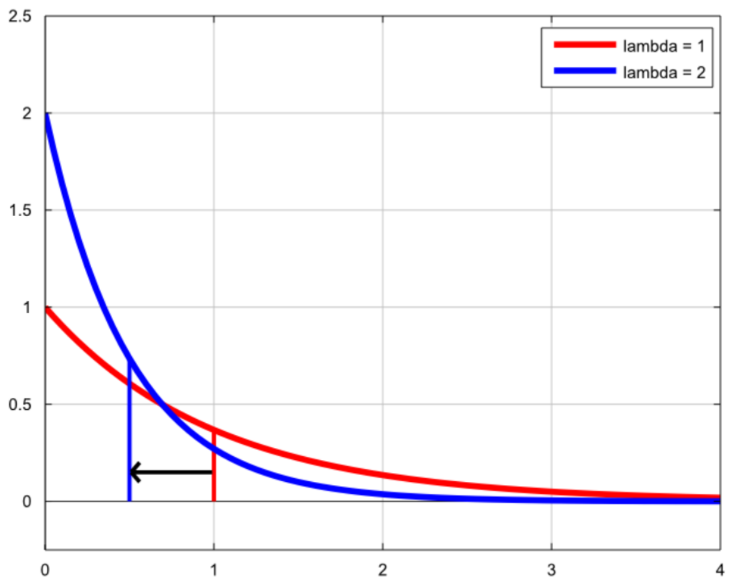


Figura: Función de densidad.

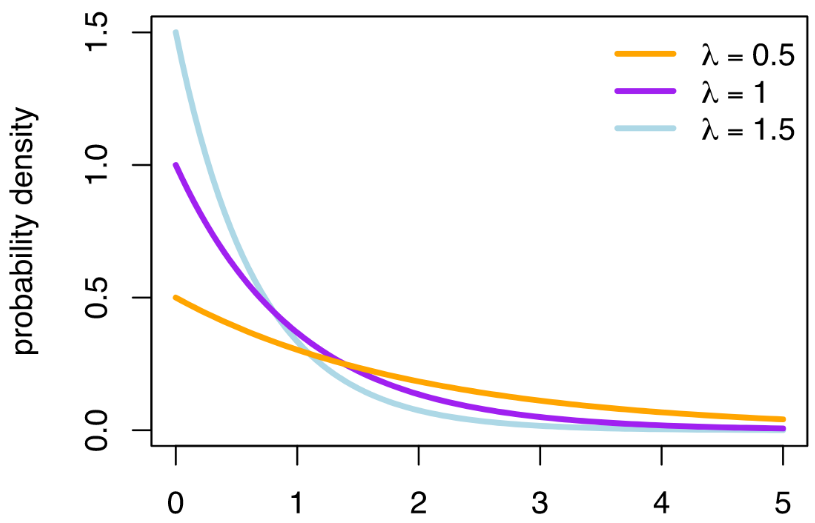


Figura: Función de densidad. λ: frecuencia media.

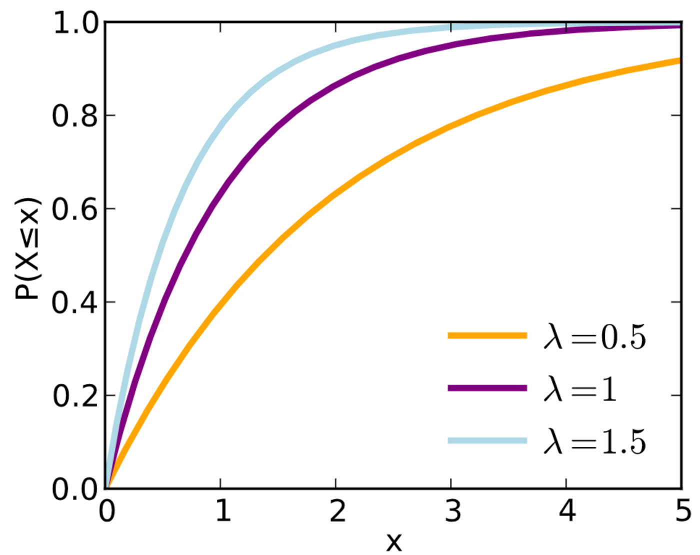


Figura: Distribución acumulada. λ: frecuencia media.

Aplicación

Se aplica en confiabilidad como modelo adecuado para la duración de vida de equipos que no "envejecen". Dado $T \sim Exp(\beta)$:

- Función de Probabilidad de falla: $\mathbb{P}[T \leq t]$.
- Función de Probabilidad de confiabilidad: $\mathbb{P}[T > t]$.

Sistema en serie

Si T_i, $i = 1, 2$, representan la vida útil de las componentes A y B respectivamente. Halle el tiempo de vida útil del sistema en serie.

Sistema en serie: Solución

- T : representa la vida útil del sistema en serie. $T = \min\{T_1, T_2\}$
- El tiempo de vida útil es otras palabras la Función de confiabilidad:

$$\begin{aligned}\mathbb{P}[T > t] &= \mathbb{P}[T_1 > t \cap T_2 > t]\\ &= \mathbb{P}[T_1 > t] \times \mathbb{P}[T_2 > t],\end{aligned}$$

donde se supone que sus vidas útiles son independientes una de la otra.

Observación 1

En el caso paralelo la vida útil del sistema es dado por $T = \max\{T_1, T_2\}$.

Ejemplo 1

Un sistema consiste de 2 componentes conectadas en serie cada una con tasa de falla constante $\lambda = 10^{-3}$. La durabilidad de las componentes se miden en horas.

a) ¿Cuántas horas promedio dura el sistema?
b) ¿Cuánto es el periodo de operación sin falla del sistema si este tiene una confiabilidad del 95 % durante t horas?

Solución

- Sea T_i, $i = 1, 2$ las vidas útiles, modelamos sus fallas con la distribución exponencial. Y T la vida útil del sistema.
- Sus funciones de falla son: $F_{T_i} = 1 - e^{-0{,}001t}$, $i = \{1, 2\}$.
- La función de confiabilidad del sistema es:

$$\mathbb{P}[T > t] = e^{-0{,}001t} \times e^{-0{,}001t} = e^{-0{,}002t}.$$

- La función de densidad de T es

$$f_T(t) = 0{,}002e^{-0{,}002t}.$$

Solución (cont..)

a) La duración media del sistema es:

$$\mathbb{E}(T) = \frac{1}{0{,}002} = 500\,h.$$

b) El periodo de operación t_0 con una confiabilidad de 95 % (representa la probabilidad), se calcula de:

$$0{,}95 = e^{-0{,}002t_0},$$

luego, el dicho periodo será

$$t_0 = -500\ln(0{,}95) \approx 25{,}65\,h.$$

Ejemplo 2

En una empresa de computadoras, el 90 % de las unidades producidas resultan no defectuosas. La vida útil en (años) del producto defectuoso se distribuye exponencialmente con media 0,2 y la del producto no defectuoso se distribuye uniformemente en el intervalo $[0, 2]$. Si del registro de computadoras vendidas se elige una muestra al azar y se revisa la garantía, se determina que una computadora al azar no dura más de un año. ¿Cómo cambia este evento los porcentajes de producción defectuosa y no defectuosa?

Solución

- Eventos: D : "Defectuoso", y
  E : "El producto no dura más de un año"
- $X_D \sim Exp(0{,}2)$ y $X_{D^c} \sim U[0, 2]$ representan las vidas útiles.
- Sus funciones de vida útil vienen dadas por:

$$F_{D^c}(t) = t/2,\ t \in [0, 2] \quad \wedge \quad F_D(t) = 1 - e^{-5t},\ t \geq 0.$$

- Por el teorema de probabilidad total para la partición $\{D, D^c\}$ de la producción, se tiene

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(D)\mathbb{P}(E|D) + \mathbb{P}(D^c)\mathbb{P}(E|D^c)\\ &= 0{,}1 \times (1 - e^{-5}) + 0{,}9 \times 0{,}5 = 0{,}0993 + 0{,}45 = 0{,}5493,\end{aligned}$$

Solución (cont...)

- luego las probabilidades actualizadas seran:

$$\mathbb{P}(D|E) = \frac{\mathbb{P}(D)\mathbb{P}(E|D)}{\mathbb{P}(E)} = \frac{0{,}0993}{0{,}5493} = 0{,}1808,$$
$$\mathbb{P}(D^c|E) = \frac{\mathbb{P}(D^c)\mathbb{P}(E|D^c)}{\mathbb{P}(E)} = \frac{0{,}45}{0{,}5493} = 0{,}8192.$$

Propiedad de pérdida de memoria

- Se utiliza para derivar la solución de los sistemas de colas.
- Nos dice que cada instante es como el comienzo de un nuevo período aleatorio, que tiene la misma distribución independientemente de cuánto tiempo haya transcurrido.
- Es decir, las probabilidades no se verán influenciadas por la historia del proceso.

- La exponencial es la única variable aleatoria continua con la propiedad de "pérdida de memoria".
- Esta propiedad nos dice que el tiempo que se debe esperar hasta el próximo evento es siempre $1/\lambda$, sin importar cuánto tiempo se haya esperado para que ocurra una nueva llegada.
- Este comportamiento es un poco contraintuitivo, porque podríamos esperar que la llegada de un evento es más probable cuanto más esperemos.

Problema (deducción)

Sea X una distribución exponencial con frecuencia esperada λ. Dado $t \in \mathbb{R}_{>0}$, supongamos que sabemos que $X > t$. ¿Cuál es la probabilidad de que X también sea mayor que $s + t$ para algún $s \in \mathbb{R}_{>0}$?

Solución

Es decir, queremos hallar $\mathbb{P}[X > s + t|X > t]$, luego

$$\begin{aligned}\mathbb{P}[X > s + t|X > t] &= \frac{\mathbb{P}[X > s + t]}{\mathbb{P}[X > t]} \\ &= \frac{e^{-\lambda(s+t)}}{e^{-\lambda t}} \\ &= \mathbb{P}[X > s].\end{aligned}$$

La probabilidad de que X exceda el valor $s + t$ dado t es la misma cuando X excedió el valor s, independientemente de t.

Ejemplo 3

Desde el punto de vista del tiempo de espera hasta la llegada de un cliente, la propiedad "pérdida de memoria" significa que no importa cuánto haya esperado hasta ahora. Si no ha observado a un cliente hasta el tiempo t_0, la distribución del tiempo de espera (desde el tiempo t_0) hasta el siguiente cliente es la misma que cuando comenzó en el tiempo cero.

Ejemplo 4

Suponga que el tiempo entre llegadas de clientes en una tienda está dado por una variable aleatoria exponencial $X \sim Exp(\beta)$, de modo que el tiempo promedio entre llegadas es de 2 minutos. Supongamos que pasas por la tienda y notas que está vacía. ¿Cuál es la probabilidad desde el momento en que pasas por la tienda, de que la tienda permanezca vacía por más de 5 minutos?

Solución

- Como $\mathbb{E}(X) = 2$, es decir el periodo esperado, se tiene que la frecuencia esperada es $\lambda = 1/2$.
- Supongamos que pasamos por la tienda en el tiempo t_0 (en minutos) y notamos que está vacía, el problema nos pide que encontremos

$$\mathbb{P}[X > t_0 + 5 | X > t_0].$$

Solución (cont...)

- Por la propiedad de "pérdida de memoria", se tiene que

$$\mathbb{P}[X > t_0 + 5 | X > t_0] = \mathbb{P}[X > 5].$$

- Por tanto,

$$\mathbb{P}[X > 5] = e^{-\lambda 5} = e^{-5/2} \approx 0{,}082084.$$

Distribución Gamma

- X : "Tiempo de espera hasta que ocurran n eventos de Poisson con frencuencia de ocurrencia media λ sobre el intervalo $[0, 1]$".
- $X \sim \Gamma(n, \lambda)$, con función de densidad

$$f(t) = \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{\Gamma(n)}, \quad \text{con } t > 0 \ \wedge \ \Gamma(n) = (n-1)!$$

- Momentos:

$$\mathbb{E}(X) = \frac{n}{\lambda} \quad \wedge \quad Var(X) = \frac{n}{\lambda^2}.$$

Observación 2

Si $n = 1$ la distribución X es una distribución exponencial. Es decir $X \sim Exp(\beta)$.

Deducción de la densidad

- Notamos que los siguiente eventos son iguales:

$$\{X \leq t\} = \{N_t \geq n\},$$

donde N_t representa el N° de eventos de Poisson en $[0, t]$.

- N_t tiene una distribución de Poisson con parámetro λt.
- Hallamos la distribución acumulada:

$$\begin{aligned} F(t) = \mathbb{P}[X \leq t] &= \mathbb{P}[N_t \geq n] \\ &= \sum_{j=n}^{+\infty} \mathbb{P}[N_t = j] \\ &= \sum_{j=n}^{+\infty} e^{-\lambda t} \frac{(\lambda t)^j}{j!}. \end{aligned}$$

Cont...

- Derivando con respecto a t, se obtiene la densidad

$$\begin{aligned} f(t) &= \sum_{j=n}^{+\infty} \left(\frac{e^{-\lambda t} j(\lambda t)^{j-1}\lambda}{j!} + \frac{(-\lambda)e^{-\lambda t}(\lambda t)^j}{j!} \right) \\ &= \sum_{j=n}^{+\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^{j-1}}{(j-1)!} - \sum_{j=n}^{+\infty} \frac{\lambda e^{-\lambda t}(\lambda t)^j}{j!} \\ &= \frac{\lambda e^{-\lambda t}(\lambda t)^{n-1}}{(n-1)!}. \end{aligned}$$

Forma General

Diremos que $X \sim \Gamma(\alpha, \lambda)$ donde sus parámetros son reales positivos, si posee la siguiente densidad

$$f(t) = \begin{cases} \dfrac{\lambda e^{-\lambda t}(\lambda t)^{\alpha-1}}{\Gamma(\alpha)} & , \ t > 0, \\ 0 & , \ \text{en otro caso.} \end{cases}$$

Con función gamma:

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x}\, dx, \quad \alpha > 0.$$

Observación 3

Si $\alpha > 1$, entonces $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1) \quad \wedge \quad \Gamma(1) = 1.$

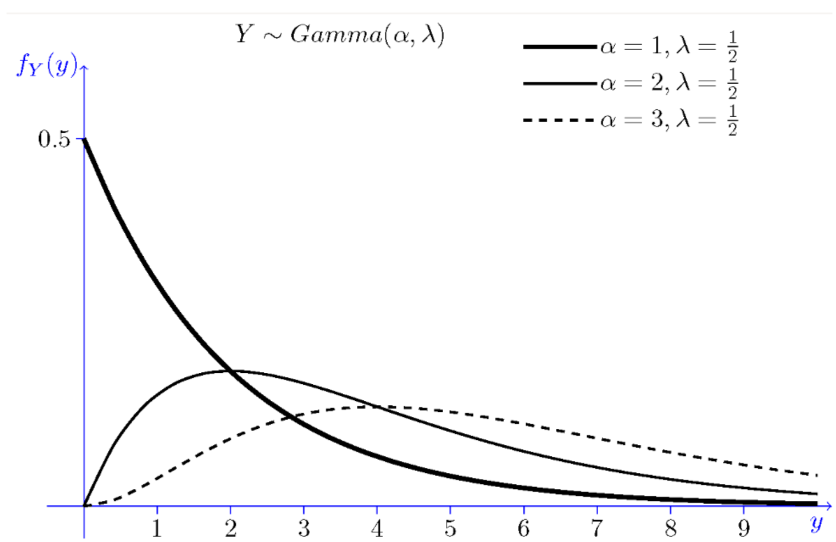


Figura: Función de densidad. α: N° de eventos. λ: frecuencia media.

Observación 4

A la distribución gamma con $\alpha \in \mathbb{Z}_{>0}$ se conoce como distribución de Erlang.

Ejemplo 5

El número de clientes, en promedio, que llegan por minuto a solicitar servicio a un banco es de 5. ¿Cuál es la probabilidad de que dos clientes tarden de 30 a 45 segundos en llegar al banco?

Solución

- El número de clientes por minuto que llegan a solicitar servicio a un banco es una variable aleatoria discreta de Poisson (con parámetro λ)
- El tiempo que tarden k clientes en llegar a solicitar servicio a un banco es una variable aleatoria continua Gamma (con parámetros $\alpha = k$ y λ).
- la frecuencia promedio esperada es

$$\lambda = \frac{5\,\text{clientes}}{1\,\text{minuto}} = \frac{5\,\text{clientes}}{60\,\text{segundos}} = \frac{1}{12}$$

- Sea X la v.a. continua que representa el tiempo, en segundos, que tardan $k = 2$ clientes en llegar a un banco. Es decir $X \sim \Gamma(\alpha = 2, \lambda = 1/12)$.

Solución (cont...)

- Nos piden calcular:

$$\begin{aligned}\mathbb{P}[30 \leq X \leq 45] &= \int_{30}^{45} f(x)\,dx \\ &= \frac{1}{12^2\Gamma(2)} \int_{30}^{45} xe^{-x/12}dx, \quad \Gamma(2) = 1.\end{aligned}$$

- Por integración por partes: $u = x$, $v' = e^{-x/12}$, se tiene

$$\begin{aligned}\int_{30}^{45} xe^{-x/12}dx &= \left[-12e^{-x/12}x - \int -12e^{-x/12}dx\right]_{30}^{45} \\ &= \left[-12e^{-x/12}x - 144e^{-x/12}\right]_{30}^{45} \\ &= \frac{-684}{e^{15/4}} + \frac{504}{e^{5/2}} = 25{,}284701.\end{aligned}$$

Solución (cont...)

- Por lo tanto la probabilidad pedida es $\frac{25{,}284701}{144} = 0{,}17559$. De esta manera, la probabilidad de que dos clientes tarden de 30 a 45 segundos en llegar al banco es de aproximadamente 17,56 %.

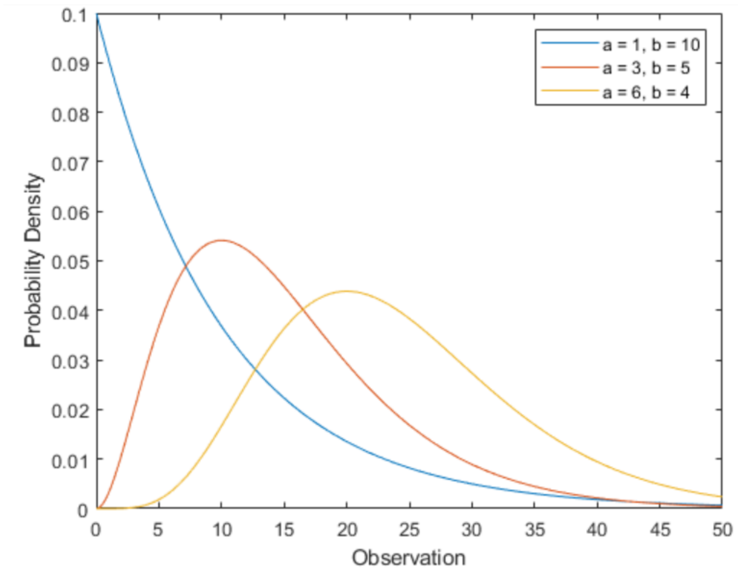


Figura: Función de densidad: $a = \alpha$ y $b = \frac{1}{\lambda}$

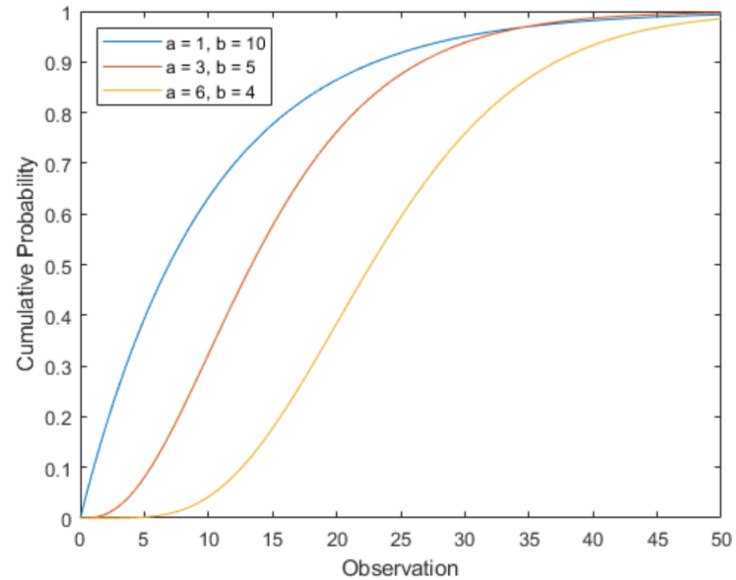


Figura: Función de distribución: $a = \alpha$ y $b = \frac{1}{\lambda}$

FIN

Octava sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

4 de noviembre de 2021

Outline

La distribucíon continua más importante es la distribución normal (o distribución de Gauss), debido a sus aplicaciones en ingeniería, en física, en la Economía y la Estadística. Por ejemplo:

- Muchos fenómenos continuos parecen seguirla o aproximarse mediante ella.
- Se puede utilizar para aproximar distribuciones discretas y de esta forma evitar cálculos engorrosos.
- El Teorema del límite central (TLC) justificará el uso de ella en las aproximaciones mencionadas.

La distribución normal fue introducida por Carl Friedrich Gauss a principios del siglo XIX, al estudiar los errores de medida en los movimientos de los cuerpos celestes.

Definición 1 (Distribución normal)

Una variable aleatoria continua Y, se dice que está distribuida normalmente, con media $\mu \in \mathbb{R}$ y varianza $\sigma^2 > 0$, $Y \sim N(\mu, \sigma^2)$, si su función de densidad esta dada por

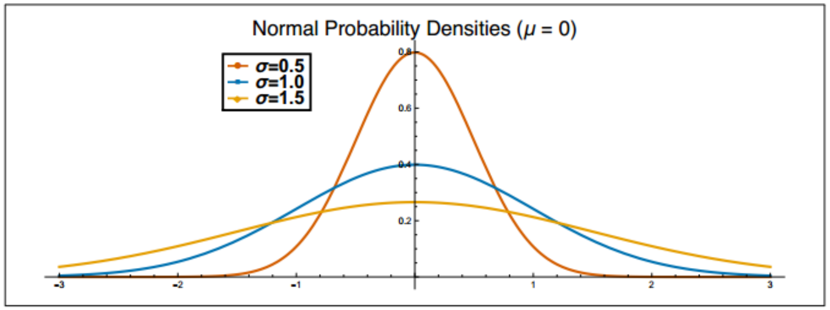
$$f_{\mu,\sigma^2}(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

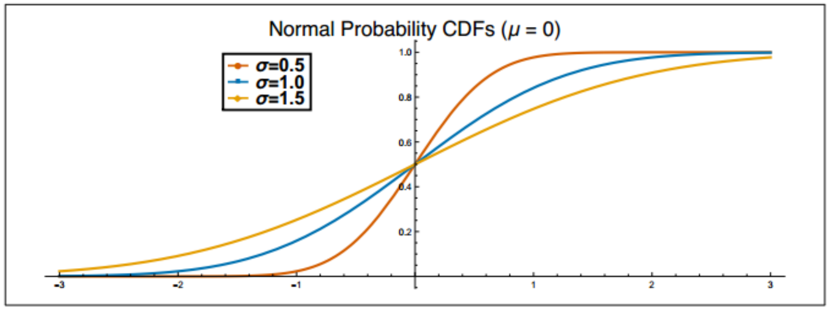
Definición 2

Dada una variable Y. Su variable aleatoria estandarizada o nomalizada se define como

$$\frac{Y-\mu}{\sigma},$$

donde $\mu = \mathbb{E}(Y)$ y $\sigma^2 = Var(Y)$.





Proposición 1

Dada X v.a. continua con función de densidad f_X. Entonces la v.a. $Y = bX + c$ donde $b > 0$ y $c \in \mathbb{R}$, tiene función de densidad

$$f_Y(y) = \frac{1}{b} f_X\left(\frac{y-c}{b}\right), \quad y \in \mathbb{R}.$$

Observación 1

"c" es llamado parámetro de localización, y "b" parámetro de escala. Multiplicar por el factor b corresponde a un cambio de escala y añadir c resulta en un cambio de localización.

Definición 1 (Normal estándar)

La v.a. Z se dice normal estándar si es una variable normal con media 0 y varianza 1, es decir $Z \sim N(0,1)$, esta tiene función de densidad

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

Se puede probar que $\int_{\mathbb{R}} f_Z = 1$. En efecto,

$$\begin{aligned}\left(\int_{\mathbb{R}} f_Z(x)\,dx\right)^2 &= \left(\int_{\mathbb{R}} f_Z(x)\,dx\right)\left(\int_{\mathbb{R}} f_Z(y)\,dy\right)\\ &= \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} e^{-\frac{x^2+y^2}{2}}\,dxdy = 1,\end{aligned}$$

esta integral se puede resolver usando cambio de coordenadas polares.

También se puede verificar sus momentos:

- Debido a que $te^{-t^2}/2$ es impar se tiene que

$$\mathbb{E}(Z) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} te^{-t^2/2}\,dt = 0.$$

- De lo anterior se tiene $\mathrm{Var}(Z) = \mathbb{E}(Z^2)$. Luego, usando el cambio de variable $u = t$ y $dv = te^{-t^2/2}$, obtenemos

$$\int_{\mathbb{R}} t^2 e^{-t^2/2}\,dt = \int_{\mathbb{R}} e^{-t^2/2}\,dt = \sqrt{2\pi}.$$

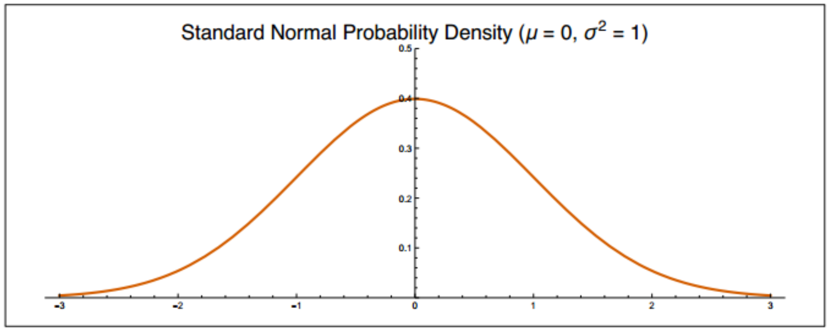
  Por tanto, $\mathrm{Var}(Z) = 1$.

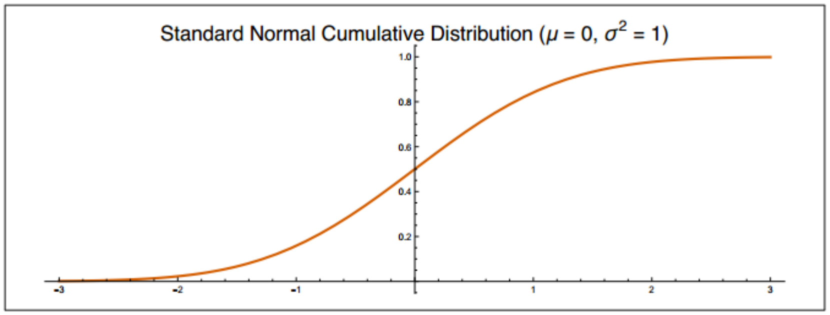
La función de distribución acumulada de esta variable aleatoria, se denota como Φ (o F_Z) y se define como:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\,dt.$$

Observación 2

Cada $Y \sim N(\mu, \sigma^2)$ se puede escribir como $Y = \sigma Z + \mu$, donde Z es normal estándar. De la Proposición 1, se puede hallar la función de densidad f_Y a partir de f_Z y verificar además que $\mathbb{E}(Y) = \mu$ y $Var(Y) = \sigma^2$.





Definición 3

Las variables que son independientes y tienen las misma distribución, se les llama independientes e idénticamente distribuidas o `i.i.d` o `I.I.D`.

Observación 3

Independiente *e* **idénticamente distribuidos** *son dos conceptos a menudo confundidos pero completamente diferentes. Las variables aleatorias son independientes si no proporcionan información entre sí; se distribuyen de forma idéntica si tienen la misma función de densidad (o equivalentemente, la misma distribución acumulada). Además, existen*

- **Independientes y no idénticamente distribuidas**.
- **Dependientes e idénticamente distribuidas**.
- **Dependientes y no idénticamente distribuidas**.

Teorema 1 (Teorema del Límite Central - TLC)

Sean $X_1, X_2, \ldots$ variables aleatorias iid cada una con media μ y varianza distinta de cero σ^2. Entonces

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{X_1 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \leq z\right) = F_Z(z) \quad \forall z \in \mathbb{R},$$

donde F_Z representa la distribución acumulada de $Z \sim N(0,1)$.

Observaciones

Se define la variable suma $S_n = X_1 + \cdots + X_n$ la cual tiene media $n\mu$ y desviación estándar $\sigma\sqrt{n}$, luego la estandarizada Z_n:

$$Z_n = \frac{S_n - \mathbb{E}(S_n)}{\sqrt{\mathrm{Var}(S_n)}}.$$

cont...

Z_n es aproximadamente normal estándar. La estimación es buena para $n \geq 30$. En el caso discreto, se usa una corrección por continuidad:

$$\mathbb{P}[S_n = k] \approx \mathbb{P}[k - 0{,}5 \leq S_n \leq k + 0{,}5],$$
$$\mathbb{P}[a \leq S_n \leq b] \approx \mathbb{P}[a - 0{,}5 \leq S_n \leq b + 0{,}5].$$

Ejemplo 1

Una distribución discreta como una suma de n V.A se aproxima bien a una normal cuando $n \geq 30$.

- *Dada* $X \sim B(n,p)$ *puede escribirse como* $X = X_1 + \cdots + X_n$ *donde* $X_i \sim Be(p)$ *son independientes. La experiencia indica una buena aproximación cuando* $np > 5$ *para* $p \leq 0{,}5$ *ó* $n(1-p) > 5$ *para* $p > 0{,}5$.
- *Dada* $X \sim \mathcal{P}(\lambda)$. *Si* $X = X_1 + \cdots + X_n$ *donde los* X_i *son independientes y tienen frecuencia esperada* λ/n *y* $\lambda > 5$, *entonces* X *puede aproximarse a una normal.*

Ejemplo 1

Cargamos en un avión 100 paquetes cuyos pesos son variables aleatorias independientes que se distribuyen uniformemente entre 5 y 50 libras. ¿Cuál es la probabilidad de que el peso total sea superior a 3000 libras?.

Solución

No es fácil calcular la función densidad acumulada del peso total y la probabilidad deseada. Una respuesta aproximada se puede obtener usando el teorema del límite central:
Calculemos $\mathbb{P}(S_{100} > 3000)$, donde S_{100} es la suma de los pesos de los 100 paquetes. En este caso la media y la varianza de los pesos de un único paquete es:

$$\mu = \frac{5+50}{2} = 27{,}5 \qquad \sigma^2 = \frac{(50-5)^2}{12} = 168{,}75.$$

basados en la fórmulas para la media y la varianza de la función densidad de una distribución uniforme, calculemos el valor normalizado :

$$z = \frac{3000 - 100 \cdot 27{,}5}{\sqrt{168{,}75 \times 100}} = \frac{250}{129{,}9} = 1{,}92$$

y usando las tablas normal estándar para obtener la aproximación tenemos:

$$\mathbb{P}(S_{100} \leq 3000) \approx \Phi(1{,}92) = 0{,}9726.$$

y así la probabilidad pedida es:

$$\mathbb{P}(S_{100} > 3000) = 1 - \mathbb{P}(S_{100} \leq 3000) \approx 1 - 0{,}97626 = 0{,}0274.$$

Teorema 1 (Propiedad reproductiva de la normal)

Sean $X \sim N(\mu_1, \sigma_1)$ y $Y \sim N(\mu_2, \sigma_2)$ independientes, entonces $Z = X + Y$ es normal. Además $\mathbb{E}(Z) = \mathbb{E}(X) + \mathbb{E}(Y)$ y $Var(Z) = Var(X) + Var(Y)$.

Demostración

Dado que X e Y son independientes, se puede probar que

$$f_{X+Y}(a) = \int_{\mathbb{R}} f_X(a-y) f_Y(y)\, dy.$$

En caso $X \sim N(0, \sigma^2)$ y $Y \sim N(0,1)$, se tiene

$$f_{X+Y}(a) = \frac{1}{2\pi\sigma} e^{-\frac{a^2}{2(1+\sigma^2)}} \int_{\mathbb{R}} e^{-c\left(y - \frac{a}{1+\sigma^2}\right)^2} dy \quad \text{donde } c = \frac{1+\sigma^2}{2\sigma^2}.$$

cont...

Haciendo el cambio de variable $x = y - \frac{a}{1+\sigma^2}$, se tiene

$$f_{X+Y}(a) = \frac{1}{\sqrt{2\pi}\sqrt{1+\sigma^2}} e^{-\frac{1}{2}\left(\frac{a}{\sqrt{1+\sigma^2}}\right)^2},$$

por tanto $X + Y \sim N(0, 1+\sigma^2)$. En general, se concluye de esto

$$X + Y = \sigma_2 \left(\frac{X - \mu_1}{\sigma_2} + \frac{Y - \mu_2}{\sigma_2} \right) + \mu_1 + \mu_2.$$

Distribución de Cauchy

Si X e Y son normales estándar independientes, entonces Y/X tiene una distribución de Cauchy. La distribución de Cauchy, es la distribución de la tangente de un ángulo aleatoriamente seleccionado desde $[-\pi, \pi]$.
La densidad es:

$$f(x) = \frac{1}{\pi(1+x^2)},$$

y el CDF:

$$F(t) = \frac{1}{\pi}\arctan(t) + \frac{1}{2}.$$

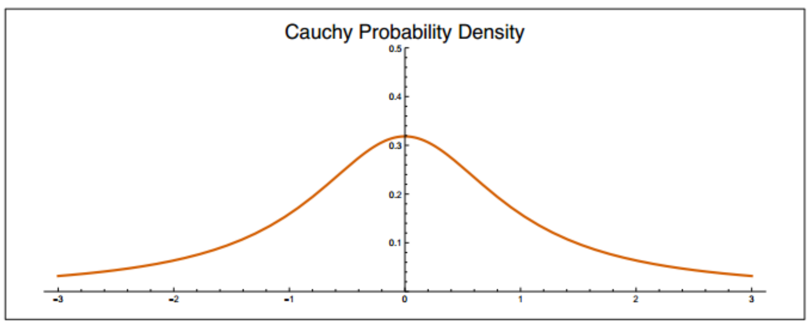
$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{dx}{(1+x^2)} = \frac{1}{\pi}\int_{-\infty}^{\infty}\frac{d\tan^{-1}(x)}{dx}$$
$$= \frac{1}{\pi}[\tan^{-1}(\infty) - \tan^{-1}(-\infty)] = \frac{1}{\pi}[\frac{\pi}{2} - (-\frac{\pi}{2})] = 1$$

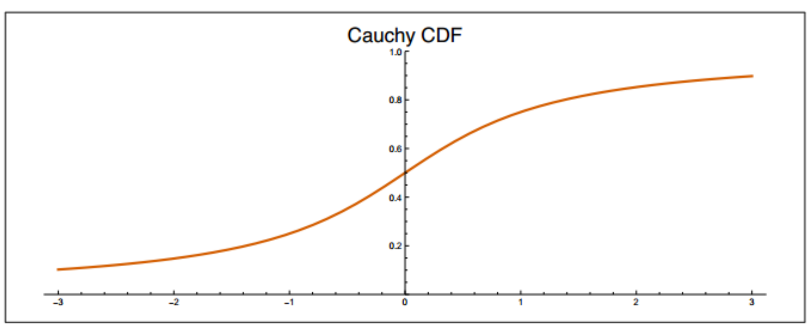
La esperanza de una distribución de Cauchy no existe :

$$\int_0^t \frac{x}{\pi(1+x^2)}dx = \frac{\ln(1+t^2)}{2\pi} \to \infty \quad \text{si} \quad t \to \infty.$$
$$\int_{-t}^0 \frac{x}{\pi(1+x^2)}dx = \frac{-\ln(1+t^2)}{2\pi} \to -\infty \quad \text{si} \quad t \to \infty.$$

Así

$$\int_{-\infty}^{\infty} \frac{x}{\pi(1+x^2)}dx \quad \text{es divergente.}$$





Definición 2 (Distribución chi-cuadrado)

Se dice que $X \sim \chi^2(n)$ con n grados de libertad, si existen $Z_1, \ldots, Z_n$ variables aleatorias normales estándar independientes tal que:

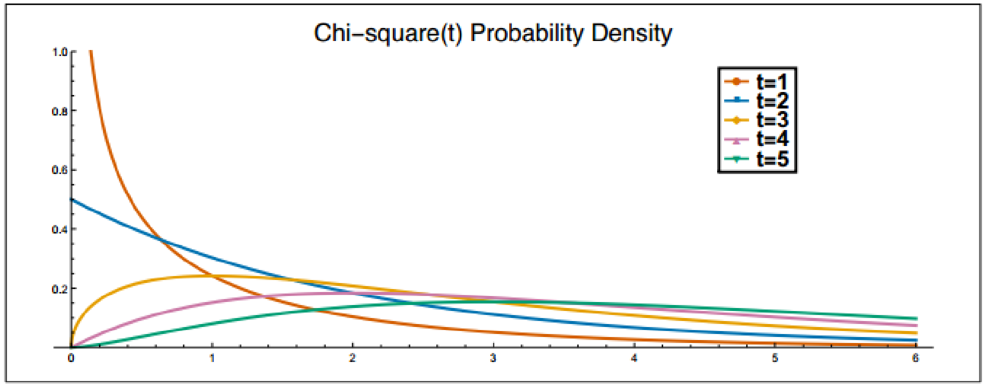
$$X = Z_1^2 + \cdots + Z_n^2.$$

La función densidad de esta distribución, está dado por:

$$f_X(t) = \frac{1}{2^{n/2}\Gamma(n/2)}t^{(n/2)-1}e^{-t/2}, \quad (t > 0),$$

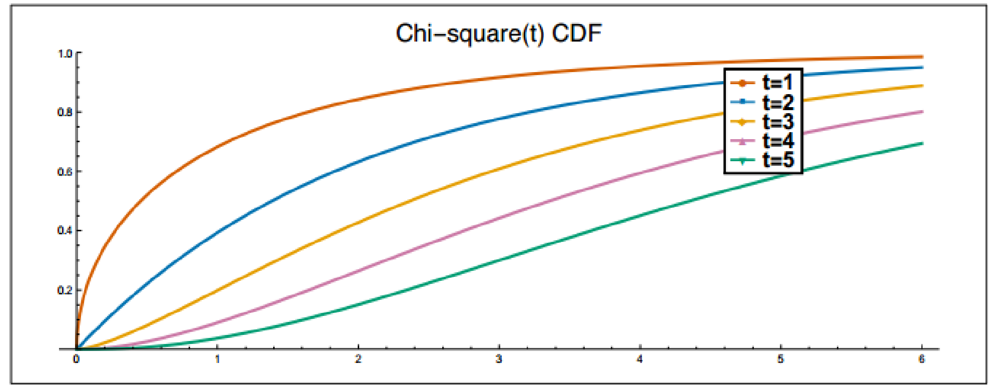
La media y varianza son caracterizadas por:

$$\mathbb{E}(X) = n, \quad Var(X) = 2n.$$



Se sigue de la definición que si X e Y son independientes y $X \sim \chi^2(n)$ y $Y \sim \chi^2(m)$, entonces:

$$(X+Y) \sim \chi^2(n+m).$$



Esta es una distribución importante en estadística inferencial para la validación de hipótesis de la distribución de observaciones en diferentes categorías. Es la base de la prueba de ajustes chi-cuadrado y el método de estimación de mínimos de chi-cuadrado.

Observación 4

La distribución chi-cuadrada $\chi^2(n)$ con n grados de libertad es una distribución Gamma $\Gamma(\alpha, \lambda)$ con parámetros $\alpha = n/2$ y $\lambda = 1/2$.

Ejemplo 2

Imagine que le interesa la relación entre los accidentes de tránsito y la edad del conductor. Podríamos obtener al azar registros de 60 accidentes de los archivos de la policía y ver cuántos de los conductores pertenecían a cada una de las siguientes categorías de edad: 17-20, 21-30, 31-40, 41-50, 51-60 y mayores de 60. Si no existe una relación entre la tasa de accidentes y la edad, entonces los conductores deberían estar igualmente distribuidos en las diferentes franjas de edad (hipótesis inicial). Sin embargo, obtenemos la siguiente tabla

cont...

	RANGO DE EDADES					
	17-20	21-30	31-40	41-50	51-60	> 60
Frecuencia observada de accidentes (O)	25	15	5	5	5	5
Frecuencia esperada de accidentes (E)	10	10	10	10	10	10

¿Esta distribución de frecuencias observadas ocurrió por casualidad, o hay algún patrón no aleatorio entre la edad y el número de accidentes?

Solución

En este caso particular, con solo mirar las frecuencias, parece bastante obvio pensar que ser más jóven implica mayor proporción de accidentes. Para responder a este tipo de preguntas, se utiliza un estadístico de la distribución chi-cuadrada con $r = 6 - 1$ grados de libertad (esta fórmula y el subsecuente desarrollo escapa del objetivo de este curso):

$$\begin{aligned}\chi^2 &= \sum \frac{(O-E)^2}{E}\\ &= 22{,}5 + 2{,}5 + 2{,}5 + 2{,}5 + 2{,}5 + 2{,}5 = 35.\end{aligned}$$

Ahora, consultaremos una tabla estadística para obtener el valor teórico $x = \chi^2 = \chi^2_{1-\alpha,r}$, tal que $\mathbb{P}[\chi^2 \leq x] = 1 - \alpha$, donde α representa el nivel de significancia. Consideramos en general $\alpha \leq 5\,\%$, obteniendo como valor crítico $\chi^2 = 11{,}07$.

cont...

Nuestro valor obtenido de 35 es mucho mayor que el valor crítico de 11,07. Por lo tanto, podemos estar relativamente seguros al concluir que nuestras frecuencias observadas son significativamente diferentes de las frecuencias que esperaríamos obtener si todas las categorías estuvieran distribuidas por igual. En otras palabras, la edad está relacionada con la cantidad de accidentes de tránsito que ocurren. Esto implica que es poco probable que esta distribución haya ocurrido por casualidad.

FIN

Novena sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

11 de noviembre de 2021

Outline

1. Función generadora de momentos
   - Función generadora de momentos

Si X es una variable aleatoria tomando valores en $\{0, 1, 2, \dots\}$. Se definió su función generadora de probabilidad como:

$$G_X(s) = \mathbb{E}(s^X) = \sum_{k=0}^{\infty} s^k \mathbb{P}(X = k) \tag{1}$$

Las funciones generadoras de probabilidad son muy útiles, pero sólo cuando las variables aleatorias son discretas y toman valores enteros no negativos. Para variables aleatorias mucho más generales, es necesario hacer una modificación de la ecuación anterior.

Definición 1

La función generadora de momentos (MGF) de la variable aleatoria de X es la función M_X definida como:

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^{tX}) \\ &= \begin{cases} \displaystyle\sum_x e^{tx} p(x) & \text{Si X es discreta con funcion de masa } p(x) \\ \displaystyle\int_{\mathbb{R}} e^{tx} f(x)\, dx & \text{Si X es continua con densidad } f(x). \end{cases} \end{aligned}$$

para todo $t \in \mathbb{R}$ para el cual la esperanza existe.

La relación con la función generadora de probabilidad para v.a. X discreta no negativa es

$$M_X(t) = \mathbb{E}(e^{tX}) = G_X(e^t)$$

Ejemplo 1

Si X tiene una distribución de Poisson con parámetro λ y función de masa de probabilidad:

$$p_X(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad , x = 0, 1, 2, \dots$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!}$$

Si realizamos el cambio de variable $a = e^t \lambda$, entonces obtenemos $M_X(t) = e^{\lambda(e^t - 1)}$.

La razón del nombre de función generadora de momentos es la siguiente expansión que muestra que $M_X(t)$ es la función generadora exponencial de los momentos de X:

$$\begin{aligned} M_X(t) = \mathbb{E}(e^{tX}) &= \mathbb{E}\left(1 + tX + \frac{1}{2!}(tX)^2 + \cdots\right) \\ &= 1 + t\mathbb{E}(X) + \frac{1}{2!}t^2\mathbb{E}(X^2) + \cdots \end{aligned}$$

Teorema 1

Si $M_X(t)$ existe en una vecindad de 0, entonces para $k = 1, 2 \ldots$

$$\mathbb{E}(X^k) = M_X^{(k)}(0)$$

la `k-ésima` derivada de $M_X(t)$ evaluada en $t = 0$.

El cálculo de la distribución de la suma de variables aleatorias, puede determinarse muy rápidamente a través de la función generadora de momento. Consideremos primero la función lineal $aX + b$ de la variable aleatoria X. Si $a, b \in \mathbb{R}$

$$M_{aX+b}(t) = \mathbb{E}(e^{t(aX+b)}) = \mathbb{E}(e^{atX}e^{tb}) = e^{tb}\mathbb{E}(e^{(at)X})$$

lo que produce que $M_{aX+b}(t) = e^{tb}M_X(at)$.

Teorema 2

Si X e Y son variables aleatorias independientes, entonces $X + Y$ tiene una función generadora de momentos:

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

Tenemos por independencia y propiedad de la esperanza para la variable aleatoria X y Y:

$$\begin{aligned} M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) &= \mathbb{E}(e^{tX}e^{tY}) \\ &= \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) \end{aligned}$$

De manera similar se tiene para n v.a independientes.

Teorema 3

Sean $Y_1, Y_2, ..., Y_n$ variablea aleatorias independientes con funciones generadoras de momento $M_{Y_1}(t), M_{Y_2}(t), ..., M_{Y_n}(t)$, respectivamente. Si $U = Y_1 + Y_2 + ... + Y_n$, entonces

$$M_U(t) = M_{Y_1}(t) \times M_{Y_2}(t) \times .... \times M_{Y_n}(t).$$

Ejemplo 2

Demostrar que la función generadora de momento de $Z \sim N(0, 1)$ es $M_Z(t) = e^{t^2/2}$.

Solución

$$\begin{aligned} M(t) &= \int_{\mathbb{R}} e^{tx} f(x) dx = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tx - x^2/2} dx \\ &= \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-t)^2/2} dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-u^2/2} du = e^{t^2/2}, \end{aligned}$$

donde hemos usado el cambio de variable $u = x - t$ y el hecho que $\int_{\mathbb{R}} e^{-u^2/2} du = \sqrt{2\pi}$.

Ejemplo 3

Demuestre que la FGM de $Y \sim N(\mu, \sigma^2)$ es $M_Y(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$.

Solución

Dado que

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

tiene una distribución `normal estándar`. Entonces

$$\begin{aligned} M_Y(t) &= \mathbb{E}[e^{t(\sigma Z + \mu)}] \\ &= e^{t\mu}\mathbb{E}[e^{(t\sigma)Z}] \\ &= e^{t\mu} M_Z(t\sigma) \\ &= e^{t\mu} e^{t^2\sigma^2/2}. \end{aligned}$$

Ejemplo 4

Demuestre que la FGM de $X \sim \Gamma(\alpha, \lambda)$ es $M_X(t) = (1 - \beta t)^{-\alpha}$ si $t < \lambda$.

Solución

Dados $\alpha > 0$, $\lambda = 1/\beta > 0$ y $x > 0$, se tiene

$$f(x) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha - 1}}{\Gamma(\alpha)} = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}.$$

Luego, factorizando se toma $1/\beta - t > 0$ para evitar que la integral diverga al infinito.

$$\begin{aligned} M(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty e^{tx} x^{\alpha-1} e^{-x/\beta}\, dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/\left(\frac{\beta}{1-\beta t}\right)}\, dx. \end{aligned}$$

Haciendo el cambio $u = x/\theta$ con $\theta = \frac{\beta}{1-\beta t} > 0$ y de que $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\, dt$, se tiene

$$\begin{aligned} M(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty u^{\alpha-1}\theta^{\alpha-1} e^{-u}\theta\, du \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha}\Gamma(\alpha)\theta^\alpha = \frac{1}{(1-\beta t)^\alpha}, \end{aligned}$$

para $t < 1/\beta$.

Ejemplo 5

Dada $Z \sim N(0,1)$. Demuestre que la FGM de Z^2 es $(1 - 2t)^{-1/2}$ para $t < 1/2$.

La función generadora de momentos para Z^2 es

$$\begin{aligned} M(t) = E(e^{tZ^2}) &= \int_{-\infty}^\infty e^{tz^2} f(z) dz = \int_{-\infty}^\infty e^{tz^2} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz \\ &= \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)(1-2t)} dz. \end{aligned}$$

Esta integral se puede evaluar ya sea consultando una tabla de integrales o tomando en cuenta que, si $1 - 2t > 0$ (de manera equivalente, $t < 1/2$), el integrando

$$\frac{\exp\left[-\left(\frac{z^2}{2}\right)(1-2t)\right]}{\sqrt{2\pi}} = \frac{\exp\left[-\left(\frac{z^2}{2}\right)\Big/(1-2t)^{-1}\right]}{\sqrt{2\pi}}$$

es proporcional a la función de densidad de una variable aleatoria normalmente distribuida con media 0 y varianza $(1-2t)^{-1}$. Para hacer del integrando una función de densidad normal (para que la integral definida sea igual a 1), multiplicamos el numerador y denominador por la desviación estándar, $(1-2t)^{-1/2}$. Entonces

$$M(t) = \frac{1}{(1-2t)^{1/2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-1/2}} \exp\left[-\left(\frac{z^2}{2}\right) \Big/ (1-2t)^{-1}\right] dz.$$

La integral es igual a 1, si $t < 1/2$.

$$M(t) = \frac{1}{(1-2t)^{1/2}} = (1-2t)^{-1/2}.$$

Por unicidad de $M(t)$ con las variables aleatorias, se tiene comparando con la función generadora de momento de la distribución gamma, que $Z^2 \sim \Gamma(\alpha = 1/2, \lambda = 1/2)$. Además se puede expresar la función de densidad de $\Gamma(1/2, 1/2)$ como una distribución χ^2 con $\nu = 1$ grado de libertad. Es decir la función de densidad para $U = Z^2$ se expresa como

$$f_U(u) = \begin{cases} \dfrac{u^{-1/2}e^{-u/2}}{\Gamma(1/2)2^{1/2}}, & u \geq 0, \\ 0 & , \text{ en cualquier otro punto} \end{cases}$$

Ejemplo 6

La cantidad de clientes que llegan a una caja para pagar en un intervalo determinado de tiempo posee aproximadamente una distribución de Poisson. Halle la distribución del tiempo de espera hasta que llegue el n-ésimo cliente. (Sug. Use FGM)

Solución

Si Y_1 denota el tiempo que transcurre hasta la primera llegada, Y_2 denota el tiempo entre la primera y la segunda llegada, . . . , y Y_n denota el tiempo entre la $(n-1)$-ésima llegada y la n-ésima llegada. Entonces se puede demostrar que $Y_1, Y_2, ..., Y_n$ son variables aleatorias independientes exponenciales de periodo medio entre llegadas θ, es decir

$$f_{Y_i}(y_i) = \begin{cases} \frac{1}{\theta}e^{-y_i/\theta}, & y_i > 0, \\ 0, & \text{en cualquier otro punto.} \end{cases}$$

Entonces la distribución del tiempo de espera desde que se abre la caja hasta la llegada del n-ésimo cliente es $U = Y_1 + Y_2 + ... + Y_n$. Del Ejemplo 4, para $\alpha = 1$, se tiene $M_{Y_i}(t) = (1-\theta t)^{-1}$, $i = 1, 2, .., n$. Luego, por el Teorema 3, se tiene

$$M_U(t) = (1-\theta t)^{-n}.$$

Ésta es la función generadora de momentos de una variable aleatoria con distribución gamma con función de densidad

$$f_U(u) = \begin{cases} \dfrac{1}{\Gamma(n)\theta^n}(u^{n-1}e^{-u/\theta}) & , u > 0, \\ 0 & , \text{ en cualquier otro punto.} \end{cases}$$

Teorema 4

Sean $Y_1, Y_2, ..., Y_n$ V.A independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $V(Y_i) = \sigma_i^2$, para $i = 1, 2, ..., n$ y sean $a_1, a_2, ..., a_n$ constantes. Entonces

$$U = \sum_{i=1}^{n} a_i Y_i \sim N(\mu, \sigma^2),$$

donde

$$E(U) = \sum_{i=1}^{n} a_i \mu_i$$

y

$$V(U) = \sum_{i=1}^{n} a_i^2 \sigma_i^2$$

La función generadora de los Y_i es

$$M_{Y_i}(t) = \exp\left(\mu_i t + \frac{\sigma_i^2 t^2}{2}\right).$$

Por tanto $a_i Y_i$ tiene función generadora de momentos

$$M_{a_i Y_i}(t) = E(e^{t a_i Y_i}) = M_{Y_i}(a_i t) = \exp\left(\mu_i a_i t + \frac{a_i^2 \sigma_i^2 t^2}{2}\right)$$

como los Y_i son independientes, entonces los $a_i Y_i$ también. Luego, se tiene

$$\begin{aligned} M_U(t) &= M_{a_1 Y_1}(t) \times M_{a_2 Y_2}(t) \times .... \times M_{a_n Y_n}(t) \\ &= \exp\left(\mu_1 a_1 t + \frac{a_1^2 \sigma_1^2 t^2}{2}\right) \times .... \times \exp\left(\mu_n a_n t + \frac{a_n^2 \sigma_n^2 t^2}{2}\right) \\ &= \exp\left(t \sum_{i=1}^{n} a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2\right). \end{aligned}$$

Teorema 5

Sean $Y_1, Y_2, ..., Y_n$ V.A independientes normalmente distribuidas con $E(Y_i) = \mu_i$ y $V(Y_i) = \sigma_i^2$, para $i = 1, 2, ..., n$ y definimos

$$Z_i = \frac{Y_i - \mu_i}{\sigma_i^2} \quad i = 1, 2, ..., n.$$

Entonces $Y = \sum_{i=1}^{n} Z_i^2 \sim \chi^2(n)$ tiene función generadora de momentos igual a $M(t) = (1 - 2t)^{-n/2}$ para $t < 1/2$.

Como Y_i está normalmente distribuida con media μ_i y varianza σ_i^2, luego los Z_i están normalmente distribuida con media 0 y varianza 1. Entonces tenemos que los Z_i^2 son variables aleatorias χ^2 con 1 grado de libertad independientes. Como

$$M_{Z_i^2}(t) = (1 - 2t)^{-1/2}.$$

Entonces $V = \sum_{i=1}^{n} Z_i^2$ tiene FGM:

$$\begin{aligned} M_V(t) &= M_{Z_1^2}(t) \times M_{Z_2^2}(t) \times .... \times M_{Z_n^2}(t) \\ &= (1 - 2t)^{-1/2} \times (1 - 2t)^{-1/2} \times ... \times (1 - 2t)^{-1/2} \\ &= (1 - 2t)^{-n/2}. \end{aligned}$$

La cual es una distribución gamma $\Gamma(n/2, 1/2)$ y que a su vez es igual a la función de densidad de una chi-cuadrada con n grados de libertad (compare!).

FIN

Décima sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

18 de septiembre de 2021

Outline

Definición 1 (Función de masa de probabilidad conjunta)

Sean Y_1 y Y_2 variables aleatorias discretas. La función de masa de probabilidad conjunta para Y_1 y Y_2 se define como

$$p(y_1, y_2) := \mathbb{P}(Y_1 = y_1; Y_2 = y_2), \quad y_1 \in R_{Y_1}, \; y_2 \in R_{Y_2}.$$

Observación 1

1. $p(y_1, y_2) \geq 0$ *para todo* y_1, y_2.
2. $\displaystyle\sum_{y_1 \in R_{Y_1}} \sum_{y_2 \in R_{Y_2}} p(y_1, y_2) = 1.$

Ejemplo 1

Consideremos el experimento de lanzar dos dados. Sea Y_i la v.a que representa el resultado de cada dado. Hallar $\mathbb{P}(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2)$.

Solución

Aplicamos el axioma de adición y equiprobabilidad

$$\mathbb{P}(2 \leq Y_1 \leq 3, 1 \leq Y_2 \leq 2)$$
$$= p(2,1) + p(2,2) + p(3,1) + p(3,2) = 4/36.$$

Para calcular $p_1(y_1)$ sumamos $p(y_1, y_2)$ para todos lo valores de Y_2 y por tanto acumulamos las probabilidades en el eje Y_1 (o margen).

Proposición 1 (Marginales)

Sean Y_1, Y_2 variables aleatorias discretas con función de probabilidad conjunta $p(y_1, y_2)$. Luego, las funciones de masa de Y_1 y Y_2 se calculan como

$$p_1(y_1) = \sum_{y_2 \in R_{Y_2}} p(y_1, y_2) \quad y \quad p_2(y_2) = \sum_{y_1 \in R_{Y_1}} p(y_1, y_2).$$

Usualmente se les llama funciones de masa de probabilidad marginal.

Ejemplo 2

De un grupo de tres republicanos, dos demócratas y uno independiente se ha de seleccionar aleatoriamente un comité de dos personas. Denote con Y_1 el número de republicanos y con Y_2 el número de demócratas del comité. Encuentre la función de probabilidad conjunta de Y_1 y Y_2 y luego encuentre la función de probabilidad marginal de Y_1.

Solución:

Veamos por ejemplo

$$P(Y_1 = 1, Y_2 = 1) = p(1,1) = \frac{\binom{3}{1}\binom{2}{1}\binom{1}{0}}{\binom{6}{2}} = \frac{3(2)}{15} = \frac{6}{15}$$

debido a que hay 15 puntos muestrales igualmente probables; para el evento en cuestión debemos seleccionar un republicano de entre los tres, un demócrata de entre los dos y cero independientes.

Para calcular $p(y_1)$ debemos sumar los valores de Y_2.

$$p_1(0) = p(0,0) + p(0,1) + p(0,2) = 0 + 2/15 + 1/15 = 3/15.$$

$Y_2 \backslash Y_1$	0	1	2	*Total*
0	0	3/15	3/15	6/15
1	2/15	6/15	0	8/15
2	1/15	0	0	1/15
Total	3/15	9/15	3/15	1

Definición 2 (Función de distribución acumulada conjunta)

Sean Y_1 e Y_2 variables aleatorias. Se define función de distribución acumulada conjunta de Y_1 e Y_2 como

$$F(y_1, y_2) = P(Y_1 \leq y_1; Y_2 \leq y_2), \quad y_1 \in R_{Y_1}, \quad y_2 \in R_{Y_2}.$$

En el caso de que Y_1 e Y_2 sean discretas se tiene que

$$F(y_1, y_2) = \sum_{t_1 \leq y_1} \sum_{t_2 \leq y_2} p(t_1, t_2) \quad \forall y_1, y_2 \in \mathbb{R}. \tag{1}$$

Ejemplo 3

Siguiendo el Ejemplo 1 del lanzamiento de dados. Hallar $F(2,3)$.

Solución

Aplicando (1) y equiprobabilidad, se obtiene

$$\begin{aligned} F(2,3) &= \mathbb{P}(Y_1 \leq 2, Y_2 \leq 3) \\ &= p(1,1) + p(1,2) + p(1,3) + p(2,1) + p(2,2) + p(2,3) \\ &= 6/36. \end{aligned}$$

Proposición 2

Sean Y_i, $i = 1, 2$ variables aleatorias. Entonces, se tiene que

$$\begin{aligned} F_{Y_1}(y_1) &= F(y_1, +\infty) := \lim_{y_2 \to +\infty} F(y_1, y_2), \\ F_{Y_2}(y_2) &= F(+\infty, y_2) := \lim_{y_1 \to +\infty} F(y_1, y_2). \end{aligned}$$

Usualmente se les llama funciones de distribución acumulada marginal.

Demostración

Se usa el hecho que $[Y_1 \leq y_1] = [Y_1 \leq y_1, Y_2 \leq +\infty]$ y la propiedad de continuidad dada en la Proposición 4 de la Sesión 3.

Proposición 3

Sean X, Y variables aleatorias con función de distribución acumulada conjunta $F(x, y)$. Si $x_1 < x_2$ y $y_1 < y_2$ entonces

$$\begin{aligned} &\mathbb{P}[x_1 < X \leq x_2, y_1 < Y \leq y_2] \\ &= F(x_2, y_2) + F(x_1, y_1) - F(x_1, y_2) - F(x_2, y_1). \end{aligned}$$

Demostración

Reescribimos el evento:

$$\begin{aligned} [x_1 < X \leq x_2, y_1 < Y \leq y_2] = &[X \leq x_2, Y \leq y_2] \\ &- [X \leq x_1, Y \leq y_2] \cup [X \leq x_2, Y \leq y_1]. \end{aligned} \tag{2}$$

cont...

Además, como

$$\mathbb{P}\big([X \le x_1, Y \le y_2] \cup [X \le x_2, Y \le y_1]\big) = F(x_1, y_2) + F(x_2, y_1) - \mathbb{P}[X \le x_1, Y \le y_1]. \tag{3}$$

Como el sustraendo es un subconjunto de minuendo en (2), se tiene

$$\mathbb{P}[x_1 < X \le x_2, y_1 < Y \le y_2] = F(x_2, y_2) - \mathbb{P}\big([X \le x_1, Y \le y_2] \cup [X \le x_2, Y \le y_1]\big),$$

luego, se concluye reemplazando (3).

Proposición 4 (Independencia)

Dados X, Y variables aleatorias. Se dice que X e Y son independientes si y solo si

$$F(x, y) = F_X(x)\, F_Y(y) \quad \forall x, y \in \mathbb{R}.$$

Demostración

De la definición de independencia:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \quad \forall A, B \text{ eventos de } \mathbb{R}^2.$$

Se prueba la equivalencia, consideramos $A = [X \le x]$ y $B = [Y \le y]$ para $x, y \in \mathbb{R}$ y los axiomas de probabilidad.

Proposición 5 (Independencia)

Sea (X, Y) una variable aleatoria bidimensional discreta con función de masa conjunta $p(x, y)$ y marginales p_X y p_Y. Se dice que X e Y son independientes si y solo si

$$p(x, y) = p_X(x)\, p_Y(y) \quad \forall (x, y) \in R_X \times R_Y.$$

Demostración

$\Rightarrow$) Basta considerar $A = \{x\}$ y $B = \{y\}$.

$\Leftarrow$) Usando el axioma de aditividad, se tiene

$$\mathbb{P}[X \in A, Y \in B] = \sum_{y \in B} \sum_{x \in A} p(x, y) = \mathbb{P}[X \in A]\mathbb{P}[Y \in B].$$

Ejemplo 1

Sean $X \sim \mathcal{P}(\lambda_1)$ y $Y \sim \mathcal{P}(\lambda_2)$ independientes. Halle el tipo de distribución de $X + Y$.

Solución

Fíjese que el evento $\{X + Y = n\}$ puede escribirse como la unión disjunta de eventos $\{X = k, Y = n - k\}$ para $0 \le k \le n$. Entonces, por la independencia se tiene

$$\begin{aligned} p_{X+Y}(n) &= \sum_{k=0}^{n} p_X(k)\, p_Y(n-k) \\ &= \sum_{k=0}^{n} \frac{e^{-\lambda_1}\lambda_1^k}{k!} \times \frac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!} \\ &= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^n}{n!}. \end{aligned}$$

Ejemplo 2

Suponga que X e Y tienen la siguiente función de probabilidad conjunta:

$y \backslash x$	2	4
1	0.10	0.15
3	0.20	0.30
5	0.10	0.15

Determine si las variables aleatorias X e Y son independientes.

Si X e Y son variables aleatorias discretas y sea $g : \mathbb{R}^2 \to \mathbb{R}$ una función medible, por ejemplo continua. Entonces $Z = g(X, Y)$ es una variable aleatoria definida como $Z(\omega) = g(X(\omega), Y(\omega))$ para $\omega \in \mathbb{R}$.

Teorema 1

Sean X, Y y $g(X, Y)$ v.a. discretas. Entonces

$$\mathbb{E}(g(X, Y)) = \sum_{x \in R_X} \sum_{y \in R_Y} g(x, y)\mathbb{P}(X = x, Y = y)$$

siempre que esta suma converga absolutamente.

Demostración

Ver Proposición 2.1 de [1].

Ejemplo 3

Dados X e Y son variables aleatorias discretas y $a, b \in \mathbb{R}$. Demuestre que

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y),$$

Solución

Siempre que $\mathbb{E}(X)$ y $\mathbb{E}(Y)$ existan. Se tiene

$$\begin{aligned}
\mathbb{E}(aX + bY) &= \sum_x \sum_y (ax + by)\mathbb{P}(X = x, Y = y) \\
&= a\sum_x x \sum_y \mathbb{P}(X = x, Y = y) + b\sum_y y \sum_x \mathbb{P}(X = x, Y = y) \\
&= a\sum_x x\mathbb{P}(X = x) + b\sum_y y\mathbb{P}(Y = y) \\
&= a\mathbb{E}(X) + b\mathbb{E}(Y).
\end{aligned}$$

Teorema 2

Dados X e Y son variables aleatorias discretas independientes, y g, h cualesquiera. Demuestre

$$\mathbb{E}[g(X)\, h(Y)] = \mathbb{E}[g(X)] \times \mathbb{E}[h(Y)].$$

Solución

$$\begin{aligned}
\mathbb{E}[g(X)\, h(Y)] &= \sum_x \sum_y g(x)h(y)p(x, y) \\
&= \sum_x \sum_y g(x)h(y)p_X(x)p_Y(y) \\
&= \sum_x \Big[g(x)p_X(x)\Big] \sum_y \Big[h(y)p_Y(y)\Big] \\
&= \mathbb{E}[g(X)] \times \mathbb{E}[h(Y)].
\end{aligned}$$

Observación 2

- *Analice el caso $g(X,Y) = X$ en el Teorema 2.*
- *Siguiendo el Ejemplo 2, demuestre*

$$\mathbb{E}[aH(X) + bG(Y)] = a\mathbb{E}[H(X)] + b\mathbb{E}[G(Y)].$$

- *Del Teorema 3, deduzca que $\mathbb{E}[XY] = \mathbb{E}(X)\mathbb{E}(Y)$.*

Ejercicio 1

Si X e Y son independientes, demuestre que

$$Var(X+Y) = Var(X) + Var(Y).$$

FIN

Onceava sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

18 de septiembre de 2021

Outline

La covarianza es una medida de cuánto varían juntas dos variables aleatorias respecto a sus valores esperados. Por ejemplo, la altura y el peso de las jirafas tienen covarianza positiva porque cuando una es grande la otra tiende también a ser grande.

Definición 1 (Covarianza)

Sean X e Y variables aleatorias (Discretas) con medias $\mu_X = \mathbb{E}(X)$, $\mu_Y = \mathbb{E}(Y)$. *La covarianza de X e Y es definida como*

$$\begin{aligned} Cov(X,Y) &= \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \\ &= \sum_{(x,y)\in R_{X\times Y}} (x-\mu_X)(y-\mu_Y)p(x,y), \end{aligned}$$

donde p representa la función de masa de probabilidad conjunta.

Propiedades

- $Cov(X,Y) = \mathbb{E}(XY) - \mu_X\mu_Y$.
- Si X e Y son independiente, entonces $Cov(X,Y) = 0$.
- Lo contrario es falso: covarianza nula no siempre implica independencia.
- Dados $a, b \in \mathbb{R}$, y siempre que exista, se tiene

$$Var(aX+bY) = a^2 Var(X) + b^2 Var(Y) + 2ab\, Cov(X,Y).$$

- Si X e Y son independientes, de lo anterior, se deduce que la varianza es aditiva.

Ejemplo 1

Se lanza una moneda 3 veces. Sea X el número de caras en los 2 primeros lanzamientos y sea Y el número de caras en los 2 últimos lanzamientos (para que haya superposición en el lanzamiento central). Calcule $Cov(X,Y)$.

Solución

De 3 lanzamientos, solo hay 8 resultados $\{CCC, CCS, \cdots\}$. Luego, se obtiene la tabla de probabilidad conjunta:

$X\backslash Y$	0	1	2	$p(x_i)$
0	1/8	1/8	0	1/4
1	1/8	2/8	1/8	1/2
2	0	1/8	1/8	1/4
$p(y_j)$	1/4	1/2	1/4	1

cont...

De las marginales, se obtiene que $\mathbb{E}(X) = 1 = \mathbb{E}(Y)$. De la definición, tenemos

$$cov(X,Y) = \frac{1}{8}(0-1)(0-1) + \frac{1}{8}(2-1)(2-1) = \frac{1}{4}.$$

Otra forma, hallamos

$$\mathbb{E}(XY) = 1\frac{2}{8} + 2\frac{1}{8} + 2\frac{1}{8} + 4\frac{1}{8} = \frac{5}{4}.$$

Así, $cov(X,Y) = \mathbb{E}(XY) - \mu_X\mu_Y = \frac{5}{4} - 1 = \frac{1}{4}$.

Ejemplo 2

Sea X una v.a. que toma los valores $-2, -1, 0, 1, 2$; y cada uno con probabilidad $1/5$. Sea $Y = X^2$. Mostrar que $Cov(X, Y) = 0$ pero X e Y son dependientes.

Solución

Hacemos una tabla de probabilidad conjunta:

$Y \backslash X$	-2	-1	0	1	2	$p(y_j)$
0	0	0	1/5	0	0	1/5
1	0	1/5	0	1/5	0	2/5
4	1/5	0	0	0	1/5	2/5
$p(x_i)$	1/5	1/5	1/5	1/5	1/5	1

Usando las marginales calculamos medias $\mathbb{E}(X) = 0$ y $\mathbb{E}(Y) = 2$.

cont...

Como $\mathbb{P}(X = -2, Y = 0) = 0$ y $\mathbb{P}(X = -2) \cdot \mathbb{P}(Y = 0) = 1/25$, entonces X e Y son dependientes. Finalmente, calculamos la covarianza

$$Cov(X, Y) = \frac{1}{5}(-8 - 1 + 1 + 8)) - \mu_X \mu_Y = 0.$$

Observación 1

- *Este ejemplo muestra que $Cov(X, Y) = 0$ no implica que X e Y sean independientes. De hecho, si conoces el valor de X, entonces conoces el valor de X^2 con un $100\,\%$ de certeza.*
- *El punto clave es que la $Cov(X, Y)$ mide la relación lineal entre X e Y. En el ejemplo anterior, X y X^2 tienen una relación cuadrática que la $Cov(X, Y)$ no identifica.*

Definición 2 (Correlación)

Se deduce como la se estandarización de la covarianza, lo que nos permite hacer comparaciones sin tener que preocuparnos por la escala en que se midan dichas variables. El coeficiente de correlación de X e Y, denotado a veces por ρ_{XY}, se define como

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}.$$

Observación 2

La covarianza no nula solo nos indica la dirección a la que tienden a moverse dos variables aleatorias. Sin embargo la correlación mide la fuerza de la relación lineal entre dichas variables.

Propiedades

- $-1 \leq \rho \leq 1$.
- $Corr(X, Y) = Corr(Y, X)$.
- Si X e Y son independientes, entonces la correlación es 0.
- Si $Corr(X, Y) = \pm 1$, entonces es posible que exista una dependencia lineal. Si es 1 es una correlación lineal positiva perfecta y si es -1 es negativa perfecta.
- Dada $Y = aX + b$, si $a > 0$ la correlación es 1 y si es $a < 0$ es -1.
- Si a, b, c y d son constantes con $a > 0$ y $b > 0$, entonces

$$Corr(aX + c, bY + d) = Corr(X, Y).$$

Ejemplo 3

Hallar la correlación entre X e Y en el Ejemplo 1.

Solución

Se obtuvo que $Cov(X, Y) = 1/4$. Ahora, como es usual, sea X_i el resultado del i-ésimo lanzamiento, $X_i \sim Bernoulli(0,5)$. Luego,

$$X = X_1 + X_2 \quad \wedge \quad Y = X_2 + X_3.$$

Se sabe que $\mathbb{E}(X_i) = 1/2$ y $Var(X_i) = 1/4$. Entonces, $Var(X) = 2Var(X_j) = 1/2$. Así que $\sigma_X = 1/\sqrt{2}$. De manera similar $\sigma_Y = 1/\sqrt{2}$. Por tanto

$$Corr(X, Y) = \frac{cov(X, Y)}{\sigma_X \sigma_Y} = \frac{1/4}{1/2} = \frac{1}{2}.$$

Variable aleatoria condicional

Dados $X, Y : \Omega \to \mathbb{R}$ v.a. discretas conjuntas, muchas veces estamos interesados en una v.a. condicionada "$X|Y = y$", donde $[Y = y]$ es un evento que suponemos que ocurrió. Por tanto, definimos su función de masa (condicional).

Definición 1 (Función de masa condicional)

Dado el vector aleatorio discreto (X, Y), se define la función de probabilidad condicional de X dado $Y = y$, como

$$\forall x \in R_X : p_{X|y}(x) := \mathbb{P}[Z = x] = \frac{p(x, y)}{p_Y(y)}, \quad p_Y(y) > 0,$$

donde $Z = [X|Y = y]$ es una v.a condicionada.

Ejemplo 1

Considere el lanzamiento de un dado no trucado. Sea $X = 1$, si el número es par (es decir, 2, 4 o 6) y $X = 0$ en otro caso. Además, sea $Y = 1$, si el número es primo (es decir, 2, 3 o 5) e $Y = 0$ en caso contrario. Calcule $p_X(1) = \mathbb{P}(X = 1)$, $p_{X|1}(1) = \mathbb{P}[X = 1|Y = 1]$ y compare.

Solución

Observamos que la probabilidad de que $X = 1$ es $3/6 = 1/2$ (ya que hay seis posibles resultados de los cuales tres son pares), mientras que la probabilidad de que $X = 1$ condicionada a $Y = 1$ es $1/3$ (dado que hay tres posibles números primos, 2, 3 y 5, de los cuales uno es par). La función de masa condicionada es menor, esta diferencia se debe a que el espacio muestral ha cambiado.

Momentos condicionados

El rango de "$X|Y = y$" está incluido en el rango de X, pero en general no es igual, debido a que nuestro espacio muestral sólo considera resultados para los cuales $[Y = y]$. Por lo tanto los momentos de "$X|Y = y$" se hallarán sobre este espacio muestral reducido.

Definición 2 (Esperanza condicional)

Dado (X, Y), $y \in R_Y$ t.q $\mathbb{P}[Y = y] > 0$. El valor esperado condicional de X dado $Y = y$, se denota por $\mu_{X|y}$ y se define como

$$\mu_{X|y} = \mathbb{E}[X|Y = y] = \sum_{x \in R_X} x\, p_{X|y}(x).$$

Observación 3

Se considera $p_{X|y}(x) = 0$ cuando los $\omega \in \Omega$ correspondiente a x no están en el espacio reducido $[Y = y]$.

Notación

Dado $X, Y, Z : \Omega \to \mathbb{R}$ v.a. discretas y las v.a condicionadas $X|Y = y$ y $Z|Y = y$. Se denotará:

$$\forall \lambda, \mu \in \mathbb{R} : \ [(\lambda X + \mu Z)|Y = y] := \lambda[X|Y = y] + \mu[Z|Y = y].$$

Definición 3 (Varianza condicional)

La varianza condicional de X dado $Y = y$, se denota por $\sigma^2_{X|y}$ y se define como

$$\sigma^2_{X|y} = \mathbb{E}\left[(X - \mu_{X|y})^2 | Y = y\right] = \sum_{x \in R_X} (x - \mu_{X|y})^2 \, p_{X|y}(x)\,.$$

Observación 4

Los momentos condicionados cumplen todas las propiedades de los momentos no condicionados: $\sigma^2_{X|y} = \mathbb{E}[X^2|Y = y] - \mathbb{E}[X|Y = y]^2$.

Observación 5

Si X, Y son independientes, entonces $\mathbb{E}[X|Y = y] = \mathbb{E}(X)$ siempre que $p_Y(y) > 0$.

Proposición 1

Dados (X, Y) vector aleatorio discreto, se cumple

$$\mathbb{E}(X) = \sum_y \mathbb{E}[X|Y = y]\mathbb{P}[Y = y].$$

Demostración

$$\begin{aligned}
\sum_y \mathbb{E}[X|Y = y]\mathbb{P}[Y = y] &= \sum_y \sum_x x\mathbb{P}[X = x|Y = y]\mathbb{P}[Y = y] \\
&= \sum_x x \sum_y \mathbb{P}[X = x|Y = y]\mathbb{P}[Y = y] \\
&= \sum_x x\mathbb{P}[X = x] = \mathbb{E}(X).
\end{aligned}$$

Ejemplo 2

Sea $Y_1, Y_2, \cdots$, una sucesión infinita de v.a. de Benoulli(p) independientes, donde $0 < p < 1$. Definimos X tal que

$$X = i \quad \text{si } Y_i = 1 \text{ pero } Y_j = 0 \text{ siempre que } j < i\,.$$

Calcule $\mathbb{E}(X)$ condicionando previamente a Y_1.

Solución

Es fácil ver que $X \sim Geom(p)$. Luego

$$\begin{aligned}
X|Y_1 = 1 &\quad \text{es la constante } 1; \\
X|Y_1 = 0 &\quad \text{es } 1 + Z \text{ donde } Z \sim Geom(p).
\end{aligned}$$

Solución (cont...)

Se tiene que $\mathbb{P}(Y_1 = 1) = p$ y $\mathbb{P}(Y_1 = 0) = 1 - p$. Además,

$$\mathbb{E}[X|Y_1 = 1] = \mathbb{E}(1) = 1 \quad \text{y}$$

$$\mathbb{E}[X|Y_1 = 0] = \mathbb{E}(1 + Z) = 1 + \mathbb{E}(Z) = 1 + \mathbb{E}(X).$$

Por lo tanto

$$\begin{aligned}
\mathbb{E}(X) &= \mathbb{E}[X|Y_1 = 1]\mathbb{P}(Y_1 = 1) + \mathbb{E}[X|Y_1 = 0]\mathbb{P}(Y_1 = 0) \\
&= 1 \cdot p + (1 + \mathbb{E}(X)) \cdot (1 - p) \\
&= 1 + (1 - p)\mathbb{E}(X) \Rightarrow \mathbb{E}(X) = \frac{1}{p}.
\end{aligned}$$

Ejemplo 4

Se elige repetidamente un número al azar entre los números $1, 2, \cdots, 10$ *con reposición. Sea* X *el número de repeticiones hasta que el número* 1 *aparece por primera vez e* Y *el número de repeticiones hasta que aparece el número* 10 *por primera vez. ¿Cuál es la función de masa de probabilidad condicional de* X *dado* $Y = y$*, es decir* $p_{X|y}(\cdot)$*?*

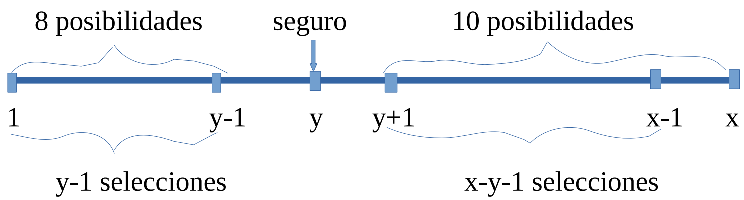
Solución

Dado $Y = y$, se tiene los siguientes casos:

- Si $x < y$ se sabe que 10 no parece en ninguna de las x primeras selecciones luego, en el experimento condicional $X \sim \text{Geom}(\frac{1}{9})$, y
$$p_{X|y}(x) = P(X = x|Y = y) = (\frac{8}{9})^{x-1}\frac{1}{9}.$$
- Si $y < x$, se sabe que en las $y - 1$ primeras selecciones no parece ni el 1 ni el 10, es decir solo se tiene 8 posibles resultados de 9 (debido a que quitamos el valor de 10), en la selección y-ésima se obtiene 10 con seguridad.

cont...

En las siguientes $x - y - 1$ selecciones, se tiene 9 posibles resultados de 10, y por último en la selección x debe ocurrir el resultado 1 con una proporción de $1/10$. Luego la probabilidad buscada es por la regla de multiplicación: $(\frac{8}{9})^{y-1} \times 1 \times (\frac{9}{10})^{x-y-1} \times \frac{1}{10}$.



Cont...

Resumiendo se tiene que la función de masa condicional (dado $y \in \{1, \cdots, 10\}$) es

$$p_{X|y}(x) = \begin{cases} (\frac{8}{9})^{x-1} \times \frac{1}{9} & , x < y \\ (\frac{8}{9})^{y-1} \times (\frac{9}{10})^{x-y-1} \times \frac{1}{10} & , x > y \\ 0 & , x = y, \end{cases}$$

para todo $x \in \{1, \cdots, 10\}$.

Referencias bibliográficas

1. Sheldon Ross. A First Course in Probability, Pearson, 9th ed. 2014.
2. `https://www.probabilitycourse.com/chapter5/5_1_2_joint_cdf.php`

FIN

Doceava sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

18 de septiembre de 2021

Outline

Teorema 1 (Fubini)

Sea $f : Q = [a, b] \times [c, d] \to \mathbb{R}$ continua. Supongamos que para cada $y \in [c, d]$, la integral $\int_a^b f(x, y)dx$ existe y además que $\int_a^b f(x, y)dx$ como una función de y es integrable en $[c, d]$. Entonces

$$\int\int_Q f = \int_c^d \int_a^b f(x, y)dxdy = \int_a^b \int_c^d f(x, y)dydx.$$

Extendemos la idea de distribuciones conjuntas de tipo discreto al de dos V.A de tipo continuo.

Definición 1 (Densidad conjunta)

Dados X e Y v.a. continuas sobre Ω. Se dice que $f(x, y)$ es la función de densidad de probabilidad conjunta de X e Y si

(a) *$f(x, y) \geq 0$ para todo $(x, y) \in \mathbb{R}^2$.*

(b) *$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dxdy = 1$.*

(c) *$\mathbb{P}[(X, Y) \in A] = \int\int_A f(x, y)dxdy$, donde $[(X, Y) \in A]$ es un evento de $\mathbb{R}^2$.*

Observación 1

La propiedad (c) implica que $\mathbb{P}[(X, Y) \in A]$ es el volumen de el sólido sobre la región A en el XY-plano y acotado por la superficie $z = f(x, y)$.

Interpretación de la función de densidad conjunta

La densidad de probabilidad conjunta de dos variables aleatorias se mide por unidad de área. Suponiendo dx y dy infinitesimales y f continua en (x, y), se tiene:

$$\mathbb{P}(x < X \leq x + dx, y < Y \leq y + dy) \approx f(x, y)dxdy.$$

Las definiciones y teoremas dados en el caso discreto se particularizan al caso continuo reemplazando las sumas por integraciones para X, Y v.a. continuas.

- Las marginales: para todo $x, y \in \mathbb{R}$

$$f_X(x) = \int_{\mathbb{R}} f(x, y)dy \ \wedge \ f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx\,.$$

- Función de distribución acumulada conjunta: para todo $(x, y) \in \mathbb{R}^2$

$$F(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(u, v)dudv,$$

- Diferenciación: Suponiendo que existen las derivadas parciales y aplicando teorema fundamental de cálculo, se obtiene que

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

Ejemplo 1

Sea el pdf conjunta de X e Y

$$f(x,y) = \left(\frac{4}{3}\right)(1-xy) \quad 0 \le x \le 1, \quad 0 \le y \le 1.$$

Hallar las pdf marginal, las medias y varianzas de X e Y y por último $\mathbb{P}(Y \le X/2)$.

Solución

Las marginales son

$$f_X(x) = \int_0^1 \left(\frac{4}{3}\right)(1-xy)dy = \left(\frac{4}{3}\right)(1-\frac{x}{2}) \quad 0 \le x \le 1,$$

y

$$f_Y(y) = \int_0^1 \left(\frac{4}{3}\right)(1-xy)dx = \left(\frac{4}{3}\right)(1-\frac{y}{2}) \quad 0 \le y \le 1,$$

cont...

La media de X es $\mu_X = E(X) = \int_0^1 x\left(\frac{4}{3}\right)(1-\frac{x}{2})dx = \left(\frac{4}{3}\right)\left(\frac{1}{2}-\frac{1}{6}\right) = \frac{4}{9}$, de manera similar la media de Y es $\mu_Y = E(Y) = \frac{4}{9}$.
La varianza de X es

$$\begin{aligned} Var(X) &= E(X^2) - [E(X)]^2 \\ &= \int_0^1 x^2\left(\frac{4}{3}\right)(1-\frac{x}{2})dx - \frac{16}{81} \\ &= \left(\frac{4}{3}\right)\left(\frac{1}{3}-\frac{1}{8}\right) - \frac{16}{81} = \frac{13}{162} \end{aligned}$$

Tambien la varianza de Y es $Var(Y) = \frac{13}{162}$.

cont...

Finalmente, $\mathbb{P}(Y \le X/2)$ es calculada por Fubini:

$$\begin{aligned} \mathbb{P}(Y \le X/2) &= \int_0^1 \int_0^{x/2} \left(\frac{4}{3}\right)(1-xy)dydx \\ &= \int_0^1 \left(\frac{4}{3}\right)(\frac{x}{2}-\frac{x^3}{8})dx \\ &= \left(\frac{4}{3}\right)(\frac{1}{4}-\frac{1}{32}) = \frac{7}{34}. \end{aligned}$$

Ejemplo 2

La función de distribución acumulada conjunta de dos v.a. continuas es dada por:

$$F_{X,Y}(x,y) = (1-e^{-\lambda x})(1-e^{-\mu y}), \; x,y \ge 0, \; \lambda,\mu > 0,$$

y es cero en otro caso. Calcular $\mathbb{P}(\mathcal{C})$, con $\mathcal{C} = \{0 \le x \le 1, 0 \le y \le 1\}$.

Solución

Aplicando la Proposición 3 de la Sesión 12, obtenemos

$$\begin{aligned} \mathbb{P}(\mathcal{C}) &= F_{X,Y}(1,1) + F_{X,Y}(0,0) - F_{X,Y}(0,1) - F_{X,Y}(1,0) \\ &= (1-e^{-\lambda})(1-e^{-\mu}) + (1-e^0)(1-e^0) \\ &- (1-e^0)(1-e^{-\mu}) - (1-e^{-\lambda})(1-e^0) \\ &= (1-e^{-\lambda})(1-e^{-\mu}). \end{aligned}$$

Ejemplo 3

Del ejemplo anterior, podemos calcular las distribuciones marginales tomando límite en $F_{X,Y}(x,y)$:

$$F_X(x) = (1 - e^{-\lambda x})(1 - 0) = (1 - e^{-\lambda x}),$$
$$F_Y(y) = (1 - 0)(1 - e^{-\mu y}) = (1 - e^{-\mu y}).$$

Proposición 1 (Independencia)

Dados X, Y v.a. continuas. Entonces, son equivalentes

i) *X e Y son independientes.*

ii) $F_{X,Y}(x,y) = F_X(x)F_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$.

iii) $f(x,y) = f_X(x)f_Y(y) \quad \forall (x,y) \in \mathbb{R}^2$.

Ejemplo 4

Sean X e Y son variables aleatorias con una función de distribución conjunta:

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{si } \ x, y \geq 0 \\ 0 & \text{en otros casos.} \end{cases}$$

Determinar si X e Y son independientes.

Solución

Si son independientes, usando (ii), solo basta calcular F_X y F_Y y multiplicarlas, por ejemplo

$$F_X(x) = \lim_{y \to \infty} F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} & \text{si } \ x \geq 0 \\ 0 & \text{en otro casos} \end{cases}$$

Ejemplo 5

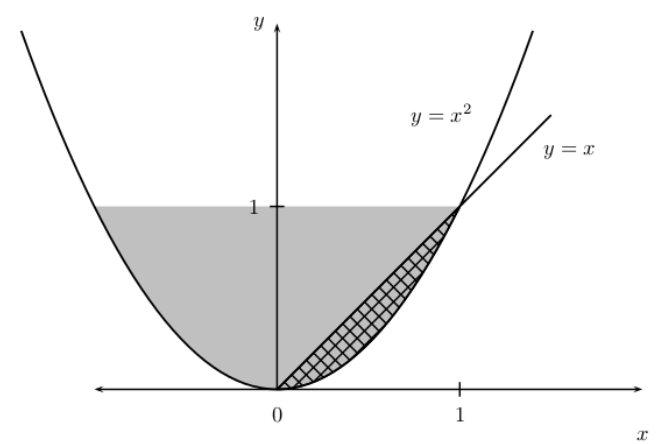
Sean X e Y variables aleatorias continuas, que tienen una función de densidad conjunta:

$$f_{X,Y}(x,y) = \begin{cases} cx^2y & \text{si } \ x^2 \leq y \leq 1 \\ 0 & \text{en otro caso.} \end{cases}$$

Hallar el valor de c y $\mathbb{P}(X \geq Y)$.

Solución

Notemos primero que $-1 \leq x \leq 1$. Para encontrar el valor de c, fijemos un valor x y dejemos que y varie en su rango, el cuál es $x^2 \leq y \leq 1$.



$$1 = \int\int f(x,y)dxdy = c\int_{-1}^{1}\int_{x^2}^{1} x^2 y dydx$$
$$= c\int_{-1}^{1} x^2\left[\int_{x^2}^{1} ydy\right]dx = c\int_{-1}^{1} x^2 \frac{1-x^4}{2}dx = \frac{4c}{21}.$$

Calculemos $\mathbb{P}(X \geq Y)$:
Esto corresponde al conjunto $A = \{(x, y); 0 \leq x \leq 1, x^2 \leq y \leq x\}$. (Se puede ver esto, en el gráfico anterior). Así:

$$\mathbb{P}(X \geq Y) = \frac{21}{4}\int_0^1 \int_{x^2}^{x} x^2 y dy dx = \frac{21}{4}\int_0^1 x^2 \Big[\int_{x^2}^{x} y dy\Big] dx$$
$$= \frac{21}{4}\int_0^1 x^2 \Big(\frac{x^2 - x^4}{2}\Big) dx = \frac{3}{20}.$$

Teorema 2

Sean X, Y v.a. continuas y $h(X, Y)$ una nueva variable aleatoria. Entonces

$$\mathbb{E}(h(X, Y)) = \int_{\mathbb{R}}\int_{\mathbb{R}} h(x, y) f_{X,Y}(x, y) dxdy,$$

siempre que $\int_{\mathbb{R}}\int_{\mathbb{R}} |h(x, y)| f_{X,Y}(x, y) dxdy < \infty$.

Ejemplo 6

Sean X, Y v.a. continuas, cuya función densidad conjunta es:

$$f_{X,Y}(x, y) = \begin{cases} 1 - (x + y)/3, & 0 \leq x \leq 1, 0 \leq y \leq 2, \\ 0 & \text{en otros casos} \end{cases}$$

Hallar $\mathbb{E}(XY)$.

Solución

$$\mathbb{E}(XY) = \int_{y=0}^{2}\Big(\int_{x=0}^{1} xy[1 - (x + y)/3] dx\Big) dy$$
$$= \int_{y=0}^{2}\Big(\frac{x^2 y}{2} - \frac{x^3 y}{9} - \frac{x^2 y^2}{6}\Big)\Big|_{x=0}^{1} dy = \int_{y=0}^{2}\Big(\frac{y}{2} - \frac{y}{9} - \frac{y^2}{9}\Big) dy$$
$$= \Big(\frac{7y^2}{36} - \frac{y^3}{18}\Big)\Big|_0^2 = \frac{12}{36} = 1/3$$

Definición 2 (Función de densidad condicional)

Dado X, Y v.a. continuas. La `función densidad condicional` de Y dado que $X = x$, denotada por $f_{Y|x}(\cdot)$ es definida como:

$$f_{Y|x}(y) = \frac{f(x, y)}{f_X(x)} \quad \forall y \in \mathbb{R},$$

para x que cumplen que $f_X(x) > 0$.

Observación 2

No tiene sentido hallar $\mathbb{P}(Y = y|X = x)$ para la v.a. continua "$Y|X = x$". Por lo tanto procedemos a hallar su función distribución acumulada condicional que denotamos por $G(y)$ para $y \in \mathbb{R}$.

cont...

Usando las identidades $\int_x^{x+\delta} f(u, v) du \approx f(x, v)\,\delta$, $\int_x^{x+\delta} f_X(t) dt \approx f_X(x)\,\delta$ y Fubini, se obtiene

$$G(y) := \lim_{\delta \to 0+} \mathbb{P}(Y \leq y | x \leq X \leq x + \delta)$$
$$= \lim_{\delta \to 0+} \frac{\mathbb{P}(Y \leq y, x \leq X \leq x + \delta)}{\mathbb{P}(x \leq X \leq x + \delta)}$$
$$= \lim_{\delta \to 0+} \frac{\int_x^{x+\delta}\int_{-\infty}^{y} f(u, v) dv du}{\int_x^{x+\delta} f_X(t) dt}$$
$$= \frac{\int_{-\infty}^{y} f(x, v)\, dv}{f_X(x)},$$

luego, derivando con respecto a "y" y usando el TFC, se obtiene una expresión para la función de densidad condicional.

Ejemplo 7

Dada la función densidad conjunta

$$f(x,y) = \begin{cases} x+y & \text{si } 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{en otros casos.} \end{cases}$$

Encontrar $\mathbb{P}(X < 1/4 | Y = 1/3)$.

Solución

Como $f_Y(y) = y + (1/2)$, tenemos que $f_{X|y}(x) = \dfrac{f(x,y)}{f_Y(y)} = \dfrac{x+y}{y+\frac{1}{2}}$.

Entonces

$$\begin{aligned}\mathbb{P}\left(X < \frac{1}{4} \Big| Y = \frac{1}{3}\right) &= \int_0^{1/4} f_{X|Y}\left(x \;\Big|\; \frac{1}{3}\right) dx \\ &= \int_0^{1/4} \frac{x+\frac{1}{3}}{\frac{1}{3}+\frac{1}{2}} dx = \frac{\frac{1}{32}+\frac{1}{12}}{\frac{1}{3}+\frac{1}{2}} = \frac{11}{80}.\end{aligned}$$

Definición 3 (Momentos condicionales)

Siempre que $f_X(x) > 0$, se tiene

- $\mathbb{E}(Y|X=x) = \displaystyle\int_{\mathbb{R}} y\, f_{Y|x}(y) dy = \frac{\int_{\mathbb{R}} y\, f(x,y)}{f_X(x)} dy,$
- $Var(Y|X=x) = \mathbb{E}[Y^2|X=x] - \mathbb{E}[Y|X=x]^2.$

Proposición 2

Dados X, Y v.a. continuas y $h(X,Y)$ variable aleatoria. Entonces

$$\mathbb{E}[h(X,Y)|Y=y] = \int_{-\infty}^{\infty} h(x,y) f_{X|y}(x) dx.$$

Teorema 3

Si X e Y son variables aleatorias continuas conjuntas, entonces:

$$\mathbb{E}(Y) = \int_{\mathbb{R}} \mathbb{E}(Y|X=x) f_X(x) dx$$

donde la integral está definida para los valores de x, tal que $f_X(x) > 0$.

Demostración

$$\begin{aligned}\mathbb{E}(Y) &= \int y\, f_Y(y) dy = \iint y\, f(x,y)\, dxdy \\ &= \iint y\, f_{Y|x}(y) f_X(x) dxdy \\ &= \int \left(\int y\, f_{Y|x}(y) dy \right) f_X(x) dx\end{aligned}$$

Observación 3

En otras palabras, al calcular $\mathbb{E}(Y)$ podemos fijar primero el valor de X y luego promediar sobre este valor más tarde.

Ejemplo 8

Sean X y Y dos variables aleatorias continuas, cuya función de probabilidad de densidad conjunta es dada por:

$$f_{X,Y}(x,y) = \begin{cases} 1-(x+y)/3, & 0 \le x \le 1,\ 0 \le y \le 2 \\ 0 & \text{en otros casos.} \end{cases}$$

Encontrar $\mathbb{E}(X|Y=0{,}5)$ y $\mathbb{E}(Y|X=0{,}5)$.

Solución

Hallamos las marginales de X e Y:

$$f_X(x) = \int_0^2 [1-(x+y)/3] dy = \frac{4}{3} - \frac{2x}{3}, \quad 0 \le x \le 1,$$

$$f_Y(y) = \int_0^1 [1-(x+y)/3] dx = \frac{5}{6} - \frac{y}{3}, \quad 0 \le y \le 2.$$

Luego, las funciones de densidad condicionales:

$$f_{X|0,5}(x) = \frac{f_{X,Y}(x, 0{,}5)}{f_Y(0{,}5)} = \frac{1 - x/3 - 0{,}5/3}{5/6 - 0{,}5/3} = 5/4 - x/2,$$
$$f_{Y|0,5}(y) = \frac{f_{X,Y}(0{,}5, y)}{f_X(0{,}5)} = \frac{1 - 0{,}5/3 - y/3}{4/3 - 1/3} = 5/6 - y/3.$$

Finalmente, hallamos las esperanzas condicionales

$$\begin{aligned}\mathbb{E}(X|Y = 0{,}5) &= \int_{-\infty}^{\infty} x\, f_{X|0,5}(x)\,dx \\ &= \int_0^1 x\left(\frac{5}{4} - \frac{x}{2}\right)dx = \left(\frac{5x^2}{8} - \frac{x^3}{6}\right)\Bigg|_0^1 = 11/24.\end{aligned}$$

De manera similar $\mathbb{E}(Y|X = 0{,}5) = 7/9$.

FIN

Treceava sesión

Cálculo de Probabilidades - CM1H2

Los profesores[1]

[1]Facultad de Ciencias
Universidad Nacional de Ingeniería

18 de septiembre de 2021

Outline

Teorema 1 (Cambio de Variable)

Supongamos que hacemos un cambio de las variables (transformación) de (x, y) *a* (u, v)*, digamos con* $x = g(u, v), y = h(u, v)$. *Entonces sobre los límites apropiados de integración, se tiene que*

$$\int\int f(x,y)dxdy = \int\int f(g(u,v),h(u,v))\left|\frac{\partial(x,y)}{\partial(u,v)}\right|dudv,$$

donde

$$\frac{\partial(x,y)}{\partial(u,v)} = det\begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}.$$

Asumimos que las derivadas parciales existen y son continuas y que el determinante es distinto de cero.

Ejemplo 1

Encontremos el área de un círculo de radio 1. Para encontrar el área de una región, sólo necesitamos integrar 1 sobre esa región (por lo que cualquier dilema proviene de los límites de la integración, la función que estamos integrando, es sólo la constante 1). Así el área es

$$\int\int_{x^2+y^2\leq 1} 1dxdy = \int_{-1}^{1}\int_{-\sqrt{1-y^2}}^{\sqrt{1-y^2}} 1dxdy = 2\int_{-1}^{1}\sqrt{1-y^2}dy.$$

Nota que los límites para la variable interna (x) de la integral doble pueden depender de la variable exterior (y), mientras que los límites externos son constantes. La última integral puede hacerse con una sustitución trigonométrica, pero en lugar de ello, vamos a simplificar el problema transformándolo por **Coordenadas polares**.

cont...

Sean

$$x = r\cos\theta, \quad y = r\sin\theta,$$

donde r es la distancia desde (x, y) al origin y $\theta \in [0, 2\pi)$ es el ángulo. La matriz Jacobiana de esta transformación es

$$\frac{d(x,y)}{d(r,\theta)} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$$

así el valor del determinante es $r(\cos^2\theta + \sin^2\theta) = r$. Esto es, $dxdy$ llega a ser $rdrd\theta$. Luego es área del círculo es

$$\int_0^{2\pi}\int_0^1 rdrd\theta = \int_0^{2\pi}\frac{1}{2}d\theta = \pi.$$

Para un círculo de radio r, se sigue que el valor del área es πr^2, ya que podemos imaginar convertir nuestras unidades de medida a la unidad para la cual el radio es 1.

Teorema 2 (Técnica del Jacobiano)

Dado $g : \mathbb{R}^2 \to \mathbb{R}^2$ *un función de clase* C^1 *con inversa* $h := g^{-1}$*, el vector aleatorio* $X = (X_1, X_2)$ *tal que* X_1, $X_2 : \Omega \to \mathbb{R}$. *Entonces el vector aleatorio* $Y = (Y_1, Y_2) := g(X)$*, tiene función de densidad conjunta:*

$$f_Y(y_1,y_2) = (f_X \circ h)(y_1,y_2)\,|\det J\,h(y_1,y_2)|\,, \tag{1}$$

donde $J\,h$ *representa el jacobiano de* h*, es decir:*

$$Jh(y_1,y_2) = \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix}(y_1,y_2)$$

donde $x_i = h_i(y_1, y_2)$, $i = 1, 2$ *y* $h = (h_1, h_2)$.

Ejemplo 2

Sean Z_1, Z_2 normales estándar independientes y $\rho \in (-1,1)$. Halle la función de densidad conjunta del vector aleaorio (X_1, X_2) dado por

$$X_1 = Z_1 \quad , \quad X_2 = \rho Z_1 + \sqrt{1-\rho^2} Z_2 \,.$$

Solución

Por la independencia, (Z_1, Z_2) tienen distribución conjunta:

$$f_{Z_1,Z_2}(z_1,z_2) = \frac{1}{2\pi} \exp\left\{ -\frac{1}{2}(z_1^2 + z_2^2) \right\}$$

Usamos para el método del jacobiano, el cambio de variable:

$$(Z_1, Z_2) = h(X_1, X_2) = \left(X_1, -\frac{\rho}{\sqrt{1-\rho^2}} X_1 + \frac{1}{\sqrt{1-\rho^2}} X_2 \right).$$

cont...

Calculamos su jacobiano:

$$Jh(x_1,x_2) = \begin{bmatrix} 1 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{bmatrix}.$$

Entonces, concluimos remmplazando

$$\begin{aligned} f_{X_1,X_2}(x_1,x_2) &= f_{Z_1,Z_2}(h(x_1,x_2))\,|\det Jh(x_1,x_2)| \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} f_{Z_1,Z_2}\left(x_1, -\frac{\rho}{\sqrt{1-\rho^2}}x_1 + \frac{1}{\sqrt{1-\rho^2}}x_2\right) \end{aligned}$$

cont...

$$\begin{aligned} &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2}\left[x_1^2 + \frac{\rho^2}{1-\rho^2}x_1^2 + \frac{1}{1-\rho^2}x_2^2 - \frac{2\rho}{1-\rho^2}x_1x_2 \right] \right\} \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{ -\frac{1}{2(1-\rho^2)}\left[(1-\rho^2)x_1^2 + \rho^2 x_1^2 + x_2^2 - 2\rho x_1 x_2 \right] \right\} \end{aligned}$$

Como $\mathbb{E}X_1 = \mathbb{E}X_2 = 0$ y $Var\,X_1 = Var\,X_2 = 1$, entonces

$$\begin{aligned} Corr(X_1,X_2) &= Cov(Z_1, \rho Z_1 + \sqrt{1-\rho^2}Z_2) \\ &= \rho Cov(Z_1,Z_1) + \sqrt{1-\rho^2}\,\underbrace{Cov(Z_1,Z_2)}_{0} \\ &= \rho\,\underbrace{Var\,Z_1}_{1} = \rho\,. \end{aligned}$$

Las variables aleatorias rara vez son independientes en la práctica. Así que se busca obtener la estructura de dependencia entre ellas. Estudiaremos las bivariadas por los sgtes motivos:

i) la estructura de correlación y otras propiedades son más fáciles de entender y la gráfica de densidad conjunta se puede mostrar más fácilmente.

ii) Una distribución bivariada se puede normalmente extender a uno multivariado a través de una representación vectorial o matricial.

Por ejemplo, podemos mencionar la Distribución normal bivariada, la cual modela muchos fenómenos naturales, al igual que en el caso univariado, véase [1, pag 272] ó [2]. A continuación, estudiaremos el modelo más simple para determinar la dependencia de una variable aleatoria Y respecto a otra X.

Regresión lineal simple

Definición 1

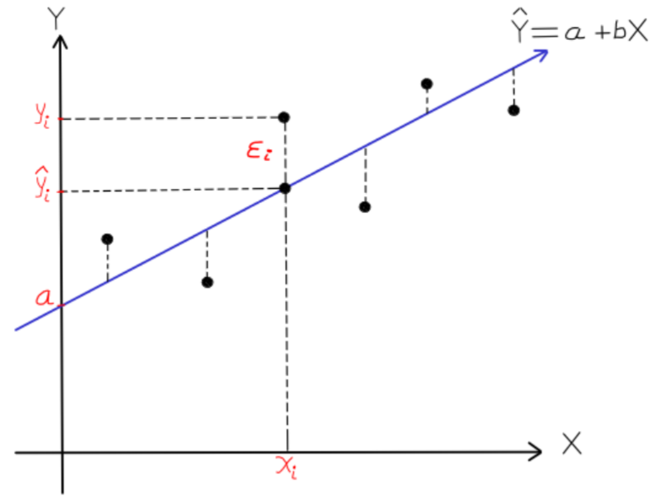
Dada una muestra de n datos $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ observados de un vector aleatorio (X, Y). Si existe una relación lineal (o aproximada) entre las v.a. X e Y, entonces a la recta $\hat{Y} = a + bX$, se le llama regresión lineal simple de la variable Y (dependiente) con respecto a X (independiente).

Observación 1

En el caso de observar una relación lineal aproximada, se calcula la recta que mejor ajuste los valores de la muestra. Gauss propuso en 1809 el método de mínimos cuadrados para calcular dicha recta: $\hat{Y} = a + bX$, donde $\hat{Y}$ representa la v.a. generada por la regresión.

Definición 2

Se llama REGRESIÓN al proceso de obtener el modelo de regresión, analizar su validez y predecir Y dado X.



Notación

- $\hat{y}_i$ representa un valor de $\hat{Y}$, cuando X tome el valor x_i, es decir $\hat{y}_i = a + bx_i$. Al valor $\hat{y}_i$ se denomina valor pronosticado.
- ε_i representa el error que se comete por la regresión, es decir $\varepsilon_i = y_i - \hat{y}_i$, donde y_i es el valor observado por la v.a. Y.

Observación 2

Se considera que las v.a. X, Y están distribuidas normalmente y por tanto los errores se distribuyen también de manera normal. Es decir $\varepsilon_i \sim N(0, \sigma^2)$, donde σ^2 indica la distancia (cuadrada) "promedio" de los valores observados a la recta de regresión. Por la normalidad de los errores, se espera que el 95 % de los puntos estén comprendidos en $[-2\sigma, 2\sigma]$.

Método de mínimo cuadrados

Consiste en minimizar la suma de los errores al cuadrado ε_i^2:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2.$$

Así, se obtiene los coeficientes a y b :

$$a = \overline{y} - b\overline{x},$$

donde $\overline{x}$ y $\overline{y}$ representa la media de los valores observados de X e Y, respectivamente y la pendiente se halla como

$$b = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - (\sum x_i)^2}.$$

Para estudiar la relación lineal existente entre dos variables continuas es necesario disponer de parámetros que permitan cuantificar dicha relación. Por ejemplo, la covarianza, la correlación y el coeficiente de determinación.

Definición 3 (Determinación)

El coeficiente de determinación de un vector aleatorio $(X;Y)$ se define como el cuadrado del coeficiente de correlación de (X,Y).

Proposición 1

Dado el conjunto de valores (x_i, y_i), $i = 1, \cdots, n$ de un vector aleatorio (X, Y), el coeficiente de correlación se aproxima como

$$\rho^2 := Corr(X,Y)^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \overline{y})^2}{\sum_{i=1}^n (y_i - \overline{y})^2}, \tag{2}$$

donde $\hat{y}_i$ denota los valores de la regresión e $\overline{y}$ el promedio de los y_i.

Interpretación

El coeficiente de determinación es una medida que se utiliza para explicar la proporción de la variación en la variable dependiente que es predecible por la variabilidad de la variable independiente. Se conoce también como "bondad de ajuste" del modelo. También podemos pensar que el $\rho^2 \times 100\,\%$ de las veces el modelo de regresión aproxima correctamente. Y $(1-\rho^2) \times 100\,\%$ representa el porcentaje de variación no explicada por la recta de regresión.

Observación 3

La identidad 2 es determinada en el siguiente Ejercicio. Véase la ecuación 3.

Deducción del método de mínimos cuadrados

Ejercicio

Sea el vector aleatorio (X,Y) aproximado por el conjunto de valores (x_i, y_i), $i \in \{1, 2, \cdots, n\}$ y sea la v.a $\hat{Y} := a + bX$. Se dice que $\hat{Y}$ es la regresión lineal de Y, si minimiza la siguiente suma:

$$S = \sum_{i=1}^n \varepsilon_i^2, \quad \varepsilon_i := y_i - \hat{y}_i.$$

a) Determine los coeficientes a y b de $\hat{Y}$, hallando los puntos críticos de la función $S = S(a,b)$.

b) Demuestre que el punto $(\bar{x}, \bar{y})$ pertenece a la recta $\mathcal{L} : \hat{Y} := a + bX$. y que $\mathbb{E}(Y) = \mathbb{E}(\hat{Y})$

Ejercicio (cont...)

c) Demuestre que $\sum_{i=1}^n \varepsilon_i x_i = 0$ y concluya que $Var(Y) = Var(\hat{Y}) + S/n$.

d) Demuestre

$$\frac{S}{n} = Var(Y) + b^2 Var(X) - 2b\, Cov(X,Y).$$

e) Dado el coeficiente de correlación $\rho := \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$, demuestre que el coeficiente de determinación se puede escribir como

$$\rho^2 = \frac{Var(\hat{Y})}{Var(Y)}. \tag{3}$$

Solución

Como todos los valores son distintos entonces las funciones de masa son

$$p_X(x_i) = p_Y(y_i) = p_{XY}(x_i y_i) = \frac{1}{n} \quad \text{para todo } i \in \{1, \cdots, n\}.$$

a) Hallamos los puntos críticos, es decir hacemos $\nabla S = 0$:

$$\begin{aligned} S_a &= \sum_1^n 2(y_i - a - bx_i)(-1) = 0 \\ &\rightarrow \sum_1^n y_i = \sum_1^n (a + bx_i) \\ &\rightarrow \sum_1^n y_i = na + b\sum_1^n x_i \\ a &= \bar{y} - b\bar{x} \end{aligned} \tag{4}$$

cont...

$$\begin{aligned} S_b &= \sum_1^n 2(y_i - a - bx_i)(-x_i) = 0 \\ &\rightarrow \sum_1^n x_i y_i = a\sum_1^n x_i + b\sum_1^n x_i^2 \\ &\rightarrow \mathbb{E}[XY] = a\mathbb{E}(X) + b\mathbb{E}[X^2] \\ &\rightarrow \mathbb{E}[XY] = (\bar{y} - b\bar{x})\bar{x} + b\mathbb{E}[X^2] \\ &\rightarrow \mathbb{E}[XY] = \bar{x}\bar{y} + bVar(X) \\ &\rightarrow b = \frac{\mathbb{E}[XY] - \bar{x}\bar{y}}{Var(X)} \\ b &= \frac{Cov(X, Y)}{Var(X)} \end{aligned} \tag{5}$$

cont...

b) Se deduce directamente de (4) que $(\bar{x}, \bar{y}) \in \mathcal{L}$. Además de $S_a = 0$, se tiene

$$\sum_i^n (y_i - a - bx_i) = 0$$

$$\sum_i^n \varepsilon_i = 0 \tag{6}$$

luego, $\mathbb{E}(Y) = \mathbb{E}(\hat{Y}) = \bar{y}$.

c) De $S_b = 0$, se deduce que

$$\sum_i^n (y_i - a - bx_i)x_i = 0$$

$$\sum_i^n \varepsilon_i x_i = 0 \tag{7}$$

cont...

Además, usando (6) y (7), se tiene

$$\begin{aligned} \sum_i^n \varepsilon_i(\hat{y}_i - \bar{y}) &= \sum_i^n \varepsilon_i \hat{y}_i - \bar{y}\cancel{\sum_i^n \varepsilon_i} \\ &= \sum_i^n \varepsilon_i (a + bx_i) \\ &= a\cancel{\sum_i^n \varepsilon_i} + b\cancel{\sum_i^n \varepsilon_i x_i} \\ &= 0\,. \end{aligned} \tag{8}$$

Ahora, de (8) y la definición de varianza, se tiene cont...

$$
\begin{aligned}
n\,Var(Y) &= \sum_1^n (y_i - \bar{y})^2 \\
&= \sum_1^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\
&= \sum_1^n (y_i - \hat{y}_i)^2 + \sum_1^n (\hat{y}_i - \bar{y})^2 + \cancel{2\sum_i \varepsilon_i(\hat{y}_i - \bar{y})} \\
&= \sum_1^n \varepsilon_i^2 + n\,Var(\hat{Y})\,.
\end{aligned}
$$

Entonces

$$
Var(Y) = Var(\hat{Y}) + \frac{S}{n}, \tag{9}
$$

donde S/n es el error cuadrático medio.

cont...

d) Como sólo se observa los pares (x_i, y_i), $i = 1, \cdots, n$, cualquier otro par no ocurre, luego la función de masa conjunta es

$$
\forall i \in \{1, \cdots, n\} : \quad p(x_i, y_i) = \frac{1}{n}\,.
$$

Esto implica: $Cov(X, Y) = \dfrac{1}{n}\displaystyle\sum_1^n (x_i - \bar{x})(y_i - \bar{y})$. Del error cuadrático, se tiene

$$
\begin{aligned}
\sum_1^n (y_i - \hat{y}_i)^2 &= \sum_1^n (y_i - a - bx_i)^2 \\
&= \sum_1^n (y_i - \bar{y} + b\bar{x} - bx_i)^2 \\
&= \sum_1^n \left[(y_i - \bar{y}) - b(x_i - \bar{x})\right]^2
\end{aligned}
$$

cont...

$$
= \sum_1^n (y_i - \bar{y})^2 + b^2 \sum_1^n (x_i - \bar{x})^2 - 2b \sum_1^n (x_i - \bar{x})(y_i - \bar{x}).
$$

Por tanto, de la masa conjunta, se tiene

$$
\frac{S}{n} = Var(Y) + b^2 Var(X) - 2b\,Cov(X, Y). \tag{10}
$$

e) De la definición del coeficiente de correlación y (5), se tiene

$$
\begin{aligned}
\rho^2 &= \frac{Cov(X, Y)}{Var(X)} \cdot \frac{Cov(X, Y)}{Var(Y)} \\
&= \frac{b\,Cov(X, Y)}{Var(Y)}. \qquad (11)
\end{aligned}
$$

cont...

De (10), (5) y (9) , se tiene

$$
\begin{aligned}
\frac{S}{n} &= Var(Y) + b \cdot b\,Var(X) - 2b\,Cov(X, Y) \\
&= Var(Y) - b\,Cov(X, Y) \\
&= Var(\hat{Y}) + \frac{S}{n} - b\,Cov(X, Y).
\end{aligned}
$$

Por tanto, se tiene

$$
Var(\hat{Y}) = b\,Cov(X, Y). \tag{12}
$$

Reemplazando (12) en (11), se concluye.

Ejemplo 3

Una fábrica de cierta marca de refrescos ha tomado al azar 10 semanas al año, observando la temperatura media correspondiente (en grados centígrados) a cada una de ellas y la cantidad de refrescos pedidos durante cada uno de dichos períodos. La información obtenida es la siguiente:

$$\begin{array}{rl} \text{T.media}(^\circ\text{C}) : & 10\ \ 28\ \ 12\ \ 31\ \ 30\ \ 19\ \ 24\ \ \ 5\ \ \ 9\ \ 15 \\ \text{Cant. de refr.} : & 21\ \ 65\ \ 19\ \ 72\ \ 75\ \ 39\ \ 67\ \ 11\ \ 12\ \ 24 \end{array}$$

Calcular:

a) La recta de ajuste, es decir el grado de dependencia de la cantidad de refrescos (Y) respecto a la temperatura (X).

b) El coeficiente de determinación y el coeficiente de correlación. ¿Con este coeficiente se podría planificar la producción?

x	y	xy	x^2	y^2
10	21	210	100	441
28	65	1820	784	4225
12	19	228	144	361
31	72	2232	961	5184
30	75	2250	900	5625
19	39	741	361	1521
24	67	1608	576	4489
5	11	55	25	121
9	12	108	81	144
15	24	360	225	576
183	405	9612	4157	22687

Solución

Usando los cálculos de la tabla para $n = 10$ semanas, se tiene:

i) La recta de regresión de Y con respecto a X, está dada por $\hat{Y} - \bar{y} = b(X - \bar{x})$, de donde

$$\bar{x} = \frac{183}{10} = 18{,}3\,, \quad \bar{y} = \frac{405}{10} = 40{,}5\,, \quad b = \frac{Cov(X,Y)}{\sigma_X^2},$$

ii) Calculamos $\sigma_X^2 = \frac{\sum x^2}{n} - \bar{x}^2 = \frac{4157}{10} - 18{,}3^2 = 415{,}7 - 334{,}89 = 80{,}81$ y

$$\begin{aligned} Cov(X,Y) &= \frac{\sum xy}{n} - \bar{x}\bar{y} = \frac{9612}{10} - 18{,}3 \times 40{,}5 \\ &= 961{,}2 - 741{,}15 = 220{,}05. \end{aligned}$$

Solución (cont...)

iii) Entonces, $b = \frac{220{,}05}{80{,}81} = 2{,}723$ y por lo tanto la recta de regresión es

$$y = -9{,}331 + 2{,}723x\,.$$

iV) El grado de relación que hay entre las variables, viene dada por el coeficiente de correlación:

$$\rho := \frac{Cov(X,Y)}{\sigma_X\,\sigma_Y}.$$

v) Hallamos $\sigma_Y^2 = \frac{\sum y^2}{n} - \bar{y}^2 = \frac{22687}{10} - 40{,}5^2 = 2268{,}7 - 1640{,}25 = 628{,}45.$ Entonces, el coeficiente de determinación será

$$\rho^2 = \frac{220{,}05^2}{80{,}81 \times 628{,}45} = 0{,}95305$$

Solución (cont...)

vi) Luego, el de correlación es $\rho = 0{,}976$.

vii) Gracias al coeficiente de determinación, se sabe que el 95,3 % de la variabilidad de la producción semanal de refrescos se explicará por sus respectivas variaciones de temperatura.

viii) Como este porcentaje es alto se podrá planificar la producción utilizando la recta de regresión.

Referencias bibliográficas

1. Sheldon Ross. A First Course in Probability, Pearson, 9th ed. 2014.
2. `https://www.probabilitycourse.com/chapter5/5_3_2_bivariate_normal_dist.php`

FIN