Modelos probabilísticos y análisis estadístico

Introducción a probabilidad

Carlos Ricardo Bojacá

carlos.bojaca@utadeo.edu.co

Departamento de Ciencias Básicas
Facultad de Ciencias Naturales e Ingeniería
Universidad de Bogotá Jorge Tadeo Lozano

Tabla de contenidos

Introducción

Modelo

En investigación es común que se requiera describir diferentes tipos de fenómenos en términos matemáticos. En este contexto, un modelo corresponde a una idealización (simplificación) matemática usada para aproximarse a un fenómeno observable. El modelo será satifactorio siempre y cuando logre representar de la manera más fiel posible la realidad. Con el fin de determinar la validez del modelo *se realizan experimentos que permitan el registro de observaciones*

Modelo determinístico

En un modelo deterministico, el resultado está dado por las condiciones que establecen al momento de realizar el experimento.

* La distancia d (km) recorrida por un automóvil en t (h) a una velocidad constante v $(km \cdot h^{-1})$ está determinada por la relación $d = v \times t \implies v$ y t determinan el valor de d.

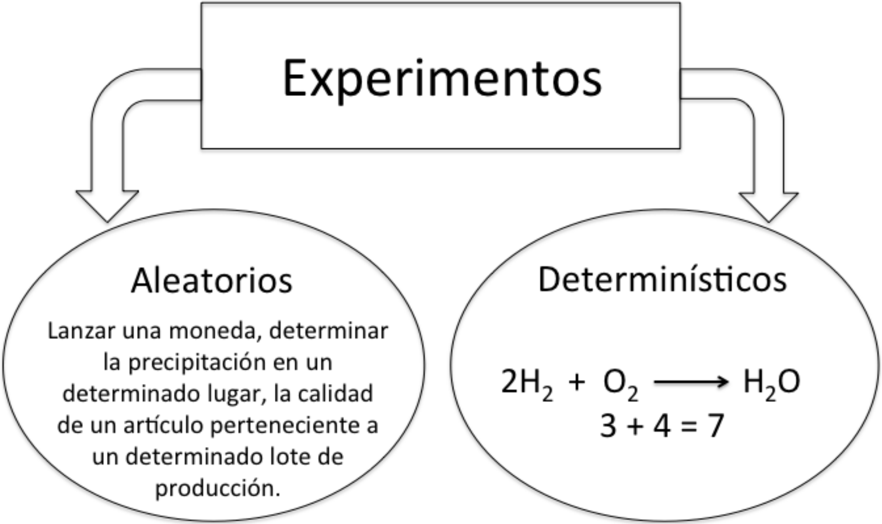
Introducción

Modelo estocástico

En un modelo estocástico, es aquel en el cual la información disponible no permite generar una relación o regla para determinar el resultado de un experimento. Muchos fenómenos (naturales o artificiales) son considerados **aleatorios** debido a que el resultado exacto no puede ser predicho; en este caso los resultados de mediciones repetidas sobre circunstancias similares generan un patrón.

* En un experimento aleatorio E, el conjunto de todos los posibles resultados de E se llama espacio muestral y se denota por la letra griega Ω. Ejemplo: en el experimento de lanzar una moneda al aire, Ω estaría compuesto por los posibles resultados: **cara** o **sello**; $\Omega = \{Ca, Se\}$. En la ejecución del experimento aletorio, aunque el resultado es impredecible, se sabe que debe corresponder a una de las opciones fijadas por el espacio muestral.

Introducción



Qué es probabilidad

Ocurrencia de un evento

- Ganar la lotería
- Tener un día soleado
- Adquirir una enfermedad
- Estimar el tiempo que tarda un operario en terminar un producto
- Evolución de un paciente sometido a un tratamiento

Medida numérica de la incertidumbre

Definiciones

Espacio Muestral (Ω): es el conjunto de todos los posibles resultados de un experimento aleatorio

Evento o suceso ($A : \ldots$): es cualquier subconjunto del espacio muestral

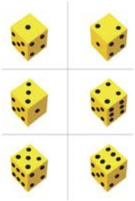
Eventos Excluyentes (Disyuntos): Sean A y B dos eventos de Ω si $A \cap B = \emptyset$, estos dos eventos se dicen excluyentes

Enfoques de probabilidad

Probabilidad

Clásica Empírica Subjetiva

Probabilidad clásica (teórica)



Probabilidad clásica (teórica)

Si un experimento aleatorio resulta de n formas *igualmente posibles* y *mutuamente excluyentes* y si n_A de dichos resultados favorecen a la situación de interés A, la probabilidad de A, denotada $P(A)$, es:

$$P(A) = \frac{n_A}{n} = \frac{\text{Casos favorables}}{\text{Casos posibles}}$$

Probabilidad clásica (teórica)

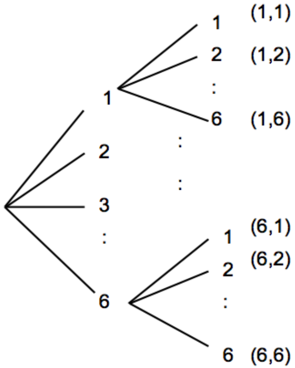
Experimento aleatorio: lanzamiento de dos dados

Espacio muestral: $\Omega = \{(1, 1)\ (1, 2)\ (1, 3)\ (1, 4)\ (1, 5)\ (1, 6)\ (2, 1)\ (2, 2)\ (2, 3)\ (2, 4)\ (2, 5)\ (2, 6)\ (3, 1)\ (3, 2)\ (3, 3)\ (3, 4)\ (3, 5)\ (3, 6)\ (4, 1)\ (4, 2)\ (4, 3)\ (4, 4)\ (4, 5)\ (4, 6)\ (5, 1)\ (5, 2)\ (5, 3)\ (5, 4)\ (5, 5)\ (5, 6)\ (6, 1)\ (6, 2)\ (6, 3)\ (6, 4)\ (6, 5)\ (6, 6)\}$

Situación de interés: A = aparece el mismo número en ambos dados

Probabilidad clásica (teórica)

Experimento aleatorio: lanzamiento de dos dados

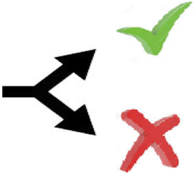


- Diagrama de árbol para el experimento de lanzar dos dados
- Resultados igualmente posibles y mutuamente excluyentes

Casos favorables: (1, 1) (2, 2) (3, 3) (4, 4) (5, 5) (6, 6) $n_A = 6$

Probabilidad de ocurrencia del evento A: $P(A) = 6/36 = 0.167$

Probabilidad como frecuencia relativa



Probabilidad como frecuencia relativa

Si un experimento se repite n veces bajo las *mismas condiciones* y si n_B de dichas veces favorecen al atributo (situación) B entonces la probabilidad de B estará dada por:

$$P(B) = lim_{n \to \infty} \frac{n_B}{n}$$

Probabilidad como frecuencia relativa

Un investigador somete a prueba un fármaco que se pretende usar en el tratamiento por intoxicación. De 190 pacientes a quienes se les administró, 150 obtuvieron algún beneficio del mismo. ¿Cuál es la probabilidad de que este fármaco sea efectivo con el siguiente paciente al cual se aplique?

Pacientes investigados: n = 190

Atributo de interés B: presencia de beneficio

$n_B = 150$

$$P(B) = \frac{150}{190} = 0.789$$

Esta probabilidad se acerca más a su verdadero valor si se investiga a un número mayor de pacientes

Probabilidad subjetiva

Representa el grado de creencia acerca de la posible ocurrencia de una situación en particular

Juicio de experto

Ejemplo: determinar la probabilidad que se presente un atentado terrorista en un determinado mes

Axiomas de probabilidad

Sea Ω un espacio muestral y A un evento cualquiera de Ω. Para todo A en Ω se asigna un *número* llamado probabilidad de A, notado $P(A)$, tal que:

i $P(A) \geq 0$

ii $P(\Omega) = 1$

iii Si $A = B_1 \cup B_2 \cup \ldots \cup B_k; B_i \cap B_j = \emptyset; \forall i \neq j; P(B_i) > 0$, entonces

$$P(A) = \sum_{i=1}^{k} P(B_i)$$

Algunos resultados importantes

Teorema 1: $P(\emptyset) = 0$

Teorema 2: $0 \leq P(A) \leq 1$

Teorema 3: $P(\bar{A}) = 1 - P(A)$

Ejemplo

Usted acaba de terminar estudios de maestría y se presenta a un concurso para asistente de investigación en el Instituto de Nacional de Ciencia; allí solicitan dos profesionales especializados. El día de la entrevista se encuentra con cuatro de sus compañeros del postgrado, entre los que se encuentra su mejor amig@, lo cual lo tranquiliza puesto que todos están en igualdad de condiciones. Si los dos asistentes se seleccionan al azar,

- Establezca los resultados posibles de este experimento
- Cual es la probabilidad de que usted sea seleccionad@?
- Encuentra la probabilidad de que Usted y su mejor amig@ sean seleccionados

Ejemplo - Solución

Sean c: usted y e: mejor amigo. Algunos posibles resultados serían:

- $ac \quad (ca)$
- bd
- ce

Espacio muestral: $\Omega = \{ab, ac, ad, ae, bc, bd, be, cd, ce, de\}$

A: usted es seleccionad@

$$A = \{ac, bc, cd, ce\} \rightarrow P(A) = \frac{4}{10} = 0.4$$

B: usted y su mejor amig@ son seleccionados $B = \{ce\}$

$$P(B) = \frac{1}{10} = 0.1$$

Ley aditiva de la probabilidad

Probabilidad para la *unión* de eventos

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

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - [P(AB) + P(AC) + P(BC)] + P(ABC)$$

Ejemplo

Las enfermedades I y II son comunes entre los individuos de cierta población. Se supone que el 10 % de la población contraerá la enfermedad I alguna vez durante su vida, 15 % contraerá eventualmente la enfermedad II, y el 3 % contraerá ambas. Encuentre la probabilidad de que un individuo elegido al azar de esta población, contraiga al menos una enfermedad

A: El individuo contrae la enfermedad I

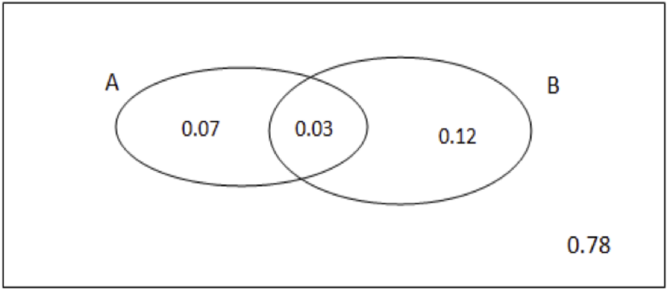
B: El individuo contrae la enfermedad II

$P(A) = 0.1,\ P(B) = 0.15,\ P(AB) = 0.03$

$P(\text{contraiga al menos una enfermedad}) = P(A \cup B)$

$P(A \cup B) = P(A) + P(B) - P(AB) = 0.1 + 0.15 - 0.03 = 0.22$

Ejemplo



Probabilidades conjunta, marginal y condicional

Caracterizar a un grupo de estudiantes según dos variables: género (H, M) y fumar (S, N)

	Fuma		
Género	Si (S)	No (N)	Total
Hombre (H)	20	40	60
Mujer (M)	10	30	40
Total	30	70	100

Probabilidad conjunta y marginal

	Fuma		
Género	Si (S)	No (N)	Total
Hombre (H)	20	40	60
Mujer (M)	10	30	40
Total	30	70	100

Probabilidad conjunta

- $P(HS) = 20/100 = 0.2$
- $P(HN) = 40/100 = 0.4$
- $P(MS) = 10/100 = 0.1$
- $P(MN) = 30/100 = 0.3$

Probabilidad marginal

- $P(H) = 60/100 = 0.6$
- $P(M) = 40/100 = 0.4$
- $P(S) = 30/100 = 0.3$
- $P(N) = 70/100 = 0.7$

Probabilidad condicional $p(A \mid B)$

	Fuma		
Género	Si (S)	No (N)	Total
Hombre (H)	20	40	60
Mujer (M)	10	30	40
Total	30	70	100

Condicionales por fila

- $P(S \mid H) = 20/60 = 0.33$
- $P(N \mid H) = 40/60 = 0.67$
- $P(S \mid M) = 10/40 = 0.25$
- $P(N \mid M) = 30/40 = 0.75$

Condicionales por columna

- $P(H \mid S) = 20/30 = 0.67$
- $P(M \mid S) = 10/30 = 0.33$
- $P(H \mid N) = 40/70 = 0.571$
- $P(M \mid N) = 30/70 = 0.429$

Probabilidad condicional $p(A \mid B)$

La relación entre las tres probabilidades viene dada por:

$$P(S \mid H) = \frac{20}{60} = \frac{\frac{20}{100}}{\frac{60}{100}} = \frac{P(SH)}{P(H)}$$

$$P(A \mid B) + P(\bar{A} \mid B) = 1$$

Ley multiplicativa de la probabilidad

Establece la forma de calcular la probabilidad de intersección de eventos

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

Ejemplo

Un embarque de 20 ambulancias contiene dos que presentan algún tipo de defecto en el funcionamiento del equipamiento médico y el comprador no lo sabe, pero aceptará el embarque si después de seleccionar, sin reposición, dos ambulancias estas pasan una rigurosa inspección. ¿Cuál es la probabilidad de que el embarque sea aceptado?

Ejemplo

Un embarque de 20 ambulancias contiene dos que presentan algún tipo de defecto en el funcionamiento del equipamiento médico y el comprador no lo sabe, pero aceptará el embarque si después de seleccionar, sin reposición, dos ambulancias estas pasan una rigurosa inspección. ¿Cuál es la probabilidad de que el embarque sea aceptado?

A: la primera ambulancia seleccionada pasa la revisión

B: la segunda ambulancia seleccionada pasa la revisión

$P(\text{aceptar}) = P(AB) = ?$

$P(A) = 18/20$ y $P(B \mid A) = 17/19$

$P(AB) = P(A) \times P(B \mid A) = (18/20) \times (17/19) = 0.805$

Independencia estadística

Dos eventos A y B son independientes si la ocurrencia de uno de ellos no afecta la probabilidad de ocurrencia del otro

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

$$P(AB) = P(A) \times P(B)$$

Ejemplo

Un contratista debe realizar tres trabajos similares. La experiencia indica que la probabilidad de que un determinado trabajo se demore mas de lo planeado es 0.4. Si el ritmo al cual trabaja es *independiente*,

a ¿cuál es la probabilidad de que los tres trabajos se terminen en el tiempo planeado?

b ¿cuál es la probabilidad de que los tres trabajos tengan retraso?

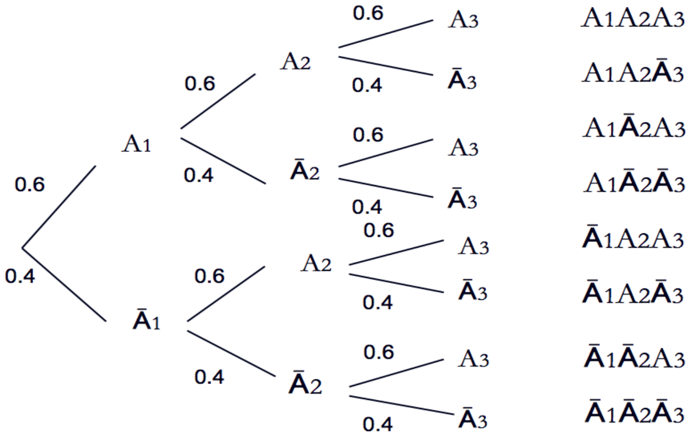
c ¿cuál es la probabilidad de que al menos uno de los trabajos tenga retraso?

Ejemplo

A_1: trabajo 1 se termina en el tiempo planeado $P(A_1) = 0.6$ $P(\bar{A}_1) = 0.4$

A_2: trabajo 2 se termina en el tiempo planeado $P(A_2) = 0.6$ $P(\bar{A}_2) = 0.4$

A_3: trabajo 3 se termina en el tiempo planeado $P(A_3) = 0.6$ $P(\bar{A}_3) = 0.4$



Ejemplo

a $P(A_1A_2A_3) = P(A_1) \times P(A_2) \times P(A_3) = 0.6 \times 0.6 \times 0.6 = 0.216$

La probabilidad de que los tres trabajos se terminen en el tiempo planeado es 0.126

b $P(\bar{A}_1\bar{A}_2\bar{A}_3) = P(\bar{A}_1) \times P(\bar{A}_2) \times P(\bar{A}_3) = 0.4 \times 0.4 \times 0.4 = 0.064$

La probabilidad de que los tres trabajos tengan retraso es 0.064

c $P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) = P(\bar{A}_1) + P(\bar{A}_2) + P(\bar{A}_3) - [P(\bar{A}_1\bar{A}_2) + P(\bar{A}_1\bar{A}_3) + P(\bar{A}_2\bar{A}_3)] + P(\bar{A}_1\bar{A}_2\bar{A}_3) = 3 \times 0.4 - 3 \times 0.16 + 0.064 = 0.784$

Otra forma: $P(\bar{A}_1 \cup \bar{A}_2 \cup \bar{A}_3) = 1 - P(A_1 \cap A_2 \cap A_3) = 1 - 0.216 = 0.784$

La probabilidad de que al menos uno de los trabajos tenga retraso es 0.784

Ley de probabilidad total

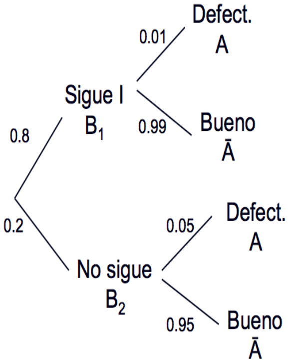
Sea Ω y $A \subset \Omega$. Si $\Omega = B_1 \cup B_2 \cup \ldots \cup B_k$ con $B_i \cap B_j = \emptyset \forall i \neq j$ y $P(B_i) > 0$, entonces:

$$P(A) = \sum_{i=1}^{k} P(B_i) P(A/B_i)$$

Ejemplo

La probabilidad que un artículo sea defectuoso es de 0.01 si el operario sigue con exactitud las instrucciones del protocolo de fabricación, y de 0.05 si el operario no las sigue. Si el operario sigue las instrucciones 80 % de las veces, ¿qué porcentaje de los productos fabricados por el operario serán defectuosos?

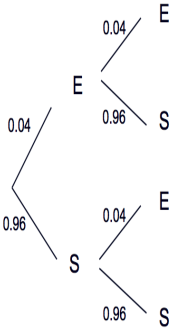
Ejemplo



B_1: el operario sigue con exactitud el protocolo
B_2: el operario no sigue con exactitud el protocolo
A: el artículo es defectuoso

$$P(B_1) = 0.8 \;\; P(A \mid B_1) = 0.01$$
$$P(B_2) = 0.2 \;\; P(A \mid B_2) = 0.05$$

Considere que se examinan dos personas y se determina si padecen cierta enfermedad (E) o se encuentran sanos (S)



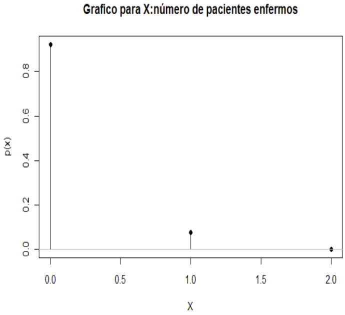
$$P(EE) = P(E) \times P(E) = 0.04 \times 0.04 = 0.0016$$
$$P(ES) = P(E) \times P(S) = 0.04 \times 0.96 = 0.0384$$
$$P(SE) = P(S) \times P(E) = 0.96 \times 0.04 = 0.0384$$
$$P(SS) = P(S) \times P(S) = 0.96 \times 0.96 = 0.9216$$

$$0.0016 + 0.0384 + 0.0384 + 0.9216 = 1$$

Función que transforma los resultados del experimento (Ω) en puntos sobre la recta real $\Re$

X: número de pacientes enfermos

Evento	x	$p(x)$
EE	2	0.0016
ES, SE	1	0.0768
SS	0	0.9216



Clasificación

- **Discreta**: toma un conjunto *contable* de valores en un intervalo de los reales ($\Re$). En general, este tipo de variables procede de conteos
  - Número de personas que compran el producto A
  - Pacientes que presentan reacción adversa a un tratamiento
- **Continua**: aquellas variables que pueden tomar cualquier valor en un intervalo de los reales ($\Re$)
  - Tiempo de vida de un producto
  - Distancia que recorre un auto por galón de combustible
  - Contenido de agua en una fruta

Función de probabilidad

La función de probabilidad de una variable es una expresión matemática mediante la cual se pueden calcular las probabilidades asociadas al rango de variación de X (Ω_x)

La función de probabilidad de una variable *discreta* se le suele denominar *función de masa de probabilidad (fmp)*

Si X es *continua* se llama *función de densidad de probabilidad* denotada con *fdb*

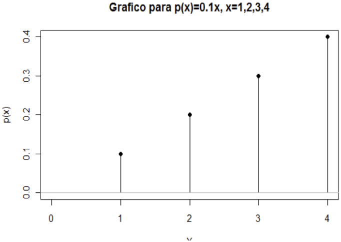
Función de probabilidad

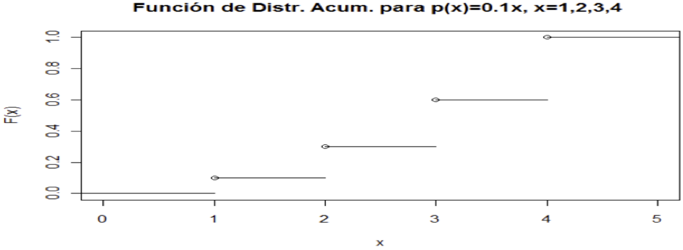
- Discreta - *fmp*
  - $0 \leq p(x) \leq 1$
  - $\sum_{x \in \Omega} p(x) = 1$
- Continua - *fdp*
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$

$f(x)$ y $F(x)$ - Variable discreta

X: número de días por semana que entrena un deportista aficionado
$f(x) = 0.1x, x = 1, 2, 3, 4$

x	$p(x)$	$F(x)$
1	0.1	0.1
2	0.2	0.3
3	0.3	0.6
4	0.4	1.0





Valor esperado y varianza

Discreta: $\mu = E(X) = \sum_{x \in \Omega} x \times p(x)$

Continua: $\mu = E(X) = \int_{-\infty}^{\infty} x \times f(x) dx$

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

Ejemplo

La vida máxima de la patente para un nuevo fármaco es 17 años. Si se resta el tiempo que requiere la FDA para probar y aprobar el fármaco se obtiene la vida real de la patente, es decir, el tiempo que la compañía farmacéutica tiene para recuperar los costos de la investigación y desarrollo y conseguir una utilidad. Considere la siguiente distribución del tiempo de vida de la patente para los nuevos fármacos

X (años)	3	4	5	6	7	8	9	10	11	12	13
$p(x)$	0.03	0.05	0.07	0.10	0.14	0.20	0.18	0.12	0.07	0.03	0.01

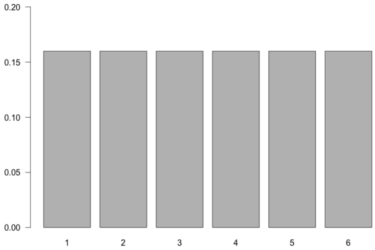
- Encuentre el número esperado de años de vida de la patente para un nuevo fármaco
- Determine la varianza y la desviación estándar de X

Distribución uniforme

Todos y cada uno de los resultados del experimento tienen la misma probabilidad de ocurrir.

Ejemplo

En el lanzamiento de un dado de 6 caras, es espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$ y la ocurrencia de cada uno de estos resultados tiene una probabilidad de 0.1666667.



x	$P(x)$
1	0.16
2	0.16
3	0.16
4	0.16
5	0.16
6	0.16

Distribución de Probabilidad Binomial

Propiedades:

1. El experimento consiste de una secuencia de n ensayos idénticos.
2. Cada ensayo tiene únicamente dos posibles resultados (éxito o fracaso). $\Omega = \{Exito, Fracaso\}$.
3. Ni la probabilidad (p) de un éxito ni la de un fracaso $(1 - p)$ cambia entre ensayos.
4. Los ensayos del experimento son independientes.

Distribución de Probabilidad Binomial

Si una variable aleatoria X sigue una distribución binomial con parámetros n y p, se define como: $X \sim B(x; n, p)$.

La probabilidad de obtener exactamente k éxitos en n pruebas está dada por la función de masa de probabilidad:

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} \qquad \text{(Función de probabilidad: Binomial)}$$

Con media $= n \times p$ y varianza $= n \times p \times (1 - p)$

Ejemplo

Considere la decisión de compra de los tres siguientes clientes que entran a una tienda de ropa. A partir de experiencias pasadas, el gerente de la tienda estima la probabilidad de compra de un cliente en 0.30. Cuál es la probabilidad de que dos de los tres siguientes clientes que entren a la tienda hagan una compra?

$$P(X = x) = \binom{n}{x} \times p^x \times (1 - p)^{n-x}$$

$$P(X = x) = \binom{3}{2} \times 0.3^2 \times (1 - 0.3)^{3-2}$$

$$P(X = x) = 3 \times 0.09 \times 0.70 = 0.189$$

La probabilidad de que dos de los siguientes tres clientes que entren a la tienda hagan una compra es de 0.189

En R sería así $\Longrightarrow$ `dbinom(x=2,size=3, prob=0.3)`

Distribución binomial en R

Se creará un vector que contenga 100 números aleatorios con una distribución binomial con parámetros $n = 7$ y $p = 0.3$ en un objeto vectorial llamado X. Donde n representa el número de ensayos, llamado size en la función de R.

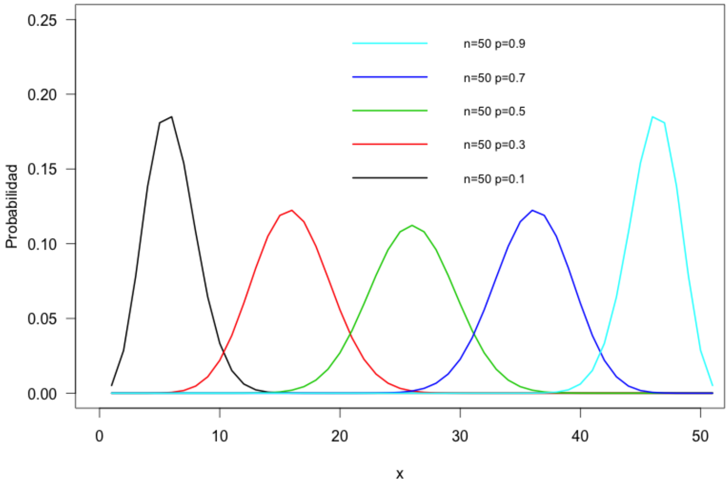
```
X <- rbinom(n=100, size=7, prob=0.3)
barplot(table(X))
```

El diagrama de barras representa una estimación de la distribución de probabilidad $P(X = x)$. El diagrama de barras es más apropiado que un histograma, ya que los datos son discretos, no continuos.

La media, la desviación estándar y la varianza se pueden calcular directamente a partir de X.

```
mean(X)
var(X)
sd(X)
```

Distribución binomial



Distribución de Probabilidad Poisson

Distribución de variables aleatorias discretas útil para estimar la ocurrencia de eventos dentro de un intervalo específico de tiempo o espacio.
$X \sim pois(x; \lambda)$

$$f(x) = \frac{\lambda^x \times e^{-\lambda}}{x!}$$

- $f(x)$: Probabilidad de x ocurrencia en un intervalo
- λ: Valor esperado o número promedio de ocurrencias dentro de un intervalo
- $e = 2.71828$

Distribución de probabilidad Poisson

Propiedades de la distribución de probabilidad Poisson

1. La probabilidad de una ocurrencia es la misma para cualquier par de intervalos de igual longitud
2. La ocurrencia o no ocurrencia en cualquier intervalo es independiente de la ocurrencia o no ocurrencia en cualquier otro intervalo.

Número de clientes atendidos en una caja de un banco durante un periodo de 15 minutos en las mañanas de días entre semana

Asumiendo que:

1 La probabilidad de que una persona llegue en cualquier intervalo de tiempo es la misma para dos intervalos de tiempo iguales y que

2 la llegada o no llegada del cliente en cualquier periodo de tiempo es independiente de la llegada o no en cualquier otro periodo de tiempo.

Ejemplo

Número de clientes atendidos en la ventanilla de un banco durante un periodo de 15 minutos, en las mañanas de días entre semana.

Suponiendo que los datos históricos indican que el promedio de clientes que llegan a la vetanilla del banco en un periodo de 15 minutos es 10; y que el gerente quiere conocer la probabilidad de que exactamente lleguen 5 clientes en un periodo de 15 minutos, entonces:

$$f(x=5) = \frac{10^5 \times e^{-10}}{5!} = 0.0378$$

En R sería así $\Longrightarrow$ `dpois(5, lambda=10)`

Valor esperado y Varianza

En la distribución de probabilidad Poisson la media y la varianza son IGUALES

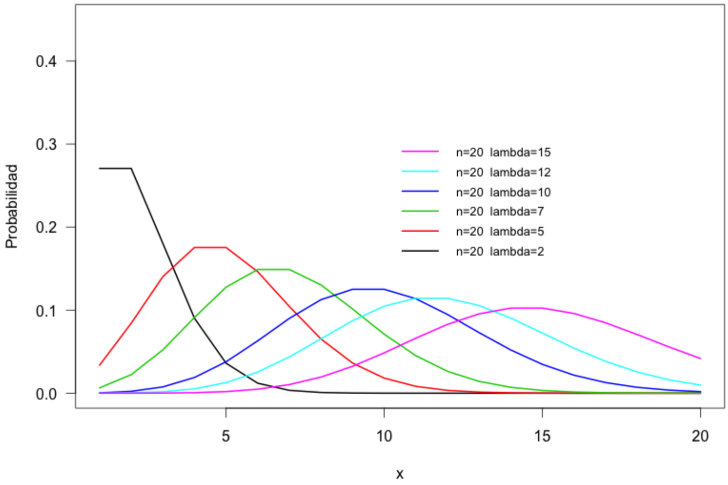
$$Media(X) = \lambda = Var(X)$$

Siguiendo con el ejemplo anterior:

$$Media(X) = \lambda = 10$$

$$\sigma = 3.16$$

Distribución Poisson



Distribución de Probabilidad Hipergeométrica

Distribución utilizada para calcular la probabilidad de que en una selección aleatoria de n elementos, seleccionados sin reemplazamiento, se obtengan x elementos calificados como éxitos y $n - x$ elementos calificados como fracasos

$$f(x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

$f(x)$: Probabilidad de x éxitos en n ensayos

x: Número de éxitos

n: Número de ensayos

N: Número de elementos en la población

r: Número de elementos en la población calificados como éxitos

Distribución de Probabilidad Hipergeométrica

$\binom{r}{x}$: Número de maneras en que x éxitos pueden ser seleccionados de un total de r éxitos en la población

$\binom{N-r}{n-x}$: Número de maneras en que $n-x$ fracasos pueden ser seleccionados de un total de $N-r$ fracasos en la población

$\binom{N}{n}$: Número de maneras en que n elementos pueden ser seleccionados de una población N

Características

La distribución de probabilidad hipergeométrica es cercana a la distribución de probabilidad binomial

- Binomial: Ensayos independientes
- Hipergeométrica: Ensayos NO son independientes

- Binomial: La probabilidad de éxito NO cambia entre ensayos
- Hipergeométrica: La probabilidad de éxito SI cambia entre ensayos

Condiciones

La distribución es válida solo cuando:

1. El número de éxitos observados es menor o igual que el número de éxitos presentes en la población

$$x \leq r$$

2. El número de fracasos es menor o igual que el número de fracasos presentes en la población

$$n - x \leq N - r$$

Ejemplo

Una compañía fabrica bombillos ahorradores de energía y los empaca en cajas de 12 unidades cada una. Un inspector elige al azar una caja para probarlos . Si la caja contiene exactamente 5 bombillos defectuosos, ¿cuál es la probabilidad de que el inspector encuentre exactamente un bombillo defectuoso al inspeccionar los tres primeros bombillos?

$$N = 12 \qquad n = 3 \qquad r = 5 \qquad x = 1$$

$$f(x = 5) = \frac{\binom{5}{1}\binom{12-5}{3-1}}{\binom{12}{3}} = \frac{5 \times 21}{220} = 0.4773$$

La probabilidad de encontrar exactamente un bombillo defectuoso al inspeccionar los primeros tres bombillos será de 0.4773.

En `R` sería así $\Longrightarrow$ `dhyper(x=1, m=5, n=7, k=3)`

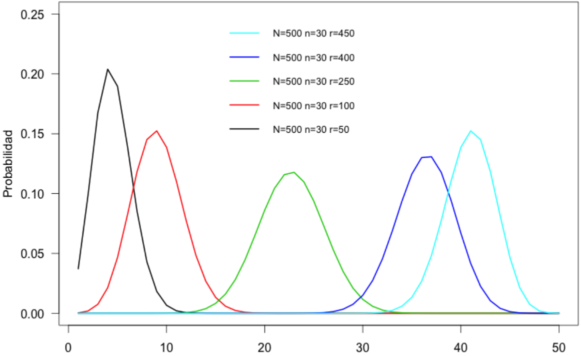
Valor Esperado y Varianza

$$Media = \mu = n\left(\frac{r}{N}\right)$$

$$Var(x) = \sigma^2 = n\left(\frac{r}{N}\right)\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Distribución Hipergeométrica

Figura: La distribución hipergeométrica describe el número de eventos k de una muestra n extraída de una población total de N sin reemplazo. En una población de N objetos de los cuales r son **defectuosos** y Nr son **no-defectuosos** (**éxito** o **fracaso**). Si tomamos muestras n elementos sin reemplazo entonces ¿cuál es la probabilidad de que exactamente k artículos en la muestra (n) sean defectuosos?.



Distribución de Probabilidad Normal

Es la distribución de probabilidad más importante para describir una variable aleatoria continua

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

μ: Promedio - media
σ: Desviación estándar

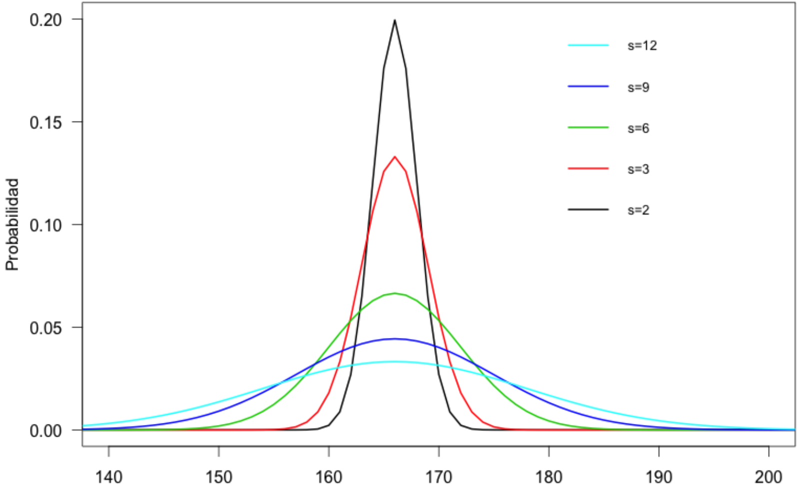
Características

- Está caracterizada por dos parámetros: la media y la desviación estándar
- El punto más alto de la curva normal representa la media, la cual también puede indicar la mediana y la moda de la distribución
- La media de una distribución normal puede ser cualquier valor: negativo, cero o positivo

Características

- La distribución normal es simétrica. Las colas de la curva normal se extienden hasta el infinito en ambas direcciones y teóricamente nunca tocan el eje horizontal
- La desviación estándar determina el ancho y la altura de la curva normal. Valores altos de desviación estándar indican curvas anchas y achatadas indicando una mayor variabilidad de los datos
- Las probabilidad de una variable aleatoria normal está dada por el área bajo la curva. El área total bajo la curva para la distribución normal es 1. Al ser simétrica, el área tanto a la derecha como a la izquierda de la curva es 0.5

Distribución Normal



Funciones en R asociadas a distribuciones de probabilidad

Cuadro: En la siguiente tabla se muestran los nombres de las funciones de R asociadas a distribuciones de probabilidad

d	Función de densidad de probabilidad (o función de masa de probabilidad).
p	Función de probabilidad acumulada (los valores están siempre en el intervalos de 0 a 1).
q	Función cuantil - el inverso (más o menos) de la función p.
r	Simulación de una muestra aleatoria de la distribución.

Distribuciones continuas más comunes

Cuadro: Las distribuciones continuas más comunes son:

Exponencial	`exp`	El parámetro es la tasa (por defecto es 1). La media de la distribución es $1/Tasa$
Normal (Gaussiana)	`norm`	Es la distribución más famosa en estadística, y se le conoce como la curva de campana. Los parámetros por defecto son media $= 0$ y la desviación estándar $= 1$.
Uniforme	`unif`	Los parámetros son `min` (por defecto es 0) y `max` (por defecto es 1).

Distribuciones discretas más comunes

Cuadro: Las distribuciones discretas más comunes son:

Binomial	`binom`	Los parámetros son `size`: el número de ensayos y `prob`: es la probabilidad de éxito en un único ensayo (No hay valores por defecto).
Geométrica	`geom`	El parámetro es `prob`, la probabilidad de éxito es cada intento independiente. Tenga en cuenta que la distribución se define en términos del número de fallos antes del primer éxito.
Poisson	`pois`	El parámetro es lambda, la media.

Cuadro: Algunas distribuciones continuas menos comunes son:

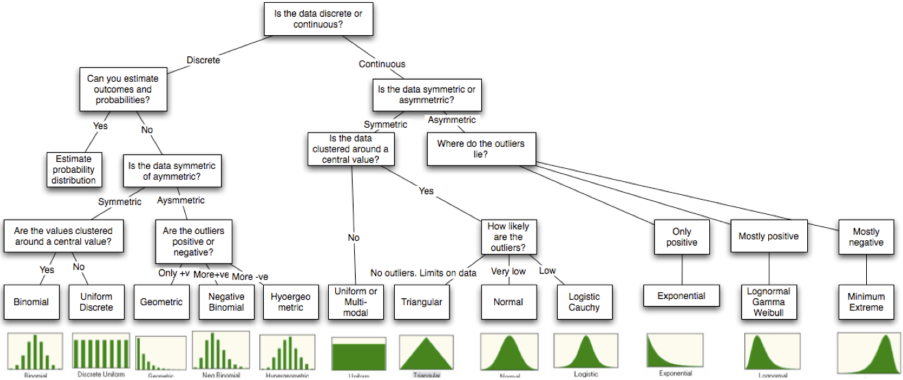
Beta	`Beta`	Los parámetros son `shape1` y `shape2` (sin valores por defecto) y `ncp` (parámetro de no-centralidad) cuyo valor predeterminado es 0, que corresponde a la beta =0.
Cauchy	`Cauchy`	Los parámetros son: `location` (0 por defecto) y `scale` (0 por defecto).
Gamma	`Gamma`	Los parámetros son `shape` y uno entre `rate` o `scale`. Los dos últimos parámetros son inversos entre si y ambos por defecto tienen un valor de `1` .
Logistica	`Logistic`	Los parámetros son `location` y `scale`.
log-normal	`Lognormal`	Los parámetros son `meanlog` y `sdlog`.
Weibull	`Weibull`	Los parámetros son `shape` y `scale` .

Distribuciones menos comunes y derivadas

Cuadro: Algunas distribuciones continuas menos comunes son:

Chi-cuadrado	`Chisquare`	Los parámetros son `df` y `ncp` parámetro de no-centralidad.
F	`FDist`	Los parámetros son `df1` y `df2`, el numerador y el denominador de los grados de libertad. Un parámetro opcional es `ncp`, el parámetro de no-centralidad. Se se omite este parámetro se asume una función `F` central.
t	`TDist`	Los parámetros son `df` y `pnc`; si se omite el parámetro de no-centralidad se asume una distribución `textbft` central.

Opciones de distribuciones, adaptado de: Probabilistic approaches to risk de Aswath Damodaran



Tomado de . . .