Estadística Aplicada

Ricardo Cao

Departamento de Matemáticas
Universidad de A Coruña

Ciencias da la Salud Curso 2007-08

Índice

Índice del Tema 1

Índice del Tema 2

Índice del Tema 3

Índice del Tema 4

Índice del Tema 5

Índice del Tema 6

Índice del Tema 7

Índice del Tema 8

Índice del Tema 9

Índice del Tema 10

Part I

Descripción estadística de una variable

Conceptos generales

- El objeto de cualquier investigación estadística es la toma de información acerca de los individuos de cierto colectivo llamado población estadística.
- Cada elemento de una población se denomina individuo o unidad estadística.
- Las poblaciones estadísticas se clasifican en poblaciones finitas e infinitas, de acuerdo con el número de individuos incluidos en las mismas.
- El proceso de toma de información acerca de los individuos de una población puede realizarse mediante la elaboración de un censo o mediante la extracción de una muestra.

Conceptos generales

- Si una muestra es representativa de una población, se pueden deducir importantes conclusiones acerca de ésta a partir del análisis de la información muestral. La parte de la Estadística que trata de las condiciones bajo las cuales tales inferencias son válidas se llama Estadística Inductiva o Inferencial.
- La parte de la Estadística que trata solamente de describir y analizar un grupo dado de datos sin sacar conclusiones o inferencias acerca de la población que los ha generado se llama Estadística Descriptiva o Deductiva.

Conceptos generales

Variables estadísticas

- Llamaremos variable estadística a cada una de las características consideradas con el propósito de describir a cada individuo de la muestra.
- Cada una de las variables estadísticas consideradas para la descripción de los individuos de la muestra puede presentar distintas modalidades o estados.
- Atendiendo a la naturaleza de las modalidades de las variables, éstas pueden clasificarse en variables cualitativas y variables cuantitativas. Estas últimas se clasifican en variables discretas, que son aquellas que toman un número finito o infinito numerable de valores distintos, y variables continuas, que son aquellas que pueden tomar cualquier valor en un intervalo de valores dado.

Distribuciones de frecuencias

- Sean $x_1, \ldots, x_n$ una muestra de una variable X que presenta las modalidades $c_1, \ldots, c_k$, y sea n_i el número de individuos de la muestra con la modalidad c_i, $i = 1, \ldots, k$.
- El número n_i se conoce como frecuencia absoluta de la modalidad c_i.
- La frecuencia relativa de c_i es $f_i = n_i/n \ (i = 1, 2, \ldots, k)$.
- La frecuencia absoluta acumulada se define como

$$N_i = n_1 + n_2 + \cdots + n_i = \sum_{j=1}^{i} n_j \quad (i = 1, 2, \ldots, k)$$

- La frecuencia relativa acumulada se define como

$$F_i = f_1 + f_2 + \cdots + f_i = \sum_{j=1}^{i} f_j = \frac{N_i}{n} \quad (i = 1, 2, \ldots, k).$$

Distribuciones de frecuencias

Propiedades

1. $0 \leq n_i \leq n; \quad 0 \leq f_i \leq 1 \ (i = 1, 2, \ldots, k).$
2. $N_k = n_1 + n_2 + \cdots + n_k = n; \quad F_k = f_1 + f_2 + \cdots + f_k = 1.$

Distribuciones de frecuencias

Modalidad	Frecuencia absoluta	Frecuencia relativa	Fr. absoluta acumulada	Fr. relativa acumulada
c_1	n_1	f_1	N_1	F_1
c_2	n_2	f_2	N_2	F_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_i	n_i	f_i	N_i	F_i
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
c_k	n_k	f_k	N_k	F_k
Total	n	1		

Construcción de las modalidades

- Si la variable es cualitativa, se tomarán como modalidades las distintas respuestas observadas en la muestra.
- Si la variable es discreta (que tome pocos valores distintos), las modalidades coincidirán con los distintos valores medidos en la muestra
- Si la variable es continua (o bien discreta, pero que toma muchos valores distintos), se tomarán como modalidades los intervalos de clase, tomándose como frecuencia absoluta de cada modalidad el número de observaciones agrupadas en el intervalo correspondiente.

Intervalos de clase

- Si $e_0 < e_1 < \cdots < e_{i-1} < e_i < \ldots < e_k$ son los extremos de los k intervalos de clase construidos, la frecuencia absoluta del i-ésimo intervalo es $n_i =$ número de observaciones x_i tales que $e_{i-1} < x_i \leq e_i$, siendo su **amplitud** $a_i = e_i - e_{i-1}$, para $i = 1, 2, \ldots, k$.
- Una vez construidos los intervalos de clase se elige un representante en cada uno de ellos, llamado marca de clase, que usualmente es tomado como el punto medio del intervalo

$$c_i = \frac{e_{i-1} + e_i}{2} \quad (i = 1, 2, \ldots, k).$$

Distribuciones de frecuencias

Intervalos de clase

- *¿Cuántos intervalos construir?* Se recomienda construir tantos intervalos como el número entero (entre 5 y 20) más próximo a $\sqrt{n}$.
- *¿Intervalos de la misma amplitud o de amplitudes distintas?* En general construiremos intervalos de la misma amplitud.
- *¿Qué valor elegir como extremo inferior del primer intervalo* (e_0)*?* Se toma como e_0 un valor "un poco menor" que el mínimo de la muestra.

Representaciones gráficas

Representación gráfica de variables cualitativas

- Diagrama de barras. En un sistema de ejes de coordenadas se representa en el eje de abscisas las modalidades de la variable y en el eje de ordenadas las frecuencias (n_i o f_i). A continuación, sobre cada modalidad se levanta un rectángulo o barra de altura igual a la frecuencia (absoluta o relativa) representada, siendo todos los rectángulos de igual base.
- Diagrama de sectores. Se trata de repartir un círculo de radio arbitrario en sectores proporcionales a la frecuencia de cada modalidad.

Representación gráfica de variables cuantitativas discretas

- Diagrama de barras. Se construye de la forma anteriormente descrita para variables cualitativas, pudiendo sustituirse las barras por segmentos de recta.
- Diagrama acumulativo de frecuencias. Su construcción se realiza representando en un sistema de ejes de coordenadas los puntos (c_i, N_i) o (c_i, F_i), según se representen frecuencias absolutas o relativas acumuladas, uniéndose a continuación de forma escalonada los puntos representados.

Representación gráfica de variables cuantitativas continuas

- Histograma. Es el equivalente continuo del diagrama de barras. Partiendo de un sistema de ejes de coordenadas, se representan en el eje de abscisas los extremos e_i de los intervalos de clase y a continuación, tomando como base de cada rectángulo la amplitud a_i del intervalo de clase correspondiente, se levanta sobre cada uno de los intervalos un rectángulo de altura $h_i = n_i/a_i$ o $h_i = f_i/a_i$, según se representen frecuencias absolutas o relativas.
- Uniendo, mediante segmentos de recta, los puntos medios de las bases superiores de cada rectángulo del histograma se obtiene la representación gráfica llamada polígono de frecuencias.

Representaciones gráficas

Representación gráfica de variables cuantitativas continuas

- histograma móvil. Para cada punto x se define

$$h(x) = \frac{\text{nº de observaciones } x_i \text{ en } [x - a/2, x + a/2]}{a \cdot n}$$

se representan gráficamente los pares $(x, h(x))$ para un número grande de valores x seleccionados, el gráfico resultante corresponde a una curva más "suave".

- Polígono acumulativo de frecuencias. Es el resultado que se obtiene al unir mediante segmentos de recta los puntos (e_i, N_i) o (e_i, F_i), según se trate de frecuencias absolutas o relativas acumuladas, representados en un sistema de ejes de coordenadas.

Diagrama de tallo y hojas

- Se utiliza a menudo para la descripción de variables cuantitativas (discretas o continuas), presenta la particularidad de permitir visualizar globalmente la distribución de frecuencias manteniendo la individualidad de los datos.
- Se redondean los datos a dos o tres cifras significativas, tomándose como tallos la primera o las dos primeras cifras de cada dato y como hojas las últimas cifras de cada dato.
- A continuación, separados por una línea vertical, se dispondrán los tallos a la izquierda y las hojas a la derecha del tallo correspondiente. De esta manera cada tallo, que se representa una sola vez, define una clase y el número de hojas representa la frecuencia de dicha clase.

Medidas de posición central

- Media aritmética. Se define la media aritmética ($\overline{x}$) de una variable cuantitativa X como el valor numérico

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

- Propiedades:

1. $\min(x_i) \leq \overline{x} \leq \max(x_i)$.
2. Si $y_i = a + bx_i$ $(i = 1, 2, \ldots, n)$, entonces $\overline{y} = a + b\overline{x}$.
3. $\frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}) = 0$
4. $\overline{x} = \arg\min \sum_{i=1}^{n} (x_i - a)^2$

Medidas de posición central

- Media truncada. Un inconveniente que presenta la media aritmética es su sensibilidad a la presencia de observaciones atípicas (outliers) en la muestra. Para evitar este tipo de problema se utilizan las medias truncadas, que consisten en medias aritméticas calculadas con un porcentaje central de los datos.
- Media recortada. Consiste en la media aritmética de una modificación de los datos originales: un porcentaje central permanece sin modificar, cada uno de los datos de los menores excluidos se sustituye por el menor de los datos del porcentaje central no modificado y cada uno de los datos de los mayores excluidos se reemplaza por el mayor de los datos del porcentaje central no modificado.

Medidas de posición central

- Media cuadrática (Q).

$$Q = \sqrt{\frac{x_1^2 + x_2^2 + \ldots + x_n^2}{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$$

.

- Media geométrica (G). Supuesto que $x_i > 0,\ 1 \leq i \leq n$, se define como $G = \sqrt[n]{x_1 \cdot x_2 \cdots x_n} = \prod_{i=1}^{n} x_i^{1/n}$.
- Media armónica (H). Si $x_i > 0,\ 1 \leq i \leq n$, es $H = \dfrac{n}{1/x_1 + \cdots + 1/x_n} = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\right)^{-1}$.

Medidas de posición central

- Mediana (M_e). Es la medida de posición central que, supuestos los datos ordenados de menor a mayor, deja igual número de valores a su izquierda que a su derecha.
- Si el número de datos es par, se toma como mediana la media aritmética de los dos valores centrales. Si el número de datos es impar, se toma como mediana el valor central.
- En el supuesto de que los datos hayan sido agrupados en intervalos de clase, se calcula el intervalo mediano (el intervalo de clase que contiene a la mediana), eligiéndose como mediana un representante de dicho intervalo (por ejemplo, la marca de clase).
- La mediana es una medida robusta, esto es, poco sensible a la presencia de observaciones atípicas en la muestra.

Medidas de posición central

- Moda (M_0). Si X es una variable discreta, se define como el valor más frecuente de la variable.
- Si X es una variable continua, no tiene sentido hablar del valor más frecuente, por lo que se introduce el concepto de intervalo modal, que se define como aquel intervalo de clase al que le corresponde mayor altura en el histograma de frecuencias. Finalmente, se toma como moda un representante del intervalo modal (por ejemplo, la marca de clase).
- Debe observarse que, de acuerdo con la definición, la moda puede no ser única, en cuyo caso tendremos distribuciones multimodales (con 2 o más modas).

Otras medidas de posición

- Cuantiles. Supuestas ordenadas las observaciones de menor a mayor, se define el cuantil de orden p $(0 < p < 1)$ como el valor q_p que deja a lo sumo np observaciones a su izquierda y a lo sumo $n(1-p)$ observaciones a su derecha.
- Cuartiles. Son los cuantiles de órdenes $1/4$ (cuartil de primer orden o primer cuartil Q_1), $1/2$ (cuartil de segundo orden o mediana) y $3/4$ (cuartil de tercer orden o tercer cuartil Q_3).
- Deciles. Son los cuantiles de órdenes $r/10$ $(r = 1, \ldots, 9)$, y dividen el conjunto de observaciones en diez partes de igual frecuencia.
- Percentiles. Son los cuantiles de órdenes $r/100$ $(r = 1, \ldots, 99)$.

Medidas características

Diagramas de caja

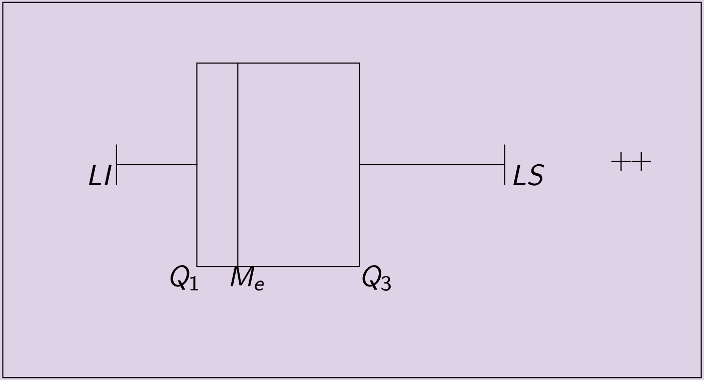
Para la construcción de esta representación, se calculan previamente la mediana, los cuartiles Q_1 y Q_3 y los valores extremos LI y LS, donde

$$\begin{aligned} LI &= \min\left\{x_i \,/ x_i \geq Q_1 - 1'5(Q_3 - Q_1)\right\} \\ LS &= \max\left\{x_i \,/ x_i \leq Q_3 + 1'5(Q_3 - Q_1)\right\} \end{aligned}$$

Las observaciones que caen fuera del intervalo (LI, LS) se consideran datos atípicos (*outliers*).

Medidas características

Diagramas de caja. Las observaciones representadas con + son outliers.



Medidas de dispersión absoluta

- Varianza. $s^2 = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2$.
- Desviación típica. $s = \sqrt{\frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2}$
- Propiedades:
  1. La varianza y la desviación típica toman siempre valores no negativos.
  2. Si $y_i = a + bx_i$ $(i = 1, 2, \ldots, n)$, se tiene que $s_y^2 = b^2 s_x^2$ y $s_y = |b| s_x$.
  3. $s^2 = \frac{1}{n}\sum_{i=1}^n x_i^2 - \overline{x}^2 = \overline{x^2} - \overline{x}^2$.

Desigualdad de Tchebychev

En el intervalo de centro la media aritmética y radio k veces la desviación típica están comprendidas como mínimo el $100(1 - 1/k^2)\%$ de las observaciones.

Variables tipificadas

Se define la variable tipificada de una variable estadística X como la variable (Z) que resulta de restarle su media aritmética y dividir por su desviación típica:

$$Z = \frac{X - \overline{x}}{s}$$

obteniéndose de esta manera una variable adimensional de media cero y desviación típica unidad $(\overline{z} = 0, \ s_z = 1)$.

Medidas características. Medidas de dispersión

Medidas de dispersión absoluta

- Recorrido o rango. Es la diferencia entre los valores extremos:

$$R = \max(x_i) - \min(x_i)$$

- Recorrido intercuartílico. Es la diferencia entre los cuartiles de tercer y primer órdenes:

$$RI = Q_3 - Q_1$$

Medidas de dispersión relativas

- Recorrido relativo. Supuesto que $\overline{x} > 0$, se define como el cociente entre el recorrido y la media

$$RR = \frac{R}{\overline{x}}$$

- Coeficiente de variación. Es la medida de dispersión relativa más popular y, supuesto que $\overline{x} > 0$, se define como el cociente entre la desviación típica y la media

$$CV = \frac{s}{\overline{x}}$$

Medidas características. Medidas de forma

Medidas de asimetría

- Coeficiente de asimetría de Pearson. Para distribuciones unimodales, mide si las observaciones están dispuestas simétricamente con respecto a la moda. $v = \frac{\overline{x} - M_0}{s}$
- Coeficiente de asimetría de Fisher. Mide si las observaciones están dispuestas simétricamente con respecto a la media y es obtenible aunque la distribución no sea unimodal. $g_1 = \frac{m_3}{s^3}$
- Si una distribución es simétrica o insesgada $v \simeq 0$ $(\overline{x} \simeq M_0)$ y $g_1 \simeq 0$. Si $v > 0$ $(\overline{x} > M_0)$ o $g_1 > 0$, diremos que la distribución es sesgada a la derecha; diremos que la distribución es sesgada a la izquierda si $v < 0$ $(\overline{x} < M_0)$ o $g_1 < 0$.

Medidas características. Medidas de apuntamiento o curtosis

Coeficiente de apuntamiento de Fisher

- Mide el grado de apuntamiento de una distribución con respecto al modelo *normal* de referencia. Es una medida adimensional definida según la expresión:

$$g_2 = \frac{m_4}{s^4} - 3$$

- Diremos que una distribución es platicúrtica (más aplastada que el modelo normal) si $g_2 < 0$, mesocúrtica si $g_2 \simeq 0$ y leptocúrtica (más apuntada que el modelo normal) si $g_2 > 0$.

Part II

Descripción estadística de dos variables

Variable estadística bidimensional

Introducción

- Si para un mismo individuo observamos simultáneamente k características obtendremos como resultado una variable estadística **k**-dimensional.
- Nos ocuparemos del estudio de las variables estadísticas bidimensionales ($k = 2$).
- Representaremos por (X, Y) la variable bidimensional estudiada, donde X e Y son las variables unidimensionales correspondientes a las primera y segunda características, respectivamente.
- El estudio de cada variable bidimensional particular (X, Y) variará según las variables unidimensionales X e Y sean cuantitativas o cualitativas y, de ser cuantitativas, según sean continuas o discretas.

Distribución de frecuencias conjunta

- Sea $(x_1, y_1), \ldots, (x_n, y_n)$ una muestra de n observaciones de una variable estadística bidimensional (X, Y), tal que que X presenta k modalidades $c_1, c_2, \ldots, c_k$ e Y presenta l modalidades $c'_1, c'_2, \ldots, c'_l$, y sea n_{ij} el número de individuos de la muestra que presentan la modalidad c_i de X y la modalidad c'_j de Y $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, l)$.
- El número n_{ij} se conoce como la frecuencia absoluta del par (c_i, c'_j).
- La frecuencia relativa de (c_i, c'_j) es el número $f_{ij} = n_{ij}/n$, que representa la proporción de individuos de la muestra que presentan el par (c_i, c'_j) $(i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, l)$.

Distribuciones de frecuencias

Propiedades

- $0 \leq n_{ij} \leq n \ (i = 1, 2, \ldots, k; \ j = 1, 2, \ldots, l).$
- $0 \leq f_{ij} \leq 1 \ (i = 1, 2, \ldots, k; \ j = 1, 2, \ldots, l).$
- $\displaystyle\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} = n_{11} + n_{12} + \cdots + n_{kl} = n.$
- $\displaystyle\sum_{i=1}^{k} \sum_{j=1}^{l} f_{ij} = f_{11} + f_{12} + \cdots + f_{kl} = 1.$

Distribución de frecuencias **conjunta**

$X \setminus Y$	c'_1	c'_2	$\ldots$	c'_j	$\ldots$	c'_l
c_1	$n_{11}\ (f_{11})$	$n_{12}\ (f_{12})$	$\ldots$	$n_{1j}\ (f_{1j})$	$\ldots$	$n_{1l}\ (f_{1l})$
c_2	$n_{21}\ (f_{21})$	$n_{22}\ (f_{22})$	$\ldots$	$n_{2j}\ (f_{2j})$	$\ldots$	$n_{2l}\ (f_{2l})$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$
c_i	$n_{i1}\ (f_{i1})$	$n_{i2}\ (f_{i2})$	$\ldots$	$n_{ij}\ (f_{ij})$	$\ldots$	$n_{il}\ (f_{il})$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\ldots$	$\vdots$
c_k	$n_{k1}\ (f_{k1})$	$n_{k2}\ (f_{k2})$	$\ldots$	$n_{kj}\ (f_{kj})$	$\ldots$	$n_{kl}\ (f_{kl})$

Distribuciones marginales

- Llamaremos distribuciones marginales a las distribuciones de frecuencias unidimensionales de las variables X e Y.
- Para la variable X, las frecuencias absoluta y relativa de la modalidad c_i $(i = 1, 2, \ldots, k)$ son, respectivamente,

$$n_{i\cdot} = n_{i1} + n_{i2} + \cdots + n_{il} \quad \text{y} \quad f_{i\cdot} = \frac{n_{i\cdot}}{n} = f_{i1} + f_{i2} + \cdots + f_{il}$$

- Para la variable Y, las frecuencias absoluta y relativa de la modalidad c'_j $(j = 1, 2, \ldots, l)$ son, respectivamente,

$$n_{\cdot j} = n_{1j} + n_{2j} + \cdots + n_{kj} \quad \text{y} \quad f_{\cdot j} = \frac{n_{\cdot j}}{n} = f_{1j} + f_{2j} + \cdots + f_{kj}$$

Distribuciones de frecuencias

Tabla de frecuencias marginales absolutas (relativas) para X

X	c_1	c_2	$\cdots$	c_i	$\cdots$	c_k
	$n_{1\cdot}(f_{1\cdot})$	$n_{2\cdot}(f_{2\cdot})$	$\cdots$	$n_{i\cdot}(f_{i\cdot})$	$\cdots$	$n_{k\cdot}(f_{k\cdot})$

Tabla de frecuencias marginales absolutas (relativas) para Y

Y	c'_1	c'_2	$\cdots$	c'_j	$\cdots$	c'_l
	$n_{\cdot 1}(f_{\cdot 1})$	$n_{\cdot 2}(f_{\cdot 2})$	$\cdots$	$n_{\cdot j}(f_{\cdot j})$	$\cdots$	$n_{\cdot l}(f_{\cdot l})$

Distribuciones condicionadas

- Para cada modalidad c'_j de Y, la frecuencia absoluta de la modalidad c_i de X condicionada a $Y = c'_j$ es

$$n_{i/j} = n(X = c_i / Y = c'_j) = n_{ij} \quad (i = 1, 2, \ldots, k)$$

- La frecuencia relativa de c_i de X condicionada a $Y = c'_j$ es

$$f_{i/j} = f(X = c_i / Y = c'_j) = \frac{n_{ij}}{n_{\cdot j}} = \frac{f_{ij}}{f_{\cdot j}} \quad (i = 1, 2, \ldots, k)$$

- Análogamente, se obtienen las frecuencias, absolutas y relativas, de cada modalidad de Y condicionada por cada modalidad de X.

Distribuciones de frecuencias

Tabla de distribuciones de X condicionada por $Y = c'_j$

$X/Y = c'_j$	c_1	$\cdots$	c_i	$\cdots$	c_k
	$n_{1/j}(f_{1/j})$	$\cdots$	$n_{i/j}(f_{i/j})$	$\cdots$	$n_{k/j}(f_{k/j})$

Tabla de distribuciones de Y condicionada por $X = c_i$

$X/Y = c'_j$	c_1	$\cdots$	c_i	$\cdots$	c_k
	$n_{1/j}(f_{1/j})$	$\cdots$	$n_{i/j}(f_{i/j})$	$\cdots$	$n_{k/j}(f_{k/j})$

Representaciones gráficas

- Diagrama de dispersión. Es la construcción gráfica que resulta de representar en un sistema de ejes de coordenadas los pares de observaciones (x_i, y_i) $(i = 1, 2, \ldots, n)$ de la variable bidimensional (X, Y).
- Histograma. En un sistema de ejes tridimensional, se construye una cuadrícula en el plano (X,Y) y sobre estas prismas rectangulares cuyos volúmenes deberán ser proporcionales a las frecuencias absolutas (relativas) de las modalidades (c_i, c_j') definidas por cada rectángulo.
- Representación gráfica de distribuciones condicionadas. Las distribuciones (unidimensionales) condicionadas pueden representarse en gráficos conjuntos, donde se incluyen los diagramas de barras, histogramas o diagramas de caja relativos a una variable condicionada a los valores particulares de otra.

Medidas características

Momentos

- Sea $\{(x_i, y_i)\}_{i=1}^n$ una muestra de una variable cuantitativa (X, Y). El momento con respecto al origen de orden (r, s) $(r \geq 0, s \geq 0)$ de la variable (X, Y) es el número $a_{rs} = \frac{1}{n} \sum_{i=1}^n x_i^r y_i^s$
- El momento con respecto a las medias de orden (r, s) $(r \geq 0, s \geq 0)$ es $m_{rs} = \frac{1}{n} \sum_{i=1}^n (x_i - \overline{x})^r (y_i - \overline{y})^s$
- Las medias marginales de X e Y son, respectivamente, $a_{10} = \overline{x}$ y $a_{01} = \overline{y}$; siendo $m_{20} = s_x^2$ y $m_{02} = s_y^2$ las varianzas marginales de X e Y, respectivamente.

Medidas características

Covarianza

En el caso particular $r = 1$, $s = 1$ se obtiene el momento m_{11}, conocido como la covarianza de (X, Y)

$$Cov(X, Y) = s_{xy} = m_{11} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})$$

que puede interpretarse como una medida de la relación lineal entre las variables X e Y.

Covarianza. Propiedades

- La covarianza de (X, Y) es igual a la de (Y, X): $s_{xy} = s_{yx}$.
- La covarianza de (X, X) es igual a la varianza marginal de X: $s_{xx} = s_x^2$.
- Si a, b, c, d son constantes cualesquiera, la covarianza de (U, V), con $U = a + bX$ y $V = c + dY$, es $s_{uv} = bds_{xy}$.
- Para el cálculo en la práctica de la covarianza de (X, Y) puede utilizarse la relación

$$s_{xy} = a_{11} - a_{10}a_{01} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x}\,\overline{y}$$

Regresión lineal. Correlación

- Es interesante investigar la posible existencia de alguna relación de dependencia entre las variables y la construcción de algún modelo matemático que permita describir dicha relación.
- Entre las variables X e Y existe una relación de dependencia exacta si conociendo el valor de una se conoce exactamente el valor de la otra.
- X e Y son variables independientes si una variable no contiene información sobre la otra.
- La variable X contiene cierta información (incompleta) acerca de la variable Y, pudiéndose predecir aproximadamente el valor de Y a partir del conocimiento del valor que ha tomado X mediante la construcción de lo que llamaremos modelos de regresión; en estas situaciones diremos que X e Y son variables dependientes.

El coeficiente de correlación lineal

- El propósito del estudio de la correlación es la construcción de medidas del grado de dependencia o asociación entre variables estadísticas, siendo la más popular de éstas el coeficiente de correlación lineal.
- Dadas observaciones $\{(x_i, y_i)\}_{i=1}^n$ de una variable bidimensional (X, Y) se define el coeficiente de correlación lineal de X e Y como el número

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

que es una medida del grado de dependencia lineal entre las variables X e Y.

El coeficiente de correlación lineal. Propiedades

- El coeficiente de correlación lineal es una medida adimensional.
- $-1 \leq r_{xy} \leq 1$.
- Si $r_{xy} = \pm 1$ significa que existe dependencia lineal exacta entre X e Y. Si $r_{xy} = 0$ significa que no existe dependencia lineal entre X e Y (incorreladas). I
- $r_{xy} = r_{yx}$. Si $r_{xy} > 0$ la recta de regresión es creciente y si $r_{xy} < 0$ Idecreciente.
- $r_{xx} = 1$.
- Si a, b, c, d son constantes, dadas nuevas variables $U = a + bX$ y $V = c + dY$ se verifica que $r_{uv} = r_{xy}$, si $bd > 0$, y $r_{uv} = -r_{xy}$, si $bd < 0$.

Part III

Probabilidad

Introducción

- El concepto de probabilidad está asociado a experimentos (procesos de observación) donde existe incertidumbre sobre el resultado final, que desde un punto de vista práctico son la mayoría de los experimentos reales.
- La Teoría de la Probabilidad es importante como soporte teórico de la Estadística (Inferencia Estadística) y como herramienta en el estudio de la mayoría de las áreas de conocimiento: Ingeniería, Economía, Sociología, Medicina, Biología, etc.
- El origen de la Teoría de la Probabilidad está ligado al estudio de los juegos de azar, siendo pioneros los trabajos realizados por G. Cardano y G. Galilei en el siglo XVI. Actualmente constituye un área científica de intensa investigación.

Experimentos y sucesos

Experimentos

- Un experimento es "un proceso por medio del cual se obtiene una observación".
- Un experimento determinista es el que al realizarse repetidas veces, en idénticas condiciones, proporciona siempre el mismo resultado y, por tanto, puede predecirse de antemano.
- Un experimento aleatorio es el que puede dar lugar a diferentes resultados, conocidos previamente, sin que sea posible predecir cuál va a ser el resultado que va a ocurrir en una determinada realización del experimento.
- La Teoría de la Probabilidad y la Estadística estudian los experimentos aleatorios que, en mayor o menor medida, son todos los experimentos reales.

Experimentos y sucesos

Álgebra de sucesos

- Suceso elemental o simple: es cada uno de los posibles resultados del experimento aleatorio.
- Espacio muestral: es el conjunto formado por todos los sucesos elementales. Lo denotaremos por $\Omega = \{\omega/\omega$ es un suceso elemental$\}$. Se clasifica en: discreto (si es finito o infinito numerable) y continuo.
- Suceso: es un subconjunto del espacio muestral. Son sucesos de interés: Ω, el suceso seguro, formado por todos los sucesos elementales y $\emptyset$, el suceso imposible, que no contiene elementos.
- Álgebra de sucesos: es el conjunto formado por todos los sucesos asociados a un experimento aleatorio. Lo denotaremos por $\mathcal{A} = \{A/A$ es un suceso$\}$.

Experimentos y sucesos

Ejemplo 1

Considérese el experimento aleatorio "lanzar un dado y observar el número de puntos obtenido". Los sucesos elementales son $\omega_i =$ "se obtienen i puntos", donde $i = 1, 2, \ldots, 6$. Son sucesos $A =$ "se obtiene un número par" $=$ "el resultado es 2, 4 o 6" y $B =$ "se obtiene un número mayor que 2" $=$ "el resultado es 3, 4, 5 o 6".

Ejemplo 2

Considérese el experimento aleatorio "tiempo de ejecución de un programa". Los sucesos elementales son $\omega_t =$ "la ejecución ha durado t segundos", con $t \in \mathbb{R}$, $t \geq 0$. Son sucesos $C =$ "el tiempo de ejecución es superior a 10 segundos" y $D =$ "el tiempo de ejecución está entre 5 y 15 segundos".

Experimentos y sucesos

Álgebra de sucesos. Operaciones

- Unión de sucesos: si $A, B \in \mathcal{A}$, se define el suceso unión, $A \cup B$, como el que ocurre si sucede A o sucede B.
- Intersección de sucesos: si $A, B \in \mathcal{A}$, se define el suceso intersección, $A \cap B$, como el que ocurre si sucede A y sucede B. Por sencillez, $A \cap B$ también se escribe AB.
- Suceso complementario o contrario: si $A \in \mathcal{A}$, se define el suceso contrario, $\bar{A}$, como el que ocurre si no sucede A.
- Inclusión de sucesos: si $A, B \in \mathcal{A}$, se dice que A está contenido en B o que A implica B, $A \subset B$, si siempre que sucede A ocurre B.

Álgebra de sucesos. Operaciones

- Diferencia de sucesos: si $A, B \in \mathcal{A}$, se define el suceso diferencia, $A \backslash B$, como el que ocurre si sucede A y no sucede B, esto es,
$$A \backslash B = A \cap \bar{B}$$
- Diferencia simétrica de sucesos: si $A, B \in \mathcal{A}$, se define el suceso diferencia simétrica, $A \nabla B$, como el que ocurre si sucede sólo A o sólo B, esto es,
$$A \nabla B = (A \cup B) \backslash (A \cap B) = (A \cap \bar{B}) \cup (B \cap \bar{A})$$
- Sucesos incompatibles: dos sucesos $A, B \in \mathcal{A}$ son incompatibles si $A \cap B = \emptyset$.

Experimentos y sucesos

Álgebra de sucesos. Operaciones

- Conjunto exhaustivo de sucesos: $\{A_1, A_2, \ldots, A_n / \ A_i \in \mathcal{A}\}$ es un conjunto exhaustivo de sucesos si $\bigcup_{i=1}^n A_i = \Omega$
- Conjunto completo de sucesos: $\{A_1, A_2, \ldots, A_n / \ A_i \in \mathcal{A}\}$ es un conjunto completo de sucesos si es exhaustivo y los sucesos son incompatibles dos a dos:
  $\bigcup_{i=1}^n A_i = \Omega$ y $A_i \cap A_j = \emptyset$ si $i \neq j$
  A un "conjunto completo de sucesos" también se le denomina partición del espacio muestral. El conjunto de los sucesos elementales es una clase completa de sucesos y la partición más fina del espacio muestral.
- El álgebra de sucesos, $\mathcal{A}$, asociada a un experimento aleatorio tiene estructura de álgebra de Boole respecto a las operaciones unión e intersección

Experimentos y sucesos

Álgebra de sucesos. Propiedades

- Conmutativa. $A \cup B = B \cup A, \qquad A \cap B = B \cap A$
- Asociativa.] $A \cup (B \cup C) = (A \cup B) \cup C,$
  $A \cap (B \cap C) = (A \cap B) \cap C$
- Elemento neutro. El suceso imposible ($\emptyset$) para la unión $A(\cup\emptyset) = A$ y el suceso seguro (Ω) para la intersección ($A \cap \Omega = A$).
- Complementario. Dado $A \in \mathcal{A}$ existe $\bar{A}$, que llamaremos suceso complementario o contrario de A, tal que $A \cup \bar{A} = \Omega$ y $A \cap \bar{A} = \emptyset$.

Álgebra de sucesos. Propiedades

- Idempotente. $A \cup A = A, \qquad A \cap A = A$
- Simplificativa. $A \cup (A \cap B) = A, \qquad A \cap (A \cup B) = A$
- Relativas al elemento neutro. $A \cup \Omega = \Omega, \qquad A \cap \emptyset = \emptyset$
- Leyes de De Morgan. $\overline{A \cup B} = \bar{A} \cap \bar{B}, \qquad \overline{A \cap B} = \bar{A} \cup \bar{B}$

Experimentos y sucesos

Ejemplo 3

Respecto al experimento del ejemplo 1 se obtienen los siguientes sucesos:

$A \cup B =$ "obtener 2, 3, 4, 5 o 6".

$A \cap B =$ "obtener 4 o 6".

$\bar{A} =$ "obtener un número impar".

$\bar{B} =$ "obtener 1 o 2".

$A \backslash B =$ "obtener el 2".

$B \backslash A =$ "obtener 3 o 5".

$A \nabla B =$ "obtener 2, 3 o 5".

Experimentos y sucesos

Ejemplo 4

Respecto al experimento del ejemplo 2 se obtienen los siguientes sucesos:

$C \cup D=$ "el tiempo de ejecución es superior a 5 segundos".

$C \cap D=$ "el tiempo de ejecución está entre 10 y 15 segundos".

$\bar{C}=$ "el tiempo de ejecución es inferior o igual a 10 segundos".

$\bar{D}=$ "el tiempo de ejecución es menor o igual que 5 segundos o mayor o igual que 15 segundos".

$C \backslash D=$ "el tiempo de ejecución es mayor o igual que 15 segundos".

$D \backslash C=$ "el tiempo de ejecución es superior a 5 segundos y menor o igual que 10 segundos".

$C \nabla D=$ "el tiempo de ejecución está en $(5, 10] \cup [15, \infty)$".

Definición de probabilidad

Definición axiomática de Kolmogorov

La probabilidad (P) asociada a un experimento aleatorio es una aplicación del álgebra de sucesos ($\mathcal{A}$) en $\mathbb{R}$

$$P : \mathcal{A} \longrightarrow \mathbb{R}$$

verificando los siguientes axiomas:

1. Para todo suceso A, $P(A) \geq 0$
2. $P(\Omega) = 1$
3. ($\sigma-$aditividad) Si $\{A_n\}_{n=1}^{\infty}$ es una sucesión de sucesos incompatibles dos a dos, entonces:

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Definición de probabilidad

Definición

Llamaremos espacio de probabilidad a la terna, $(\Omega, \mathcal{A}, \mathcal{P})$ formada por el espacio muestral (Ω), el álgebra de sucesos $(\mathcal{A})$ y la aplicación (P) verificando los anteriores axiomas.

Definición de probabilidad

Ejemplo 5

En relación con el experimento del ejemplo 1, puede definirse la función de probabilidad a partir de la probabilidad de los sucesos elementales, $A_i =$ "obtener el número i", de la siguiente forma: $P(A_i) = \frac{1}{6}, \quad i = 1, 2, \ldots, 6$

Ejemplo 6

En relación con el experimento del ejemplo 2, puede definirse la función de probabilidad a partir de la probabilidad de sucesos de la forma $A_t =$ "la duración de la ejecución del programa es inferior a t segundos", como $P(A_t) = 1 - e^{-t}$ $(t > 0)$.

Definición de probabilidad

Propiedades

1. $P(\emptyset) = 0$.
2. Si $\{A_i\}_{i=1}^n$ es un conjunto de sucesos incompatibles dos a dos entonces, $P\left(\bigcup_{j=1}^{n} A_j\right) = \sum_{j=1}^{n} P(A_j)$
3. $P(\bar{A}) = 1 - P(A)$.
4. Para cualquier suceso A, $0 \leq P(A) \leq 1$.
5. Si $A \subset B$ entonces $P(A) \leq P(B)$ y $P(B \backslash A) = P(B) - P(A)$.
6. Para dos sucesos cualesquiera A y B se verifica que, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Definición de probabilidad

Ejemplo 7

La probabilidad de que el estudiante A apruebe un examen es $0'5$, la probabilidad de que apruebe B es $0'3$ y la probabilidad de que aprueben los dos es $0'2$.

- La probabilidad de que al menos uno de los dos apruebe es $P(A \cup B) = 0'5 + 0'3 - 0'2 = 0'6$.
- La probabilidad de que exactamente uno de los dos apruebe es $P(A \nabla B) = P(A) + P(B) - 2P(A \cap B) = 0'4$.
- La probabilidad de que no apruebe ni A ni B es $P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - 0'6 = 0'4$.
- La probabilidad de que apruebe A pero no B es $P(A \cap \bar{B}) = P(A \backslash B) = P(A) - P(A \cap B) = 0'5 - 0'2 = 0'3$.

Ejemplo 8

Supóngase que la probabilidad de obtener el número i al lanzar un dado es inversamente proporcional a dicho número. Calcular la probabilidad de obtener un número par en una tirada.

Llamamos $p_i = P(\text{“obtener el número } i\text{”}) = k/i$, $i = 1, 2, \ldots, 6$, con k una constante por determinar, que obtenemos de la siguiente igualdad

$$\sum_{i=1}^{6} p_i = k \sum_{i=1}^{6} \frac{1}{i} = P(\Omega) = 1 \implies k = \frac{60}{147}$$

Por tanto, $P(\text{“obtener un número par”}) = \dfrac{55}{147}$.

Definición de probabilidad

Asignación de probabilidades

- Método de las frecuencias: Definir la probabilidad del suceso como el límite de las frecuencias realtivas.
- Método clásico: En los espacios muestrales finitos equiprobables, podemos calcular la probabilidad del suceso A como el cociente entre el número de *"casos favorables"* en que sucede A y el número de *"casos posibles"* que se pueden dar. Esta regla se conoce como definición clásica o Ley de Laplace.
- Método subjetivo: en el que una determinada persona asigna de forma subjetiva probabilidades a cada uno de los posibles resultados de un proceso según su propio juicio sobre la verosimilitud de cada resultado.

Ejemplo 9

En un curso de Estadística de 80 estudiantes aprobaron 50, de los que 35 eran chicas. La probabilidad de que haya aprobado un alumno elegido al azar es: $P(\text{aprobar}) = \frac{50}{80} = 0'625$

Pero si el número de chicas que participaron en el curso fue de 45, entonces la probabilidad de que haya aprobado un alumno elegido al azar *sabiendo que es una chica*, es:

$$\begin{aligned} P(\text{aprobar}/\text{ser chica}) &= \frac{P(\text{aprobar y ser chica})}{P(\text{ser chica})} \\ &= \frac{35/80}{45/80} = 0'777 \end{aligned}$$

Probabilidad condicionada e independencia de sucesos

Definición

Sean A y B dos sucesos cualesquiera con $P(B) > 0$. Se define la probabilidad del suceso A condicionada al suceso B y se representa por $P(A/B)$ como:

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

Probabilidad condicionada. Comentarios

1. La probabilidad condicionada es muy importante en la práctica, ya que, en muchas situaciones, pequeñas modificaciones en la información básica producen cambios sustanciales en las probabilidades condicionadas.
2. Con la definición anterior, es fácil probar que la probabilidad condicionada a un suceso B verifica la axiomática de la probabilidad dada en la definición.
3. Es importante diferenciar entre $P(AB)$ y $P(A/B)$: la primera indica la probabilidad de ocurrencia de A y B conjuntamente, por tanto siempre es menor o igual que $P(A)$; y la segunda indica la probabilidad de ocurrencia de A cuando es conocido que ha ocurrido el suceso B y puede ser menor, igual o mayor que $P(A)$.

Ejemplo 10

En un almacén se dispone de diez motores de los cuales tres son defectuosos. Si se eligen dos motores al azar y Denominando por D_i al suceso "el motor elegido en lugar i-ésimo es defectuoso" y N_i al suceso "el motor elegido en lugar i-ésimo es no defectuoso", se pueden calcular las siguientes probabilidades condicionadas

1. $P(D_2/N_1) = \frac{P(N_1 \cap D_2)}{P(N_1)} = \frac{7/10 \cdot 3/9}{7/10} = \frac{3}{9}$
2. $P(D_2/D_1) = \frac{P(D_1 \cap D_2)}{P(D_1)} = \frac{3/10 \cdot 2/9}{3/10} = \frac{2}{9}$
3. $P(D_2) = \frac{3}{10}$

Probabilidad condicionada e independencia de sucesos

Ejemplo 11

En una encuesta realizada en La Coruña se ha determinado que el 40% de los encuestados lee el periódico *La Voz de Galicia*, el 15% lee *El Ideal Gallego* y el 3% lee ambos periódicos.

1. Seleccionado al azar un lector de *El Ideal Gallego*, calcular la probabilidad de que lea *La Voz de Galicia*.
   Sea V el suceso "lee *La Voz de Galicia*", e I el suceso "lee *El Ideal Gallego*", entonces $P(V/I) = \frac{P(V \cap I)}{P(I)} = \frac{3}{15} = 0'2$
2. Si se ha elegido un lector de *La Voz de Galicia*, calcular la probabilidad de que no lea *El Ideal Gallego*.
   $P(\bar{I}/V) = 1 - \frac{P(I \cap V)}{P(V)} = 1 - \frac{3}{40} = 0'925$

Ejemplo 12

En un centro de secundaria el 50% de los alumnos aprueba el Bachillerato. Se estima que si se presentasen todos los alumnos a las pruebas de Selectivo sólo suspenderían el 40% y que un 30% de los alumnos que aprobarían el Selectivo suspenden el Bachillerato. Con estos datos calcular la probabilidad de que un alumno que apruebe el Bachillerato apruebe el Selectivo.

Sea C el suceso "aprueba el Bachillerato" y S el suceso "aprueba el Selectivo", por tanto, $P(C) = 0'50$, $P(\bar{S}) = 0'40$ $P(\bar{C}/S) = 0'30$.

La probabilidad pedida es

$$P(S/C) = \frac{P(S \cap C)}{P(C)} = \frac{P(C/S)P(S)}{P(C)} = \frac{0'70\ 0'60}{0'50} = 0'84$$

Regla del producto

Sean $A_1, A_2, \ldots, A_n$ sucesos tales que $P\left(\bigcap_{i=1}^{n-1} A_i\right) > 0.$

Entonces:

$$P\left(\bigcap_{i=1}^{n} A_i\right) =$$

$$= P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \cdots P\left(A_n \Big/ \left(\bigcap_{i=1}^{n-1} A_i\right)\right)$$

Ejemplo 13

En relación con el ejemplo 10, si se eligen cuatro motores al azar, sin reemplazamiento, calcular la probabilidad de que el primer y el tercer motores elegidos sean defectuosos y los otros dos no.

$$P(D_1 N_2 D_3 N_4) =$$

$$= P(D_1)P(N_2/D_1)P(D_3/D_1 N_2)P(N_4/D_1 N_2 D_3) =$$

$$= \frac{3}{10}\,\frac{7}{9}\,\frac{2}{8}\,\frac{6}{7} = \frac{1}{20} = 0'05$$

Definición

Dos sucesos A y B se dicen independientes si

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

o, equivalentemente, $P(A/B) = P(A)$, si $P(B) > 0$, o bien $P(B/A) = P(B)$, si $P(A) > 0$.

Probabilidad condicionada e independencia de sucesos

Independencia de sucesos. Comentarios

1. La independencia de sucesos puede suponerse en algunas situaciones y deducirse del contexto del problema pero, en general, debe comprobarse experimentalmente.
2. No debe confundirse sucesos independientes con sucesos incompatibles.
3. Si A y B son sucesos independientes también lo son A y $\bar{B}$, $\bar{A}$ y B y $\bar{A}$ y $\bar{B}$.
4. Los sucesos $A_1, A_2, \ldots, A_n$ son mutuamente independientes si $P\left(\bigcap_{h=1}^{k} A_{j(h)}\right) = \prod_{h=1}^{k} P\left(A_{j(h)}\right)$ para cualesquiera índices $1 \leq j(1) < j(2) < \cdots < j(k) \leq n$.

Teorema de Bayes

Teorema de las probabilidades totales

Sea $A_1, A_2, \ldots, A_n$ un sistema completo de sucesos, con $P(A_i) > 0$ $(i = 1, \ldots, n)$, y sea B un suceso cualquiera. Entonces

$$P(B) = \sum_{i=1}^{n} P(B/A_i)P(A_i)$$

Ejemplo 15

En una escuela técnica el 50% de los alumnos es de primer curso, el 30% es de segundo y el 20% de tercero. De la encuesta de evaluación de profesorado se sabe que el 60% de los alumnos de primero tiene buena opinión del profesorado, al igual que el 70% de los de segundo y el 75% de los de tercero. Elegido un alumno al azar ¿cuál es la probabilidad de que tenga una buena opinión del profesorado?

Teorema de Bayes

Ejemplo 15

Si consideramos el suceso $B =$ "tener buena opinión del profesorado" y el sistema completo de sucesos formado por $I =$ "ser de primero", $S =$ "ser de segundo" y $T =$ "ser de tercero", la probabilidad pedida es:

$$
\begin{aligned}
P(B) &= P(B \cap \Omega) = P(B \cap (I \cup S \cup T)) \\
&= P(B \cap I) + P(B \cap S) + P(B \cap T) \\
&= P(B/I)P(I) + P(B/S)P(S) + P(B/T)P(T) \\
&= 0'6 \cdot 0'5 + 0'7 \cdot 0'3 + 0'75 \cdot 0'2 = 0'66
\end{aligned}
$$

Ejemplo 16

En una estación de ITV (Inspección Técnica de Vehículos) hay dos equipos de inspección, el equipo A rechaza el 30% de los coches inspeccionados y el equipo B no rechaza ningún coche. Si llegan tres coches a la estación y cada uno elige al azar uno de los dos equipos de inspección, ¿cuál es la probabilidad de que los tres coches superen la inspección?

Teorema de Bayes

Ejemplo 16

Sean los sucesos $A =$ "elegir equipo A", $B =$"elegir equipo B" y $S =$ "superar la inspección", por el teorema de las probabilidades totales se obtiene

$$P(S) = P(S/A)P(A) + P(S/B)P(B) = 0'7 \cdot 0'5 + 1 \cdot 0'5 = 0'85$$

Denominemos S_i al suceso "el coche i supera la inspección", con $i = 1, 2, 3$. Por la independencia de estos sucesos, la probabilidad pedida es

$$P(S_1 \cap S_2 \cap S_3) = 0'85^3 = 0'6141$$

Teorema de Bayes

Teorema de Bayes

Sea $A_1, A_2, \ldots, A_n$ un sistema completo de sucesos, con $P(A_i) > 0$ para $i = 1, \ldots, n$, (probabilidades a priori) y sea B un suceso cualquiera, con $P(B) > 0$. Entonces, para $j = 1, 2, \ldots, n$,

$$P(A_j/B) = \frac{P(A_j B)}{P(B)} = \frac{P(B/A_j)P(A_j)}{\sum_{i=1}^n P(B/A_i)P(A_i)},$$

llamadas probabilidades a posteriori.

Teorema de Bayes

Ejemplo 18

En un examen tipo test con cinco posibles respuestas, la probabilidad de que Juan sepa la respuesta es $0'6$, la probabilidad de que responda al azar es $0'2$ y la probabilidad de que no responda es $0'2$. Si el estudiante respondió correctamente ¿cuál es la probabilidad de que realmente sepa la respuesta?

Ejemplo 18

Sean los sucesos $S =$ "Juan sabe la respuesta", $A =$ "Juan responde al azar" y $N =$ "Juan no responde", con probabilidades: $P(S) = 0'6$, $P(A) = 0'2$ y $P(N) = 0'2$. Sea C el suceso "Juan respondió correctamente", se verifica que $P(C/S) = 1$, $P(C/A) = 1/5 = 0'2$ y $P(C/N) = 0$. Por el teorema de Bayes se obtiene:

$$P(S/C) = \frac{0'6 \cdot 1}{0'6 \cdot 1 + 0'2 \cdot 0'2 + 0'2 \cdot 0} = \frac{0'6}{0'64} = 0'9375$$

análogamente, $P(A/C) = \frac{0'04}{0'64} = 0'0625$ y, $P(N/C) = \frac{0}{0'64} = 0$

Análisis combinatorio

Definición

Sean n y k dos números naturales tales que $k \leq n$, se define el número combinatorio $\binom{n}{k}$ como,

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!} = \frac{n^{(k)}}{k!} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)(k-2)\cdots 1}$$

- El número combinatorio $\binom{n}{k}$ también se notará $C_{n,k}$.
- Este número se conoce como *coeficiente binomial*, por aparecer en el *teorema binomial o binomio de Newton*,

$$(a+b)^n = \sum_{j=0}^{n} \binom{n}{j} a^j b^{n-j}$$

Análisis combinatorio

Propiedades

1. $\binom{n}{n} = \binom{n}{0} = 1.$
2. $\binom{n}{1} = n.$
3. $\binom{n}{k} = \binom{n}{n-k}.$
4. $\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = (1+1)^n = 2^n.$

Part IV

Variables aleatorias unidimensionales

Definición de variable aleatoria

Definición

Dado un experimento aleatorio, con espacio muestral asociado Ω, una variable aleatoria es cualquier función, X,

$$X : \Omega \longrightarrow \mathbb{R}$$

que asocia a cada suceso elemental un número real, verificando que, para cualquier número real r, es un suceso el conjunto

$$\{w \in \Omega \text{ tales que } X(w) \leq r\} = X^{-1}\left((-\infty, r]\right)$$

Definición de variable aleatoria

Ejemplo 1

Se examinan las piezas producidas por una máquina. Representando por D el resultado "la pieza producida es defectuosa", sobre el espacio muestral $\Omega = \{D, \overline{D}\}$ puede definirse la variable aleatoria X :

$$X(w) = \begin{cases} 1 & \text{si} \quad w = D \\ 0 & \text{si} \quad w = \overline{D} \end{cases}$$

X es una variable aleatoria ya que $X^{-1}((-\infty, r])$ es un suceso:

$$X^{-1}((-\infty, r]) = \begin{cases} \phi & r < 0 \\ \{\overline{D}\} & 0 \leq r < 1 \\ \Omega & r \geq 1 \end{cases}$$

Función de distribución de una variable aleatoria

La función de distribución de una variable aleatoria X es una función real que a cada número real x le asocia la probabilidad de que la variable tome valores menores o iguales que dicho número, esto es:

$$F(x) = P(X \leq x) = P\left(\{w \in \Omega \text{ tales que } X(w) \leq x\}\right)$$

$$= P(X^{-1}\left((-\infty, x]\right))$$

La definición de F está garantizada porque el conjunto $X^{-1}\left((-\infty, x]\right)$ es un suceso.

Definición de variable aleatoria

Propiedades de la función de distribución

1. $0 \leq F(x) \leq 1$.
2. F es no decreciente $(x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2))$.
3. $F(+\infty) = 1 \left(\lim_{x \to \infty} F(x) = 1 \right)$.
4. $F(-\infty) = 0 \left(\lim_{x \to -\infty} F(x) = 0 \right)$.
5. F es continua por la derecha $\left(F(x^+) = \lim_{h \to 0^+} F(x+h) = F(x) \right)$.

Variables aleatorias discretas

Definición

Una variable aleatoria discreta es aquella que sólo puede tomar valores dentro de un conjunto finito o infinito numerable.

Ejemplos

- En un proceso de control de calidad se analiza el porcentaje de piezas defectuosas fabricadas, asociando "pieza defectuosa" con 1 y "pieza no defectuosa" con 0.
- Lanzar dos dados de seis caras equiprobables, se considera como variable la correspondencia que asocia a cada resultado la "suma de los valores aparecidos".
- "el número de conexiones a Internet a lo largo de un mes", "el número de piezas producidas antes de la primera defectuosa", etc.

Variables aleatorias discretas

Función de masa de probabilidad de una variable aleatoria discreta

- Sea X una v.a. discreta que toma los valores x_i con $p_i = P(X = x_i)$, con $\sum_i p_i = 1$. Se denomina función de masa de probabilidad o función de probabilidad de X a la función que asigna a cada x_i su probabilidad p_i.
- La función de distribución de una variable discreta es: $F(x) = P(X \leq x) = \sum_{x_i \leq x} p_i$
- La función F es escalonada, no decreciente, con saltos de discontinuidad en los puntos x_i. El valor del salto en x_i coincide con la probabilidad, p_i, de dicho valor.

Ejemplo 1

Sea $X =$ "suma de los puntos obtenidos al lanzar dos dados", la función de masa de probabilidad es:

$$\begin{aligned}
p_1 &= P(X=2) = P(\{(1,1)\}) = \frac{1}{36} \\
p_2 &= P(X=3) = P(\{(1,2),(2,1)\}) = \frac{2}{36} \\
p_3 &= P(X=4) = P(\{(2,2),(1,3),(3,1)\}) = \frac{3}{36} \\
&\ \ \vdots \qquad\qquad\qquad \vdots \\
p_{11} &= P(X=12) = P(\{(6,6)\}) = \frac{1}{36}
\end{aligned}$$

Ejemplo 1

Su función de distribución es:

$$F(x)=P(X\leq x)=\begin{cases} 0 & \text{si} \quad x<2 \\ 1/36 & \text{si} \quad 2\leq x<3 \\ 3/36 & \text{si} \quad 3\leq x<4 \\ 6/36 & \text{si} \quad 4\leq x<5 \\ 10/36 & \text{si} \quad 5\leq x<6 \\ 15/36 & \text{si} \quad 6\leq x<7 \\ 21/36 & \text{si} \quad 7\leq x<8 \\ 26/36 & \text{si} \quad 8\leq x<9 \\ 30/36 & \text{si} \quad 9\leq x<10 \\ 33/36 & \text{si} \quad 10\leq x<11 \\ 35/36 & \text{si} \quad 11\leq x<12 \\ 1 & \text{si} \quad x\geq 12 \end{cases}$$

Variables aleatorias continuas

Definición

Una variable aleatoria continua es aquella que toma valores en uno o varios intervalos de la recta real.

Ejemplos

- "la hora de llegada de un profesor a su despacho".
- "la duración de una llamada telefónica".
- "la cantidad de lluvia caída por metro cuadrado en una determinada zona".
- "la superficie de planchas metálicas producidas en una factoría".

Función de densidad de una variable aleatoria continua

- Dada una variable aleatoria continua X su función de densidad es la función real de variable real:

$$f(x) = \lim_{h \to 0^+} \frac{P(x - h \leq X \leq x + h)}{2h}$$

- De la definición anterior, se deduce que $f(x) = F'(x) = \frac{dF(x)}{dx}$
- Dada una variable continua X, con función de densidad f, su función de distribución es $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$

Propiedades de la función de densidad

1. $f(x) \geq 0$, $-\infty < x < \infty$. El conjunto de valores donde $f(x) > 0$ se llama soporte de X.
2. $\int_{-\infty}^{\infty} f(x)dx = F(+\infty) = 1$
3. $P(a \leq X \leq b) = \int_a^b f(t)dt = F(b) - F(a)$
4. $F(x_0) = P(X \leq x_0)$ mide el área de la región limitada por la función de densidad, el eje de abscisas y la recta $x = x_0$.
5. $P(X = x_0) = \int_{x_0}^{x_0} f(t)dt = 0$
6. En general, cualquier función real que verifica las propiedades 1 y 2 es la función de densidad de alguna variable aleatoria continua X.

Ejemplo 2

Un profesor llega cada día a su despacho con igual probabilidad entre las 8 y las 9 horas. ¿Cuál es la función de densidad de la variable $X =$ "hora de llegada al despacho"? ¿Cuál es la probabilidad de que llegue antes de las ocho y media?

Se trata de una variable continua con función de densidad de X es:

$$f(x) = \begin{cases} 1 & \text{si} \quad 8 < x < 9 \\ 0 & \text{si} \quad x \notin (8,9) \end{cases}$$

La probabilidad de que llegue antes de las ocho y media es:

$$P(X \leq 8'5) = \int_8^{8'5} 1 dx = 0'5$$

Variables aleatorias continuas

Ejemplo 3

Dada una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 3x^2/8 & \text{si } 0 \leq x \leq 2 \\ 0 & \text{si } x \notin (0,2) \end{cases}$$

su función de distribución es:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ x^3/8 & \text{si } 0 \leq x \leq 2 \\ 1 & \text{si } x > 2 \end{cases}$$

La probabilidad de que la variable tome valores entre $1'5$ y 2:
$P(1'5 < X < 2) = \int_{1'5}^{2} \frac{3t^2}{8} dt = 0'578$

Estadística Aplicada
Ricardo Cao
Definición de variable aleatoria
Variables aleatorias discretas
Variables aleatorias continuas
Transformación de variables aleatorias
Características de una variable aleatoria

Ejemplo 4

En el caso de la variable "duración de una llamada de teléfono", se puede proponer como función de densidad:

$$f(x) = \begin{cases} e^{-x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Su función de distribución es:

$$F(x) = \begin{cases} 1 - e^{-x} & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases}$$

Variables aleatorias mixtas

Definición

Una variable que tome unos valores puntuales con probabilidad dada, y el resto de los valores los tome dentro de uno o varios intervalos siguiendo una determinada función de densidad, se dirá que es una variable aleatoria con distribución mixta.

Comentario

Por simplicidad, en la definición anterior se está usando el concepto de función de densidad en un sentido amplio, pues, si una variable sigue un modelo mixto de distribución, es evidente que la probabilidad acumulada en los intervalos en que ésta se comporta de modo continuo (igual al valor de la integral de la función de densidad en dichos intervalos) es inferior a la unidad.

Ejemplo 5

Se define la variable mixta X que toma los valores -1 y 0, con probabilidades respectivas $0'1$ y $0'2$, y que toma valores en el intervalo $(1,2)$ de acuerdo con la función de densidad

$$f(x) = \begin{cases} k(2x-1) & \text{si } x \in (1,2) \\ 0 & \text{en otro caso} \end{cases}$$

El valor de $k = 0'35$ verifica que la suma de todas las probabilidades es 1, obtenido resolviendo:

$$\begin{aligned} 1 &= P(X=-1) + P(X=0) + P\left(X \in (1,2)\right) \\ &= 0'1 + 0'2 + \int_1^2 k(2x-1)dx \end{aligned}$$

Razón de fallo de una variable continua

Definición

Sea T la variable aleatoria "tiempo de vida de una componente o sistema", y sea $F(t)$ su función de distribución. La función de fiabilidad o de supervivencia de la variable aleatoria T es: $R(t) = P(T > t) = 1 - F(t), \qquad t > 0$

Definición

La razón de fallo de una componente o sistema es la proporción de unidades que fallan en un intervalo de tiempo $(t, t + dt)$, con respecto a las que siguen funcionando en t. Esto es:

$$h(t) = \frac{f(t)}{R(t)} = -\frac{R'(t)}{R(t)}$$

Razón de fallo de una variable continua

Ejemplo 6

Dada la variable aleatoria con función de densidad

$$f(t) = \begin{cases} 24t^2 \exp(-(2t)^3) & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases}$$

sus funciones de fiabilidad y razón de fallo son:

$$R(t) = 1 - F(t) = \begin{cases} \exp(-(2t)^3) & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases}$$

$$h(t) = \frac{f(t)}{R(t)} = \begin{cases} 24t^2 & \text{si } t > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Razón de fallo de una variable continua

Ejemplo 7

Si la duración de una componente electrónica es una variable aleatoria T, con función de densidad

$$f(t) = \begin{cases} \alpha e^{-\alpha t} & \text{si } t > 0 \\ 0 & \text{en el resto} \end{cases}$$

sus funciones de fiabilidad y razón de fallo son:

$$R(t) = 1 - F(t) = e^{-\alpha t}$$

$$h(t) = \frac{f(t)}{R(t)} = \alpha$$

La razón de fallo es constante y la variable no tiene memoria.

Características de una variable aleatoria

Definición

Sea X una variable aleatoria discreta que toma los valores x_i con probabilidades p_i. Supuesto que $\sum_i |x_i| p_i < \infty$, se define la media, valor esperado o esperanza matemática de la variable X como el número real: $\mu = E(X) = \sum_i x_i p_i$

Definición

Sea X una variable aleatoria continua con función de densidad f. Supuesto que $\int_{-\infty}^{\infty} |x| f(x)dx < \infty$, se define la media, valor esperado o esperanza matemática de la variable X como el número real: $\mu = E(X) = \int_{-\infty}^{\infty} xf(x)dx$

Características de una variable aleatoria

Propiedades de la esperanza

1. $E(aX+b) = aE(X) + b$
2. $E(X+Y) = E(X) + E(Y)$
3. Si X es discreta con valores x_i, y probabilidades p_i, la media de $Y = g(X)$ es: $E(Y) = E(g(X)) = \sum_i g(x_i)p_i$
4. Si X es una variable continua con densidad f, la media de $Y = g(X)$ es: $E(Y) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx$

Ejemplo 9

Valor esperado de la variable $X =$ "suma de los resultados obtenidos al lanzar dos dados":

$$p_i = \begin{cases} (6-|6-i|)/36 & \text{si} \quad i = 1, 2, \ldots, 11 \\ 0 & \text{en otro caso} \end{cases}$$

$$E(X) = \sum_{i=1}^{11} x_i \cdot p_i = 2\frac{1}{36} + 3\frac{2}{36} + \cdots + 12\frac{1}{36} = 7$$

Ejemplo 10

Media de la variable aleatoria continua cuya función de densidad está dada por:

$$f(x) = \begin{cases} 1 & \text{si } 8 \leq x \leq 9 \\ 0 & \text{si } x \notin [8, 9] \end{cases}$$

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{8}^{9} xdx = 8'5$$

La media de $Y = g(X) = X^3$ es

$$E(Y) = E(g(X)) = \int_{8}^{9} x^3 dx = 616'25$$

Características de una variable aleatoria

Definición

Sea X una variable aleatoria con media $\mu = E(X)$. Se define la varianza de X como el valor esperado de los cuadrados de las diferencias con la media: $\sigma^2 = Var(X) = E[(X - E(X))^2]$

Definición

Se define la desviación típica de la variable X como la raíz positiva de la varianza: $\sigma = +\sqrt{E[(X - E(X))^2]}$

Propiedades

1. $Var(aX + b) = a^2 Var(X)$.
2. $Var(X) = E(X^2) - [E(X)]^2$.

Características de una variable aleatoria

Tipificacíón de una variable aleatoria

- Una variable aleatoria se dirá que está estandarizada o tipificada, si su media es 0 y su varianza es 1.
- Para transformar una variable X con media μ y desviación típica σ en otra tipificada, basta con aplicar la transformación lineal, $Y = \dfrac{X-\mu}{\sigma}$.
- La nueva variable, Y, tiene media 0 y varianza 1.

Características de una variable aleatoria

Otras medidas características de una variable aleatoria

- Se define el coeficiente de variación de una variable X, con media $\mu > 0$ y desviación típica σ, como : $CV(X) = \frac{\sigma}{\mu}$
- Se define el momento con respecto al origen de orden r, α_r, de una variable aleatoria X como $\alpha_r = E(X^r)$
- Se define el momento central o con respecto a la media de orden r, μ_r, de X como $\mu_r = E\left((X - E(X))^r\right)$
- $\alpha_1 = E(X)$ y $\mu_2 = Var(X)$.
- Haciendo uso del *binomio de Newton*, se tiene que:

$$\mu_r = \sum_{k=0}^{r} (-1)^k \binom{r}{k} \alpha_{r-k} \alpha_1^k$$

Características de una variable aleatoria

Otras medidas características de una variable aleatoria

- La mediana de una variable aleatoria es la medida de centralización que divide la distribución en dos partes de igual probabilidad. Es el valor M_e que, para una variable con distribución F, verifica: $M_e = \inf\{x/F(x) \geq 1/2\}$
- Cuantiles de orden p, siendo $0 < p < 1$, son los valores Q_p, que dejan un $100p\%$ de la distribución de probabilidad a su izquierda. Esto es: $Q_p = \inf\{x/F(x) \geq p\}$
- Los cuantiles más usados son $Q_{\frac{1}{4}}$, $Q_{\frac{1}{2}}$ y $Q_{\frac{3}{4}}$ que se denominan también primer, segundo y tercer cuartil.
- Recorrido intercuartílico $RI = Q_{\frac{3}{4}} - Q_{\frac{1}{4}}$

Características de una variable aleatoria

Otras medidas características de una variable aleatoria

- La moda de una variable es el valor M_0 que maximiza la función de probabilidad o la función de densidad, según se trate de una variable discreta o continua, respectivamente.
- El coeficiente de asimetría es el cociente $\gamma_1 = \frac{\mu_3}{\sigma^3}$
- Si γ_1 es 0 la distribución es simétrica. En otro caso presentará asimetría positiva o negativa de acuerdo con el signo de γ_1.
- El coeficiente de apuntamiento o curtosis es el número $\gamma_2 = \frac{\mu_4}{\sigma^4} - 3$
- Si γ_2 es 0 la distribución presenta una forma similar a la distribución normal.

Ejemplo 13

Vamos a calcular algunas de las medidas características de la distribución continua que tiene por función de densidad

$$f(x) = \begin{cases} 0'1 & \text{si } 0 \leq x \leq 10 \\ 0 & \text{en otro caso} \end{cases}$$

- Media y varianza:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^{10} \frac{1}{10} x dx = 5$$

$$\sigma^2 = E\left((x-5)^2\right) = \int_0^{10} \frac{1}{10} (x-5)^2 dx = 8'3333$$

Características de una variable aleatoria

Ejemplo 13

- Coeficiente de variación: $CV = \frac{\sigma}{\mu} = 0'57735$
- Momentos con respecto al origen:

$$\alpha_k = \int_0^{10} \frac{1}{10} x^k dx = \left[\frac{x^{k+1}}{(k+1)\,10} \right]_0^{10} = \frac{10^k}{k+1}$$

- Momentos centrales o respecto a la media:

$$\mu_k = \int_0^{10} \frac{1}{10} (x-5)^k dx = \left[\frac{(x-5)^{k+1}}{10(k+1)} \right]_0^{10}$$

en particular, $\mu_3 = 0$ y $\mu_4 = 125$.

Ejemplo 13

- Al tratarse de una variable continua, el cuartil Q_p es el valor de la variable que verifica que $F(Q_p) = p$, siendo F la función de distribución de la variable:

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \dfrac{x}{10} & \text{si } 0 \leq x \leq 10 \\ 1 & \text{si } x > 10 \end{cases}$$

- La mediana es el valor M_e que verifica que $F(M_e) = 1/2$:

$$F(M_e) = \frac{1}{2} \Rightarrow \frac{M_e}{10} = \frac{1}{2} \Rightarrow M_e = 5$$

Ejemplo 13

- Los cuantiles de orden p serán aquellos valores Q_p tales que

$$F(Q_p) = p \Rightarrow \frac{Q_p}{10} = p \Rightarrow Q_p = 10p$$

- Recorrido intercuartílico: $RI = Q_{\frac{3}{4}} - Q_{\frac{1}{4}} = 5$
- Como $\gamma_1 = 0$ y $\gamma_2 = -1'2$, la distribución es simétrica y platicúrtica (más aplastada que la normal).

Part VI

Distribuciones notables

Distribuciones unidimensionales discretas

El proceso de Bernoulli

- En cada observación se clasifica el elemento de la población en una de las dos posibles categorías, correspondientes a la ocurrencia o no de un suceso. Llamaremos éxito (E) a la ocurrencia y fracaso (F) a la no ocurrencia.
- La proporción de éxitos en la población es constante y no depende del número de elementos de ésta. Representaremos por p $(0 < p < 1)$ a la probabilidad de éxito y por q $(q = 1 - p)$ a la probabilidad de fracaso.
- Las observaciones son independientes. Es decir, la probabilidad de éxito es siempre la misma y no se modifica dependiendo de los elementos observados.

Distribución de Bernoulli

- La variable con distribución de Bernoulli se define como:

$$X = \begin{cases} 0 & \text{si ocurre F} \\ 1 & \text{si ocurre E} \end{cases}$$

- Función de probabilidad:

$$P(X=1) = p \text{ y } P(X=0) = q$$

- Características

$$E(X) = p, \qquad Var(X) = pq$$

Distribución binomial

- En un proceso de Bernoulli la variable aleatoria

  $X =$ "número de éxitos en n realizaciones del experimento"

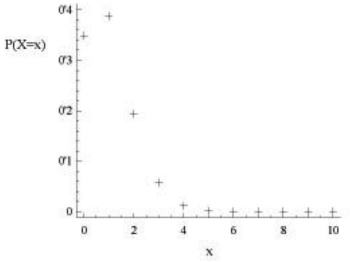
  sigue una distribución binomial de parámetros n y p. La denotaremos $X \in B(n, p)$.
- Función de probabilidad:

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad \text{si } x = 0, 1, 2, \ldots, n$$

- Características:

$$E(X) = np, \qquad Var(X) = npq$$

Función de probabilidad de una variable $B(10, 0'1)$



Distribuciones unidimensionales discretas

Distribución Binomial. Comentarios

1. La distribución de Bernoulli es la distribución binomial con $n = 1$.
2. Los valores de esta distribución están tabulados.
3. La distribución binomial se utiliza en los procesos de control de calidad y en el muestreo con reemplazamiento.
4. El número de fracasos en n pruebas tiene distribución binomial de parámetros n y q.
5. Reproductividad de la distribución binomial: Si $X \in B(n_1, p)$ e $Y \in B(n_2, p)$, entonces $X + Y \in B(n_1 + n_2, p)$.
6. Si $X \in B(n, p)$ entonces $X = \sum_{i=1}^{n} X_i$, donde $X_i \in B(1, p)$ independientes.

Ejemplo

Todos los días se seleccionan, de manera aleatoria, 15 unidades de un proceso de manufactura con el propósito de verificar el porcentaje de unidades defectuosas en la producción. Con base en información pasada, la probabilidad de tener una unidad defectuosa es $p = 0'05$. La gerencia ha decidido detener la producción cada vez que una muestra de 15 unidades tenga dos o más defectuosas. ¿Cuál es la probabilidad de que, un día determinado, el proceso se detenga, supuesto que funciona correctamente? ¿Cuál es la probabilidad de que no se detenga si $p = 0'1$?

Distribuciones unidimensionales discretas

El proceso de Poisson

- Llamaremos proceso de Poisson a aquel experimento aleatorio que consiste en observar la aparición de sucesos puntuales sobre un soporte continuo (generalmente el tiempo), de manera que:
  - El proceso sea estable. Es decir, a largo plazo el número medio de sucesos por unidad de medida, λ ($\lambda > 0$), es constante.
  - Los sucesos ocurren aleatoriamente de forma independiente.
- Es la generalización a un soporte continuo del proceso de Bernoulli.
- Ejemplos: presencia de glóbulos rojos en una gota de sangre, usuarios de Internet que acceden a un servidor, espectadores que llegan a la cola de un cine, etc.

Distribución de Poisson

- En un proceso de Poisson, la variable

$$X = \text{“número de sucesos ocurridos en un intervalo”}$$

es una variable aleatoria con distribución de Poisson de parámetro λ. La denotaremos por $X \in P(\lambda)$.

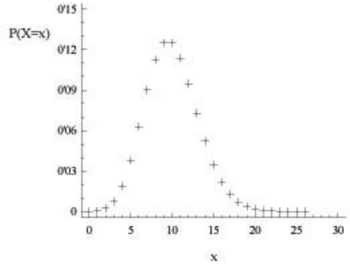
- Función de probabilidad:

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!} \qquad \text{si } x = 0, 1, 2, \ldots$$

- Características:

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

Función de probabilidad de una variable $P(10)$



Distribuciones unidimensionales discretas

Distribución de Poisson. Comentarios

- En la práctica se va a tomar como intervalo (de definición) el conjunto donde se observe la variable. Tendremos que corregir el valor del parámetro λ si variamos el intervalo inicial.
- Reproductividad de la distribución de Poisson: Sean $X \in P(\lambda_1)$ e $Y \in P(\lambda_2)$ dos variables aleatorias independientes, entonces $X + Y \in P(\lambda_1 + \lambda_2)$.
- La distribución de Poisson está tabulada.
- La distribución de Poisson se obtiene como límite de la distribución binomial cuando $n \to \infty$ y $p \to 0$.

Distribuciones unidimensionales discretas

Ejemplo

En un comercio se registra una media de 25 clientes cada hora. ¿Cuál es la probabilidad de que no haya ningún cliente en los próximos 12 minutos? ¿Y de que haya más de 10?

Supuesto que la variable $X =$ "número de clientes que hay en el comercio en 12 minutos" tiene distribución de Poisson de parámetro $\lambda = 5$ (el intervalo considerado es la quinta parte de la hora).

$$P(X = 0) = \frac{e^{-5}\ 5^0}{0!} = 0'0067$$

$$P(X > 10) = 1 - P(X \leq 10) = 1 - 0'983 = 0'0137$$

Distribución uniforme discreta

- Una variable aleatoria X se dice que tiene una distribución uniforme discreta sobre n valores $\{x_1, x_2, \ldots, x_n\}$ si todos ocurren con la misma probabilidad.
- Función de probabilidad:

$$P(X = x) = \frac{1}{n} \qquad \text{si } x = x_1, x_2, \ldots, x_n$$

- Características:

$$\mu = E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad Var(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$$

Ejemplo

Calcular la probabilidad de que el resultado obtenido al lanzar un dado sea mayor o igual que tres. La variable que a cada resultado del lanzamiento le asigna la puntuación obtenida tiene una distribución uniforme discreta sobre el conjunto $\{1, 2, \ldots, 6\}$:

$$P(X = i) = \frac{1}{6}, \quad i = 1, 2, \ldots, 6$$

$$P(X \geq 3) = P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = \frac{4}{6}$$

Distribución uniforme

- X tiene distribución uniforme en (a, b), $X \in U(a, b)$, si su función de densidad es:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in (a, b) \\ 0 & \text{en otro caso} \end{cases}$$

- Función de distribución:

$$F(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

Distribución uniforme

- Características:

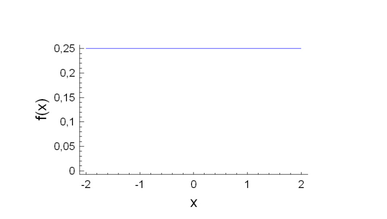
$$E(X) = \frac{a+b}{2}$$

$$Var(X) = \frac{(b-a)^2}{12}$$

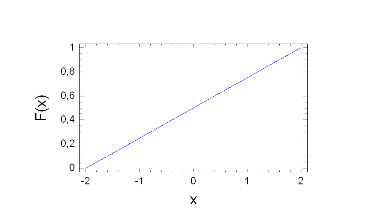
- El nombre de esta distribución se deriva del hecho de que la probabilidad se reparte de modo uniforme sobre el intervalo (a, b) (densidad constante en dicho intervalo), siendo el modelo de distribución uniforme el utilizado para describir el resultado del experimento consistente en elegir un valor al azar en un intervalo.

Distribuciones unidimensionales continuas

Función de densidad de una $U(-2,2)$



Función de distribución de una $U(-2,2)$



Ejemplo

Un estudiante acude todos los días a clase andando. Suponiendo que tarda en llegar 15 minutos, que sale de casa, en un instante aleatorio, entre las ocho cuarenta y las ocho cincuenta y que la clase comienza a las nueve, calcular la probabilidad de que un día determinado llegue tarde.
Sea X= "hora a la que sale de casa", X sigue una distribución $U(40, 50)$.
Llega tarde cuando la hora a la que sale de casa está en el intervalo (45,50]:

$$P(45 < X < 50) = \int_{45}^{50} \frac{1}{50-40}\, dx = \frac{50-45}{50-40} = \frac{1}{2}$$

Distribuciones unidimensionales continuas

Distribución normal

- La distribución normal fue considerada por primera vez por De Moivre en 1733 como límite de la distribución binomial, $B(n, p)$, cuando $n \to \infty$.
- Este descubrimiento quedó en el olvido y fue redescubierta en el siglo XIX por Gauss y Laplace al estudiar la distribución de los errores accidentales en astronomía y geodesia.
- La distribución normal o gaussiana es la más importante y de mayor uso de todas las distribuciones continuas de probabilidad. Existen multitud de experimentos cuyo resultado se ajusta a esta distribución: el peso de un individuo y su talla, datos meteorológicos, errores de medición, calificaciones de pruebas de aptitud, etc.

Distribución normal

- Diremos que una variable aleatoria tiene distribución normal de parámetros μ, σ $(-\infty < \mu < \infty, \sigma > 0)$, que denotaremos por $X \in N(\mu, \sigma)$, si su función de densidad es:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{si } -\infty < x < \infty$$

- Características:

$$E(X) = \mu$$

$$Var(X) = \sigma^2$$

Distribución normal. Comentarios

- La variable aleatoria $N(0,1)$ se denomina normal estándar, tipificada o estandarizada; la representaremos usualmente por la letra Z y a su función de distribución, que está tabulada, por Φ ($\Phi(z) = P(Z \leq z)$). Su función de densidad viene dada por:

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

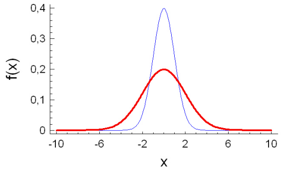
- Para calcular las probabilidades relativas a una variable $X \in N(\mu, \sigma)$ se tipifica la variable. La nueva variable $Z = \dfrac{X - \mu}{\sigma}$ tendrá distribución $N(0,1)$.

Distribución normal. Comentarios

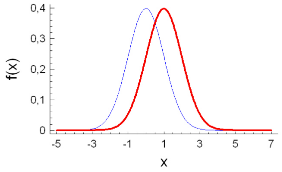
- Los valores del coeficiente de asimetría y la curtosis de una variable normal son 0.
- Sean $X \in N(\mu_1, \sigma_1)$ e $Y \in N(\mu_2, \sigma_2)$. Si dichas variables son independientes $X + Y \in N\left(\mu_1 + \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right)$.
- Dadas variables normales independientes $X_i \in N(\mu_i, \sigma_i)$ y constantes cualesquiera a_i $(i = 1, 2, \ldots, n)$, la variable aleatoria $\sum_{i=1}^n a_i X_i$ tiene distribución $N\left(\sum_{i=1}^n a_i \mu_i, \sqrt{\sum_{i=1}^n a_i^2 \sigma_i^2}\right)$.

Distribuciones unidimensionales continuas

Trazo fino densidad de $N(0,1)$ y trazo grueso densidad de $N(0,2)$



Trazo fino densidad de $N(0,1)$ y trazo grueso densidad de $N(1,1)$



Distribuciones unidimensionales continuas

Ejemplo

Una universidad espera recibir, para el siguiente año escolar, 16.000 solicitudes de ingreso para el primer año de licenciatura. Se supone que las calificaciones obtenidas por los aspirantes en la prueba selectiva siguen una distribución N(950,100). Si la universidad decide admitir al 25% de todos los aspirantes que obtengan las calificaciones más altas en la prueba selectiva, ¿cuál es la mínima calificación que es necesario obtener en esa prueba para ser admitido por la universidad?

Ejemplo

Sea $X =$ "calificación obtenida", $X \in N(950, 100)$. Nos interesa calcular a tal que:

$$P(X > a) = P\left(Z > \frac{a - 950}{100}\right) = 0'25$$

con $Z \in N(0, 1)$.

En las tablas puede comprobarse que el valor de Z que deja a su derecha probabilidad $0'25$ es aproximadamente $0'675$, obteniéndose por tanto:

$$\frac{a - 950}{100} = 0'675 \Rightarrow a = 950 + 67'5 = 1.017'5$$

Ejemplo

Sea $X \in N(\mu, \sigma)$, ¿cuáles son las probabilidades de que el valor de X se encuentre a una, dos y tres veces la desviación típica de la media?

Usando las tablas de la distribución normal estándar, las probabilidades pedidas son:

$$P(\mu - \sigma \leq X \leq \mu + \sigma) = P(-1 \leq Z \leq 1) = 0'6826$$

$$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) = P(-2 \leq Z \leq 2) = 0'9544$$

$$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) = P(-3 \leq Z \leq 3) = 0'9974$$

Distribución exponencial

- En un proceso de Poisson, consideremos la variable $X =$ "tiempo entre dos sucesos consecutivos". Diremos que X tiene distribución exponencial de parámetro λ, que denotaremos por $X \in Exp(\lambda)$ siendo $\lambda > 0$ el número medio de sucesos que ocurren por unidad de tiempo.
- Función de densidad:

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

Distribución exponencial

- Función de distribución:

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

- Características:

$$E(X) = \frac{1}{\lambda}$$

$$Var(X) = \frac{1}{\lambda^2}$$

Distribución exponencial. Comentarios

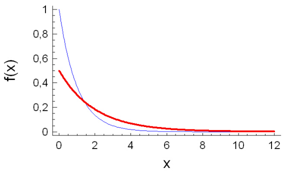
- La distribución exponencial "carece de memoria", es decir,

$$P(X > x + t / X > x) = P(X > t)$$

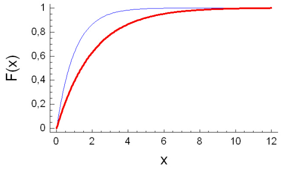
- La distribución exponencial es la generalización al caso continuo de la distribución geométrica.
- La distribución exponencial aparece, en ocasiones, caracterizada utilizando como parámetro la media, $\mu = 1/\lambda$.

Distribuciones unidimensionales continuas

Trazo fino densidad de $Exp(1)$ y trazo grueso densidad de $Exp(2)$



Trazo fino distribución de $Exp(1)$ y trazo grueso distribución de $Exp(2)$



Ejemplo

Un Departamento de la Universidad decide adquirir una estación de trabajo. Se sabe que los avances tecnológicos se producen de forma aleatoria de tal manera que, por término medio, surge un nuevo modelo que deja anticuados a los ya existentes cada 7 meses.

¿Cuál es la probabilidad de que la nueva máquina no se quede anticuada durante un período de tiempo comprendido entre 6 meses y 1 año?

Calcular la probabilidad de que la máquina se quede anticuada después de un año, sabiendo que no se ha quedado anticuada durante los primeros 7 meses.

Ejemplo

Supuesto que la variable $X =$ "tiempo, en meses, hasta que la máquina queda anticuada" sigue una distribución Exp$(1/7)$

1. $$\begin{aligned} 1 - P(6 < X < 12) &= 1 - \int_6^{12} \frac{1}{7} e^{-x/7} dx \\ &= 1 - e^{-6/7} + e^{-12/7} = 0'7557 \end{aligned}$$

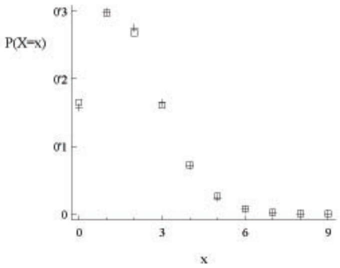
2. $$\begin{aligned} P(X > 12 \,/ X > 7) &= \frac{P(X > 12)}{P(X > 7)} = e^{-(12-7)/7} \\ &= e^{-5/7} = P(X > 5) \end{aligned}$$

Relaciones entre distribuciones

Relación entre las distribuciones hipergeométrica, binomial y de Poisson

- Si $X \in H(N,n,k)$ y el tamaño, n, de la muestra es muy pequeño comparado con el tamaño de la población, N, podremos aproximar la distribución de X por la distribución $B(n,p)$.
  En la práctica, convendremos en utilizar esta aproximación si $\frac{n}{N} < 0'1$.
- Sea $X \in B(n,p)$. Si $np = \lambda$ es constante, $n \to \infty$ y $p \to 0$, entonces la distribución de X se puede aproximar por la distribución $P(\lambda)$.
  En la práctica, utilizaremos esta aproximación si $p < 0'1$ y $n > 30$.

Comparación de las funciones de probabilidad de una variable $B(36, 0'05)$ $(+)$ y su aproximación por la variable $P(1'8)$ $(\square)$



Relaciones entre distribuciones

El teorema central del límite

Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes e idénticamente distribuidas con media μ y desviación típica σ. Para n grande, $S_n = X_1 + X_2 + \cdots + X_n$ tiene distribución aproximadamente $N\left(n\mu, \sigma\sqrt{n}\right)$.

Equivalentemente, para n grande, la distribución de $\dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ se puede aproximar por la distribución $N(0,1)$.

Comentario

Este teorema establece que cuando los resultados de un experimento son debidos a un conjunto muy grande de causas independientes que actúan sumando sus efectos, dichos resultados siguen una distribución normal.

Ejemplo

¿Qué probabilidad de ganar tiene un individuo que apuesta a que el número de caras en cien tiradas de una moneda no sesgada difiere de cincuenta en menos de cuatro?

Ejemplo

Sea X_i la variable que toma el valor 1 si se obtiene cara en el lanzamiento i y 0 si se obtiene una cruz en dicho lanzamiento y sea $X = \sum_{i=1}^{100} X_i$ la variable "número de caras en los 100 lanzamientos de la moneda".

Las variables $X_1, X_2, \ldots, X_{100}$ son variables independientes con distribución de Bernoulli de parámetro $p = 1/2$; todas con $\mu = 1/2$ y $\sigma = 1/2$. Aplicando el Teorema Central del Límite, con $n = 100$, la distribución de $Z = \dfrac{X - 50}{5}$ se puede aproximar por la distribución $N(0,1)$.

Utilizando la aproximación anterior, la probabilidad pedida es

$$P\left(|X - 50| < 4\right) \simeq P(|Z| < 4/5) = 0'5762$$

Relación entre la distribución normal y las distribuciones binomial, de Poisson y gamma

- Si $X \in B(n, p)$ entonces $X = \sum_{i=1}^{n} X_i$, donde cada $X_i \in B(1, p)$. Por el Teorema Central del Límite X se puede aproximar por una $N(np, \sqrt{npq})$. Esta aproximación es válida si $n > 30$ y $0'1 < p < 0'9$.
- Por el Teorema Central del Límite (y la reproductividad) la Poisson de parámetro λ se puede aproximar mediante la distribución $N(\lambda, \sqrt{\lambda})$, para valores grandes de λ ($\lambda > 5$).
- La propiedad de reproductividad de la distribución gamma, conjuntamente con el Teorema Central del Límite, permiten aproximar, para p grande ($p > 100$), la distribución $\Gamma(\lambda, p)$ mediante la distribución $N(p/\lambda, \sqrt{p}/\lambda)$.

Relaciones entre distribuciones

Comentario

Con el fin de mejorar la calidad de las aproximaciones obtenidas al aproximar una distribución discreta mediante una distribución continua, se recomienda utilizar lo que se conoce como corrección por continuidad.

Así, si se trata de aproximar una distribución discreta, que toma sólo valores enteros, por una continua, cometeremos un menor error en la aproximación si al calcular la $P(a \leq X \leq b)$, con X distribución discreta y a, b dos valores enteros de la variable, la aproximamos por la $P(a - 0'5 \leq X \leq b + 0'5)$, con X distribución continua.

Estadística Aplicada
Ricardo Cao
Principales distribuciones unidimensionales discretas
Principales distribuciones unidimensionales continuas
Relaciones entre distribuciones
Distribuciones multidimensionales
Distribuciones asociadas a la normal

Distribución χ^2 de Pearson

- Sean $Z_1, Z_2, \ldots, Z_n$ variables aleatorias independientes con distribución normal estándar. La variable aleatoria $\chi_n^2 = Z_1^2 + Z_2^2 + \cdots + Z_n^2$ tiene distribución χ^2 de Pearson con n grados de libertad.
- Función de densidad:

$$f(x) = \begin{cases} \dfrac{\left(\frac{1}{2}\right)^{\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}}{\Gamma\left(\frac{n}{2}\right)} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

- Características:

$$E(\chi_n^2) = n, \quad Var(\chi_n^2) = 2n$$

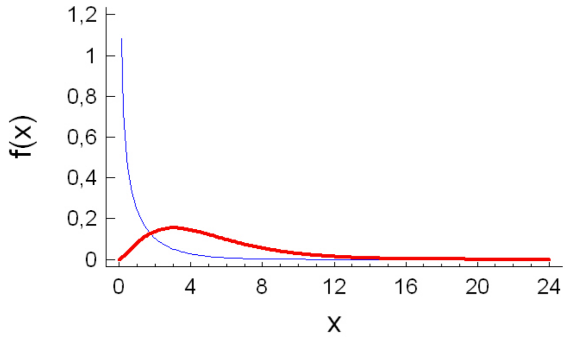
Distribuciones asociadas a la normal

Distribución χ^2 de Pearson. Comentarios

- La distribución χ^2 está tabulada.
- La distribución χ_n^2 coincide con la distribución $\Gamma(1/2, n/2)$.
- La $\sqrt{2\chi_n^2}$ se aproxima a la $N(\sqrt{2n-1}, 1)$, cuando n es grande ($n > 100$).
- Reproductividad de la distribución χ^2: Si $X \in \chi_n^2$ e $Y \in \chi_m^2$ son independientes, $X + Y \in \chi_{n+m}^2$.

Distribuciones asociadas a la normal

Función de densidad de la distribución χ^2. Trazo fino χ^2_1, trazo grueso χ^2_5



Distribución **t** de Student

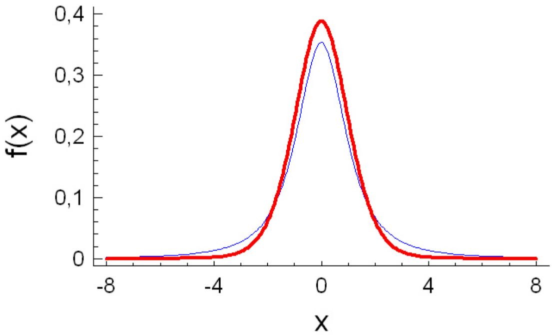
- Sean $Z \in N(0,1)$ e $Y \in \chi_n^2$ variables aleatorias independientes. La distribución t de Student con n grados de libertad es la distribución de la variable $t_n = \dfrac{Z}{\sqrt{Y/n}}$.
- Función de densidad:

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}$$

- Características:

$$E(t_n) = 0 \text{ si } n > 1, \quad Var(t_n) = \frac{n}{n-2} \text{ si } n > 2$$

Función de densidad de la distribución t. Trazo fino t_2, trazo grueso t_{10}.



Distribución **t** de Student. Comentarios

- Para n grande ($n > 30$), la distribución t de Student se aproxima por la distribución normal estándar.
- La distribución t de Student presenta mayor dispersión que la $N(0,1)$ y menor curtosis.
- La distribución de t_1 se conoce con el nombre de distribución de Cauchy y presenta la característica de no tener media.
- La distribución de t_n está tabulada.

Distribución **F** de Fisher-Snedecor

- Sea $X \in \chi_m^2$ e $Y \in \chi_n^2$, independientes. La distribución *F* de Fisher-Snedecor con m y n grados de libertad es la distribución de la variable $F_{m,n} = \dfrac{X/m}{Y/n}$.
- Función de densidad:

$$f(x) = \begin{cases} \dfrac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} \dfrac{n^{\frac{n}{2}}(mx)^{\frac{m}{2}}}{x(mx+n)^{\frac{m+n}{2}}} & \text{si } x > 0 \\ 0 & \text{en otro caso} \end{cases}$$

Distribuciones asociadas a la normal

Distribución **F** de Fisher-Snedecor

- Características:

$$E(F_{m,n}) = \frac{n}{n-2} \quad \text{si } n > 2$$

$$Var(F_{m,n}) = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)} \quad \text{si } n > 4$$

- La distribución de $F_{m,n}$ está tabulada.
- Si $X \in F_{m,n}$ entonces $\frac{1}{X} \in F_{n,m}$.

Part VII

Estimación puntual

Introducción a la inferencia estadística

Introducción

- La inferencia estadística puede definirse como el conjunto de métodos mediante los cuales podemos extraer información sobre distintas características de interés de cierta distribución de probabilidad de la cual se ha observado una serie de datos.
- La inferencia estadística comprende las fases de recogida y depuración de los datos, estimación, contrastes de simplificación, diagnosis y validación del modelo.
- Según el objeto de estudio, la inferencia estadística se clasifica en inferencia paramétrica y no paramétrica.

Introducción a la inferencia estadística

Introducción

- La inferencia paramétrica se ocupa de aquellos casos en los que la distribución de probabilidad de la población objeto de estudio se supone conocida salvo los valores que toman ciertos coeficientes, llamados parámetros. El objetivo es estimar, dar intervalos de confianza o contrastar hipótesis sobre dichos parámetros
- La inferencia no paramétrica trata problemas similares cuando se tiene una distribución poblacional totalmente desconocida, sobre la cual tan sólo se realizan suposiciones muy generales (como, por ejemplo, que es una distribución continua, que tiene una única moda, etc.).

Introducción

- El enfoque clásico trata los parámetros poblacionales desconocidos como valores fijos o constantes.
- El enfoque bayesiano considera que los parámetros desconocidos del modelo son variables aleatorias, para las cuales debe fijarse una distribución inicial, llamada distribución a priori. Utilizando la información muestral junto con la mencionada distribución a priori, los métodos bayesianos hacen uso de la regla de Bayes para ofrecer una distribución a posteriori sobre los parámetros.

Conceptos generales

- Se denomina población a un conjunto homogéneo de individuos sobre los que se estudian una o varias características que son, de alguna forma, observables.
- Una muestra es un subconjunto de la población. El tamaño muestral es el número de elementos de la muestra.
- Un método de muestreo no es más que el procedimiento empleado para la obtención de la muestra.
- Un parámetro es cualquier característica medible (normalmente numérica) de la población.

Muestreo en poblaciones finitas

- Muestreo aleatorio simple: Cada muestra posible tiene la misma probabilidad de ser elegida.
- Muestreo sistemático: Se usa muy frecuentemente cuando los individuos de la población están ordenados en listas.
- Muestreo estratificado: Si somos capaces de dividir la población en estratos, se actua dentro de cada estrato según un muestreo aleatorio simple.
- Muestreo por conglomerados: Si la población puede dividirse en conglomerados, que son homogéneos entre sí, se eligen aleatoriamente algunos conglomerados y, en cada uno de ellos, una muestra representativa.
- Muestreo polietápico: Utilizar distintos tipos de muestreo en sucesivas etapas.

Estimación puntual

Introducción

- Matemáticamente, podemos pensar que estamos interesados en el estudio de una variable aleatoria X, cuya distribución, F, es en mayor o menor grado desconocida.
- Supondremos que la familia $\mathcal{F}$ de distribuciones a la que pertence F es de la forma: $\mathcal{F} = \{F_\theta \mid \theta \in \Theta\}$, siendo $\Theta \subset \mathbb{R}^k$.
- Para tratar de disminuir el desconocimiento de la distribución teórica F de la variable aleatoria X en estudio, tomamos una muestra.

Estimación puntual

Muestra aleatoria simple

- Una muestra aleatoria simple, de tamaño n, de una variable aleatoria X, con distribución teórica F, son n variables aleatorias, $X_1, X_2, \ldots, X_n$, independientes e igualmente distribuidas, con distribución común F.
- Una realización de la muestra son los valores particulares, $x_1, x_2, \ldots, x_n$, observados para las variables $X_1, X_2, \ldots, X_n$.
- Como consecuencia, la función de distribución conjunta de $X_1, X_2, \ldots, X_n$ es:

$$F(x_1, x_2, \ldots, x_n) = F(x_1) \cdot F(x_2) \cdots F(x_n)$$

Estimación puntual

Otros conceptos

- Llamaremos espacio muestral al conjunto de muestras posibles, de tamaño determinado, que pueden obtenerse al seleccionar una muestra aleatoria de una población.
- Llamaremos estadístico a cualquier función T de la muestra.
- Un estadístico $T(X_1, X_2, \ldots, X_n)$, como función de variables aleatorias, es también una variable aleatoria. Tiene, por tanto, una distribución de probabilidad, llamada distribución en el muestreo de T.
- Los estadísticos independientes del parámetro (que son función únicamente de la muestra) se denominan estimadores ($\widehat{\theta} = T(X_1, X_2, \ldots, X_n)$).

Estimación puntual

Ejemplos de estadísticos

- $T(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i.$
- $T(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$
- $T(X_1, X_2, \ldots, X_n) = \frac{1}{n-1}\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2.$
- $T(X_1, X_2, \ldots, X_n) = \min(X_1, X_2, \ldots, X_n).$
- $T(X_1, X_2, \ldots, X_n) = \max(X_1, X_2, \ldots, X_n).$

Estimación puntual

Ejemplo

Supongamos que pretendemos "estimar" la altura media de los alumnos de la Facultad de Informática de La Coruña. Para ello seleccionamos una muestra de 20 alumnos, los medimos y tomamos la media muestral como "estimador" de la media desconocida (más adelante veremos que la elección de la media muestral es razonable).

Si este trabajo lo realizan 4 personas diferentes, los resultados obtenidos podrían ser los siguientes (en centímetros):

Persona	Muestra	Media muestral
1	170, 175, 166, . . ., 180	176
2	166, 172, 180, . . ., 177	174
3	190, 165, 185, . . .,175	178
4	175, 156, 178, . . ., 177	175

Estimación puntual

Ejemplo

Cada una de las muestras obtenidas es una realización de un vector aleatorio (en este caso de dimensión 20). Al estudiar los problemas de inferencia, consideraremos la muestra como el vector aleatorio en sí (con cierta distribución de probabilidad que dependerá de la distribución de la población y del tamaño de la muestra).

En cuanto a la media muestral, al ser una función de variables aleatorias, también es una variable aleatoria. En la tabla anterior disponemos de cuatro valores de dicha variable. Si dispusiésemos de un gran número de ellos, podríamos aproximar con precisón su distribución en el muestreo (que dependerá de nuevo de la distribución de la población y del tamaño de la muestra).

Propiedades deseables de los estimadores

Insesgadez

- Se denomina sesgo del estimador $\widehat{\theta}$ como estimador de θ a

$$Sesgo\left(\widehat{\theta}\right) = E\left(\widehat{\theta}\right) - \theta$$

- Si $E\left(\widehat{\theta}\right) = \theta$, para cada $\theta \in \Theta$, el estimador $\widehat{\theta}$ se dice centrado o insesgado para θ.
- Pueden existir muchos estimadores centrados para un parámetro. Por ejemplo, para estimar la media μ de una distribución cualquiera, todos los estimadores del tipo $\widehat{\mu} = a_1X_1 + a_2X_2 + \cdots + a_nX_n$ con $\sum_{i=1}^n a_i = 1$ son centrados.

Ejemplo

Obtener un estimador centrado para el parámetro p de una distribución binomial.

Sea $X \in B(m,p)$ y sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de X. Un estimador insesgado para p es

$$\widehat{p} = \frac{\overline{X}}{m} = \frac{X_1 + X_2 + \cdots + X_n}{nm}$$

ya que su media coincide con p:

$$E\left(\widehat{p}\right) = \frac{1}{m} E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \frac{1}{m} \frac{(mp)n}{n} = p$$

Insesgadez asintótica

Un estimador, $\widehat{\theta}_n$, se dice asintóticamente insesgado para θ si:

$$\lim_{n \to \infty} E\left(\widehat{\theta}_n\right) = \theta, \text{ para cada } \theta \in \Theta$$

Estadística Aplicada
Ricardo Cao
Introducción a la inferencia estadística
Estimación puntual
Propiedades deseables de los estimadores
Estimación de la media de una población
Estimación de la varianza de una población
Estimación de una proporción
Procedimientos para la

Ejemplo

Sea $X_1, X_2, \ldots, X_n$, una muestra aleatoria simple de una variable aleatoria X con $E(X) = \mu$.

El estimador $\widehat{\mu}_n = \dfrac{2X_1 + X_2 + \cdots + X_n}{n}$ es asintóticamente insesgado para μ:

Teniendo en cuenta las propiedades de la esperanza,

$$\lim_{n\to\infty} E\left(\widehat{\mu}_n\right) = \lim_{n\to\infty} \frac{n+1}{n}\mu = \mu$$

y, por tanto, $\widehat{\mu}_n$ es asintóticamente insesgado para μ.

Propiedades deseables de los estimadores

Error cuadrático medio

- Se define el error cuadrático medio de un estimador $\widehat{\theta}$ de un parámetro θ como el número

$$ECM\left(\widehat{\theta}\right) = E\left[\left(\theta - \widehat{\theta}\right)^2\right]$$

- De la definición anterior, se deduce fácilmente que

$$ECM\left(\widehat{\theta}\right) = \left[Sesgo\left(\widehat{\theta}\right)\right]^2 + Var\left(\widehat{\theta}\right)$$

Ejemplo

Obtener el error cuadrático medio del estimador $\widehat{p}$ de la binomial del ejemplo anterior.

Puesto que ya conocemos que $\widehat{p}$ es insesgado, su error cuadrático medio coincide con su varianza:

$$\begin{aligned} ECM\left(\widehat{p}\right) &= Var\left(\widehat{p}\right) = \frac{1}{m^2} Var\left(\overline{X}\right) = \frac{1}{m^2}\frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) \\ &= \frac{1}{m^2}\frac{1}{n^2} nmpq = \frac{pq}{mn} = \frac{p(1-p)}{mn} \end{aligned}$$

Eficiencia

- Llamaremos precisión o eficiencia de $\widehat{\theta}$, como estimador de θ, al inverso de su error cuadrático medio:

$$efic\left(\widehat{\theta}\right)=\frac{1}{ECM\left(\widehat{\theta}\right)}$$

- Diremos que $\widehat{\theta}_2$ es más eficaz o más preciso que $\widehat{\theta}_1$ si para cualquier tamaño muestral:

$$efic\left(\widehat{\theta}_2\right)\geq efic\left(\widehat{\theta}_1\right)$$

Es decir, si $ECM\left(\widehat{\theta}_2\right)\leq ECM\left(\widehat{\theta}_1\right)$.

Propiedades deseables de los estimadores

Ejemplo

Calcular la eficiencia del estimador insesgado de p de la binomial del ejemplo anterior.

Teniendo en cuenta el ejemplo anterior, la eficiencia es

$$efic\left(\widehat{p}\right) = \frac{1}{Var\left(\widehat{p}\right)} = \frac{mn}{p(1-p)}$$

Obsérvese cómo ganamos eficiencia al aumentar el tamaño muestral (a mayor información, mayor precisión).

Consistencia

- Un estimador $\widehat{\theta}_n$ se dice consistente en media cuadrática para estimar un parámetro θ si

$$\lim_{n\to\infty} ECM\left(\widehat{\theta}_n\right) = 0$$

- En adelante, cuando hablemos de consistencia de un estimador nos referiremos a la consistencia en media cuadrática.
- Una condición necesaria y suficiente para que $\widehat{\theta}_n$ sea consistente es que sea asintóticamente insesgado y que $\lim_{n\to\infty} Var\left(\widehat{\theta}_n\right) = 0$.

Ejemplo

Estudiar la consistencia del estimador insesgado de p de la binomial del ejemplo anterior.
Por tratarse de un estimador insesgado, será consistente siempre que su varianza tienda a cero:

$$\lim_{n\to\infty} Var\left(\widehat{p}\right) = \lim_{n\to\infty} \frac{p(1-p)}{mn} = 0$$

Estimación de la media de una población

Media muestral

- Es un estimador consistente de μ. Sólo hay que probar que su varianza tiende a cero.

$$Var\left(\overline{X}\right) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var\left(X_i\right) = \frac{\sigma^2}{n}$$

- La distribución exacta de $\overline{X}$ depende de la distribución de la población. Si $X \in N(\mu, \sigma)$, entonces $\overline{X} \in N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$. Por el TCL, para n grande ($n \geq 30$), la distribución de $\overline{X}$ puede aproximarse por $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

Estimación de la varianza de una población

Varianza muestral

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria simple de una variable aleatoria X con $E(X) = \mu$ y $Var(X) = \sigma^2$. Se define la varianza muestral como:

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

Estadística Aplicada
Ricardo Cao
Introducción a la inferencia estadística
Estimación puntual
Propiedades deseables de los estimadores
Estimación de la media de una población
Estimación de la varianza de una población
Estimación de una proporción
Procedimientos para la

Varianza muestral

Es un estimador asintóticamente insesgado de σ^2. Teniendo en cuenta que

$$\sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2 + n \left(\mu - \overline{X}\right)^2$$

obtenemos que

$$E\left(S^2\right) = \sigma^2 - \frac{\sigma^2}{n} = \sigma^2 \frac{n-1}{n}$$

y, por tanto,

$$\lim_{n\to\infty} E\left(S^2\right) = \sigma^2$$

Estimación de la varianza de una población

Varianza muestral

- La varianza de S^2 puede calcularse también, pero es bastante complicada (depende de los coeficientes de asimetría y de curtosis de la población).
- Es un estimador consistente de σ^2.
- La distribución de S^2 es, en general, asimétrica y su forma depende del tamaño muestral y de la población base. Usando el TCL, puede aproximarse la distribución de S^2 por la distribución normal, pero la aproximación es muy lenta y sólo se manifiesta en tamaños muestrales muy grandes.

Varianza muestral

- Si $X \in N(\mu, \sigma)$, se puede probar que $\frac{nS^2}{\sigma^2} \in \chi^2_{n-1}$.
- Dado que $E\left(\chi^2_n\right) = n$, se tiene que $E\left(\frac{nS^2}{\sigma^2}\right) = n - 1$ y, por tanto,

$$E\left(S^2\right) = \frac{n-1}{n}\sigma^2$$

- Como $Var\left(\chi^2_n\right) = 2n$, se tiene que $Var\left(\frac{nS^2}{\sigma^2}\right) = 2(n-1)$ y, por tanto,

$$Var\left(S^2\right) = \frac{2(n-1)}{n^2}\sigma^4$$

Estimación de la varianza de una población

Varianza muestral corregida o cuasivarianza

- Con el objeto de corregir el sesgo de la varianza muestral, se define la varianza muestral corregida o cuasivarianza como

$$\widehat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{n}{n-1}S^2$$

- Es un estimador insesgado de σ^2.
- Es un estimador consistente de σ^2.
- Si $X \in N(\mu, \sigma)$, las variables aleatorias $\overline{X}$ y $\widehat{S}^2$ son independientes y $\frac{(n-1)\widehat{S}^2}{\sigma^2} \in \chi^2_{n-1}$. Entonces $E\left(\widehat{S}^2\right) = \sigma^2$ y $Var\left(\widehat{S}^2\right) = \frac{2}{n-1}\sigma^4$.

Comentarios

- Si se conoce la media poblacional, μ, utilizaremos el estimador que resulta de reemplazar $\overline{X}$ por μ en la expresión de S^2. El estimador resultante es insesgado y tiene menor varianza que S^2 y $\widehat{S}^2$.
- En los programas informáticos de estadística suele utilizarse la cuasivarianza muestral, debido a su carácter insesgado.

Estimación de una proporción

Proporción muestral

- Se desea estimar la proporción, p, de individuos de una población que poseen determinada característica. Tomamos una m.a.s. de n elementos anotando un 1 si el individuo elegido posee la característica y un 0 si carece de ella. Un estimador razonable de p es:

$$\widehat{p} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

- $\widehat{p}$ es la media muestral de n variables i.i.d. $B(p)$.
- $E\left(\widehat{p}\right) = p, \; Var\left(\widehat{p}\right) = \frac{p(1-p)}{n}$ y su distribución en el muestreo, para $(n > 30)$ es aprox. $N\left(p, \sqrt{(p(1-p))/n}\right)$.

Estimación de una proporción

Ejemplo. Método de captura-recaptura

Estimar el número, N, de peces que hay en un embalse
Se capturan n de ellos, se marcan con una pequeña placa y se devuelven al embalse. A continuación se recapturan r peces, contando cuántos de ellos están marcados (supongamos que x lo están).
Utilizando la proporción muestral (x/r) como estimador de la proporción poblacional (n/N) de peces marcados, podemos estimar N por rn/x, sin más que igualar ambas proporciones y despejar N.