

Introducción al Diseño de Experimentos para el Reconocimiento de Patrones

Capítulo 2: Modelos Estadísticos

Curso de doctorado impartido por
Dr. Quiliano Isaac Moro
Dra. Aranzazu Simón Hurtado
Marzo 2004

Contenido

1. Introducción
2. Estadística Descriptiva.
3. Nociones de Probabilidad.
4. Distribución de las Características Muestrales.
5. Inferencia Estadística
 1. Ejemplo de Clasificadores Estadísticos.
6. Procesos Estocásticos

2

Introducción

- **Estadística:**
ciencia cuyo objetivo es la obtención y el análisis de datos mediante el uso de medios matemáticos y herramientas informáticas.
 - El interés es el uso de DATOS, no de variables aleatorias, ni probabilidades.
- Estadística Descriptiva:
 - Generación y recopilación de datos que contengan información relevante sobre un determinado problema.
- Inferencia Estadística:
 - Análisis de esos datos con el fin de extraer dicha información.

3

Introducción. Definiciones

- **Población:**
conjunto de todos los individuos o entes que constituyen el objeto de un determinado estudio y sobre los que se desea obtener ciertas conclusiones.
- **Experimento aleatorio:**
Los individuos pueden ser generados mediante un proceso que en sucesivas realizaciones puede producir distintos individuos.
- **En toda población real existe VARIABILIDAD.**
- **Característica aleatoria:**
 - cualquier característica que puede constatarse en cada individuo de una población.
 - Si se trata de un dato numérico, se llama Variable Aleatoria.
 - Valores no numéricos → se pueden codificar numéricamente.
 - Discreta / continua.
 - Característica aleatoria K-dimensional.
 - Cuando sobre cada individuo se estudian K características diferentes.

4

Estadística Descriptiva

- **Muestras. Datos Estadísticos**
Generalmente no se puede estudiar TODA la población
 - **Muestra:** subconjunto de individuos de la población.
 - Para que los resultados sean válidos, la muestra ha de ser representativa.
 - **Datos estadísticos:** valores observados de una variable aleatoria sobre una muestra.
- **Objetivo de la Estadística Descriptiva:**
poner de manifiesto las características más relevantes y de los datos y sintetizarlas en unos pocos parámetros o mediante las gráficas adecuadas.

5

Estadística Descriptiva

- **Tablas de frecuencias.**
 - Monodimensionales
 - <valor, nº de veces que aparece>
 - Si es una variable continua se crean intervalos.
 - Bidimensionales → Tablas de contingencia.
 - Frecuencias absolutas / relativas.
 - Frecuencias Marginales.
 - Frecuencias relativas condicionales.

Variable 2 →	u_1	...	u_n	Total filas
Variable 1 ↓				
v_1				
...				
v_m				
Total columnas				Total

6

Estadística Descriptiva

- **Histogramas**
 - Diagramas de Frecuencias acumuladas.
- **Parámetros de posición**
 - Media: $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$
 - Mediana:
 - Percentiles.
 - Cuartiles:
 - Primer cuartil (percentil del 25%)
 - Tercer cuartil (percentil del 75%).
- **Parámetros de dispersión**
 - Recorrido: $V_{\max} - V_{\min}$
 - Varianza: $S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$
 - Desviación típica: $S = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$

7

Estadística Descriptiva

- **Estadística descriptiva bidimensional**
 - Tabla de contingencia
 - Distribuciones marginales y frecuencias relativas condicionales.
 - Diagramas de dispersión (scatterplot)
 - Covarianza

$$\text{COV}_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$
 - Coeficiente de correlación lineal

$$r_{xy} = \frac{\text{COV}_{xy}}{S_x S_y}$$
 - Interpretación de la covarianza como un producto escalar
 - Que exista una relación entre dos vars. (constatada por su coef. de correlación lineal), no quiere decir que haya una relación de causalidad.
 - Recta de regresión. Relaciona dos variables aleatorias de forma lineal.
 - Análisis de residuos

8

Nociones de Probabilidad

- Sea E el conjunto de valores que puede tomar una variable aleatoria. Cualquier subconjunto $A \subseteq E$ se denomina **SUCESO**.
 - Suceso imposible \rightarrow conjunto vacío (\emptyset).
 - Suma de sucesos \rightarrow unión de los subconjuntos.
 - Producto de sucesos \rightarrow intersección de los subconjuntos.
 - Sucesos excluyentes: su intersección (producto) es el suceso imposible.
 - Suceso contrario A_c

9

Nociones de Probabilidad

- **Probabilidad: número real comprendido entre 0 y 1 que se asocia a cada suceso.**
 - Informalmente: proporción de individuos de la población que verifica dicho suceso.
- **Propiedades:**
 - $P(A) \geq 0$ y $P(A) \leq 1$.
 - $P(E) = 1$ y $P(\emptyset) = 0$
 - $P(A_c) = 1 - P(A)$
 - Si A y B son sucesos excluyentes: $P(A+B) = P(A) + P(B)$
 - Si A y B no son excluyentes: $P(A+B) = P(A) + P(B) - P(AB)$

10

Nociones de Probabilidad

- **Probabilidad condicional**

Intuitivamente: Probabilidad condicional del suceso A dado el suceso B, $P(A/B)$, probabilidad de que se haya presentado el suceso A sabiendo que ha ocurrido B.

 - $P(A/B)$ sería el cociente entre el número de individuos que verifican tanto A como B dividido por el número de individuos que verifican B. Es decir: $P(A/B) = P(AB)/P(B)$
 - En el caso de que A y B sean independientes: $P(A/B) = P(A)$
- **Teorema de la probabilidad total**
 - Si A_1, \dots, A_N son sucesos mutuamente excluyentes que particionan E, dado otro suceso B, se tiene $P(B) = \sum P(B/A_i)P(A_i)$
- **Teorema de Bayes**

$$P(A_k / B) = \frac{p(A_k) p(B / A_k)}{\sum_i p(A_i) p(B / A_i)}$$

11

Nociones de Prob. Funciones de Distribución

- **Función de distribución**
 - Para una variable aleatoria X, $F(x) = \text{Prob}(X \leq x)$
 - Distribuciones discretas
 - Distribuciones continuas. Función densidad de probabilidad.
 - Se puede relacionar los conceptos de función densidad de probabilidad y el histograma de frecuencias.
- **Independencia**
- **Distribuciones marginales.**
- **Distribuciones condicionales.**

12

Nociones de Probabilidad. Momentos

Esperanza matemática

- Generaliza la idea de media.

$$E(h(X)) = \sum h(x_i)P(X = x_i) \quad \forall i$$

$$E(h(X)) = \int_{-\infty}^{\infty} h(x_i)P(X = x_i)dx$$

Momentos

- Momento de X respecto al origen de orden n: $E(X^n)$
- Momentos centrales. El origen se toma en la media.

$$\text{Varianza} = \sigma^2 = E[(x - m)^2]$$

- La varianza σ^2 . Propiedades:
 - $\sigma^2(aX) = a^2\sigma^2(X)$
 - Si X e Y son vars. independientes
 - $\sigma^2(X+Y) = \sigma^2(X) + \sigma^2(Y)$

13

Distribuciones de Probabilidad: Ej. distribuciones discretas

Binomial.

- Un suceso con dos resultados {x, y}, de probabilidades {p, 1-p}, que se repite n veces (resultados independientes). El resultado del experimento es contar el número de veces que aparece el valor x.

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

$$E(X) = np$$

$$\sigma^2(X) = np(1-p)$$

Poisson.

- Es una binomial cuando el valor de n es muy elevado, y p muy pequeño, tal que np tiende a un valor constante λ .

$$P(\text{Poisson}(\lambda) = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

$$E(X) = \lambda$$

$$\sigma^2 = \lambda$$

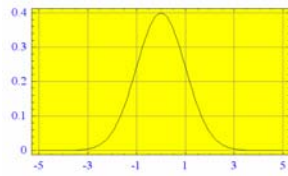
14

Función de Densidad Normal

- Es muy importante.
- Cualquier combinación lineal de variables normales, también es normal.
- Tipificación a $N(0,1)$
- Aproximaciones normales. Teorema central del límite:
 - la suma de variables aleatorias independientes tiende a distribuirse como una Normal a medida que aumenta el número de sumandos.
 - Una binomial con más de 10 experimentos puede ser considerada una Normal.
 - Una de Poisson con $\lambda > 10$.

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

$$x = \frac{a - m}{\sigma}$$



15

Distrib. de las características muestrales

Sea una población a cuyos individuos va asociada una variable aleatoria X. Para **obtener conclusiones** se obtiene una muestra aleatoria de N individuos

Muestra aleatoria simple:

- todos los individuos han tenido la misma probabilidad de ser incluidos en la muestra,
- dichos individuos han sido seleccionados de manera independiente unos de otros.

Consideremos la población de todas las posibles Muestras extraíbles de la población original.

- Ahora a cada muestra (individuo de esta nueva población) se le puede asociar valores estadísticos (media, varianza...)
- Cualquier función de los valores muestrales se denomina **estadístico**.
 - Todo estadístico es una variable aleatoria cuya distribución dependerá en general de la distribución de la población y tamaño de la muestra.

16

Distribución de las características muestrales

Sea X es la población de partida, de media m y varianza σ^2 . N es el tamaño de la muestra.

Distribución de la media muestral

- La media de la media muestral es la media poblacional: $E(x)=m$
- La varianza de la media muestral = σ^2 / N

Distribución de la varianza muestral

- La media de la varianza muestral es la varianza poblacional.
- La varianza de la varianza muestral tiende a cero cuando el tamaño de muestra tiende a infinito.

17

Muestreo de Poblaciones Normales

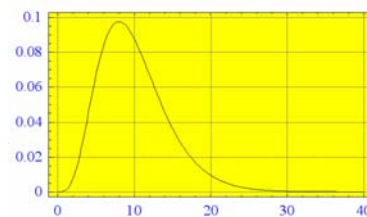
Distribución GI-dos (χ^2).

Sean X_i $i=1...,\mu$ variables Normales $N(0,1)$

$$Y = X_1^2 + X_2^2 + \dots + X_\mu^2$$

La media de Y es μ y su varianza es 2μ .

- Ejemplo: con 10 grados de libertad



18

Muestreo de Poblaciones Normales

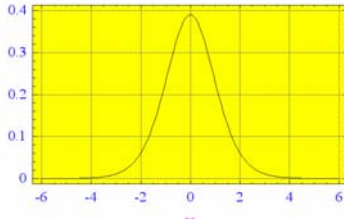
Distribución t de Student

$X \sim N(0,1)$, y una variable Y Gi-dos con μ grados de libertad.

$$t = \frac{X}{\sqrt{Y/\mu}}$$

– Tiene importancia para el estudio de una variable tipificada.

- Ejemplo con 10 grados de libertad



19

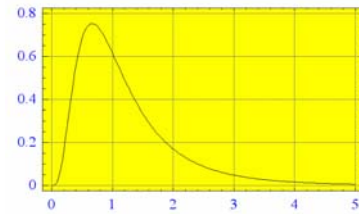
Muestreo de Poblaciones Normales

F de Snedecor

– comparación de las varianzas muestrales.

$$F = \frac{Y_1/\mu_1}{Y_2/\mu_2}$$

- Ej. con 10 grados de libertad en el numerador y en el denominador



20

Inferencia estadística

- Procedimiento que permita obtener conclusiones sobre el valor de una variable aleatoria en la población a partir de la información que hemos obtenido en la muestra.

– Muchas técnicas asumen distribuciones normales en las variables aleatorias a estudiar.

- Primero hay que ver si de verdad las poblaciones muestreadas se ajustan a la normalidad.
 - Tests gráficos:
 - » **Histogramas.** Se necesitan al menos 40 ó 50 datos.

21

Inferencia estadística: un caso sencillo

¿Se puede decir si el valor de la media observada un proceso se aleja de un valor esperado?

Pasos:

1. Determinar si la muestra es Normal por medio del Análisis descriptivo:
 - calcular diferentes parámetros de la muestra (media, desviación estándar, coeficiente de asimetría, coeficiente de curtosis...).
2. Realizar el Contraste de Hipótesis:
 - Hipótesis de partida $\mu=m$ (se la llama **hipótesis nula H0**).
 - Se usa el estadístico *t de Student*,

22

Inferencia estadística: un caso sencillo

- Fórmula para contrastar $m=m_0$ (t de Student)

$$\left| \frac{\bar{x} - m_0}{S/\sqrt{N}} \right| > t_{n-1}(\alpha)$$

Asume que las varianzas de las muestras son iguales (ver test F)

- Intervalo de confianza para m

$$\left[\bar{x} - t_{N-1}(\alpha) \frac{S}{\sqrt{N}}, \bar{x} + t_{N-1}(\alpha) \frac{S}{\sqrt{N}} \right]$$

- Intervalo de confianza para σ^2 (usar Gi-2)

$$\left[\sqrt{\frac{(N-1)S^2}{g_2}}, \sqrt{\frac{(N-1)S^2}{g_1}} \right]$$

23

Fases del estudio mediante modelos

1. Definición del problema

- Es indispensable para poder precisar los objetivos.

2. Formulación del modelo

- Definir las variables dependientes y las explicativas.
 - Mejor si se puede interpretar el significado de cada parámetro del modelo

3. Recogida de datos

4. Estimación del modelo

- datos \rightarrow valores de parámetros \rightarrow modelo
- Estimación de la precisión de esos parámetros.

5. Validar el modelo

6. Explotación del modelo.

24

Ej. de clasificador estadístico: K vecinos más próximos

La "proximidad" da una idea de la densidad de probabilidad.

- **Dados m ejemplos etiquetados, determinar la etiqueta de un nuevo dato.**
 - "El vecino más próximo". Comparar el dato con TODOS los utilizados como ejemplos. La etiqueta que se le asocia es la del más cercano.
 - "Los K vecinos más próximos". Comparar el dato nuevo con TODOS los ejemplos. Asignarle a la clase que tenga K ejemplos más próximos.
- Requiere memorizar todos los datos.

25

Ej. de clasificador estadístico: Regresión y Modelos Lineales para Series Temporales

- **Regresión lineal.**
- **Modelos lineales sencillos para Series Temporales.**
 - Modelos MA (media móvil): el valor presente está influido por los valores de las entradas en instantes anteriores.
 - Modelos AR (autorregresión): valor presente como combinación lineal de los valores anteriormente generados.
 - Idea: síntesis de una onda mediante el filtrado de ruido blanco.
 - Modelos ARMA: mezcla de los dos anteriores.

26

Métodos Bayesianos

27

Procesos Estocásticos

- **Concepto**
 - Proceso en el que una o más variables aleatorias fluctúan a lo largo del tiempo.
 - **Realización del proceso estocástico X(t)**: la secuencia de valores observados sobre un individuo de una variable aleatoria a lo largo del tiempo.
- **Según la naturaleza temporal:**
 - Continuo – se puede observar en cualquier instante.
 - Discreto – se observa a instantes específicos (no necesariamente equiespaciados)
- **Sobre X(t) se puede definir: valores medios, varianzas...**
- **Sobre los pares (X(t1), X(t2)) se puede definir la covarianza o valores de correlación.**

$$\text{COV}_{t1,t2} = \frac{\sum (x_{t1} - \bar{x})(x_{t2} - \bar{x})}{N - 1} \quad r_{xy} = \frac{\text{COV}_{t1,t2}}{s_x^2}$$

28

Procesos Estocásticos

- **Proceso estacionario: sus pautas de comportamiento no se modifican a lo largo del tiempo.**
 - Medias y varianzas de cualquier n-tuplas de variables se mantienen constantes a lo largo del tiempo.
- **Proceso no estacionario.**
- **Hipótesis de ergodicidad:**
 - Se pueden obtener datos sobre la población observando a un solo individuo a lo largo del tiempo.
 - La mayor parte de los procesos estacionarios cumplen esta hipótesis.

29

Procesos de Markov

- **En un proceso estocástico de Markov se cumple que**
$$p(X(t+u)/X(t) = a, X(t-1) = b, \dots)$$
depende sólo del valor más reciente →
$$p(X(t+u)/X(t) = a, X(t-1) = b, \dots) = p(X(t+u)/X(t) = a)$$
- **Cadenas de Markov: procesos de Markov en los que la variable estudiada es de tipo discreto.**
 - De parámetro discreto en el tiempo.
 - De parámetro continuo en el tiempo.

30

Procesos Estocásticos: Cadenas de Markov

- Estado del sistema: valor que toma la variable aleatoria $X(t) = E_i$.
- Cada estado lleva asociada una probabilidad: $p_i(t) = p(X(t) = E_i)$.
- Matriz de Transición:
 $P = p_{ij}(t) = p(X(t+1) = E_j | X(t) = E_i)$
 - Matriz homogénea en el tiempo: $p_{ij}(t) = \text{cte} \forall t, i, j$
 - A través de la matriz y la probabilidad de cada estado en el $t=0$, se puede calcular la probabilidad de cada estado en cualquier tiempo.

31

Procesos Estocásticos: Cadenas de Markov

- Cadenas de Markov como clasificadores.
 - Modelar la generación de los elementos de una serie (patrón o clase a detectar) mediante un conjunto de cadenas de Markov, un modelo para cada patrón.
 - Se dice que el patrón pertenece al modelo que lo puede generar con mayor probabilidad.
 - Problemas:
 - Calcular la probabilidad de que un modelo genere una determinada secuencia de estados es costoso en cálculos.
 - Construir los modelos (definir su matriz de transición = probabilidades de transición p_{ij}).
 - Ambos puntos se pueden resolver por métodos estadísticos o por ejemplo, con redes neuronales artificiales.
 - Se usan modelos simplificados, como por ejemplo los modelos izquierda – derecha.

32