

Modelado y Simulación (505103009)

Tema 7. Contrastes poblacionales

Javier Vales Alonso

Grado en Ingeniería Telemática

2020

Universidad Politécnica de Cartagena

Introducción

Histograma

Contraste χ^2

¿Cómo estudiar esta unidad?

1. Haga una primera lectura de la unidad. Concéntrese en ver las ideas generales y el uso del contraste χ^2 .
2. Haga una revisión más a fondo de los ejemplos presentados.
3. En caso de dudas, puede consultar los libros de referencia, o contactar con el profesor.

Introducción

En el tema anterior, hemos visto técnicas que nos permiten calcular intervalos de confianza para los parámetros de una distribución. **Otro aspecto de interés es determinar si éstos tienen una “forma” concreta.** Por ejemplo, ¿el tiempo de respuesta en una cola G/G/k es una variable aleatoria exponencial? ¿las tiradas de un dado son uniformes?

En este tema estudiaremos cómo realizar **contrastes sobre poblaciones** mediante el uso del test estadístico χ^2 .

Introducción (II)

Formalmente, se tratará de determinar si existe evidencia suficiente para rechazar la hipótesis nula:

$$\mathcal{H}_0 : X \sim \mathcal{D}_\theta$$

frente a la hipótesis alternativa trivial, para un nivel de significación α dado.

Dado un conjunto de muestras $\{X_1, \dots, X_n\}$ **se aproximará la distribución de X mediante un histograma**, y **se calculará la divergencia con la forma teórica** (la propuesta por la hipótesis nula). **Si la divergencia supera un umbral** determinado, consideraremos que \mathcal{H}_0 **debe ser rechazada**.

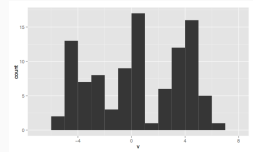
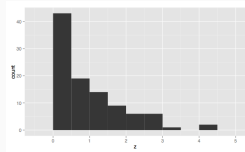
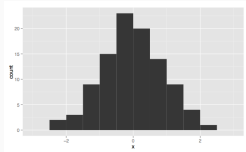
Histograma

Un histograma es una representación aproximada de la distribución de una variable aleatoria. Se construye dividiendo el rango de posibles valores de la variable en **categorías** (*bins*) disjuntas. Típicamente se escogen K categorías equiprobables.

Por ejemplo, si $X \sim \mathcal{U}(0, 1)$, puede dividirse en $K=10$ intervalos: $\{(0, 0, 1), [0, 1, 0, 2), \dots, [0, 9, 1)\}$, cada uno de probabilidad $\frac{1}{10}$.

Determinadas las categorías C_k , para $k=1, \dots, K$, el histograma simplemente es el número de muestras que caen dentro de cada categoría, $o_k = \sum_{i=1}^n \mathbf{1}_{X_i \in C_k}$.

Histogramas (II)



Ejemplos de histogramas

Contraste χ^2

El número de muestras esperado en cada categoría C_k bajo $\mathcal{H}_0 : X \sim \mathcal{D}_\theta$, es $e_k = np(X \in C_k) = n\mathcal{D}_\theta(C_k)$.

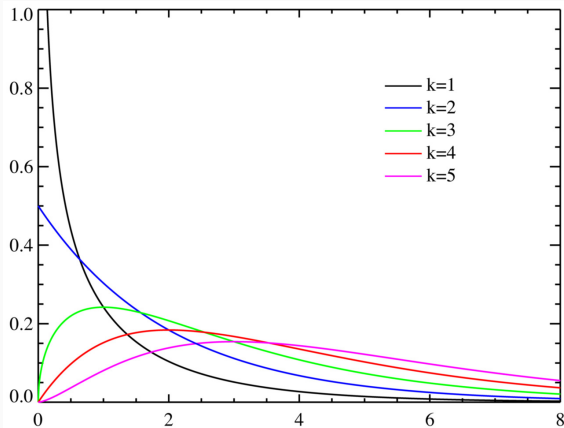
El contraste:

$$\varepsilon^2 = \sum_{k=1}^K \frac{(o_i - e_i)^2}{e_i}$$

tiene distribución χ^2 de $K - 1$ grados de libertad, si $n \rightarrow \infty$.

La hipótesis nula se rechaza si la discrepancia $\varepsilon^2 > \varepsilon_{max}^2$, siendo ε_{max}^2 el valor de discrepancia máximo para el nivel de significación α dado. Éste se obtiene de las tablas de la distribución χ^2 .

Contraste χ^2 (II)



Función de densidad de una variable χ^2 para distintos grados de libertad

Contraste χ^2 (III)

Ejemplo 1: Se desea saber si para un proceso aleatorio, del que se ha obtenido un histograma con número de observaciones $O = (11, 14, 8, 5, 12)$ para las categorías $\{(0, 0, 4), (0, 4, 0, 8), (0, 8, 1, 2), (1, 2, 1, 6), (1, 6, 2)\}$, se puede rechazar o no $\mathcal{H}_0 : X \sim \mathcal{U}(0, 2)$ para un nivel de significación del 10%.

Ejemplo 2: Determinar, para los datos del ejemplo anterior, el nivel de significación máximo para el que no se rechazaría \mathcal{H}_0 .