Contraste de hipótesis, teoría de decisión

Antonio Sala Piqueras

Notas sobre control de sistemas multivariables/complejos

Dept. Ing. Sistemas y Automatica (DISA)
Universitat Politècnica de València (UPV)

Video-presentación disponible en:

personales.upv.es/asala/videos/**thge**.html

Presentación

Motivación:

En muchas aplicaciones se desea comprobar si algún parámetro relacionado con la distribución de una variable aleatoria tiene un valor “normal” o no. Si las medidas son ruidosas, hay que tomar varias muestras en los cálculos.

Objetivos:

Comprender qué es el contraste de hipótesis, y el nivel de confianza.
Comprender el concepto de coste de errores y la optimización asociada.

Contenidos: Contraste de hipótesis: definiciones básicas. Nivel de confianza. Ejemplo: detección de fallos. Problema de optimización. Conclusiones.

Contraste estadístico de hipótesis

Test de hipótesis: determinada aserción $\mathcal{H}_0$ sobre una vble aleatoria (o varias) debe ser verificada con n muestras.

- La media de y es 7.
- Las variables x e y son independientes.
- El modelo $y = 3x + v$ es cierto.
- La desv. típica de y es 2.5.

Las muestras nunca darán media exacta 7, ni correlación cero, ni ......

Formalmente: La hipótesis no puede ser nunca "*validada*"... simplemente "*no es rechazada (por el momento)*".

Ejemplo: Con tres datos $\{6.1, 8.3, 6.7\}$ no rechazamos que la media sea 7... con trescientos datos $\{6.1, 8.3, 6.6, 6.9, 6.2, \ldots, 5.5, 5.9, \ldots, 2.2\}$ de los que el 90% están por debajo de 7, sí lo rechazamos... con una cierta probabilidad de equivocarnos.

Nivel de confianza (1-α) o significación (α)

Se calcula un **estadístico** (función Z de los datos, especialmente elegida) de modo que si $\mathcal{H}_0$ es cierta, entonces el $(1-\alpha) = 99\%$ de los experimentos deberían dar $Z \in [z_{low}, z_{up}]$.

- Si, con unos datos, $Z \notin [z_{low}, z_{up}]$, rechazaré hipótesis (considero que no se ha salido "por puro azar"), pero un $\alpha =$1% de las veces rechazaré por **error**.

*Obviamente, debería elegirse Z tal que, conforme aumenta el número de datos, la anchura del intervalo se reduzca (a cero en el límite con infinitos datos).

- Si $Z \in [z_{low}, z_{up}]$, $\mathcal{H}_0$ puede no ser cierta... más datos podrían desmentirla. Compromiso:
  - α **alto:** Intervalo grande, casi nunca rechazo, tardo mucho (muchos datos) a detectar fallo, pero estoy "segurísimo" cuando detecto.
  - α **bajo:** detecto cambios enseguida, pero muchas "falsas alarmas".

Nivel de confianza (1-α) o significación (α)

Se calcula un **estadístico** (función Z de los datos, especialmente elegida) de modo que si $\mathcal{H}_0$ es cierta, entonces el $(1-\alpha) = 99\%$ de los experimentos deberían dar $Z \in [z_{low}, z_{up}]$.

- Si, con unos datos, $Z \notin [z_{low}, z_{up}]$, rechazaré hipótesis (considero que no se ha salido "por puro azar"), pero un $\alpha =$1% de las veces rechazaré por **error**.

*Obviamente, debería elegirse Z tal que, conforme aumenta el número de datos, la anchura del intervalo se reduzca (a cero en el límite con infinitos datos).

- Si $Z \in [z_{low}, z_{up}]$, $\mathcal{H}_0$ puede no ser cierta… más datos podrían desmentirla. Compromiso:
  - α **alto:** Intervalo grande, casi nunca rechazo, tardo mucho (muchos datos) a detectar fallo, pero estoy "segurísimo" cuando detecto.
  - α **bajo:** detecto cambios enseguida, pero muchas "falsas alarmas".

Aplicación a detección de fallos en procesos

$\mathcal{H}_0$: Bien por un modelo teórico, bien por recolección de datos anteriores, se estima que determinada variable y (o conjunto de ellas) tiene una cierta distribución de probabilidad (p.ej., normal media 2, desv.típ. 0.4).

Objetivo: Tomar, periódicamente, muestras de y para comprobar que *la distribución de probabilidad de y no ha cambiado*.

Objetivos (menos ambiciosos) pero más sencillos:

- Comprobar que la media de y no ha cambiado o, si lo ha hecho, que no ha superado ciertos umbrales.
- Comprobar que la varianza de y no ha cambiado.

El objetivo sería detectar los cambios "lo más pronto posible" (evitar no detectar un fallo presente, error tipo 2), sin cometer "falsas alarmas" (error tipo 1).

Generalización: teoría de decisión

En un problema práctico, tenemos:

1. Coste de **adquirir información**.

   Realizar un TAC craneal; desmontar inyección coche para comprobar su estado

   A esto hay que añadir el ¡coste de **"retrasar" la decisión***! (si Pedro tiene un tumor y me espero dos meses al resultado de las pruebas, es más difícil de tratar)*

2. Coste de procesar información. *Consulta de médico: 60 EUR.*

3. Coste de descartar $\mathcal{H}_0$ siendo cierto.

   Descarto "Pedro no tiene tumor cerebral", o sea, diagnostico "sí tiene" y le meto radioterapia estando sano.

   Cambio inyección (EUR 2500) cuando lo único que pasaba es que la batería estaba descargada.

4. Coste de afirmar $\mathcal{H}_0$ siendo falso.

   Diagnosticar "Pedro no tiene nada grave"... y fallece 48 horas después.

   "La inyección del coche está bien"... y el cliente se queda tirado en un camino de Bielorusia.

Por tanto, el problema general es de **optimización**: minimizar coste.

Teoría de decisión (II)

Existen dos tipos de soluciones al problema de optimización:

- **Feedforward (fuera de linea):** tomar una muestra semanal de leche, analizar la densidad presencia de bacterias, y comercializar 15000 litros si (dens$\geq$1.04 & bact: negativo); tirar por el desagüe en caso contrario.

- **Feedback:** En función de la información hasta el momento, decidir si:
  1. Diagnóstico: todo **normal**,
  2. Diagnóstico: "**fallo xxx**",
  3. **hago pruebas adicionales** (adquiero información para actualizar probabilidades) antes de decidir.

Hay soluciones explícitas teóricas para casos sencillos, pero por ahora siguen haciendo falta médicos, mecánicos e ingenieros.

Conclusiones

- En muchas aplicaciones interesa saber si determinada afirmación sobre una variable aleatoria podría ser desmentida.
- Dicha afirmación suele ser una situación considerada "normal", "no pasa nada" (hipótesis nula $\mathcal{H}_0$).
- Se propone una función de los datos Z (*estadístico*) y su distribución de probabilidad si $\mathcal{H}_0$ fuera cierta.
- Se **rechaza** si el valor de Z para un experimento concreto es "muy improbable si $\mathcal{H}_0$ fuera cierta".
- El grado de confianza (1-probabilidad de falsa alarma) alto aumenta el número de muestras necesario para detectarlo...
- Decidir con qué probabilidad aceptar/rechazar es, en general, un problema de optimización de costes (teoría de decisión).