Estadística No Paramétrica

Clase 2

Joaquin Cavieres G.

Ingeniería en Estadística

Facultad de Ciencias, Universidad de Valparaíso

Estadística paramétrica vs no paramétrica

- La idea principal de la estadística no paramétrica es hacer inferencia sobre cantidades desconocidas pero sin recurrir a un conocimiento previo del problema.

- La idea principal de la estadística no paramétrica es hacer inferencia sobre cantidades desconocidas pero sin recurrir a un conocimiento previo del problema.

  Considere $X \sim F$ y deseamos estimar la esperanza de X ($\mathbb{E}(X)$).
- Dentro de la estadística paramétrica se asume que F viene de algún tipo de función de distribución la cual está gobernada por un pequeño número de parámetros.
- Por ejemplo: Considere la distribución Normal:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

aquí los parámetros son estimados y podemos hacer inferencia sobre las cantidades en que estamos interesados ($\mathbb{E}(X)$) bajo el supuesto de que $X \sim N(\hat{\mu}, \hat{\sigma^2})$.

Función de distribución empírica

Comencemos con la estimación de la función de distribución acumulada (CDF en inglés).

Supongamos que $X \sim F$, donde $F(x) = \mathbb{P}(X \leq x)$ es una función de distribución. Dado lo anterior, la función de distribución empírica $\hat{F}$ es la CDF $1/n$ en cada punto de los datos x_i:

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I\left(x_i \leq x\right) \tag{1}$$

donde I es la función indicadora.

Propiedades de $\hat{F}$

- En cualquier valor fijo de x

$$\mathbb{E}\{\hat{F}(x)\} = F(x) \tag{2}$$

$$\mathbb{V}\{\hat{F}(x)\} = \frac{1}{n}F(x)(1 - F(x)) \tag{3}$$

- Lo anterior implica que :

$$\hat{F}(x) \xrightarrow{\mathrm{P}} F(x) \tag{4}$$

para cualquier x

Teorema de Glivenko-Cantelli

El teorema de Glivenko-Cantelli es una prueba aún más fuerte de convergencia ya que:

- Supongamos $X_1, ..., X_n$ son variables aleatorias *i.i.d* con CDF F, por lo tanto:

$$\sup_x |\hat{F}(x) - F(x)| \xrightarrow{\text{a.s.}} 0 \tag{5}$$

- Este teorema ha sido llamado el teorema fundamental de la estadística no paramétrica.

Teorema de Glivenko-Cantelli

El teorema de Glivenko-Cantelli es una prueba aún más fuerte de convergencia ya que:

- Supongamos $X_1, ..., X_n$ son variables aleatorias *i.i.d* con CDF F, por lo tanto:

$$\sup_x |\hat{F}(x) - F(x)| \xrightarrow{\text{a.s.}} 0 \tag{5}$$

- Este teorema ha sido llamado el teorema fundamental de la estadística no paramétrica.

Pregunta: ¿Cual tipo de convergencia es más fuerte, en probabilidad o casi segura?

Estadística paramétrica vs no paramétrica

El termino “no paramétrica" es un tanto impreciso para la función de distribución. El mejor significado esta relacionado con una distribución no conocida a priori y sin parámetros que la gobiernen. En inglés esta es conocida como *distribution-free*.

Propiedades

Ventajas de los métodos no paramétricos:

- Requieren pocos supuestos desde la población en que se obtienen los datos.
- Se pueden obtener p-values exactos para los test, intervalos de confianza, tasas de errores en comparaciones, etc.
- Son generalmente fáciles de aplicas.
- Son generalmente fáciles de comprender.
- Son relativamente insensibles a datos atípicos.
- El método de Bootstrap dispone de aproximaciones para utilizar en escenarios donde la teoría paramétrica es casi impracticable

Propiedades

Desventajas de los métodos no paramétricos:

- Parecen sacrificar demasiada información básica de las muestras.
- Un poco menos eficiente que la teoría de los métodos basadas en la distribución Normal.

Propiedades de la función de distribución

La función de distribución no esta gobernada por parámetros conocidos a priori y su estructura esta determinada por los datos observados. Generalmente se le conoce como *Distribución-free*.

Test Binomial

Ejemplo: Queremos calcular la probabilidad de éxito de cierto experimento (p) sobre n muestras. Los resultados se obtiene en forma dicotómica/binaria.

Test Binomial

Ejemplo: Queremos calcular la probabilidad de éxito de cierto experimento (p) sobre n muestras. Los resultados se obtiene en forma dicotómica/binaria.

Supuestos

- Los resultados de cada experimento pueden ser éxito/fracaso, si/no, 0/1, etc.
- La probabilidad de éxito (p) se mantiene constante ensayo tras ensayo.
- Las muestras son independientes.

Test Binomial

Consideremos lo siguiente:

$$H_0 : p = p_0 \tag{6}$$

donde p_0 es algún número entre $0 < p_0 < 1$ y $B =$ número de éxitos

Prueba de cola superior

$$H_0 : p = p_0 \tag{7}$$

vs

$$H_1 : p > p_0 \tag{8}$$

con un nivel de signifcancia α. Entonces rechazamos H_0 si $B \geq b_\alpha$, de otra manera no rechazamos H_0.

b_α es el punto del percentil α superior de la distribución Binomial con tamaño de muestra n y probabilidad de éxito p_0.

Test Binomial

$$H_0 : p = p_0$$

donde p_0 es algún número entre $0 < p_0 < 1$ y $B =$ número de éxitos

Prueba de cola inferior

$$H_0 : p = p_0 \tag{9}$$

vs

$$H_1 : p < p_0 \tag{10}$$

con un nivel de signifcancia α. Entonces rechazamos H_0 si $B \leq c_\alpha$, de otra manera no rechazamos H_0.

c_α es el punto del percentil α inferior de la distribución Binomial con tamaño de muestra n y probabilidad de éxito p_0. Para el caso especial del test $p = 0.5 \rightarrow c_\alpha = n - b_\alpha$.

$$H_0 : p = p_0$$

donde p_0 es algún número entre $0 < p_0 < 1$ y $B =$ número de éxitos

Prueba de cola ambos lados

$$H_0 : p = p_0 \tag{11}$$

vs

$$H_1 : p \neq p_0 \tag{12}$$

con un nivel de signifcancia α. Entonces rechazamos H_0 si $B \geq b_{\alpha_1}$ o $B \leq c_{\alpha_2}$, de otra manera no rechazamos H_0.

b_{α_1} es el punto del percentil α_1 superior de la distribución, c_{α_2} es el punto del percentil α_2 inferior y $\alpha_1 + \alpha_2 = \alpha$.

Comentarios adicionales:

- Los supuestos del test Binomial son los generales que se asumen en este tipo de experimentos.
- La constante b_α es elegida tal que la probabilidad de rechazar H_0, cuando H_0 es verdadera, es α.
- Podemos controlar este error tipo I en función de los supuestos y de la especificación de p determinan la distribución de probabilidad de B.

Ejemplo

Supogamos que $n = 8$ y deseamos probar que $H_0 : p = 0.4$ versus $H_1 : p > 0.4$.

b	0	1	2	3	4	5	6	7	8
$P_{.4}(B \geq b)$	1	0.9832	0.8936	0.6846	0.4059	0.1737	0.0498	0.0085	0.0007

Desarollo: Podemos encontrar la constante b_α que satisface la ecuacion $P_{0.4}\{B \geq b_\alpha\} = \alpha$ para ciertos valores de α. Para un $\alpha = 0.0085$, $b_{0.0085} = 7$. Para un $\alpha = 0.1737$, $b_{0.1737} = 5$. Por lo tanto, a medida que aumenta α la constante b_α disminuye, en consecuencia tambien decrese la probabilidad de tener un error tipo II en nuestro test.

Ejemplo

¿Como lo haría para generar el test si queremos evaluar $H_0 : p = 0.4$ versus $H_1 : p > 0.4$?

Ejemplo

¿Como lo haría para generar el test si queremos evaluar $H_0 : p = 0.4$ versus $H_1 : p > 0.4$?

Desarollo: n = 8 y queremos evaluar que $p = 0.4$ versus la hipótesis alternatica de $p > 0.4$ (¿Cual test deberiamos utilizar?).

Ejemplo

¿Como lo haría para generar el test si queremos evaluar $H_0 : p = 0.4$ versus $H_1 : p > 0.4$?

Desarollo: n $= 8$ y queremos evaluar que $p = 0.4$ versus la hipótesis alternatica de $p > 0.4$ (¿Cual test deberiamos utilizar?).Supongamos que $\alpha = 0.1064$ entonces $P_{0.4}\{B \geq 2\} = 0.8936$ y $P_{0.4}\{B \leq 1\} = 1 - 0.8936 = 0.1064$. Por lo tanto, $c_{0.1064} = 1$ y produce que $\alpha = 0.1064$. Dado lo anterior rechazamos H_0 si $B \leq 1$ y rechazamos H_1 si $B > 1$.

Ejemplo

¿Como lo haría para generar el test de ambos lados si queremos evaluar $H_0 : p = 0.4$ versus $H_1 : p > 0.4$?

Ejemplo

¿Como lo haría para generar el test de ambos lados si queremos evaluar $H_0 : p = 0.4$ versus $H_1 : p > 0.4$?

Desarollo: Note que 6 es el valor superior del percentil 0.0498 de la distribución nula de B y 1 es el punto inferior del percentil 0.1064. Por lo tanto, el test que rechaza H_0 cuando $B \geq 6$ o cuando $B \leq 1$ o cuando o rechaza H_0 cuando 1 < B < 6 es un $\alpha = 0.0498 + 0.1064 = 0.1562$.

Aproximación mediante n grande

La aproximación mediante un n grande esta basada en la normalidad asintótica de B (estandarizada adecuadamente). ¿Que significa que sea adecuadamente estandarizada?

Aproximación mediante n grande

La aproximación mediante un n grande esta basada en la normalidad asintótica de B (estandarizada adecuadamente). ¿Que significa que sea adecuadamente estandarizada? La media y varianza de B deben ser conocidas cuando la hipótesis nula es verdadera.

Si H_0 no se rechaza (verdadera)

$$\mathbb{E}_{p_0}(B) = np_0 \tag{13}$$

$$\mathbb{VAR}_{p_0}(B) = np_0\,(1 - p_0) \tag{14}$$

La versión estandarizada de B es:

$$B^* = \frac{B - E_{p_0}(B)}{\{\mathrm{var}_{p_0}(B)\}^{1/2}} = \frac{B - np_0}{\{np_0\,(1 - p_0)\}^{1/2}} \tag{15}$$

Cuando H_0 es verdadera, B^* tiene una distribución Normal (0,1) asintótica cuando n tiende al infinito.

Aproximación mediante n grande

z_α es el punto del percentil α superior de la distribución Normal (0.1). Para encontrar z_α podemos utilizar la función $\texttt{qnorm}(1-\alpha, 0, 1)$ en R. Por ejemplo, para encontrar $z_{0.5}$ escribimos $\texttt{qnorm}(0.95, 0, 1)$ y obtenemos $z_{0.5} = 1.645$. La aproximación Normal para:

H_0 si $B \geq b_\alpha$ (superior)

- Rechazamos H_0 si $B^* \geq z_\alpha$, de otra forma no se rechaza

H_0 si $B \leq c_\alpha$ (inferior)

- Rechazamos H_0 si $B^* \leq -z_\alpha$, de otra forma no se rechaza

H_0 si $B \geq b_{\alpha_1}$ o $B \leq c_{\alpha_2}$ (ambos lados)

- Rechazamos H_0 si $|B^*| \geq z_{\alpha/2}$, de otra manera no se rechaza

Ejemplo

Dickinson, Putz y Canham (1993). Los arbustos a menudo forman grupos densos donde la abundancia de árboles se ha mantenido baja en forma artificial (por ejemplo, en los derechos de paso de las líneas eléctricas). Digamos que tenemos éxito si mueren más ramas de las que viven en los huecos entre los grupos. Sea p la correspondiente probabilidad de éxito. Suponemos que la probabilidad de éxito para sitios ricos en nutrientes con suelo húmedo ha sido establecida por estudios previos en 15%.

¿Los sitios de suelo seco y pobre en nutrientes tienen la misma probabilidad de éxito que los sitios de suelo húmedo y rico en nutrientes?

Ejemplo

Desarrollo:

$$H_0 : p = 0.15$$

vs

$$H_1 : p > 0.15$$

Tenemos una muestra de tamaño $n = 7$ y existen $B = 6$ casos de éxito.

x	0	1	2	3	4	5	6	7
$P_{.15}(B > x)$	.6794	.2834	.0738	.0121	.0012	.0001	.0000	.0000

Ejemplo

Desarollo: Desde `R` escribimos: `round(pbinom(0:7,7,.15,lower.tail=F),4)` y obtenemos las probabilidades $P_{0.15}(B > x)$, parar $x = 0, ..., 7$. Para encontrar $P_{0.15}(B > x)$ note que $P_{0.15}(B \geq x) = P_{0.15}(B > x - 1)$.