Testes não paramétricos para famílias de distribuições

O teste de aderência de Kolmogorov é adequado para verificar se um conjunto de dados está de acordo com uma função distribuição <u>totalmente especificada</u>, ou seja, quando não existirem parâmetros desconhecidos que devem ser estimados a partir da amostra.

O teste qui-quadrado de aderência é flexível nesse aspecto, pois permite que alguns parâmetros sejam estimados a partir da amostra. Subtrai-se um grau de liberdade para cada parâmetro estimado. No entanto, este teste tem duas limitações, exige que os dados sejam agrupados e esse procedimento de agrupamento é arbitrário. Além disso, a distribuição da estatística de teste é aproximada.

Os testes de Lilliefors são modificações do teste de Kolmogorov, que permitem testar a aderência a uma distribuição de probabilidades sem especificar seus parâmetros.

As estatísticas de testes são as mesmas do teste de Kolmogorov, mas as tabelas com a distribuição da estatística sob a hipótese nula não são as mesmas. Mudam de uma distribuição hipotetizada para outra.

Teste de Lilliefors para normalidade

Testa se um conjunto de dados é proveniente de uma distribuição normal.

Dados: $X_1, X_2 \ldots X_n$ amostra aleatória de uma função distribuição desconhecida $F(x)$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i = \hat{\mu} \quad \text{e}$$

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2} = \hat{\sigma}$$

$$Z_i = \frac{X_i - \bar{X}}{\qquad} \quad i = 1, 2, \ldots n,$$ variáveis padronizadas.

H_0: Os dados provem de uma distribuição normal
H_a: A função distribuição dos X_i não é normal.

Estatística de Teste:

$$T_1 = \sup_x |F^*(x) - F_n(x)| \quad \text{onde}$$

$F^*(x)$ - função distribuição acumulada da distribuição $N(0,1)$.

$$F^*(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \qquad \forall x.$$

$F_n(x)$ - função distribuição empírica das variáveis padronizadas.

Regra de Decisão:

Rejeita-se H_0 ao nível α se $T_1 > w_{1-\alpha}$, sendo que $w_{1-\alpha}$ é dado na Tabela A.14, Conover, 3ª edição [Quantis do teste de Lilliefors para a distribuição normal]. Fornece o quantil de ordem p para $p = 0{,}80,\ 0{,}85,\ 0{,}90,\ 0{,}95,\ 0{,}99$.

Ver Exemplo 1, pag. 444, Conover

Teste de Lilliefors para a distribuição exponencial

Pag 448 e Tabela A.15, Conover

Teste de Smirnov para duas amostras independentes

Dispomos de duas amostras independentes de duas populações. Desejamos testar a hipótese de que as funções distribuição associadas à variável de interesse nestas populações são idênticas.

O teste de Smirnov é uma versão para duas amostras do teste de Kolmogorov.

Dados: $X_1, X_2 \dots X_n \rightarrow F(x)$

$Y_1, Y_2 \dots Y_m \rightarrow G(x)$

Suposições:

1- As amostras são aleatórias e independentes entre si.

2- A escala de medidas é pelo menos ordinal.

3- Para que o teste seja exato, as variáveis X e Y devem ser contínuas.

Hipóteses:

A) $H_0: F(x) = G(x)$ para todo x.

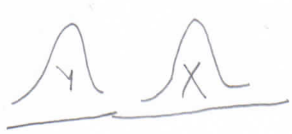
$H_a: F(x) \neq G(x)$ para algum x

B) $H_0: F(x) \leq G(x) \quad \forall x$

$H_a: F(x) > G(x)$ para algum x [os valores de X tendem a ser menores que os de Y]

C) $H_0: F(x) \geq G(x) \quad \forall x$

$H_a: F(x) < G(x)$ para algum x [os valores de X tendem a ser maiores que os de Y]



Sejam

$F_n(x)$ - função distribuição empírica com base em $X_1, X_2 \ldots X_n$

$G_m(x)$ - " " " $Y_1, Y_2 \ldots Y_m$

Estatísticas de Teste:

A) $T_1 = \sup_x |F_n(x) - G_m(x)|$

B) $T_1^+ = \sup_x [F_n(x) - G_m(x)]$

C) $T_1^- = \sup_x [G_m(x) - F_n(x)]$

Regra de Decisão:

Rejeitamos H_0 ao nível α se T_1, T_1^+ ou T_1^- excedem o quantil $w_{1-\alpha}$ dado pela Tabela A.19 se $n = m$ ou Tabela A.20 se $n \neq m$. Por que só todos unicaudais?

Exemplo:

$H_0: F(x) = G(x) \quad \forall x$

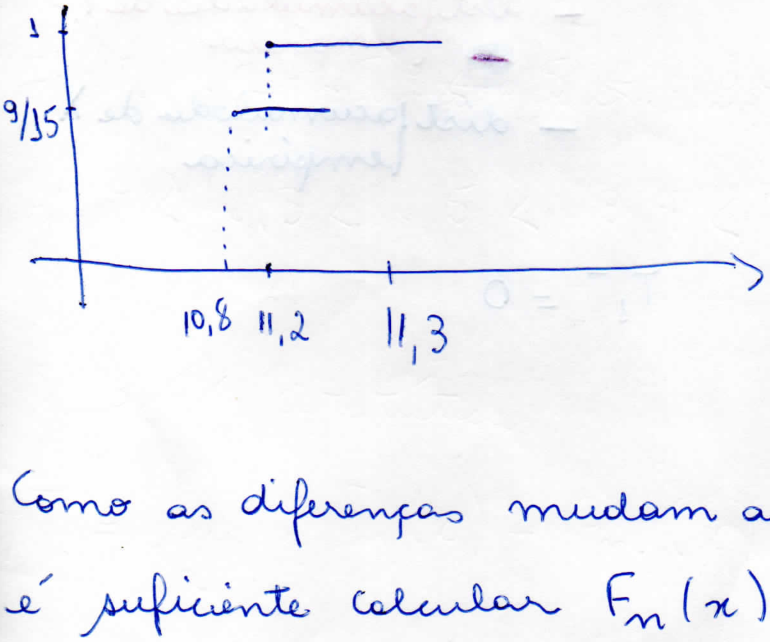
$H_a: F(x) \neq G(x)$

$n = 9 \qquad m = 15$

$X: \quad 7{,}6 < 8{,}4 < 8{,}6 < 8{,}7 < 9{,}3 < 9{,}9 < 10{,}1 < 10{,}6 < 11{,}2$

$Y: \quad 5{,}2 < 5{,}7 < 5{,}9 < 6{,}5 < 6{,}8 < 8{,}2 < 9{,}1 < 9{,}8 < 10{,}8 < 11{,}3 < 11{,}5 < 12{,}3 < 12{,}5 < 13{,}4 < 14{,}6$

X_i	Y_i	$F_n(x) - G_m(x)$
	5,2	$0 - 1/15 = -1/15$
	5,7	$0 - 2/15 = -2/15$
	5,9	$0 - 3/15 = -1/5$
	6,5	$0 - 4/15 = -4/15$
	6,8	$0 - 5/15 = -1/3$
7,6		$1/9 - 5/15 = -2/9$
	8,2	$1/9 - 6/15 = -13/45$
8,4		$-8/45$
8,6		$-1/15$
8,7		$2/45$
	9,1	$-1/45$
9,3		$4/45$
	9,8	$1/45$
9,9		$2/15$
10,1		$11/45$
10,6		$16/45$
	10,8	$13/45$
11,2		$1 - 9/15 \quad 2/5$
	11,3	$1/3$
	11,5	$4/15$
	12,3	$1/5$
	12,5	$2/15$
	13,4	$1/15$
	14,6	0



$n = 9 \quad m = 15$

Tabela A.20

$\alpha = 0{,}05 \quad w_{0,95} = \frac{8}{15}$

$T_1 = \frac{2}{5} = \frac{6}{15}$

Como as diferenças mudam apenas nos valores X_i e Y_i, é suficiente calcular $F_n(x) - G_m(x)$ apenas para os valores da amostra. [Isto pode ser obs. graficamente]

$n = 9 \quad m = 15$ tabela 17

$\alpha = 0{,}05 \quad \ldots \quad 8 \quad T = 2 = 6$

não rejeitamos H_0

Distribuição exata de T_1^+ sob $H_0: F(x) = G(x)$

T_1, T_1^+ e T_1^- dependem dos X's e Y's apenas através da ordem e não de seus valores numéricos.

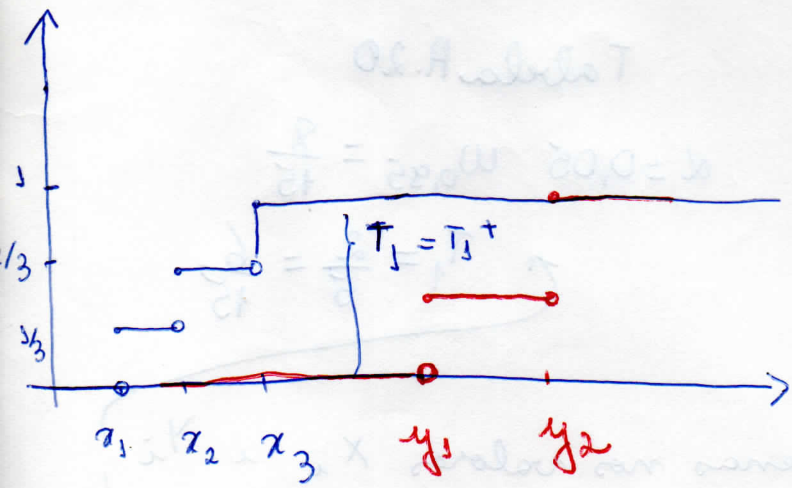
$n = 3 \quad X_1, X_2, X_3 \qquad m = 2 \quad Y_1, Y_2$

Arranjo	T_1	T_1^+	T_1^-
$X<X<X<Y<Y$	1	1	0
$X<X<Y<X<Y$	$2/3$	$2/3$	0
$X<Y<X<X<Y$	$1/2$	$1/2$	$1/6$
$Y<X<X<X<Y$	$1/2$	$1/2$	$1/2$
$X<X<Y<Y<X$	$2/3$	$2/3$	$1/3$

Arranjo	T_1	T_1^+	T_1^-
$X<Y<X<Y<X$	$1/3$	$1/3$	$1/3$
$Y<X<X<Y<X$	$1/2$	$1/6$	$1/2$
$X<Y<Y<X<X$	$2/3$	$1/3$	$2/3$
$Y<X<Y<X<X$	$2/3$	0	$2/3$
$Y<Y<X<X<X$	1	0	1

— dist acumulada de Y empírica

— dist acumulada de X empírica



$T_1^- = 0$

Sob H_0, cada arranjo tem prob: $\dfrac{1}{\binom{n+m}{n}} = \dfrac{1}{10}$

$\therefore$ sob H_0 : $F(x) = G(x)$

T_1	$P(T_1)$
1/3	1/10
1/2	3/10
2/3	2/5
1	1/5

T_1^+	$P(T_1^+)$
0	1/5
1/6	1/10
1/3	1/5
1/6	1/5
2/3	1/5
1	1/10

T_1^-	$P(T_1^-)$
0	1/5
1/6	1/10
1/3	1/5
1/2	1/5
2/3	1/5
1	1/10

sob H_0 T_1^+ e T_1^- são ident. distrib $\forall$ n e m.

Com estas distrib, obtemos os quantis.

Para testes unicaudais o procedimento é o mesmo pois

$$\sup P(\text{rej } H_0 \mid F(x) \leq G(x)) = P(\text{rej } H_0 \mid F(x) = G(x)$$