UN NOUVEAU TEST STATISTIQUE POUR LA COMPARAISON DE PROPORTIONS

Éric Taillard, Ph. Wälti, J. Zuber

EIVD

Haute École spécialisée de Suisse occidentale,

Yverdon-les-Bains, Suisse

EXEMPLE TYPIQUE DE RÉSULTATS NUMÉRIQUES :

Repris de Kim[3], JoH 9 (3), Juin 2003

Exemple de problème	Nombre d'exécutions	TCC	RSC	FSC	SSC	SHC	TMC
Sorting network design, $n = 7$	10	7	5	**8**	**8**	**8**	5
Sorting network design, $n = 7$	10	3	2	3	4	3	2
Sorting network design, $n = 13$	10	0	0	0	0	0	0
2DTTTgame	10	6	8	4	**9**	6	6
Nim(3,4,5,4)	10	3	2	**6**	**6**	4	3
Nim(5,7,11,6)	10	0	0	**1**	**1**	0	0

Conditions de l'expérience :

Un certain nombre d'exécutions de méthodes non-déterministes sur le même exemple, ou

Des méthodes (non-) déterministes exécutées sur des exemples de problèmes choisis aléatoirement

Question typique:

Est-ce que SSC est significativement meilleure que FSC pour 2DTTT game ?

i. e. est-ce qu'un taux de succès de 9/10 est significativement meilleur qu'un taux de 4/10 ?

TEST STANDARD, COMPARAISON DE PROPORTIONS

Conditions de l'expérience :

n_a exécutions de la méthode A, a succès, probabilité de succès de A (inconnue) : p_a $(= 1 - q_a)$

n_b exécutions de la méthode B, b succès, probabilité de succès de B (inconnue) : p_b $(= 1 - q_b)$

Les exécutions sont toutes supposées indépendantes

Soient X_A (resp. X_B) la variable aléatoire associée au succès ($X_A = 1$) de A (resp. B)

Si n_a et n_b sont suffisamment grand alors :

$D = X_A / n_a - X_B / n_b$ est approximativement normalement distribuée

Moyenne : $d = p_a - p_b$

Variance : $\sigma_D^2 = \frac{p_a \cdot q_a}{n_a} + \frac{p_b \cdot q_b}{n_b}$

TEST STANDARD UNILATÉRAL

Hypothèse nulle :

$p_a \leq p_b$

Hypothèse alternative :

$p_a > p_b$

Le paramètre p peut être approché par l'estimateur groupé : $\hat{p} = 1 - \hat{q} = \frac{a + b}{n_a + n_b}$

Calculer : $\hat{d} = \frac{a}{n_a} - \frac{b}{n_b}$ **et** $\hat{s} = \sqrt{\frac{\hat{p} \cdot \hat{q}}{n_a} + \frac{\hat{p} \cdot \hat{q}}{n_b}}$

Rejeter l'hypothèse nulle (et accepter l'hypothèse alternative) au seuil α si :

$\Phi\left(\frac{\hat{d}}{\hat{s}}\right) > 1 - \alpha$ (Φ : fonction de répartition d'une var. aléatoire normale centrée réduite)

($\frac{\hat{d}}{\hat{s}} > 1.645$ pour $\alpha = 5\%$ et $\frac{\hat{d}}{\hat{s}} > 2.326$ pour $\alpha = 1\%$)

TEST STANDARD BILATÉRAL

(Test du χ^2 pour comparer des proportions dans une table de contigence 2×2)

Hypothèse nulle :

$p_a = p_b = p$ (i.e. $d = 0$)

Hypothèse alternative :

$p_a \neq p_b$

Calculer : $T = \dfrac{(n_a + n_b) \cdot (a \cdot n_b + b \cdot n_a)^2}{n_a \cdot n_b \cdot (a+b) \cdot (n_a + n_b - a - b)}$ **(i.e.** $\left(\dfrac{\ddot{d}}{\hat{s}}\right)^2$**)**

Sous l'hypothèse nulle, la statistique de test *T* suit approximativement un loi du χ^2 à 1 degré de liberté

Rejeter l'hypothèse nulle (et accepter l'hypothèse alternative) au seuil α si : $T > \chi^2_{1-\alpha}[1]$

$T > 3{,}841$ pour $\alpha = 5\%$

$T > 6{,}635$ pour $\alpha = 1\%$

NOUVEAU TEST UNILATÉRAL

Problème : Comparer les proportions (inconnues) p_a et p_b de succès dans deux échantillons

Échantillons : n_a exécutions de la méthode A, a succès,

n_b exécutions de la méthode B, b succès, sans être restrictif, on suppose $a/n_a > b/n_b$

Hypothèse nulle : $p_a \leq p_b$

Hypothèse alternative : $p_a > p_b$

Méthode :

La probabilité $S(a, n_a, b, n_b, p_a, p_b)$ d'observer :

a succès ou plus pour la méthode A

b succès ou moins pour la méthode B

est donnée par $S(a, n_a, b, n_b, p_a, p_b) = \displaystyle\sum_{i=a}^{n_a} \sum_{j=0}^{b} \binom{n_a}{i} \cdot p_a^i \cdot (1-p_a)^{n_a - i} \cdot \binom{n_b}{j} \cdot p_b^j \cdot (1-p_b)^{n_b - j}$

Rejeter l'hypothèse nulle au seuil α si :

le maximum de $S(a, n_a, b, n_b, p_a, p_b) < \alpha$**, sous contrainte** $p_a \leq p_b$

MAXIMUM DE $S(A, N_A, B, N_B, P_A, P_B)$ AVEC $P_A \le P_B$

1) Le maximum aura lieu pour

$p_a = p_b = p$ car on a supposé $a/n_a > b/n_b$ (sinon, on ne peut refuser l'hypothèse nulle $p_a \le p_b$)

2) Le calcul algébrique du maximum n'est que rarement possible

Poser $\hat{p} = \dfrac{a+b}{n_a + n_b}$, effectuer quelques pas de la méthode de Newton sur la fonction $S(\ldots, p, p)$

3) Calcul de *S* malcommode

Tabulation de valeurs de a et b pour $n_a \le 15$ et $n_b \le 13$ respectant le seuil $\alpha = 5\%$ dans les actes

Le lecteur d'un article n'a qu'à compter n_a, a, n_b, b et consulter la table (pas de calculs !)

Le calcul de S est disponible sur le web http://ina.eivd.ch/projects/stamp

Les codes sources sont disponibles

EXEMPLES DE CALCULS ALGÉBRIQUES DE S

1) Tous les n_a exécutions de A sont des succès, tous les n_b exécutions de B sont des échecs.

$$S(n_a, n_a, 0, n_b, p, p) = 1 \cdot p^{n_a} \cdot (1-p)^0 \cdot 1 \cdot p^0 \cdot (1-p)^{n_b} = p^{n_a} \cdot (1-p)^{n_b}$$

Maximum de S sur p : $\dfrac{dS}{dp} = 0 = n_a \cdot p^{n_a - 1} \cdot (1-p)^{n_b} - n_b \cdot p^{n_a} \cdot (1-p)^{n_b - 1} \Rightarrow p = \dfrac{n_a}{n_a + n_b} = \dfrac{a+b}{n_a + n_b}$

Pour $n_a = 3$, $n_b = 2$, max $S(n_a, n_a, 0, n_b, p, p) = 108/3125 < 5\%$

On peut supposer, au seuil de confiance de 95% qu'un taux de succès de 3/3 soit significativement plus élevé qu'un taux de succès de 0/2.

2) 3 succès de *A* sur 4 exécutions, 0 succès de *B* sur 3 exécutions

a) Estimateur groupé $\hat{p} = \dfrac{3}{7}$, $S(3, 4, 0, 3, 3/7, 3/7) < 4\%$

Au seuil de confiance de 96% on peut supposer que A soit meilleur que B

b) $S\left(3, 4, 0, 3, \dfrac{6-2\sqrt{2}}{7}, \dfrac{6-2\sqrt{2}}{7}\right) > 4\%$ On ne peut pas supposer A meilleur que B (4%)

Nouveau test non paramétrique

Problème : Comparer les proportions p_a et p_b de succès dans deux échantillons

Échantillons : n_a exécutions de la méthode A, a succès,

n_b exécutions de la méthode B, b succès

Hypothèse nulle : $p_a = p_b = p$

Hypothèse alternative : $p_a \neq p_b$

Calcul direct :

La probabilité P d'observer a/n_a et b/n_b succès sous l'hypothèse nulle est donnée par :

$$P = \frac{\binom{n_a}{a} \cdot p^a \cdot (1-p)^{n_a - a} \cdot \binom{n_b}{b} \cdot p^b \cdot (1-p)^{n_b - b}}{\binom{n_a + n_b}{a + b} \cdot p^{a+b} \cdot (1-p)^{n_a + n_b - a - b}} = \frac{\binom{n_a}{a} \cdot \binom{n_b}{b}}{\binom{n_a + n_b}{a + b}}$$

Rejeter l'hypothèse nulle au seuil α si $P < \alpha$

Test de McNemar

Variante du test de signes

Comparaison de données appariées (typiquement : avant-après) :

Les méthodes A et B sont exécutées sur les **mêmes** exemples de problèmes.

Les égalités (succès, succès) ou (échec, échec) sont éliminées.

a succès pour la méthode A, b succès pour B sur $a + b$ exemples de problèmes

Hypothèse nulle : $P(X_a = 0, X_b = 1) = \mathrm{P}(X_a = 1, X_b = 0) = 0{,}5$

Hypothèse alternative : $P(X_a = 0, X_b = 1) \neq \mathrm{P}(X_a = 1, X_b = 0)$ (Test bilatéral)

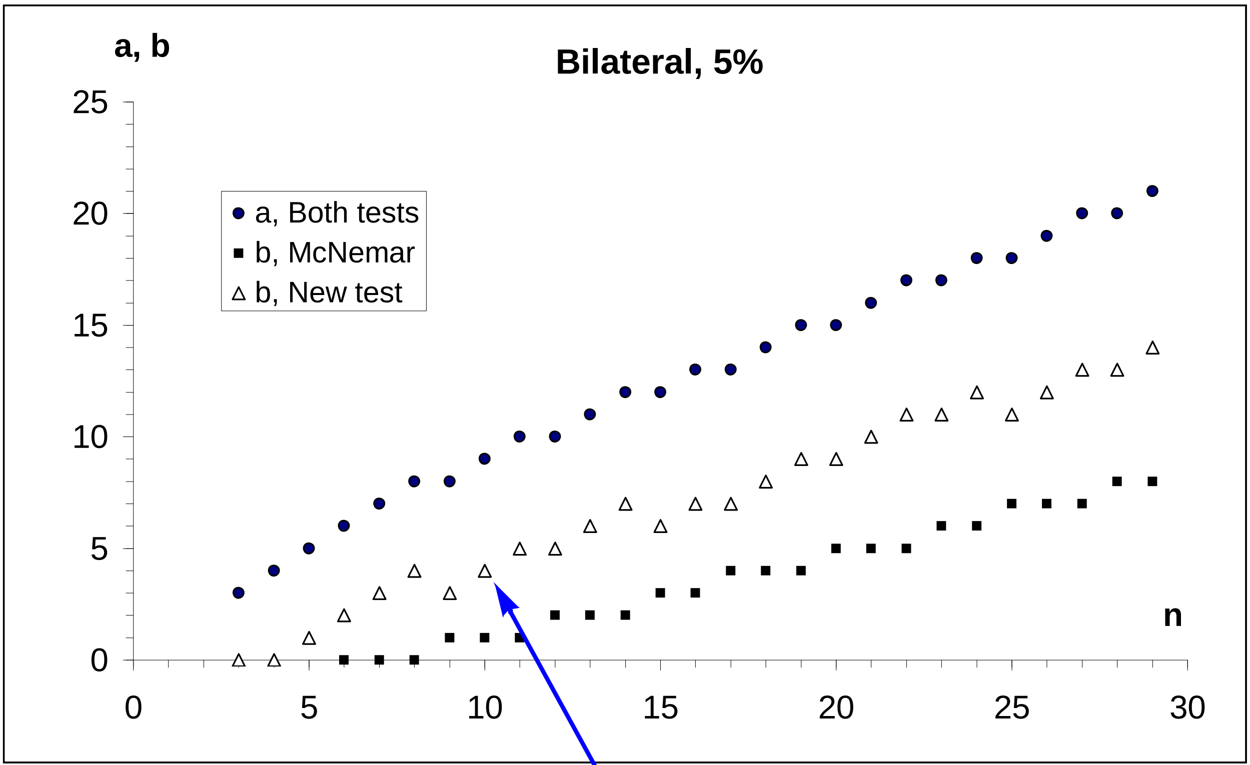
Calcul direct :

La probabilité P d'observer b succès ou moins sur $a + b$ observations sous l'hypothèse nulle est :

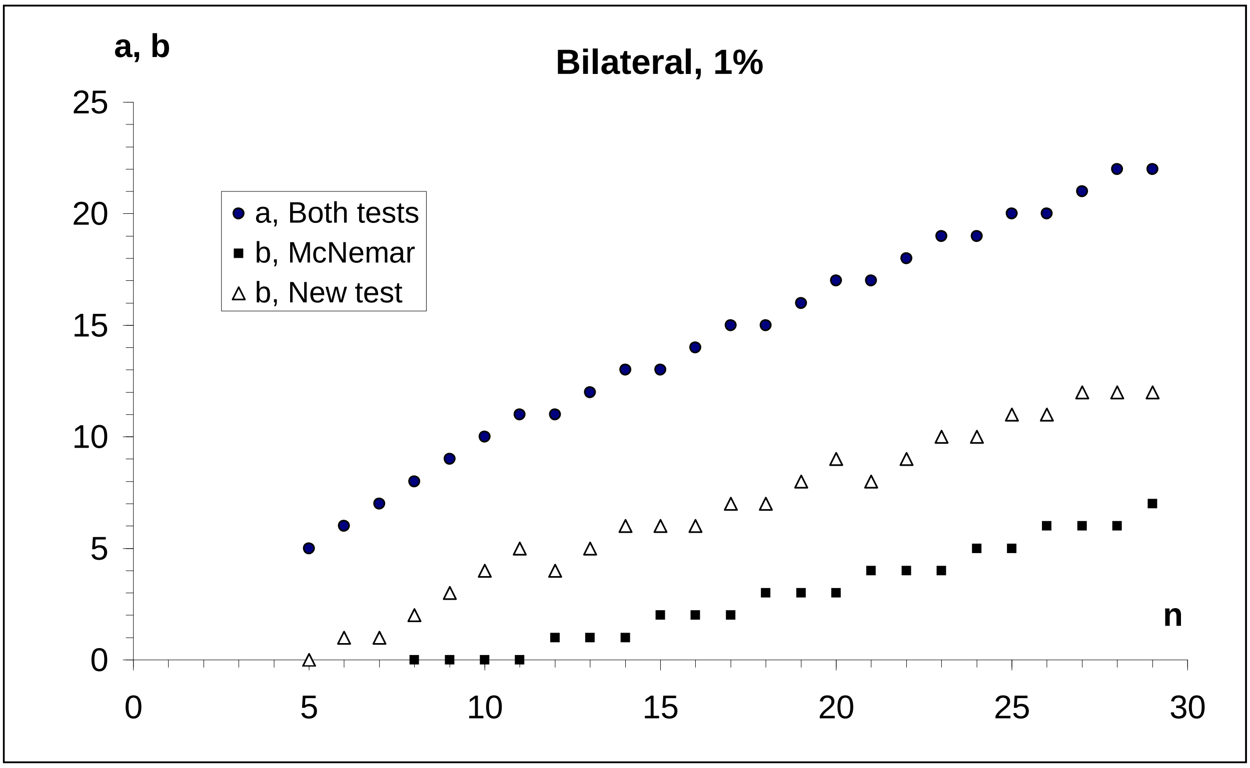
$$P = \sum_{i=0}^{b} \binom{a+b}{i} \cdot 0.5^i \cdot (1-0.5)^{a+b-i}$$

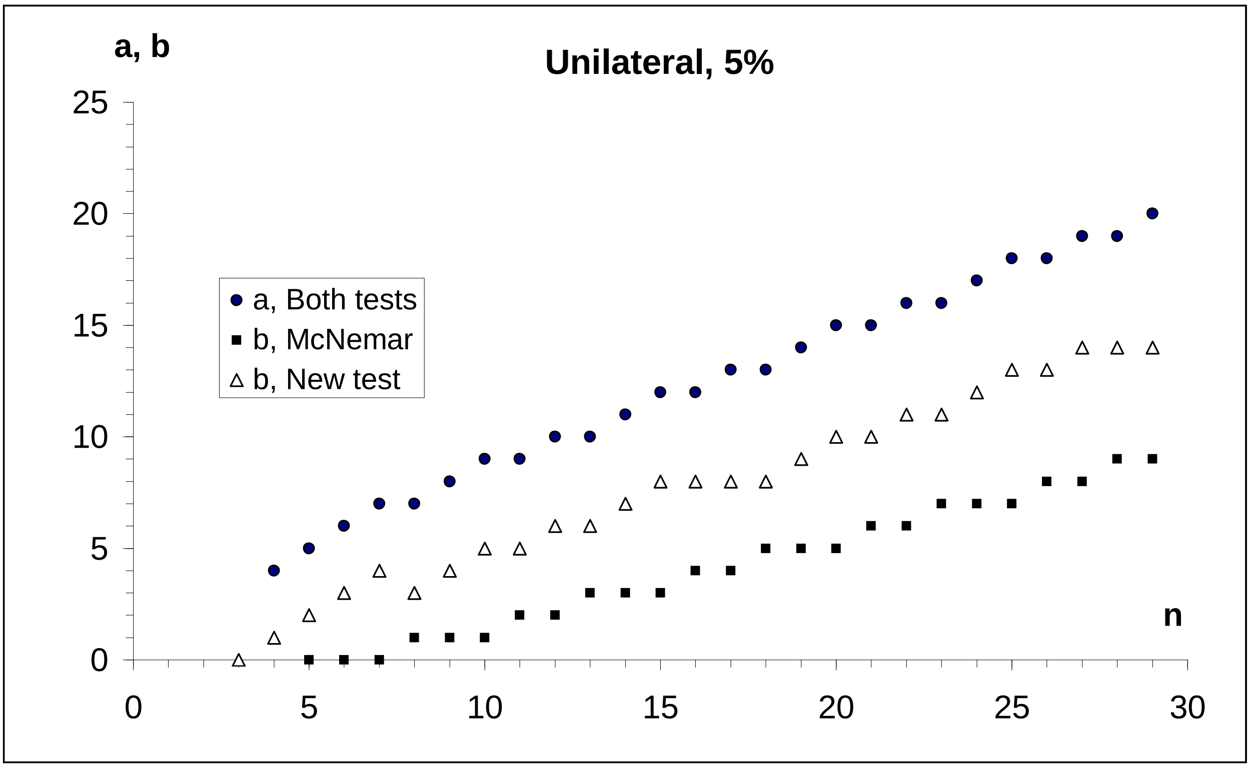
Rejeter l'hypothèse nulle au seuil α si $2P < \alpha$ ou si $2P > 1 - \alpha$

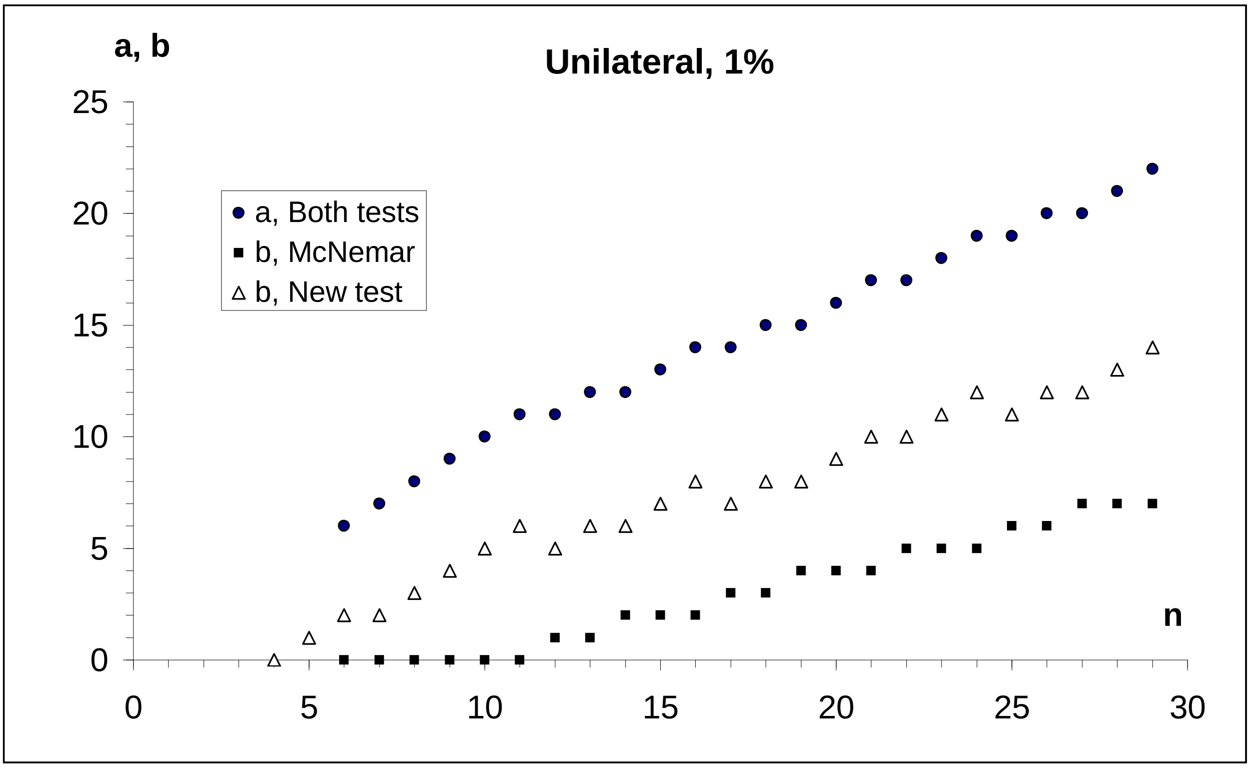
COMPARAISON MCNEMAR, NOUVEAU TEST



9/10 significativement ≠ 4/10




a, b
Unilateral, 5%
a, Both tests
b, McNemar
b, New test
n
0
5
10
15
20
25
30


a, b
Unilateral, 1%
a, Both tests
b, McNemar
b, New test
n
0
5
10
15
20
25
30

CONCLUSIONS

Un nouveau test non paramétrique puissant a été développé

Permet de comparer des taux de succès de deux méthodes lorsque

Les méthodes sont non déterministes (un ou plusieurs exemples de problèmes)

Les méthodes sont déterministes, les problèmes choisis aléatoirement

Peut être appliqué à de très petits échantillons, pas nécessairement de la même taille

Permet de réduire l'effort de calcul pour l'ajustement de paramètres

Disponible en ligne.

Nettement plus puissant que le test de McNemar