Theorem 5. *The Hamming distance for χ^2-statistics based on a 3×2 contingency table in GWAS is*

$$\begin{aligned}& d\left(T_{3\times 2}(p,q,r,s,t,u), T_{3\times 2}(p',q',r',s',t',u')\right) \\ = \;& \frac{|p-p'|+|q-q'|+|r-r'|+|s-s'|+|t-t'|+|u-u'|}{2},\end{aligned}$$

where $T_{3\times 2}(p,q,r,s,t,u)$ represents the following table data:

		Disease Status		Total
		0	1	
Genotype	0	p	q	$p+q$
	1	r	s	$r+s$
	2	t	u	$t+u$
Total		$p+r+t$	$q+s+u$	N

Proof. We consider the case where $(p>p')\wedge(q\le q')\wedge(r\le r')\wedge(s\le s')\wedge(t\le t')\wedge(u\le u')$. Since $p+q+r+s+t+u=p'+q'+r'+s'+t'+u'=N$,

$$p-p'=(q'-q)+(r'-r)+(s'-s)+(t'-t)+(u'-u).$$

Thus, when we move $p-p'$ elements out of p to $q,r,\ldots,u$ by $(q'-q),(r'-r),\ldots,(u'-u)$, respectively, $T_{3\times 2}(p,q,r,s,t,u)$ changes to $T_{3\times 2}(p',q',r',s',t',u')$. Therefore, the Hamming distance *dist* satisfies

$$\begin{aligned} dist &\le p-p' \\ &= \frac{|p-p'|+|q-q'|+|r-r'|+|s-s'|+|t-t'|+|u-u'|}{2}.\end{aligned}$$

The similar discussions can be made for the other cases.

Here, when one element in $T_{3\times 2}(p,q,r,s,t,u)$ moves and the table changes to $T_{3\times 2}(\tilde{p},\tilde{q},\tilde{r},\tilde{s},\tilde{t},\tilde{u})$, the following inequality holds:

$$|p-\tilde{p}|+|q-\tilde{q}|+\cdots+|u-\tilde{u}|\le 2.$$

Therefore, even when we move

$$k<\frac{|p-p'|+|q-q'|+|r-r'|+|s-s'|+|t-t'|+|u-u'|}{2}$$

elements and obtain $T_{3\times 2}(p'',q'',r'',s'',t'',u'')$, the table $T_{3\times 2}(p',q',r',s',t',u')$ never appears because

$$\begin{aligned} & |p-p''|+|q-q''|+\cdots+|u-u''| \\ \le\;& 2k<|p-p'|+|q-q'|+\cdots+|u-u'|.\end{aligned}$$

Consequently, we can show that

$$dist=\frac{|p-p'|+|q-q'|+|r-r'|+|s-s'|+|t-t'|+|u-u'|}{2}.$$

□

Theorem 6. *The Hamming distance for χ^2-statistics based on a 2×2 contingency table in GWAS is*

$$\begin{aligned} & d\left(T_{2\times 2}(a,b,c,d), T_{2\times 2}(a',b',c',d')\right) \\ = & \left\lceil \frac{|a-a'|+|b-b'|+|c-c'|+|d-d'|}{4} \right\rceil, \end{aligned}$$

where $T_{2\times 2}(a,b,c,d)$ represents the following table data:

		Disease Status		Total
		0	1	
Allele	0	a	b	$a+b$
	1	c	d	$c+d$
Total		$a+c$	$b+d$	$2N$

Proof. We consider the case of $(a > a') \wedge (b \leq b') \wedge (c \leq c') \wedge (d \leq d')$. Since $a+b+c+d = a'+b'+c'+d'$, $a-a' = (b'-b)+(c'-c)+(d'-d)$. Then, when we move $a-a'$ elements out of a to b, c, and d by $(b'-b)$, $(c'-c)$, and $(d'-d)$, respectively, $T_{2\times 2}(a,b,c,d)$ changes to $T_{2\times 2}(a',b',c',d')$. Therefore, the Hamming distance *dist* satisfies

$$\begin{aligned} dist & \leq \left\lceil \frac{a-a'}{2} \right\rceil \\ & = \left\lceil \frac{|a-a'|+|b-b'|+|c-c'|+|d-d'|}{4} \right\rceil \end{aligned}$$

because a change in one individual of the dataset causes a change in two elements of the table. The similar discussions can be made for the other cases.

Here, in a similar manner to the proof of Theorem 5, we can show that when we move

$$k < \left\lceil \frac{|a-a'|+|b-b'|+|c-c'|+|d-d'|}{4} \right\rceil \tag{1}$$

elements from $T_{2\times 2}(a,b,c,d)$, the table $T_{2\times 2}(a',b',c',d')$ never appears, because $k \in \mathbb{N}_{\geq 0}$ and (7) stands the following inequality:

$$k < \frac{|a-a'|+|b-b'|+|c-c'|+|d-d'|}{4}.$$

Consequently, the Hamming distance is

$$dist = \left\lceil \frac{|a-a'|+|b-b'|+|c-c'|+|d-d'|}{4} \right\rceil.$$

□