Exact *p*-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers

by Eisinga, Heskes, Pelzer & Te Grotenhuis, *BMC Bioinformatics*, 2017

THEOREM 1: *For n mutually independent integer-valued rankings, each with equally likely rank scores ranging from 1 to k, the exact probability to obtain pairwise difference d for any two rank sums equals*

$$P(D=d;k,n)=\left\{k(k-1)\right\}^{-n}W(D=d;k,n),$$

where

$$W(D=d;k,n)=\left\{k(k-1)\right\}^{n}\sum_{h=0}^{n}\binom{n}{h}k^{-h}(1-k)^{-n}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h-1}{k(j-i)-d-h},$$

is the number of distinct ways a rank sum difference of d can arise, with d having support on $d=[-n(k-1),n(k-1)]$.

Proof: Our approach exploits the observation that the difference among a pair of rank sums is the sum of rank pair differences. Problem-solving tool is the probability generating function

$$f(t)=\sum_{r=1}^{\infty}p_r t^r,$$

where the probabilities $p_r=P(R=r)$ of the integer rank values r are the coefficients of t^r. Denote two distinct ranks as r_i and r_j, with $r_i\neq r_j$. For a single ranking (i.e., one block), the generating function of the sequence of probabilities of the random rank differences is given by the finite power series

$$f(t;k)=\sum_{r_i=1}^{k-1}\sum_{r_j=r_i+1}^{k}\left\{p_{r_i}p_{r_j}t^{(r_i-r_j)}+p_{r_i}p_{r_j}t^{(r_j-r_i)}\right\}.$$

According to the null hypothesis each ranking within a block is equally likely. Consequently, rank r_i has probability $P(R=r_i)=p_{r_i}=k^{-1}$, and rank r_j has probability $P(R=r_j)=p_{r_j}=(k-1)^{-1}$, for $1\leq r_i,r_j\leq k.$ The generating function of the probabilities may therefore be expressed as

$$f(t;k)=\sum_{r_i=1}^{k-1}\sum_{r_j=r_i+1}^{k}\left\{\frac{1}{k(k-1)}t^{(r_i-r_j)}+\frac{1}{k(k-1)}t^{(r_j-r_i)}\right\}=\left\{\frac{1}{k(k-1)}\sum_{r_i=1}^{k}\sum_{r_j=1}^{k}t^{(r_i-r_j)}\right\}-\frac{1}{k-1}.$$

This function can be put in more compact form by summing the geometric progressions:

$$f(t;k) = -\frac{1}{k(k-1)}\left\{\frac{(t-t^k)}{(1-t)}\frac{(t-t^{-k})}{(1-t)} - (k-1)\right\} = \frac{1}{k(1-k)}\frac{t(1-t^k)(1-t^{-k})}{(1-t)^2} - \frac{1}{k-1}.$$

The form of the first equation on the right demonstrates that the generating function is symmetric in t around zero, that is $f(t^{-1};k) = f(t;k)$. From the convolution theorem (e.g., Feller [49]), the probability generating function of the sum of n mutually independent rankings, each assuming the rank differences $\{-(k-1),\ldots,-1,1,\ldots,(k-1)\}$, is given by the nth power of $f(t;k)$:

$$f(t;k,n) = \left\{\frac{1}{k(1-k)}\frac{t(1-t^k)(1-t^{-k})}{(1-t)^2} - \frac{1}{k-1}\right\}^n,$$

which, according to the binomial theorem, can be rewritten as

$$f(t;k,n) = \sum_{h=0}^{n}\binom{n}{h}\left\{\frac{1}{k(1-k)}\right\}^h\left\{\frac{t(1-t^k)(1-t^{-k})}{(1-t)^2}\right\}^h\left\{\frac{1}{1-k}\right\}^{n-h}.$$

Applying the binomial expansions of $(1-t^k)^h$ and $(1-t^{-k})^h$, the power series expansion of $(1-t)^{-2h}$, and the -1 transformation and the upper negation identity transformation of the binomial coefficient, i.e.,

$$\sum_{l=0}^{\infty}\binom{-2h}{l}(-t)^l = \sum_{l=0}^{\infty}(-1)^l\binom{-2h}{l}t^l = \sum_{l=0}^{\infty}\binom{l+2h-1}{l}t^l,$$

the expression becomes

$$f(t;k,n) = \sum_{h=0}^{n}\binom{n}{h}\left\{\frac{1}{k^h(1-k)^n}\right\}t^h\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}t^{k(i-j)}\sum_{l=0}^{\infty}\binom{l+2h-1}{l}t^l.$$

To calculate the probability $P(D=d;k,n)$ of rank sum difference d as the coefficient of the power of t, we first collect all terms involving powers of t, and subsequently make the change of variable $h-k(j-i)+l=d$, yielding

$$f(t;k,n)=\sum_{h=0}^{n}\binom{n}{h}\left\{\frac{1}{k^h(1-k)^n}\right\}\sum_{i=0}^{h}\sum_{j=0}^{h}\sum_{l=0}^{\infty}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{l+2h-1}{l}t^{h+k(i-j)+l}$$

$$=\sum_{d=-n(k-1)}^{n(k-1)}\left[\sum_{h=0}^{n}\binom{n}{h}\left\{\frac{1}{k^h(1-k)^n}\right\}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)+d+h-1}{k(j-i)+d-h}\right]t^d,$$

$$=\sum_{d=-n(k-1)}^{n(k-1)}\left[\sum_{h=0}^{n}\binom{n}{h}\left\{\frac{1}{k^h(1-k)^n}\right\}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h-1}{k(j-i)-d-h}\right]t^{-d},\quad |t|<1\wedge t\neq 0,$$

where the second equation follows by noting that whenever $h-k(j-i)+l=d$ and $l\geq 0$, we have $l+2h-1=k(j-i)+d+h-1$ and $k(j-i)+d-h\geq 0$, and the third equation follows from the property of symmetry of the generating function. The probability of rank sum difference d is therefore

$$P(D=d;k,n)=\{k(k-1)\}^{-n}W(D=d;k,n),$$

where

$$W(D=d;k,n)=\{k(k-1)\}^{n}\sum_{h=0}^{n}\binom{n}{h}k^{-h}(1-k)^{-n}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h-1}{k(j-i)-d-h},$$
$$d=-n(k-1),\ldots,n(k-1),$$

represents the number of composition of d into n parts, where each part is restricted to the interval $[-(k-1),k-1]$. Hence $W(D=d;k,n)$ gives the number of different ways a rank sum difference of d can arise, for n mutually independent rankings, each with equally likely integer-valued ranks $\{1,2,\ldots,k\}$. This completes the proof.

The cumulative distribution of rank sum difference d has generating function $f(t;k,n)/(1-t)$ [see, 50-52], and by repeated application of Pascal's identity

$$\binom{m}{r}+\binom{m}{r-1}=\binom{m+1}{r},\qquad \text{for } 1\leq r\leq m+1,$$

the expression for the exact *p*-value is obtained as

$$P(D\geq d;k,n)=\sum_{h=0}^{n}\binom{n}{h}k^{-h}(1-k)^{-n}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h}{k(j-i)-d-h},\quad d=-n(k-1),\ldots,n(k-1).$$

To illustrate the derivations presented above, Additional file 4 offers a small-sized numerical example (k=3, n=2). Additional file 5 tabulates the number of compositions of d for combinations of n=k=2,…,6, for inclusion in the OEIS [53].

Exact *p*-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers

by Eisinga, Heskes, Pelzer & Te Grotenhuis, *BMC Bioinformatics*, 2017

THEOREM 2: *For nonnegative integers d and k*

$$\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h}{k(j-i)-d-h}=\sum_{s=0}^{h}(-1)^{s}\binom{2h}{h+s}\binom{ks-d+h}{ks-d-h}.$$

Proof: Note that, for nonnegative integers *d* and *k,* the right-most binomial coefficient on the left-hand side equals 0, whenever $j<i.$ We set $j-i=s,$ and rewrite the left side as follows

$$\sum_{s=0}^{j}\sum_{j=0}^{h}(-1)^{s}\binom{h}{j-s}\binom{h}{j}\binom{ks-d+h}{ks-d-h}=\sum_{s=0}^{h}(-1)^{s}\binom{ks-d+h}{ks-d-h}\sum_{j=0}^{h}\binom{h}{j}\binom{h}{j-s}.$$

The upper bound of the outer summation over *s* in the first equation is *j*. However, the same summation result is obtained if the upper bound is *h,* as in the second equation, since

$$\binom{h}{j-s}$$

equals 0, whenever $s>j.$ The second equation additionally takes both the alternating sign term and the right-most binomial coefficient of the first equation out of the summation over *j,* as they are conditioned by *s*.

We subsequently simplify the summation

$$\sum_{j=0}^{h}\binom{h}{j}\binom{h}{j-s}$$

to a closed-form expression, using a special case of Vandermonde's convolution identity, stating

$$\sum_{j=0}^{r}\binom{h}{j}\binom{h}{r-j}=\binom{2h}{r}.$$

We find that

$$\sum_{j=0}^{h}\binom{h}{j}\binom{h}{j-s}=\sum_{j=0}^{h}\binom{h}{j}\binom{h}{h+s-j}=\sum_{j=0}^{h+s}\binom{h}{j}\binom{h}{h+s-j}=\binom{2h}{h+s},$$

where the third equation follows by noting that the binomial coefficient with lower index *j* equals 0, whenever $j > h.$ It is therefore immaterial to the result whether the upper bound of the summation is h or $h+s.$ Inserting the closed-form expression and rearranging yields

$$\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(j-i)}\binom{h}{i}\binom{h}{j}\binom{k(j-i)-d+h}{k(j-i)-d-h}=\sum_{s=0}^{h}(-1)^{s}\binom{2h}{h+s}\binom{ks-d+h}{ks-d-h},$$

as claimed. This completes the proof.

We additionally note that an elegant alternative expression for the summation on the right hand side of Theorem 2 is

$$\sum_{s=0}^{h}(-1)^{s}\binom{2h}{h+s}\binom{ks-d+h}{ks-d-h}=\sum_{s=0}^{h}(-1)^{s}\binom{2h}{h+s}\binom{ks-d+h}{2h},$$

and that the integer sequence

$$(-1)^{s}\binom{2h}{h+s}$$

is a signed version of the elements in the even-numbered rows of Pascal's triangle (see Sloane's [52] series A062344 and A034870). For $k = 2$, the sequence

$$\sum_{s=0}^{h}(-1)^{s}\binom{2h}{h+s}\binom{ks-d+h}{2h}$$

is a signed version of the OEIS [52] triangle A008949.

Exact *p*-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers

by Eisinga, Heskes, Pelzer & Te Grotenhuis, *BMC Bioinformatics*, 2017

The following is a small numerical example to illustrate the calculation. Suppose we want to obtain the rank sum difference distribution for $k = 3$ groups and $n = 2$ blocks. Assuming no ties, the ranks are $r = 1, 2, 3,$ and the rank differences are $-2, -1, 1, 2$. The permutation distribution assigns equal probabilities to the three ranks, i.e., $p_r = 1/3,$ for $r = 1, 2, 3,$ and equal probabilities to the four pairwise rank differences, i.e., $p_{r_i} p_{r_j} = 1/3 \times 1/2 = 1/6,$ for $i, j = 1, 2, 3,$ with $i \neq j$. The outcomes are not equally likely however. There are two different ways to obtain rank sum difference value 1, namely 2-1 and 3-2, but there is only one way to obtain a rank sum difference of 2, namely 3-1.

The generating function of the rank sum differences for a single block generates the probabilities as

$$f(t;3) = \frac{1}{6}t^{-2} + \frac{2}{6}t^{-1} + \frac{2}{6}t^{1} + \frac{1}{6}t^{2} = \frac{1}{6}\left(t^{-2} + 2t^{-1} + 2t^{1} + t^{2}\right).$$

The coefficients of the powers of *t* in the first equation on the right are the probabilities assigned to the rank sum differences, indicated by the powers -2,-1,1,2 of *t*. The coefficients of the powers of *t* within brackets in the second equation are the number of different ways to obtain the rank sum difference in question. For example, rank sum difference -1 has 2 configurations (1-2 and 2-3) and the associated probability is 2/6. The probability generating function of the sum of the rank differences for two blocks, assuming independence of the blocks, is the product of the two generation functions, i.e.,

$$f(t;3,2) = \left(\frac{1}{6}t^{-2} + \frac{2}{6}t^{-1} + \frac{2}{6}t^{1} + \frac{1}{6}t^{2}\right)^{2},$$

which is written compactly as

$$f(t;3,2) = \left\{\frac{1}{6}\frac{t(1-t^{3})(1-t^{-3})}{(1-t)^{2}} - \frac{1}{2}\right\}^{2}.$$

Expanding and collecting terms yields the generating function

$$f(t;3,2)=\sum_{d=-4}^{4}\left[\sum_{h=0}^{2}\binom{2}{h}\frac{1}{3^h(-2)^2}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(i+j)}\binom{h}{i}\binom{h}{j}\binom{3(j-i)-d+h-1}{3(j-i)-d-h}\right]t^{-d}$$

$$=\sum_{d=-4}^{4}\left[\left\{\frac{1}{6}\right\}^2 W(D=d;3,2)\right]t^{-d}$$

$$=\sum_{d=-4}^{4}\left[(P=d;3,2)\right]t^{-d},$$

where we can read the probability $(P=d;3,2)$ as the coefficient of t^{-d}. The probability mass distribution is obtained as

$$P(D=d;3,2)=\sum_{h=1}^{2}\binom{2}{h}(-3)^{(2-h)}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(i+j)}\binom{h}{i}\binom{h}{j}\binom{3(j-i)-d+h-1}{3(j-i)-d-h},$$

and the cumulative distribution as

$$P(D\geq d;3,2)=\sum_{h=0}^{2}\binom{2}{h}\frac{1}{3^h(-2)^2}\sum_{i=0}^{h}\sum_{j=0}^{h}(-1)^{(i+j)}\binom{h}{i}\binom{h}{j}\binom{3(j-i)-d+h}{3(j-i)-d-h} \qquad \text{for } d=-4,\ldots,4.$$

As explained in the main text, the *p*-value of non-negative *d* may also be obtained from the simplified expression

$$P(D\geq |d|;3,2)=\begin{cases}2\sum_{h=0}^{2}\binom{2}{h}\frac{1}{3^h(-2)^2}\sum_{s=0}^{h}(-1)^s\binom{2h}{h+s}\binom{3s-d+h}{3s-d-h} & \text{for } d=1,\ldots,4\\ 1 & \text{for } d=0.\end{cases}$$

Table S1 summarizes the results for $k=3$ groups and $n=2$ blocks. In this setting, there are a total of 9 possible (positive and negative) values for the differences in rank sums.

Table S1. Rank sum difference distribution for $k = 3$ and $n = 2$

d	-4	-3	-2	-1	0	1	2	3	4
$W(D = d;3,2)$	1	4	4	4	10	4	4	4	1
$W(D \geq d;3,2)$	36	35	31	27	23	13	9	5	1
$P(D = d;3,2)$	1/36	4/36	4/36	4/36	10/36	4/36	4/36	4/36	1/36
$P(D \geq d;3,2)$	1	35/36	31/36	27/36	23/36	13/36	9/36	5/36	1/36
$W(D = \|d\|;3,2)$					10	8	8	8	2
$W(D \geq \|d\|;3,2)$					36	26	18	10	2
$P(D = \|d\|;3,2)$					10/36	8/36	8/36	8/36	2/36
$P(D \geq \|d\|;3,2)$					1	.722	.500	.278	.056

In practical settings we are testing absolute differences and, therefore, we use only the right tail of the symmetric distribution. It is immediately clear from the figures in the bottom line of the table that an absolute rank sum difference of 4 would be significant at the (unadjusted) $.10$ level, as $P(D \geq |4|;3,2) < .10,$ but not at an alpha level of .05. The R code to compute the results presented in the table is included in Additional file 3. The R script `pexactfrsd` computes the exact *p*-value presented in the bottom row of Table S1 and, optionally, all statistics presented in the bottom panel of the table.

Exact *p*-values for pairwise comparison of Friedman rank sums, with application to comparing classifiers

by Eisinga, Heskes, Pelzer & Te Grotenhuis, *BMC Bioinformatics*, 2017

The entries in the tables below are the number of compositions of *d*, referred to in the main text as $W(D=d;k,n)$, for $k,n=2,\ldots,6$. The tabulated figures are for the right half of the distribution (i.e., non-negative *d* values) only. The entries for positive *d* (i.e., $d>0$ only) should be doubled if the number of compositions of the absolute *d* values are required.

k=2

			n		
d	2	3	4	5	6
0	2	0	6	0	20
1	0	3	0	10	0
2	1	0	4	0	15
3		1	0	5	0
4			1	0	6
5				1	0
6					1

k=3

			n		
d	2	3	4	5	6
0	10	24	198	880	5380
1	4	36	152	940	4920
2	4	27	136	810	4440
3	4	14	120	600	3832
4	1	12	68	480	2799
5		6	40	312	1980
6		1	24	165	1324
7			8	90	732
8			1	40	366
9				10	172
10				1	60
11					12
12					1

k=4

			n		
d	2	3	4	5	6
0	28	180	2268	23200	260500
1	16	210	2064	23075	252960
2	15	180	1896	21230	236718
3	12	144	1616	18565	211336
4	10	96	1327	15160	180864
5	4	72	936	12003	146040
6	1	44	628	8810	112644
7		21	392	6005	82488
8		6	226	3760	57192
9		1	104	2215	36936
10			36	1182	22260
11			8	545	12456
12			1	200	6400
13				55	2904
14				10	1113
15				1	340
16					78
17					12
18					1

k=5

			n		
d	2	3	4	5	6
0	60	726	14020	246820	4555950
1	40	786	13360	244100	4479924
2	38	711	12580	231050	4282038
3	32	616	11280	211200	3967972
4	25	501	9670	186025	3564561
5	20	366	8000	157194	3105240
6	10	276	6120	128665	2613194
7	4	186	4480	100720	2128368
8	1	111	3105	75380	1674237
9		56	2024	53760	1269836
10		21	1232	36353	926880
11		6	664	23370	648084
12		1	310	14110	433288
13			120	7890	276012
14			36	4005	166716
15			8	1792	94820
16			1	690	50139
17				220	24288
18				55	10582
19				10	4056
20				1	1335
21					364
22					78
23					12
24					1

k=6

	n				
d	2	3	4	5	6
0	110	2136	59634	1592100	43909940
1	80	2241	57904	1576230	43409280
2	77	2088	55388	1515700	42016245
3	68	1883	51216	1420855	39780200
4	56	1632	45821	1297540	36836370
5	44	1347	39688	1153375	33353784
6	35	1036	33384	996220	29526143
7	20	801	26744	838750	25519272
8	10	576	20704	684960	21539496
9	4	384	15408	542250	17737880
10	1	234	10976	415524	14238543
11		126	7456	307525	11129856
12		56	4803	219220	8461490
13		21	2864	150445	6243984
14		6	1564	98860	4465953
15		1	768	61828	3089816
16			330	36520	2062644
17			120	20160	1324488
18			36	10260	814832
19			8	4755	477516
20			1	1972	264804
21				715	137852
22				220	66726
23				55	29700
24				10	12004
25				1	4332
26					1365
27					364
28					78
29					12
30					1