

Supplementary material for A novel high-power test for continuous outcomes truncated by death

Andreas Kryger Jensen and Theis Lange
Section of Biostatistics, Department of Public Health, University of Copenhagen

05 February, 2023

Simulation results

Table 1: Power simulation - scenarios 1 and 2

	Scenario 1				Scenario 2			
	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT
50	6.29	7.80	25.00	26.58	14.85	15.61	14.41	14.54
75	6.86	9.31	34.59	35.33	19.89	20.59	18.95	19.04
100	7.69	11.12	44.42	44.79	24.89	25.73	23.21	23.22
125	8.36	12.83	53.87	54.08	30.00	31.27	28.04	28.05
150	9.44	14.75	62.18	62.27	34.44	35.88	32.95	33.00
175	9.67	16.08	69.50	69.69	39.32	40.92	37.19	37.20
200	10.84	18.35	75.54	75.57	43.95	45.43	42.09	42.11
225	11.87	19.63	80.26	80.33	48.46	49.88	46.87	46.88
250	12.30	21.25	84.75	84.80	52.62	54.09	50.93	50.95
275	13.58	23.54	88.08	87.94	56.39	58.00	54.56	54.57
300	13.20	24.69	90.61	90.63	60.75	62.39	59.52	59.49
325	14.59	26.72	93.03	92.99	63.52	65.27	62.30	62.31
350	15.05	27.57	94.39	94.36	67.65	69.38	66.82	66.78

Table 2: Power simulation - scenarios 3 and 4

	Scenario 3				Scenario 4			
	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT
50	10.81	7.96	35.49	36.86	10.22	25.92	67.83	66.49
75	13.09	9.07	49.56	50.17	12.99	36.36	85.00	83.98
100	15.89	10.28	62.05	62.28	15.58	46.55	93.58	93.16
125	18.55	11.19	72.90	72.96	18.30	55.95	97.39	97.22
150	21.66	13.08	80.62	80.68	21.25	64.08	98.89	98.82
175	24.24	14.24	86.44	86.42	24.30	70.44	99.66	99.64
200	27.34	15.40	90.92	90.98	27.57	76.91	99.83	99.82
225	29.29	15.95	93.79	93.75	29.48	80.86	99.97	99.97
250	32.26	17.59	95.90	95.90	32.09	85.62	99.98	99.98
275	34.95	18.92	97.18	97.18	35.69	88.84	100.00	100.00
300	37.46	19.68	98.17	98.15	37.69	90.89	100.00	100.00
325	40.34	21.55	98.97	98.97	40.36	92.81	100.00	100.00
350	42.88	22.48	99.36	99.36	43.28	94.73	100.00	100.00

Table 3: Type I error simulation - scenario 1

	$\pi = 0.2$				$\pi = 0.4$				$\pi = 0.6$				$\pi = 0.8$			
	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT
100	5.00	5.16	5.70	6.11	4.80	4.86	5.18	5.24	5.03	5.08	5.37	5.40	5.16	5.14	5.44	5.39
200	5.14	5.17	5.49	5.51	5.05	5.14	5.36	5.35	4.91	4.84	5.22	5.22	4.66	4.90	4.93	4.96
300	5.08	4.92	5.45	5.50	4.93	4.98	4.97	4.98	5.07	5.03	5.06	5.07	5.05	5.05	5.07	5.08
400	5.00	4.98	5.12	5.10	4.79	4.81	5.00	4.98	5.07	4.90	4.92	4.90	4.93	4.95	4.92	4.92
500	4.85	4.88	5.00	4.99	5.08	5.02	5.07	5.08	5.02	4.85	4.98	4.97	5.04	5.10	4.93	4.91
600	4.79	4.89	4.79	4.77	5.12	5.07	5.27	5.27	5.40	5.29	5.22	5.24	5.10	5.16	4.72	4.74
700	5.10	4.86	5.04	5.08	4.84	4.78	5.00	5.01	5.12	5.18	5.13	5.14	5.18	5.15	5.18	5.17
800	5.24	5.07	5.01	4.99	4.98	4.93	5.15	5.16	5.12	5.04	5.11	5.12	4.89	4.92	4.76	4.77
900	4.81	4.98	4.92	4.92	5.34	5.26	5.11	5.10	5.06	4.92	5.12	5.12	4.95	5.15	5.14	5.15
1000	4.88	5.06	5.00	5.01	4.98	4.92	5.03	5.01	5.00	5.14	5.04	5.04	5.00	5.10	5.23	5.23

Table 4: Type I error simulation - scenario 2

	$\pi = 0.2$				$\pi = 0.4$				$\pi = 0.6$				$\pi = 0.8$			
	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT	Wilcoxon	t-test	LRT	SPLRT
100	5.24	5.32	5.76	5.73	4.95	4.82	5.26	5.18	4.98	5.12	5.25	5.19	5.01	4.95	5.24	5.20
200	5.02	4.92	5.34	5.27	4.88	4.98	5.10	5.08	5.03	4.99	5.03	5.02	4.90	5.08	4.91	4.90
300	5.02	5.01	5.07	5.05	4.98	4.88	4.96	4.94	4.92	4.96	4.84	4.82	4.86	4.93	4.98	4.96
400	4.88	4.96	4.98	4.95	4.83	4.67	4.98	4.95	5.26	5.23	5.35	5.32	5.01	5.09	5.23	5.22
500	4.72	4.88	4.86	4.84	5.14	5.20	5.27	5.26	4.70	4.86	4.90	4.90	4.91	4.91	4.88	4.87
600	5.00	5.18	5.18	5.20	4.94	4.85	5.13	5.09	5.03	4.99	4.96	4.94	4.82	4.84	5.14	5.12
700	5.13	5.13	5.00	4.97	4.93	5.04	5.12	5.12	4.82	4.90	4.84	4.83	5.06	5.10	5.12	5.11
800	5.12	4.92	5.00	5.00	4.99	5.05	5.12	5.10	4.92	4.92	5.09	5.09	5.33	5.24	5.27	5.26
900	5.34	5.08	5.03	5.02	4.92	4.84	5.07	5.05	4.64	4.61	5.08	5.07	5.04	5.00	4.94	4.94
1000	4.96	4.89	4.88	4.86	5.07	5.17	5.22	5.21	4.96	4.76	4.96	4.96	5.00	5.05	4.93	4.93

Derivation of semi-parametric test statistic

The empirical likelihood ratio function that compares the empirical maximum likelihood under a constraint set \mathcal{C} to an unconstrained maximum likelihood is given by

$$R_{1,n}^E(\mathcal{C}) = \frac{\sup_{\{p_i\}} \left(\prod_{i \in \mathcal{I}_0} p_i \mid \mathcal{C} \right) \sup_{\{q_j\}} \left(\prod_{j \in \mathcal{I}_1} q_j \mid \mathcal{C} \right)}{\sup_{F \in \mathcal{F}} \prod_{i \in \mathcal{I}_0} F(\{Y_i\}) \sup_{G \in \mathcal{F}} \prod_{j \in \mathcal{I}_1} G(\{Y_j\})} \quad (1)$$

where \mathcal{F} is the family of cadlag functions. This empirical likelihood ratio is simply a comparison between the non-parametric unconstrained maximum likelihood values and a null-model where the maximum likelihoods in the two treatment groups are constrained according to a constraint set \mathcal{C} . The sets of weights, $\{p_i\}$ and $\{q_j\}$, form a constrained multinomial distribution over the observations.

It is well-known that the solutions to the two unconstrained maximizations in the denominator are given by the empirical distribution functions that put equal weight on each observation. From here it follows that the likelihood ratio function can be written as

$$R_{1,n}^E(\mathcal{C}) = \sup_{\{p_i\}, \{q_j\}} \left(\prod_{i \in \mathcal{I}_0} |\mathcal{I}_0| p_i \prod_{j \in \mathcal{I}_1} |\mathcal{I}_1| q_j \mid \mathcal{C} \right) \quad (2)$$

where $|\cdot|$ denotes the cardinality of the index set.

The constraint set is chosen so that $\{p_i\}$ and $\{q_j\}$ are bona fide multinomial distributions, and so that the expected value of the observations with indices in \mathcal{I}_0 have expectation β and the difference between the expectations comparing the observations with indices in \mathcal{I}_1 to those with indices in \mathcal{I}_0 is given by β_δ similar to the structure of the linear model in Equation (??). This yields the following set of constraints:

$$p_i \geq 0, \quad \sum_{i \in \mathcal{I}_0} p_i = 1, \quad \sum_{i \in \mathcal{I}_0} p_i (Y_i - \beta) = 0 \quad (3)$$

$$q_j \geq 0, \quad \sum_{j \in \mathcal{I}_1} q_j = 1, \quad \sum_{j \in \mathcal{I}_1} q_j (Y_j - \beta - \beta_\delta) = 0 \quad (4)$$

To find the values of $\{p_i\}$ and $\{q_j\}$ in Equation (2) that satisfies the constraint set we solve the constrained optimization problem through the following objective function

$$\begin{aligned} O(\{p_i\}, \{q_j\}) = & \sum_{i \in \mathcal{I}_0} \log(|\mathcal{I}_0| p_i) + \sum_{j \in \mathcal{I}_1} \log(|\mathcal{I}_1| q_j) + \gamma_1 \left(\sum_{i \in \mathcal{I}_0} p_i - 1 \right) + \gamma_2 \left(\sum_{j \in \mathcal{I}_1} q_j - 1 \right) \\ & - |\mathcal{I}_0| \lambda_1 \sum_{i \in \mathcal{I}_0} p_i (Y_i - \beta) - |\mathcal{I}_1| \lambda_2 \sum_{j \in \mathcal{I}_1} q_j (Y_j - \beta - \beta_\delta) \end{aligned} \quad (5)$$

where γ_1 , γ_2 , λ_1 and λ_2 are Lagrange multipliers.

Calculating the partial derivatives of $O(\{p_i\}, \{q_j\})$ with respect to p_i and q_j and setting them equal to zero we have that

$$0 = \frac{\partial}{\partial p_i} O(\{p_i\}, \{q_j\}) = \frac{1}{p_i} - |\mathcal{I}_0| \lambda_1 (Y_i - \beta) + \gamma_1 \quad (6)$$

$$0 = \frac{\partial}{\partial q_j} O(\{p_i\}, \{q_j\}) = \frac{1}{q_j} - |\mathcal{I}_1| \lambda_2 (Y_j - \beta - \beta_\delta) + \gamma_2 \quad (7)$$

and by applying the constraints it follows that

$$\begin{aligned} 0 = & \sum_{i \in \mathcal{I}_0} p_i \frac{\partial O(\{p_i\}, \{q_j\})}{\partial p_i} \\ = & \sum_{i \in \mathcal{I}_0} p_i \left(\frac{1}{p_i} - |\mathcal{I}_0| \lambda_1 (Y_i - \beta) + \gamma_1 \right) \\ = & \sum_{i \in \mathcal{I}_0} 1 - |\mathcal{I}_0| \lambda_1 \sum_{i \in \mathcal{I}_0} p_i (Y_i - \beta) + \gamma_1 \sum_{i \in \mathcal{I}_0} p_i \\ = & |\mathcal{I}_0| + \gamma_1 \end{aligned} \quad (8)$$

Therefore $\gamma_1 = -|\mathcal{I}_0|$ and $\gamma_2 = -|\mathcal{I}_1|$ by similar calculation for q_j . From Equations (6) and (7) and the previous results we obtain the following solutions

$$p_i = \frac{1}{|\mathcal{I}_0|(1 + \lambda_1(Y_i - \beta))}, \quad q_j = \frac{1}{|\mathcal{I}_1|(1 + \lambda_2(Y_j - \beta - \beta_\delta))} \quad (9)$$

and by inserting these expressions into the expression for the empirical likelihood ratio we obtain

$$\begin{aligned} \log R_{1,n}^E(\beta, \beta_\delta) &= \sum_{i \in \mathcal{I}_0} \log(|\mathcal{I}_0|p_i) + \sum_{j \in \mathcal{I}_1} \log(|\mathcal{I}_1|q_j) \\ &= \sum_{i \in \mathcal{I}_0} \log \frac{1}{1 + \lambda_1(Y_i - \beta)} + \sum_{j \in \mathcal{I}_1} \log \frac{1}{1 + \lambda_2(Y_j - \beta - \beta_\delta)} \\ &= - \sum_{i \in \mathcal{I}_0} \log(1 + \lambda_1(Y_i - \beta)) - \sum_{j \in \mathcal{I}_1} \log(1 + \lambda_2(Y_j - \beta - \beta_\delta)) \end{aligned} \quad (10)$$

The values of λ_1 and λ_2 can then be found by combining Equations (6) and (7) with the constraints leading to the following set of equations

$$|\mathcal{I}_0|^{-1} \sum_{i \in \mathcal{I}_0} \frac{Y_i - \beta}{1 + \lambda_1(Y_i - \beta)} = 0, \quad |\mathcal{I}_1|^{-1} \sum_{j \in \mathcal{I}_1} \frac{Y_j - \beta - \beta_\delta}{1 + \lambda_2(Y_j - \beta - \beta_\delta)} = 0 \quad (11)$$

that in practice can be solved for λ_1 and λ_2 using a numerical root finding method.

The empirical LRT statistic is therefore

$$\begin{aligned} W_{1,n}^E(\beta, \beta_\delta) &= -2 \log R_{1,n}^E(\beta, \beta_\delta) \\ &= 2 \left(\sum_{i \in \mathcal{I}_0} \log(1 + \lambda_1(Y_i - \beta)) + \sum_{j \in \mathcal{I}_1} \log(1 + \lambda_2(Y_j - \beta - \beta_\delta)) \right) \end{aligned} \quad (12)$$

where $W_{1,n}^E(\beta_\delta^*) = \sup_{\beta} W_{1,n}^E(\beta, \beta_\delta^*)$ the profile version testing the difference between treatments.