

---

## Supplementary material

### A Bayesian nonparametric procedure for comparing algorithms

---

#### 1. Proof of Theorem 1

A probability measure sampled from the posterior DP (9) has the following form:

$$P = w_0 P_0 + \sum_{j=1}^n w_j \delta_{\mathbf{x}_j},$$

with  $P_0 \sim Dp(s, \alpha^*)$ . Let us consider  $E[R(X_m)]$  w.r.t.  $P$ , we have that

$$E[R(X_i)] = w_0 R_{m0} + \sum_{j=1}^n w_j R_{mj} = w_0 R_{m0} + \mathbf{R}_m^T \mathbf{w},$$

where  $R_{m0} = \int \sum_{k=1}^{m-1} I_{\{X_m > X_k\}}(\mathbf{X}) dP_0(\mathbf{X})$ ,  $R_{mj} = \sum_{k=1}^{m-1} I_{\{X_{mj} > X_{kj}\}} + 1$  and  $\mathbf{R}_m^T = [R_{m1}, \dots, R_{mn}]$ . Assuming  $P_0 = \alpha^* = \delta_{\mathbf{x}}$ , we have that  $R_{m0} = \int \sum_{k=1}^{m-1} I_{\{X_m > X_k\}}(\mathbf{X}) d\alpha^*(\mathbf{X})$ . By defining the vector  $\mathbf{R}_0 = [R_{1l}, \dots, R_{m0}]^T$  and the matrix  $\mathbf{R}$  whose rows are the vectors  $\mathbf{R}_m^T$ , we obtain (15).

Since the weights are Dirichlet distributed  $(w_0, \mathbf{w}^T) \sim Dir(s, 1, 1, \dots, 1)$ , it follows that

$$\mathcal{E}[E[R(X_1), \dots, R(X_m)]^T] = E_{Dir}[w_0 \mathbf{R}_0 + \mathbf{R} \mathbf{w}]$$

where the r.h.s. expected value is taken w.r.t. the Dirichlet distribution above. From  $E_{Dir}[w_0] = s/(s+n)$  and  $E_{Dir}[w_i] = 1/(s+n)$  for  $i > 0$ , we obtain the mean vector  $\mu$  in (16). The covariance matrix can be obtained in a similar way by computing  $E_{Dir}[w_0^2]$ ,  $E_{Dir}[\mathbf{w}^T w_0]$ ,  $E_{Dir}[\mathbf{w} \mathbf{w}^T]$  etc..

#### 2. Proof of Theorem 2

From (13) we can write

$$\begin{aligned} & \mathcal{P}[P(X_2 > X_1) + \frac{1}{2}P(X_2 = X_1) > \frac{1}{2}] \\ &= P_{Dir}[w_0 (P_0(X_2 > X_1) + \frac{1}{2}P_0(X_2 = X_1)) \\ &\quad + \sum_{i=1}^n w_i H(X_{2i} - X_{1i}) > \frac{1}{2}] \\ &= P_{Dir}\left[\frac{w_0}{2} + \sum_{i=1}^n w_i H(X_{2i} - X_{1i}) > \frac{1}{2}\right], \end{aligned} \tag{1}$$

where we have used the fact that  $\alpha^* = \delta_{X_1=X_2}$  and thus  $P_0(X_2 = X_1) = 1$ . By denoting with  $\theta_t = w_0 +$

$\sum_{i=1}^n w_i I_{\{X_{2i} = X_{1i}\}}$ ,  $\theta_g = \sum_{i=1}^n w_i I_{\{X_{2i} > X_{1i}\}}$  and  $\theta_l = 1 - \theta_t - \theta_g$ , and considering a partition of the space  $\mathbb{X}$  of the form  $B_0 = \{(X_2, X_1) : X_2 = X_1\}$ ,  $B_g = \{(X_2, X_1) : X_2 > X_1\}$  and  $B_l = \{(X_2, X_1) : X_2 < X_1\}$  it can easily be verified that  $\boldsymbol{\theta} = (\theta_t, \theta_g, \theta_l)$  has a Dirichlet distribution with parameters  $(s + n_t, n_g, n - n_t - n_g)$ . Then we have

$$\begin{aligned} & P_{Dir}\left[\frac{w_0}{2} + \sum_{i=1}^n w_i H(X_{2i} - X_{1i}) > \frac{1}{2}\right] \\ &= P_{Dir}\left[\frac{1}{2}(w_0 + \sum_{i=1}^n w_i I_{\{X_{2i} = X_{1i}\}}) + \sum_{i=1}^n w_i I_{\{X_{2i} > X_{1i}\}} > \frac{1}{2}\right] \\ &= \int I_{\{0.5\theta_t + \theta_g > 0.5\}}(\boldsymbol{\theta}) Dir(\boldsymbol{\theta}; s + n_t, n_g, n - n_t - n_g) d\theta_t d\theta_g \\ &= K_1 \int_0^1 d\theta_t \int_{0.5(1-\theta_t)}^1 \theta_t^{s+n_t-1} \theta_g^{n_g-1} \\ &\quad (1 - \theta_t - \theta_g)^{n-n_t-n_g-1} d\theta_g, \end{aligned} \tag{2}$$

where  $K_1 = \frac{\Gamma(n+s)}{\Gamma(s+n_t)\Gamma(n_g)\Gamma(n-n_t-n_g)}$ . By the change of variables  $\theta'_g = \frac{\theta_g}{1-\theta_t}$  we obtain

$$\begin{aligned} & P_{Dir}\left[\frac{w_0}{2} + \sum_{i=1}^n w_i H(X_{2i} - X_{1i}) > \frac{1}{2}\right] \\ &= K_1 \int_0^1 \theta_t^{s+n_t-1} (1 - \theta_t)^{n-n_t-1} d\theta_t \\ &\quad \int_{0.5}^1 (\theta'_g)^{n_g-1} (1 - \theta'_g)^{n-n_t-n_g-1} d\theta'_g \\ &= K_1 K_2 K_3 \int_{0.5}^1 Beta(\theta; n_g, n - n_t - n_g) d\theta, \end{aligned} \tag{3}$$

where  $K_2 = \frac{\Gamma(n-n_t)\Gamma(s+n_t)}{\Gamma(s+n)}$  and  $K_3 = \frac{\Gamma(n_g)\Gamma(n-n_t-n_g)}{\Gamma(n)}$ . This proves the theorem, since  $K_1 K_2 K_3 = 1$ .

#### 3. Matrix of the ranks of Example 1

Table 1 gives the ranks in  $n = 30$  datasets of the four algorithms in Example 1.

000  
001  
002  
003  
004  
005  
006  
007  
008  
009  
010  
011  
012  
013  
014  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030  
031  
032  
033  
034  
035  
036  
037  
038  
039  
040  
041  
042  
043  
044  
045  
046  
047  
048  
049  
050  
051  
052  
053  
054  
055  
056  
057  
058  
059  
060  
061  
062  
063  
064  
065  
066  
067  
068  
069  
070  
071  
072  
073  
074  
075  
076  
077  
078  
079  
080  
081  
082  
083  
084  
085  
086  
087  
088  
089  
090  
091  
092  
093  
094  
095  
096  
097  
098  
099  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109

110	165
111	166
112	167
113	168
114	169
115	170
116	171
117	172
118	173
119	174
120	175
121	176
122	177
123	178
124	179
125	180
126	181
127	182
128	183
129	184
130	185
131	186
132	187
133	188
134	189
135	190
136	191
137	192
138	193
139	194
140	195
141	196
142	197
143	198
144	199
145	200
146	201
147	202
148	203
149	204
150	205
151	206
152	207
153	208
154	209
155	210
156	211
157	212
158	213
159	214
160	215
161	216
162	217
163	218
164	219

*Table 1.* Ranks of the algorithms in Example 1.

$X_1$	4	4	4	4	4	3	3	3	3	4	4	4	4	4	4	3	3	3	3	3	2	2	2	2	4	3	3	4	3	3
$X_2$	3	3	3	3	3	4	4	4	4	4	4	2	2	2	2	2	2	2	2	2	2	2	2	2	3	3	4	4	3	3
$X_3$	2	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3	4	4	4	4	4	4	4	3	3	1	1	1	1	1
$X_4$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	3	4		