SUPPLEMENT TO "EXACT CALCULATIONS FOR FALSE DISCOVERY PROPORTION WITH APPLICATION TO LEAST FAVORABLE CONFIGURATIONS"

BY ETIENNE ROQUAIN AND FANNY VILLERS

UPMC University of Paris 6

We provide some proofs for the paper (1). We use the notation and the equation numbering defined in (1).

1. Proof of (31)-(1). To prove (31)-(1), remark that we may assumed that $G(x) = x$ up to replace $(t_j)_j$ by $(G(t_j))_j$ (because G is continuous increasing). Next, assume $k < k' < m$ and denote $L(r) = \sum_{i=1}^m \mathbf{1}\{p_i \leq t_r\}$. By definition of a step-down procedure,

$$
\begin{aligned}
&\mathbb{P}\left[|\mathrm{SD}(\mathbf{t})| = k, |\mathrm{SD}(\mathbf{t}')| = k'\right] \\
=& \mathbb{P}[\forall j \leq k, L(j) \geq j, L(k+1) = k, \\
&\forall j, k+1 \leq j \leq k', L(j+1) \geq j, L(k'+2) = k'] \\
=& \binom{m}{k}\binom{m-k}{k'-k}\ \mathbb{P}[\forall j \leq k, L(j) \geq j, \forall j, k+1 \leq j \leq k', L(j+1) \geq j, \\
&p_1, ..., p_k \leq t_{k+1} < p_{k+1}, ..., p_{k'} \leq t_{k'+2} < p_{k'+1}, ..., p_m] \\
=& \binom{m}{k}\binom{m-k}{k'-k}\ \mathbb{P}\Big[p_1, ..., p_k \leq t_{k+1} < p_{k+1}, ..., p_{k'} \leq t_{k'+2} < p_{k'+1}, ..., p_m\,, \\
&\forall j \leq k, \sum_{i=1}^{k} \mathbf{1}\{p_i \leq t_j\} \geq j,\ \ \forall j, k+1 \leq j \leq k', \sum_{i=k+1}^{k'} \mathbf{1}\{p_i \leq t_{j+1}\} \geq j-k\Big] \\
=& \binom{m}{k}\binom{m-k}{k'-k}\mathbb{P}\left[p_{(1)} \leq t_1, ..., p_{(k)} \leq t_k\right](1-t_{k'+2})^{m-k'} \\
&\times \mathbb{P}\left[t_{k+1} < p_{(1)} \leq t_{k+2}, ..., p_{(k'-k)} \leq t_{k'+1}\right],
\end{aligned}
$$

AMS 2000 subject classifications: Primary 62J15; secondary 62G10; tertiary 60C05

Keywords and phrases: False discovery rate, false discovery proportion, multiple testing, least favorable configuration, power, equicorrelated multivariate normal distribution, step-up, step-down

where we used that the p-values are i.i.d. Simple computations give that

$$\mathbb{P}\left[t_{k+1} < p_{(1)} \le t_{k+2}, ..., p_{(k'-k)} \le t_{k'+1}\right]$$
$$= \mathbb{P}\left[p_{(1)} \le t_{k+2} - t_{k+1}, ..., p_{(k'-k)} \le t_{k'+1} - t_{k+1}\right].$$

This leads to (31)-(1). The cases $k < k' = m$ and $k = k'$ are similar.

2. Proof of (16)-(1), (17)-(1) and (18)-(1). Let us state the following general expression: for $1 \le j \le k \le m$,

(1)
$$\mathbb{P}[|\mathcal{H}_0(H) \cap \mathrm{SD}(\mathbf{t})| = j \mid |\mathrm{SD}(\mathbf{t})| = k] = \binom{k}{j} \pi_0^j \pi_1^{k-j} \frac{\Psi_{k,j,F_1}(t_1, ..., t_k)}{\Psi_k(G(t_1), ..., G(t_k))},$$

where we let $\Psi_{k,j,F_1}(t_1, ..., t_k) = \mathbb{P}[U_{(1)} \le t_1, ..., U_{(k)} \le t_k]$, in which the probability is taken over $(U_i)_{1 \le i \le k}$ such that $(U_i)_{1 \le i \le j}$ are i.i.d. uniform, independently of $(U_i)_{j+1 \le i \le k}$ i.i.d. of c.d.f. F_1.

Applying (1) in the particular cases $j = k$, $j = 0$ and $F_1 = 1$ leads to (16)-(1), (17)-(1) and (18)-(1), respectively. To state (1), we use that $(H_i, p_i)_i$ is a i.i.d. sequence:

$$\begin{aligned}
&\mathbb{P}[|\mathcal{H}_0(H) \cap \mathrm{SD}(\mathbf{t})| = j, |\mathrm{SD}(\mathbf{t})| = k] \\
&= \binom{k}{j}\binom{m}{k} \mathbb{P}[\mathrm{SD}(\mathbf{t}) = \{1, ..., k\}, H_1 = ... = H_j = 0, H_{j+1} = ... = H_k = 1] \\
&= \binom{k}{j}\binom{m}{k} \mathbb{P}\Big[\forall \ell \le k, \sum_{i=1}^k \mathbf{1}\{p_i \le t_\ell\} \ge \ell, \forall i \ge k+1, p_i > t_{k+1}, \\
&H_1 = ... = H_j = 0, H_{j+1} = ... = H_k = 1\Big] \\
&= \binom{k}{j} \pi_0^j \pi_1^{k-j} \mathbb{P}\Big[\forall \ell \le k, \sum_{i=1}^k \mathbf{1}\{p_i \le t_\ell\} \ge \ell \mid H_1 = ... = H_j = 0, \\
&H_{j+1} = ... = H_k = 1\Big] \binom{m}{k} (1 - G(t_{k+1}))^{m-k},
\end{aligned}$$

which leads to (1) by applying (19)-(1).

3. Proof of Proposition 3.8 of (1). We focus on the step-up case, when $\rho \in (-1, 1)$ and $m_0 = 1$ (the remaining cases are left to the reader). Without loss of generality, we may assume that the first coordinate corresponds to the true null, that is, $H = (0, 1)$. In this context, the FDP takes one of the three values: $0, \frac{1}{2}, 1$, according to the location of the test statistics $Y_i = X_i + \mu H_i$ with respect to the critical values z_1 and z_2. From the definition of a step-up procedure, we may define the region $\mathcal{D}_i = \{(y_1, y_2) \in \mathbb{R}^2 \,|\, \mathrm{FDP}(y_1, y_2) = i/2\}$, for $i \in \{1, 2\}$, where $\mathrm{FDP}(y_1, y_2)$ denotes the FDP of $\mathrm{SU}(\mathbf{t})$ taken in the p-values $\mathbf{p} = (\overline{\Phi}(y_1), \overline{\Phi}(y_2))$. The regions $\mathcal{D}_i$ are represented on Figure 1(**a**). Next, since (Y_1, Y_2) follows the EMN model, we may write for $i \in \{1, 2\}$,

$$
\begin{aligned}
&\mathbb{P}(\mathrm{FDP}(Y_1, Y_2) = i/2) \\
&= (2\pi)^{-1}(1-\rho^2)^{-1/2} \int_{\mathcal{D}_i} \exp\left\{ -\frac{\big(y_1 - \rho(y_2 - \mu)\big)^2}{2(1-\rho^2)} - \frac{1}{2}(y_2 - \mu)^2 \right\} dy_1 dy_2 \\
&= (2\pi)^{-1} \int_{\widetilde{\mathcal{D}}_i} \exp\{-(u^2 + v^2)/2\}\, du dv,
\end{aligned}
$$

by using the substitution $u = (1-\rho^2)^{-1/2}(y_1 - \rho(y_2 - \mu))$ and $v = y_2 - \mu$, and where the resulting integration domain $\widetilde{\mathcal{D}}_i$ is represented on Figure 1(**b**). Therefore, we obtain

$$
\mathbb{P}(\mathrm{FDP} = 1/2) = (2\pi)^{-1/2} \int_{z_1 - \mu}^{\infty} \exp\{-v^2/2\} \overline{\Phi}\left((z_1 - \rho v)/\sqrt{1-\rho^2} \right) dv,
$$

and

$$
\mathbb{P}(\mathrm{FDP} = 1) = (2\pi)^{-1/2} \int_{-\infty}^{z_1 - \mu} \exp\{-v^2/2\} \overline{\Phi}\left((z_2 - \rho v)/\sqrt{1-\rho^2} \right) dv,
$$

and the final expression results by using the substitution $w = \overline{\Phi}(v)$.

4. Proof that the linear threshold satisfies condition (A)-(1). Let us prove that the threshold $t_j = \alpha j/m$, $1 \le j \le m$, satisfies (A)-(1). For any $m \ge 2$ and $1 \le k \le m$, let us denote

$$
S_{m,k} = \sum_{i=0}^{m-k} \frac{k}{k+i} \widetilde{\mathcal{D}}_{m-k} \left(\left(\frac{\alpha j/m}{1 - \alpha k/m} \right)_{1 \le j \le m-k}, i \right)
$$

and prove that $k \in \{1, ..., m\} \mapsto S_{m,k}$ is nondecreasing. Letting $a_{m,k} = \frac{k}{m}\left(1 - \alpha \, \frac{m-k}{m-\alpha k}\right)$ (increasing in k) and $b_{m,k} = \frac{m-\alpha}{m} \frac{m-k}{m-\alpha k}$ (decreasing in k), we may prove the following recursion (see the proof below):

$$
S_{m,k} = a_{m,k} + b_{m,k} S_{m-1,k}. \tag{2}
$$

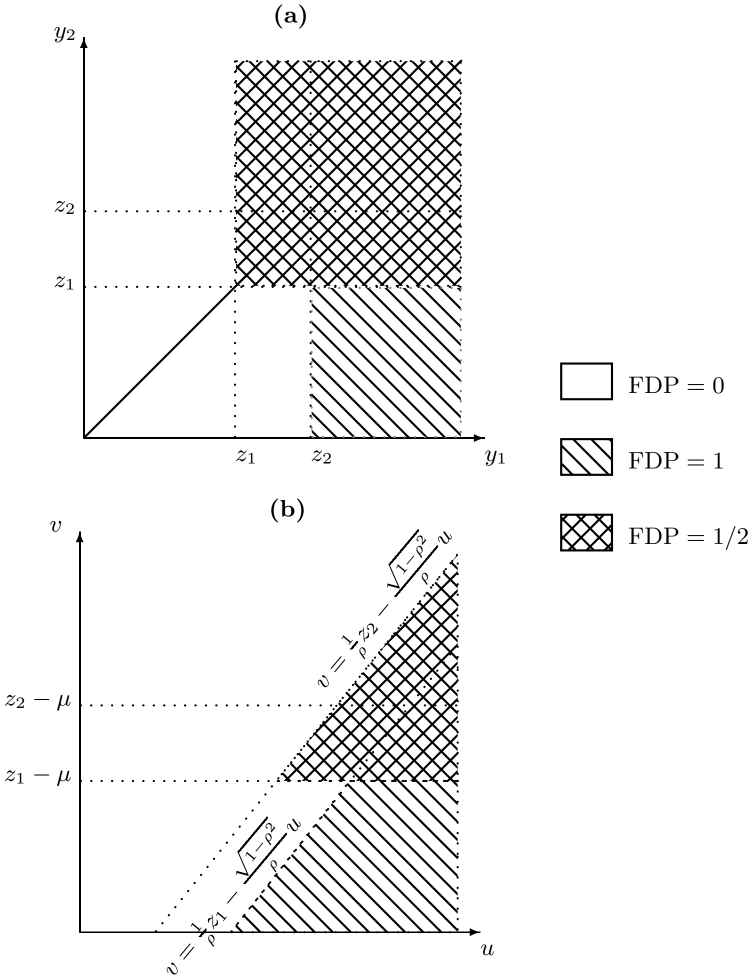

FIG 1. *Top:* $\mathcal{D}_i$. *Bottom:* $\widetilde{\mathcal{D}}_i$. *For* $i = 1$*(double cross-hatch) and* $i = 2$ *(simple cross-hatch). Graph for* $\rho < 0$.

Using (2), we can state that $k \in \{1, ..., m\} \mapsto S_{m,k}$ is nondecreasing by a recurrence on $m \geq 2$: the property is obviously true for $m = 2$. Assuming the property true for $m-1$, we obtain for any $2 \leq k \leq m-1$,

$$\begin{aligned} S_{m,k} - S_{m,k-1} &= a_{m,k} - a_{m,k-1} + b_{m,k} S_{m-1,k} - b_{m,k-1} S_{m-1,k-1} \\ &= a_{m,k} - a_{m,k-1} + (b_{m,k} - b_{m,k-1}) S_{m-1,k} \\ &\quad + b_{m,k-1}(S_{m-1,k} - S_{m-1,k-1}) \\ &\geq (a_{m,k} + b_{m,k}) - (a_{m,k-1} + b_{m,k-1}), \end{aligned}$$

because $S_{m-1,k} \leq 1$. Hence, since $a_{m,k} + b_{m,k} = 1 - \frac{\alpha}{m} \frac{m-k}{m-\alpha k}$, the quantity $a_{m,k} + b_{m,k}$ is increasing in k and $S_{m,k} - S_{m,k-1} \geq 0$. Also, we obviously have $S_{m,m} = 1 \geq S_{m,m-1}$, and the recurrence is completed.

We now finally state (2). Let for $1 \leq j \leq m-k$, $t_j = \frac{\alpha j/m}{1-\alpha k/m}$ and $t'_j = 1 - t_{m-k-j+1}$, so that

$$S_{m,k} = \mathbb{E}_0 \left[\frac{k}{k + |\mathrm{SD}((t_j)_{1 \leq j \leq m-k})|} \right] = \mathbb{E}_0 \left[\frac{k}{m - |\mathrm{SU}((t'_j)_{1 \leq j \leq m-k})|} \right],$$

where $\mathbb{E}_0$ denotes the expectation with respect to i.i.d. uniform p-values. Hence, denoting $t'_j = \beta + j\gamma$ with $\beta = \frac{m - m\alpha - \alpha}{m - \alpha k}$ and $\gamma = \frac{\alpha}{m - \alpha k}$, we obtain

$$\begin{aligned} \frac{S_{m,k}}{k} =& \frac{1}{m} + \frac{1}{m} \mathbb{E}_0 \left[\frac{|\mathrm{SU}((t'_j)_{1 \leq j \leq m-k})|}{m - |\mathrm{SU}((t'_j)_{i \leq j \leq m-k})|} \right] \\ =& \frac{1}{m} + \frac{1}{m} \sum_{j=1}^{m-k} \frac{j}{m-j} \binom{m-k}{j} (t'_j)^j \Psi_{m-k-j}(1 - t'_{m-k}, ..., 1 - t'_{j+1}) \\ =& \frac{1}{m} + \frac{m-k}{m} \sum_{j=1}^{m-k} \frac{t'_j}{m-j} \binom{m-k-1}{j-1} (t'_j)^{j-1} \\ & \times \Psi_{m-k-j}(1 - t'_{m-k}, ..., 1 - t'_{j+1}). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{S_{m,k}}{k} =& \frac{1}{m} - \gamma \frac{m-k}{m} + \frac{m-k}{m} (\beta + m\gamma) \\ & \times \sum_{j=0}^{m-1-k} \frac{(t'_{j+1})^j}{m-1-j} \binom{m-1-k}{j} \Psi_{m-1-k-j}(1 - t'_{m-k}, ..., 1 - t'_{j+2}) \\ =& \frac{1}{m} - \gamma \frac{m-k}{m} + \frac{m-k}{m} (\beta + m\gamma) \mathbb{E}_0 \left[\frac{1}{m - 1 - |\mathrm{SU}((t'_{j+1})_{1 \leq j \leq m-1-k})|} \right] \\ =& \frac{1}{m} - \gamma \frac{m-k}{m} + \frac{m-k}{m} (\beta + m\gamma) \frac{S_{m-1,k}}{k}, \end{aligned}$$

and the recursion (2) is proved.

5. Proof of (29)-(1).

For any $0 \le i \le m-1$, let us denote

$$u_i = \mathbb{E}_0[(|\mathrm{SU}((t_j)_{i+1\le j\le m})| + i + 1)^{-1}],$$

(where $\mathbb{E}_0$ denotes the expectation with respect to i.i.d. uniform p-values) and $u_m = 1/(m+1)$, $u_{m+1} = 0$, so that u_0 equals the quantity in (29)-(1). We may prove the following recursion relation: for any $0 \le i \le m$,

$$(i+1)u_i = (1-(m-i)\gamma) - (m-i)(\beta-\gamma)u_{i+1}. \tag{3}$$

Expression (3) is proved as follows: for $i < m$,

$$\begin{aligned}
(i+1)u_i =& 1 - \mathbb{E}_0\left[\frac{|\mathrm{SU}((t_j)_{i+1\le j\le m})|}{|\mathrm{SU}((t_j)_{i+1\le j\le m})| + i + 1}\right] \\
=& 1 - \sum_{k=1}^{m-i} \frac{k}{k+i+1}\binom{m-i}{k}(t_{k+i})^k \Psi_{m-i-k}(1-t_m, ..., 1-t_{k+i+1}) \\
=& 1 - (m-i)\sum_{k=1}^{m-i} \frac{t_{k+i}}{k+i+1}\binom{m-i-1}{k-1}(t_{k+i})^{k-1} \\
& \times \Psi_{m-i-k}(1-t_m, ..., 1-t_{k+i+1}) \\
=& 1 - (m-i)\sum_{k=0}^{m-(i+1)} \left(\gamma + \frac{\beta-\gamma}{k+(i+1)+1}\right)\binom{m-(i+1)}{k}(t_{k+(i+1)})^k \\
& \times \Psi_{m-(i+1)-k}(1-t_m, ..., 1-t_{k+(i+1)+1}).
\end{aligned}$$

Next, we obtain that the solution of the recursion (3) is given by

$$u_i = \sum_{j=0}^{m-i} \frac{1-(m-(i+j))\gamma}{m-(i+j)} \; \frac{(m-i)\times\cdots\times(m-(i+j))}{(i+1)\times\cdots\times(i+j+1)}(\gamma-\beta)^j,$$

which leads to $u_0 = \sum_{j=0}^m \frac{1-(m-j)\gamma}{j+1}\binom{m}{j}(\gamma-\beta)^j$ and (29)-(1) results.

6. Proofs of the lemmas of Section 7 in (1).

6.1. *Proof of Lemma 7.1 of (1).* First note that since $\hat{k}$ is nondecreasing in each coordinate of $(p_1, ..., p_m)$, we always have $\hat{k} \le \hat{k}'_{(\ell)} + \ell$. Second, since $p_{(k)} \le t_k$ is equivalent to $|\{1 \le j \le m \mid p_j \le t_k\}| \ge k$, the rejection number of $\mathrm{SU}(\mathbf{t})$ can be defined as $\widehat{k} = \max\{k \in \{0, 1, ..., m\} \mid |\{1 \le j \le m \mid p_j \le t_k\}| \ge k\}$. Hence, $\forall 1 \le i \le \ell$, $p_i \le t_{\hat{k}'_{(\ell)}+\ell}$ is equivalent to $|\{1 \le j \le m \mid p_j \le t_{\hat{k}'_{(\ell)}+\ell}\}| \ge \ell + |\{\ell+1 \le j \le m \mid p_j \le t_{\hat{k}'_{(\ell)}+\ell}\}|$ which is equivalent to $|\{1 \leq j \leq m \mid p_j \leq t_{\hat{k}'_{(\ell)}+\ell}\}| \geq \ell + \hat{k}'_{(\ell)}$, by definition of $\hat{k}'_{(\ell)}$. As a consequence, since $\hat{k}$ is a maximum and since $\hat{k} \leq \hat{k}'_{(\ell)} + \ell$, the latter is equivalent to $\hat{k}'_{(\ell)} + \ell = \hat{k}$. This establishes the second equivalence. The first equivalence easily comes from the second equivalence and using that $t_{\hat{k}} \leq t_{\hat{k}'_{(\ell)}+\ell}$ because $(t_k)_k$ is a nondecreasing sequence.

6.2. *Proof of Lemma 7.2 of (1).* Similarly to the step-up case, the rejection number of $\mathrm{SD}(\mathbf{t})$ can be defined as $\tilde{k} = \max\{k \in \{0, 1, ..., m\} \mid \forall k' \leq k,\ |\{1 \leq j \leq m \mid p_j \leq t_{k'}\}| \geq k'\}$. Also remark that we always have $\tilde{k}'_{(1)} + 1 \geq \tilde{k}$ and, by definition of $\tilde{k}$, for any j we have $p_j \leq t_{\tilde{k}} \Leftrightarrow p_j \leq t_{\tilde{k}+1}$. First prove that $p_1 \leq t_{\tilde{k}} \Leftrightarrow \tilde{k} = \tilde{k}'_{(1)} + 1$: using the definitions of $\tilde{k}$ and $\tilde{k}'_{(1)}$ we obtain $p_1 \leq t_{\tilde{k}} \Leftrightarrow |\{2 \leq j \leq m \mid p_j \leq t_{\tilde{k}+1}\}| < \tilde{k} \Leftrightarrow \tilde{k}'_{(1)} < \tilde{k} \Leftrightarrow \tilde{k}'_{(1)} + 1 = \tilde{k}$. Second, we prove $p_1 > t_{\tilde{k}} \Leftrightarrow p_1 > t_{\tilde{k}_{(1)}+1}$: since we obviously have $\tilde{k} \geq \tilde{k}_{(1)}$, we get $p_1 > t_{\tilde{k}} \Rightarrow p_1 > t_{\tilde{k}+1} \Rightarrow p_1 > t_{\tilde{k}_{(1)}+1}$. Conversely, if $p_1 > t_{\tilde{k}_{(1)}+1}$, we get $|\{1 \leq j \leq m \mid p_j \leq t_{\tilde{k}_{(1)}+1}\}| = |\{2 \leq j \leq m \mid p_j \leq t_{\tilde{k}_{(1)}+1}\}| < \tilde{k}_{(1)} + 1$ (by definition of $\tilde{k}_{(1)}$), hence $\tilde{k}_{(1)} + 1 > \tilde{k}$ (by definition of $\tilde{k}$), which implies $\tilde{k}_{(1)} = \tilde{k}$, thus $p_1 > t_{\tilde{k}+1}$ and finally $p_1 > t_{\tilde{k}}$.

References.

[1] Roquain, E. and Villers, F. (2011). Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.* **39** 584–612.

UPMC Univ Paris 6, LPMA,
4, Place Jussieu, 75252 Paris cedex 05, France
E-mail: etienne.roquain@upmc.fr

UPMC Univ Paris 6, LPMA,
4, Place Jussieu, 75252 Paris cedex 05, France
E-mail: fanny.villers@upmc.fr