fff

smth

Помимо количества отвергнутых алгоритмом гипотез D введём величину D_i — общее количество отвергнутых гипотез, если гипотезу номер i принудительно отвергнуть вне зависимости от её p-значения. Конечно, величина D_i либо равна D, либо на единицу больше.

Величина D_i нужна нам для тонкого разложения на сомножители. Заметим, что даже при независимых p-значениях p_i величины p_i и D зависимы. Например, если $D = M$, то значит все гипотезы были отвергнуты и, согласно алгоритму, все p_i были меньше некоторого своего порога. Однако, при независимых p_i величины p_i и D_i независимы. Отсюда,

$$\mathbb{P}(p_i \leq q, D = d) = \mathbb{P}(p_i \leq q, D_i = d) = \mathbb{P}(p_i \leq q)\,\mathbb{P}(D_i = d).$$

Разложим общее число ошибочно отвергнутых гипотез в сумму индикаторов,

$$FDR = \mathbb{E}\left(\frac{FD}{\max\{1, D\}}\right) = \sum_{i \in I_0} \mathbb{E}(I(p_i \leq q(D))/D).$$

В силу $\sum_d I(D = d) = 1$ получим дальнейше разрезание формулы:

$$FDR = \sum_{i \in I_0} \sum_{d=1}^{M} \mathbb{E}(I(D = d) I(p_i \leq q(D))/D) = \sum_{i \in I_0} \sum_{d=1}^{M} \mathbb{P}(p_1 \leq q(d), D = d)/d.$$

Применяем трюк с подменой D на D_i и вспоминаем о равномерном распределении p_i при $i \in I_0$:

$$\mathbb{P}(p_i \leq q(d), D = d) = \mathbb{P}(p_i \leq q(d), D_i = d) = \mathbb{P}(p_i \leq q(d))\,\mathbb{P}(D_i = d) = \frac{d\alpha}{M}\,\mathbb{P}(D_i = d).$$

Тождество $\sum_d \mathbb{P}(D_i = d) = 1$ позволяет нам завершить вычисление FDR,

$$FDR = \sum_{i \in I_0} \sum_{d=1}^{M} \frac{\alpha}{M}\,\mathbb{P}(D_i = d) = \sum_{i \in I_0} \frac{\alpha}{M} = \frac{\alpha M_0}{M} \leq \alpha.$$

Разное:

$FWER \leq \alpha$ если и только если для любого подмножества $J \subset I_0$ вероятность отвергнуть хотя бы одну гипотезу из подмножества J не превосходит α.

Доказательство. Если из множества I_0 удалить часть гипотез до получения множества J, то вероятность отвержения хотя бы одной из них не увеличивается.

Интуитивное почти-доказательство. Можно ли спасти?

Навесим условие внутри ожидания:

$$\mathbb{E}\left(\frac{FD}{\max\{1, D\}}\right) = \mathbb{E}\left(\mathbb{E}\left(\frac{FD}{\max\{1, D\}} \mid D\right)\right)$$

Рассмотрим условное ожидание:

$$\mathbb{E}\left(\frac{FD}{\max\{1, D\}} \mid D\right) = \mathbb{E}\left(\frac{FD}{\max\{1, D\}} \mid h(D)\right)$$

Вот здесь опасный переход:

$$\mathbb{E}\left(\frac{FD}{\max\{1, D\}} \mid h(D)\right) = \frac{M_0 h(D)}{D} = \frac{M_0 \alpha D}{MD} = \frac{M_0 \alpha}{M}$$