Note: Rejection probability of the Fisher's exact test in a sequential design

Jérôme Pasquier

5 December 2025

1 Rejection probability of the Fisher's exact test

We consider two independent binomial random variables

$$X_1 \sim \mathrm{Bin}(n_1, p_1), \quad X_2 \sim \mathrm{Bin}(n_2, p_2).$$

Conditionally on the total number of observed events $M = X_1 + X_2$, and under the null hypothesis $H_0 : p_1 = p_2$, the random variable X_1 follows a hypergeometric distribution:

$$X_1 \mid M = m \ \sim \ \mathrm{Hypergeometric}(N, m, n_1),$$

where $N = n_1 + n_2$, with probability mass function

$$\mathsf{hg}(x_1 \mid m) = \frac{\binom{m}{x_1}\binom{N-m}{n_1-x_1}}{\binom{N}{n_1}}.$$

Note that

$$\frac{\binom{m}{x_1}\binom{N-m}{n_1-x_1}}{\binom{N}{n_1}} = \frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}, \quad \text{where } m = x_1 + x_2.$$

Define the test statistic

$$T_{n_1,n_2}(x_1, x_2) \ = \ -\log\left(\frac{\binom{n_1}{x_1}\binom{n_2}{x_2}}{\binom{N}{m}}\right).$$

Hence, ordering contingency tables (x_1, x_2) by increasing hypergeometric probability is equivalent to ordering them by decreasing T_{n_1,n_2}.

For each total count $m \in 0, \ldots, N$, define the critical threshold corresponding to a nominal significance level α:

$$t_\alpha^{(n_1,n_2)}(m) \ = \ \inf\Big\{t : \ \mathbb{P}_{H_0}\big(T_{n_1,n_2}(X_1, m - X_1) \geqslant t \mid M = m\big) \leqslant \alpha\Big\}.$$

The two-sided Fisher rejection region at level α is

$$\mathcal{R}(n_1, n_2; \alpha) = \Big\{(x_1, x_2) : \ 0 \leqslant x_j \leqslant n_j, \ T_{n_1,n_2}(x_1, x_2) \ \geqslant \ t_\alpha^{(n_1,n_2)}(x_1 + x_2)\Big\}.$$

For given (p_1, p_2) any fixed (n_1, n_2, α), the rejection probability is

$$\pi(n_1, n_2; \alpha \mid p_1, p_2) = \sum_{(x_1,x_2)\in\mathcal{R}(n_1,n_2;\alpha)} \binom{n_1}{x_1} p_1^{x_1}(1-p_1)^{n_1-x_1} \ \binom{n_2}{x_2} p_2^{x_2}(1-p_2)^{n_2-x_2}.$$

Under H_1, the probability π is the power of the test, whereas under H_0 it is the achieved Type I error rate (which is, in general, less than the nominal level α).

Remark on discreteness and ties. Because the test is based on a discrete distribution, the actual attained significance level satisfies

$$\mathbb{P}_{H_0}\big((X_1, X_2) \in \mathcal{R}(n_1, n_2; \alpha)\big) \leqslant \alpha,$$

and is often strictly less than α. This discreteness leads to conservative p-values and to the possibility of multiple tables having the same probability (ties) at the critical boundary.

2 Group sequential design

In a group sequential design where several planned analyses are performed, the trial may be stopped early for efficacy or futility at any interim analysis. The k-th analysis is performed only if the null hypothesis has not been rejected at the previous analysis $(k-1)$.

In what follows, we use the following notation:

- Cumulative sample sizes:
$$n_{1k},\, n_{2k}, \quad k = 1, \ldots, K,$$
where K is the number of analyses (interims + final).
- Sample size increments:
$$a_{1k} := n_{1k} - n_{1,k-1}, \quad a_{2k} := n_{2k} - n_{2,k-1},$$
with $n_{10} = 0$ and $n_{20} = 0$.
- Additional count distributions:
$$A_{1k} \sim \mathrm{Bin}(a_{1k}, p_1),\; A_{2k} \sim \mathrm{Bin}(a_{2k}, p_2),$$
mutually independent across arms and across analyses.
- Cumulative count distributions:
$$X_{1k} = \sum_{j=1}^{k} A_{1j}, \quad X_{2k} = \sum_{j=1}^{k} A_{2j}.$$

Let
$$\mathcal{R}_k := \mathcal{R}(n_{1k}, n_{2k}; \alpha_k)$$
be the rejection region for the k-th interim analysis, where α_k denote the pre-specified alpha spends. The corresponding *efficacy stopping* region is therefore $\mathcal{R}_k$. In addition, a group sequential design may include *futility stopping*. We denote by $\mathcal{F}_k$ the futility region at analysis k, consisting of outcomes for which continuing the trial cannot plausibly lead to rejection of H_0. Given these two types of early stopping, the *continue* region at interim k is defined as
$$\mathcal{C}_k := (\mathcal{R}_k \cup \mathcal{F}_k)^{\complement}.$$

The specific construction of the futility regions $\mathcal{F}_k$ depends on the chosen design and will be detailed later.

Let $f_k(x_1, x_2)$ denote the (unconditional) probability mass of observing the cumulative counts (x_1, x_2) at the k-th analysis and having continued up to that stage (i.e. all earlier outcomes were in their respective continue regions). At the first analysis, no previous stopping is possible, so f_1 is just the joint binomial pmf,
$$f_1(x_1, x_2) = \binom{n_{11}}{x_1} p_1^{x_1} (1-p_1)^{n_{11}-x_1} \binom{n_{21}}{x_2} p_2^{x_2} (1-p_2)^{n_{21}-x_2}.$$

The joint pmf of the increments A_{1k}, A_{2k} is
$$g_k(i, j) = \binom{a_{1k}}{i} p_1^{i} (1-p_1)^{a_{1k}-i} \binom{a_{2k}}{j} p_2^{j} (1-p_2)^{a_{2k}-j}.$$

Given cumulative counts $(X_{1,k-1}, X_{2,k-1}) = (x_1', x_2')$ entering analysis k, the conditional pmf of the updated totals is
$$h_k(x_1, x_2 \mid x_1', x_2') = \sum_{i,j} \mathbf{1}\{x_1' + i = x_1, x_2' + j = x_2\}\, g_k(i, j).$$

Equivalently, $h_k(x_1, x_2 \mid x_1', x_2')$ equals $g_k(x_1 - x_1', x_2 - x_2')$ when $0 \leqslant x_1 - x_1' \leqslant a_{1k}$ and $0 \leqslant x_2 - x_2' \leqslant a_{2k}$, and is zero otherwise.

The unconditional pmf at analysis k is then obtained by summing over all states that continue past stage $k-1$:

$$f_k(x_1, x_2) = \sum_{(x_1', x_2') \in \mathcal{C}_{k-1}} f_{k-1}(x_1', x_2')\, h_k(x_1, x_2 \mid x_1', x_2').$$

Define

$$\tilde{f}_k(x_1, x_2) := \mathbf{1}\{(x_1, x_2) \in \mathcal{C}_k\} f_k(x_1, x_2),$$

so after the change of variables $(x_1', x_2') \mapsto (x_1 - i, x_2 - j)$, the recursion can be written compactly as

$$\begin{aligned} f_k(x_1, x_2) &= \sum_{i=0}^{a_{1k}} \sum_{j=0}^{a_{2k}} \tilde{f}_{k-1}(x_1 - i, x_2 - j) g_k(i, j) \\ &= (g_k * \tilde{f}_{k-1})(x_1, x_2). \end{aligned}$$

The (unconditional) probability of stopping the trial for efficacy at analysis k is

$$\pi_k = \sum_{(x_1, x_2) \in \mathcal{R}_k} f_k(x_1, x_2),$$

and the probability of stopping for futility at analysis k is

$$\phi_k = \sum_{(x_1, x_2) \in \mathcal{F}_k} f_k(x_1, x_2).$$

The probability that the trial is still ongoing after analysis k (i.e. it has not stopped early for either efficacy or futility) is

$$\rho_k = \sum_{(x_1, x_2) \in \mathcal{C}_k} f_k(x_1, x_2) = 1 - \sum_{s=1}^{k} \big(\pi_s + \phi_s\big).$$

The conditional probability to reject at some stage after the k-th interim is

$$\frac{\sum_{s=k+1}^{K} \pi_s}{\rho_k}.$$

The overall rejection probability is

$$\pi_{\text{overall}} = \sum_{k=1}^{K} \pi_k.$$

The final sample size N is a random variable with expected value

$$\mathbb{E}[N] = \sum_{k=1}^{K-1} (n_{1k} + n_{2k})(\pi_k + \phi_k) + (n_{1K} + n_{2K})\rho_{K-1}.$$

Equivalently, because the increment at stage k is taken only if the study has not stopped by stage $k-1$, the expectation can be written as

$$\mathbb{E}[N] = \sum_{k=1}^{K} (a_{1k} + a_{2k})\, \rho_{k-1},$$

with $\rho_0 = 1$; this is algebraically identical to the expression above and makes explicit that the final sample size is weighted by ρ_{K-1}.

All rejection probabilities above are written for arbitrary p_1 and p_2, so the same formulas deliver the type I error when we plug in $p_1 = p_2$, or the power under an alternative when $p_1 \neq p_2$.

Unlike efficacy boundaries, which depend only on the distribution at each interim, futility boundaries depend on the future potential to reach rejection if the trial continues and are generally calculated under a specific alternative, not under H_0. Thus the conditional rejection probabilities below should be

understood as being computed with a (potentially different) pair of planning values p_1^*, p_2^* chosen for the futility calculation, even if the forward rejection probabilities are evaluated at $p_1 = p_2$. These planning probabilities can be specified independently of the p_1, p_2 used when evaluating π_k or π_{overall}.

Let

$$\pi_{\text{cond},k}(x_1, x_2)$$

denote the conditional probability, given the cumulative counts at interim k, that the trial will eventually reject the null at some later analysis $s > k$.

If $(x_1, x_2) \in \mathcal{R}_k$ the trial would already have stopped for efficacy at analysis k, so only values in $\mathcal{R}_k^{\complement}$ are relevant when constructing futility rules.

At the final analysis, no further rejection is possible after failing to reject at stage K, so we set the terminal condition

$$\pi_{\text{cond},K}(x_1, x_2) \equiv 0.$$

For $k < K$, conditioning on the increments between stages k and $k+1$ yields the backward recursion

$$\pi_{\text{cond},k}(x_1, x_2) = \sum_{(x_1', x_2') \in \mathcal{R}_{k+1}} h_{k+1}(x_1', x_2' \mid x_1, x_2) \; + \sum_{(x_1', x_2') \in \mathcal{C}_{k+1}} h_{k+1}(x_1', x_2' \mid x_1, x_2)\, \pi_{\text{cond},k+1}(x_1', x_2'),$$

where h_{k+1} is calculated using the planning values p_1^* and p_2^*.

Given a pre-specified futility cutoff γ_k, the futility region at interim k is

$$\mathcal{F}_k = \{(x_1, x_2) \notin \mathcal{R}_k : \pi_{\text{cond},k}(x_1, x_2) < \gamma_k\}, \qquad \mathcal{F}_K = \varnothing.$$

Because $\pi_{\text{cond},k}(\cdot, \cdot)$ depends on the stopping rules at all future analyses, the futility regions are constructed recursively backwards from $k = K$ to $k = 1$. Once the $\mathcal{F}_k$ are known, the continue regions

$$\mathcal{C}_k = (\mathcal{R}_k \cup \mathcal{F}_k)^{\complement}$$

are fully specified, closing the recursion loop with the forward computation of the f_k.

Appendix

A FFT implementation of the recursions

The forward and backward recursions are discrete two-dimensional convolutions and cross-correlations. Implementing them directly with nested loops scales quadratically in the grid dimensions. Using the Fast Fourier Transform (FFT) to evaluate these convolutions reduces the complexity to $O(n \log n)$.

Forward recursion as a convolution

Define the joint increment kernel

$$G_k(i, j) := g_{1k}(i)\, g_{2k}(j), \qquad 0 \leqslant i \leqslant a_{1k},\ 0 \leqslant j \leqslant a_{2k}.$$

Ignoring early stopping for the moment, the unconditional mass at stage k is the discrete convolution

$$f_k(x_1, x_2) = (f_{k-1} * G_k)(x_1, x_2) = \sum_{i=0}^{a_{1k}} \sum_{j=0}^{a_{2k}} f_{k-1}(x_1 - i, x_2 - j)\, G_k(i, j),$$

where terms that fall outside the support of f_{k-1} are treated as zero. After forming f_k, we set to zero the masses in $\mathcal{R}_k \cup \mathcal{F}_k$ to enforce early stopping; the remaining mass corresponds to the continue region $\mathcal{C}_k$.

To compute $(f_{k-1} * G_k)$ efficiently, both arrays are zero-padded to dimensions $(n_{1,k-1} + a_{1k} + 1) \times (n_{2,k-1} + a_{2k} + 1)$, the elementwise product of their FFTs is taken, and an inverse FFT returns the full convolution. The relevant $(n_{1k} + 1) \times (n_{2k} + 1)$ block corresponds exactly to the support of f_k.

Backward recursion as a valid cross-correlation

For the backward step, define the combined surface at stage $k+1$

$$M_{k+1}(x_1', x_2') := \mathbf{1}_{\mathcal{R}_{k+1}}(x_1', x_2') + \pi_{\mathrm{cond},k+1}(x_1', x_2')\, \mathbf{1}_{\mathcal{C}_{k+1}}(x_1', x_2').$$

The conditional rejection probability satisfies

$$\pi_{\mathrm{cond},k}(x_1, x_2) = \sum_{i=0}^{a_{1,k+1}} \sum_{j=0}^{a_{2,k+1}} G_{k+1}(i,j)\; M_{k+1}(x_1 + i, x_2 + j),$$

which is a 2D cross-correlation of M_{k+1} with the kernel G_{k+1}, evaluated only where both arrays overlap (the "valid" part), producing an $(n_{1k}+1) \times (n_{2k}+1)$ grid. In the FFT implementation, M_{k+1} and a flipped version of G_{k+1} are zero-padded, multiplied in the frequency domain, and inverse-transformed; extracting the valid block yields $\pi_{\mathrm{cond},k}$.

This FFT-based approach eliminates explicit loops over x_1, x_2, i, j, while preserving the exact finite-sample probabilities defined above.