Note: Score Process and Fisher Information

Jérôme Pasquier
ETH Zürich

7 Novemeber 2025

1 Score Process and Fisher Information

Consider a regular parametric model with parameter of interest θ (e.g. a mean difference $\mu_1 - \mu_2$). Let $\ell_k(\theta)$ denote the log-likelihood based on all data accrued up to interim analysis k.

We define the *score process* as

$$S_k(\theta) \equiv \frac{\partial \ell_k(\theta)}{\partial \theta},$$

and the (expected) Fisher information as

$$I_k(\theta) \equiv \mathbb{E}_\theta\left[-\frac{\partial^2 \ell_k(\theta)}{\partial \theta^2}\right].$$

Under standard regularity conditions,

$$\operatorname{Var}(S_k(\theta)) = I_k(\theta).$$

Assuming independent batches of data between steps, both the score and information are additive:

$$S_k(\theta) = \sum_{i=1}^{k} \big(S_i(\theta) - S_{i-1}(\theta)\big), \qquad I_k(\theta) = \sum_{i=1}^{k} \big(I_i(\theta) - I_{i-1}(\theta)\big),$$

with $S_0(\theta) = 0$ and $I_0(\theta) = 0$.

The variance of each increment satisfies

$$\operatorname{Var}\big(S_i(\theta) - S_{i-1}(\theta)\big) = I_i(\theta) - I_{i-1}(\theta).$$

2 Standardized Statistics

Define the standardized cumulative score and its standardized increments:

$$Z_k(\theta) \equiv \frac{S_k(\theta)}{\sqrt{I_k(\theta)}}, \qquad Z_i^*(\theta) \equiv \frac{S_i(\theta) - S_{i-1}(\theta)}{\sqrt{I_i(\theta_0) - I_{i-1}(\theta_0)}},$$

so that

$$Z_k(\theta) = \sum_{i=1}^{k} w_{ik}\, Z_i^*(\theta), \qquad Z_i^*(\theta) \overset{\text{i.i.d.}}{\sim} N(0,1),$$

with weights

$$w_{ik} = \sqrt{\frac{I_i(\theta) - I_{i-1}(\theta)}{I_k(\theta)}}.$$

The covariance between Z_j and Z_k ($j \leqslant k$) is then

$$\operatorname{Cov}(Z_j, Z_k) = \sqrt{\frac{I_j}{I_k}}.$$

3 Application: Sequential t-Tests

In a two-sample setting, consider

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2), \qquad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2),$$

and the parameter of interest

$$\theta = \mu_1 - \mu_2.$$

At each interim look k, we observe n_{1k} and n_{2k} samples, where $n_{sk} \leqslant n_{s,k+1}$ for $s = 1, 2$.

Reparameterize the means as

$$\mu = \frac{\mu_1 + \mu_2}{2}, \qquad \mu_1 = \mu + \frac{\theta}{2}, \quad \mu_2 = \mu - \frac{\theta}{2}.$$

The log-likelihood is

$$\ell_k = -\frac{n_{1k}}{2}\log(2\pi\sigma_1^2) - \frac{1}{2\sigma_1^2}\sum_{i=1}^{n_{1k}}\big(X_{1i} - \mu - \tfrac{\theta}{2}\big)^2 - \frac{n_{2k}}{2}\log(2\pi\sigma_2^2) - \frac{1}{2\sigma_2^2}\sum_{i=1}^{n_{2k}}\big(X_{2i} - \mu + \tfrac{\theta}{2}\big)^2.$$

The score functions are

$$\begin{aligned}
\frac{\partial \ell_k}{\partial \theta} &= \frac{n_{1k}}{2\sigma_1^2}\big(\bar{X}_{1k} - \mu - \tfrac{\theta}{2}\big) - \frac{n_{2k}}{2\sigma_2^2}\big(\bar{X}_{2k} - \mu + \tfrac{\theta}{2}\big),\\
\frac{\partial \ell_k}{\partial \mu} &= \frac{n_{1k}}{\sigma_1^2}\big(\bar{X}_{1k} - \mu - \tfrac{\theta}{2}\big) + \frac{n_{2k}}{\sigma_2^2}\big(\bar{X}_{2k} - \mu + \tfrac{\theta}{2}\big),
\end{aligned}$$

where

$$\bar{X}_{1k} = \frac{1}{n_{1k}}\sum_{i=1}^{n_{1k}} X_{1i}, \qquad \bar{X}_{2k} = \frac{1}{n_{2k}}\sum_{i=1}^{n_{2k}} X_{2i}.$$

Profiling out μ (by solving $\partial\ell_k/\partial\mu = 0$), we obtain

$$S_k(\theta) = \frac{n_{1k}n_{2k}}{\sigma_2^2 n_{1k} + \sigma_1^2 n_{2k}}\big(\bar{X}_{1k} - \bar{X}_{2k} - \theta\big).$$

The Fisher information is

$$I_k = \frac{n_{1k}n_{2k}}{\sigma_2^2 n_{1k} + \sigma_1^2 n_{2k}},$$

and hence the standardized score statistic is

$$Z_k(\theta) = \sqrt{\frac{n_{1k}n_{2k}}{\sigma_2^2 n_{1k} + \sigma_1^2 n_{2k}}}\big(\bar{X}_{1k} - \bar{X}_{2k} - \theta\big).$$

4 Sequential Monitoring

Under the null hypothesis H_0, the vector of standardized scores $\boldsymbol{Z}_k := (Z_1, \ldots, Z_k)$ follows a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{I}_k$, where

$$[\boldsymbol{I}_k]_{ij} = \sqrt{\frac{\min(I_i, I_j)}{\max(I_i, I_j)}}.$$

Given an α-spending function $\alpha(t_k^*)$ with $t_k^* = I_k/I_K$ (where I_K is the total information at the final analysis), we choose thresholds z_k^* such that

$$\mathbb{P}\big(|Z_k| > z_k^* \mid |Z_1| \leqslant z_1^*, \ldots, |Z_{k-1}| \leqslant z_{k-1}^*,\, H_0\big) = a_k,$$

where $a_k = \alpha(t_k^*) - \alpha(t_{k-1}^*)$ and $\alpha(t_0^*) = 0$.

The thresholds are computed iteratively at each interim look k:

- For $k = 1$: closed form
$$\mathbb{P}(|Z_1| \geqslant z_1^*) = a_1 \implies z_1^* = \Phi^{-1}\Big(1 - \frac{a_1}{2}\Big).$$

- For $k \geqslant 2$: find z_k^* by one-dimensional root finding on
$$f(z) = \frac{\mathbb{P}\big(|Z_1| < z_1^*, \ldots, |Z_{k-1}| < z_{k-1}^*,\ |Z_k| \geqslant z\big)}{\mathbb{P}\big(|Z_1| < z_1^*, \ldots, |Z_{k-1}| < z_{k-1}^*\big)} - a_k,$$
using multivariate normal probabilities with covariance matrix $\boldsymbol{I}_k$.

Note: When variances are unknown, the current estimate of Fisher information
$$\widehat{I}_k \;=\; \frac{n_{1k}\,n_{2k}}{s_{2k}^2\,n_{1k} \;+\; s_{1k}^2\,n_{2k}}.$$
is used to compute t_k^*, yielding an *information-based monitoring* procedure that preserves type-I error asymptotically.

Appendix

A Variance of Score Increments

We show by induction that
$$\operatorname{Var}(S_i - S_{i-1}) = I_i - I_{i-1}.$$
The base case $i = 1$ is immediate since $\operatorname{Var}(S_1) = I_1$. Assuming the result holds up to i, we have
$$I_{i+1} = \operatorname{Var}(S_{i+1}) = \sum_{j=1}^{i+1} \operatorname{Var}(S_j - S_{j-1}) = \operatorname{Var}(S_{i+1} - S_i) + \sum_{j=1}^{i} (I_j - I_{j-1}) = \operatorname{Var}(S_{i+1} - S_i) + I_i,$$
so
$$\operatorname{Var}(S_{i+1} - S_i) = I_{i+1} - I_i.$$

B Weights of the Standardized Decomposition

From
$$Z_k = \frac{1}{\sqrt{I_k}} \sum_{i=1}^{k} (S_i - S_{i-1}),$$
we can write
$$Z_k = \sum_{i=1}^{k} \frac{\sqrt{I_i - I_{i-1}}}{\sqrt{I_k}} \frac{S_i - S_{i-1}}{\sqrt{I_i - I_{i-1}}} = \sum_{i=1}^{k} w_{ik} Z_i^*,$$
where $w_{ik} = \sqrt{(I_i - I_{i-1})/I_k}$.

C Covariance Between Standardized Scores

Using independence of increments,
$$\begin{aligned}
\operatorname{Cov}(Z_j, Z_k) &= \operatorname{Cov}\left(\sum_{i=1}^{j} w_{ij} Z_i^*, \sum_{i=1}^{k} w_{ik} Z_i^*\right) \\
&= \sum_{i=1}^{j} w_{ij} w_{ik} \operatorname{Var}(Z_i^*) \\
&= \frac{1}{\sqrt{I_j I_k}} \sum_{i=1}^{j} (I_i - I_{i-1}) = \sqrt{\frac{I_j}{I_k}}.
\end{aligned}$$

D Profiling Out μ

We first rewrite $\partial\ell_k/\partial\theta$ as

$$\frac{\partial \ell_k}{\partial \theta} = \frac{n_{1k}}{2\sigma_1^2}\big(\bar{X}_{1k} - \frac{\theta}{2}\big) - \frac{n_{2k}}{2\sigma_2^2}\big(\bar{X}_{2k} + \frac{\theta}{2}\big) - \underbrace{\mu\big(\frac{n_{1k}}{2\sigma_1^2} - \frac{n_{2k}}{2\sigma_2^2}\big)}_{=:\xi},$$

Solving $\partial\ell_k/\partial\mu = 0$ yields

$$\mu = \frac{\sigma_2^2 n_{1k}(\bar{X}_{1k} - \theta/2) + \sigma_1^2 n_{2k}(\bar{X}_{2k} + \theta/2)}{\sigma_2^2 n_{1k} + \sigma_1^2 n_{2k}}.$$

Plugging this into ξ gives

$$\begin{aligned}
\xi &= \left(\frac{\sigma_2^2\, n_{1k}(\bar{X}_{1k} - \theta/2) + \sigma_1^2\, n_{2k}(\bar{X}_{2k} + \theta/2)}{\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k}}\right)\left(\frac{\sigma_2^2\, n_{1k} - \sigma_1^2\, n_{2k}}{2\sigma_1^2\sigma_2^2}\right)\\
&= \left(\frac{\sigma_2^2\, n_{1k} - \sigma_1^2\, n_{2k}}{\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k}}\right)\left(\frac{n_{1k}}{2\sigma_1^2}\big(\bar{X}_{1k} - \frac{\theta}{2}\big) + \frac{n_{2k}}{2\sigma_2^2}\big(\bar{X}_{2k} + \frac{\theta}{2}\big)\right).
\end{aligned}$$

Finaly, factoring out $\bar{X}_{1k} - \theta/2$ and $\bar{X}_{2k} + \theta/2$ in $\partial\ell_k/\partial\theta$, we get

$$\begin{aligned}
\frac{\partial \ell_k}{\partial \theta} &= \frac{n_{1k}}{2\sigma_1^2}\left(1 - \frac{\sigma_2^2\, n_{1k} - \sigma_1^2\, n_{2k}}{\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k}}\right)\big(\bar{X}_{1k} - \frac{\theta}{2}\big) - \frac{n_{2k}}{2\sigma_2^2}\left(1 + \frac{\sigma_2^2\, n_{1k} - \sigma_1^2\, n_{2k}}{\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k}}\right)\big(\bar{X}_{2k} + \frac{\theta}{2}\big)\\
&= \frac{n_{1k}\, 2\,\sigma_1^2\, n_{2k}}{2\,\sigma_1^2(\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k})}\big(\bar{X}_{1k} - \frac{\theta}{2}\big) - \frac{n_{2k}\, 2\,\sigma_2^2\, n_{1k}}{2\,\sigma_2^2(\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k})}\big(\bar{X}_{2k} + \frac{\theta}{2}\big)\\
&= \frac{n_{1k}\, n_{2k}}{\sigma_2^2\, n_{1k} + \sigma_1^2\, n_{2k}}\big(\bar{X}_{1k} - \bar{X}_{2k} - \theta\big).
\end{aligned}$$