The Statistics of Sequential Testing

Patrick Le

Abstract

Keywords:

1. Introduction

2. Group Sequential Testing

2.1. Setting

Data arrives over the course of K periods, each period containing N observations. At the end of each period k, $k = 1, ..., K$, a z-test is performed on the data which has accumulated thus far (so using kN observations) to determine if the mean is statistically different from zero. The z-statistic is compared to some critical value, and the null is rejected if the critical value is exceeded. The tests are performed sequentially until the null is rejected, or until all K tests have been performed.

If the critical value is the usual 1.96, then the fact that multiple tests are performed on the same data implies that Type I error will be inflated to be more than the desired 5%. The goal of sequential testing techniques is to control for Type I error. While Bonferonni corrections can be used, they are overly conservative (leading to Type I errors which are too small) and reduce the power of the test.

The one way to control for Type I error in sequential testing is to control the critical value in each test. Depending on the application, one might want the critical value to be the same for each test, or vary depending on the timing of the test. In this specific note, we show how to compute the shared critical value that can be used for each test.

This setting can be generalized to include applications like A/B testing.

2.2. The correlation structure of z-statistics

The fundamental insight of sequential testing over Bonferonni corrections is that there is significant correlation between the test at some period j and the test at some later period k. This is because both tests utilize some shared data. I now derive the correlation between the z-statistic at a period j, called z^j with the z-statistic at some other, later, period k, called z^k.

Suppose that the null is that the data, denoted by $X = (x_i)_{i=1,...,KN}$, has a mean of zero, and variance σ^2. Then using the first kN observations, the sample mean is $\bar{x}^k = \frac{\sum_{i=1}^{kN} x_i}{kN}$, which has standard deviation $\frac{\sigma}{\sqrt{kN}}$. The resulting z-statistic is given by

$$z^k = \frac{\bar{x}^k}{\frac{\sigma}{\sqrt{kN}}} = \frac{\sum_{i=1}^{kN} x_i}{kN\frac{\sigma}{\sqrt{kN}}}. \tag{1}$$

Similarly, the z-statistic for the test done at period j is given by

$$z^j = \frac{\bar{x}^j}{\frac{\sigma}{\sqrt{jN}}} = \frac{\sum_{i=1}^{jN} x_i}{jN\frac{\sigma}{\sqrt{jN}}}. \tag{2}$$

It is a standard result that the z-statistics are, under the null hypothesis, distributed asymptotically as a standard Normal (0,1) random variable.

March 23, 2020

Using the fact that these variables have mean zero, the covariance between z^j and z^k is given by:

$$Cov(z^j, z^k) = E(z^j, z^k) = E\left(\frac{\sum_{i=1}^{kN} x_i}{kN\frac{\sigma}{\sqrt{kN}}} \frac{\sum_{i=1}^{jN} x_i}{jN\frac{\sigma}{\sqrt{jN}}}\right). \tag{3}$$

Consider only the numerator for now, and rewrite the second sum as

$$\begin{aligned}
E\left(\sum_{i=1}^{jN} x_i \sum_{i=1}^{kN} x_i\right) &= E\left(\sum_{i=1}^{jN} x_i \left(\sum_{i=1}^{jN} x_i + \sum_{i=jN+1}^{kN} x_i\right)\right)\\
&= E\left(\sum_{i=1}^{jN} x_i \sum_{i=1}^{jN} x_i\right)\\
&= E\left(\sum_{i=1}^{jN} x_i^2\right)\\
&= jN\sigma^2,
\end{aligned}$$

where the second and third line use the fact that $E(x_i x_{i'}) = 0$ for any $i \neq i'$ (because the x_i's are i.i.d. mean zero), and the last line uses the fact that variance of x is σ^2.

Plugging the numerator into the expression in (3) gives

$$Cov(z^j, z^k) = E(z^j, z^k) = \frac{kN\sigma^2}{kN\frac{\sigma}{\sqrt{kN}} jN\frac{\sigma}{\sqrt{jN}}} = \sqrt{\frac{j}{k}}.$$

Therefore, $Cov(z^j, z^k) = \sqrt{\frac{j}{k}}$.

More generally, let Σ be a $K \times K$ variance covariance matrix for the z-statistics, then entries in Σ in the are given by

$$\Sigma(j,k) = \sqrt{\frac{\min(j,k)}{\max(j,k)}}, \tag{4}$$

for row j, column k.

2.3. *Computation of constant critical Z value*

Assume that an overall Type I error of α is desired, then the critical Z value, Z_c, needs to satisfy:

$$P(|z^1| \geq Z_c, \text{ or } |z^2| \geq Z_c, \text{ or ... or } |z^K| \geq Z_c) = \alpha.$$

In other words, the probability that any of the z-statistics from any of the K tests is greater or equal to Z_c is α.

The critical Z value is determined via the following simulation procedure.

- Draw a sample of size M of multivariate normal variables with mean zero and covariance matrix Σ, to obtain a matrix Z of z-statistics, with shape $M \times K$. This is essentially simulating the distribution of the z-statistics for each of the K sequential z-tests.
- For each draw, compute the maximum absolute value, to obtain a matrix Z_{abs} of maximum absolute z-statistics, with shape $M \times 1$
- Compute the $(1-\alpha)$-percentile from the empirical distribution of Z_{abs}. This will be the critical value Z_c.

By construction, the above procedure yields a critical value Z_c that guarantees that the probability of the z-statistic from any of the z-tests exceeding the critical value Z_c is exactly α.