1. The van der Waerden test

This test we’re looking at this time (and the Wilcoxon rank sum test from Tuesday) are actually instances of linear rank tests, whose test statistics are of the form

\[ T = \sum_{i=1}^{m+n} a_i Z(i) \]

for some specified \( a_i \), where \( Z(i) = 1[\text{the } i\text{th pooled, ordered datum is a } Y] \). The two-sample testing situation, null/alternative hypotheses and assumptions are the exact same as the previous homework (Tuesday), so see that for more details.

Let \( N = n + m \), then the van der Waerden test arises from the following choice of \( a_i \):

\[ a_i = \Phi^{-1} \left( \frac{i}{N+1} \right) \]

Now, under the null, it is true that \( T \), with this choice of \( a_i \), is asymptotically distributed as

\[ T \sim \mathcal{N} \left( 0, \frac{mn}{N(N-1)} \sum_{i=1}^{N} \Phi^{-1} \left( \frac{i}{N+1} \right)^2 \right) \]

In other words

\[ \text{Var } T = \frac{mn}{N(N-1)} \sum_{i=1}^{N} \Phi^{-1} \left( \frac{i}{N+1} \right)^2 \]

So the idea of the test is to reject if \( T \) ends up too far in the left or right tails of this normal approximation.

Your task is the following:

- Show that \( \text{Var } T \) is of the given form.
- Simplify the given expression for \( \text{Var } T \), asymptotically. Use the fact that it’s a Riemann sum approximation to some integral, and write it as that integral. Then simplify/evaluate this integral so there are no \( \Phi \)'s left in the expression.