1. **Wald-Wolfowitz test (exact)**

Let $X_1 \ldots X_n$ and $Y_1 \ldots Y_m$ be two data samples, where the null hypothesis says they come from the same distribution. Suppose they are put into a single dataset and ordered. Then the Wald-Wolfowitz test statistic $R$ is the number of “runs” of either variable—going along the data by their order, $R$ counts the number of times that either an $X$ datum is followed by a $Y$ datum or a $Y$ is followed by an $X$ (plus 1 for the first “run”).

Through a combinatorial argument, in class we showed that for when $r$ is even (so you can write it as $r = 2s$ for some $s$), we have under the null

$$P(R = r) = P(R = 2s) = \frac{2(m-1)(n-1)}{(m+n)}$$

Now you show that when $r$ is odd (so you can write it as $r = 2s - 1$ for some $s$), we have

$$P(R = r) = P(R = 2s - 1) = \frac{(m-1)(n-2) + (m-2)(n-1)}{(m+n)}$$

2. **Wald-Wolfowitz test (asymptotic)**

Provided $m$ and $n$ go to infinity, the runs test statistic $R$ from the previous problem has an asymptotic normal distribution. (It’s sufficiently close to use if $m$ and $n$ are both greater than about 10). You need not prove this—it’s not quite clear (to me, at least) how to write $R$ as a sum of iid random variables, anyway, so to use the CLT. Nevertheless, it is true that

$$R \sim \mathcal{N}(\mathbb{E}R, \text{Var}R) = \mathcal{N} \left( 1 + \frac{2mn}{m+n}, \frac{2mn(2mn - (m+n))}{(m+n)^2(m+n-1)} \right)$$

when $m$ and $n$ are very large.

It is true, however, that whether $m$ and $n$ are big or small, we have the exact quantities:

- $\mathbb{E}R = 1 + \frac{2mn}{m+n}$
- $\text{Var}R = \frac{2mn(2mn - (m+n))}{(m+n)^2(m+n-1)}$

Your task is to prove the first bullet point. As an optional challenge you may also prove the second.
By the way,

The Wald-Wolfowitz test can also be used to test whether some ordered data sample is actually a mix of two distributions, or generally as a test of randomness—just let the $X$’s be all those above some threshold and the $Y$’s the ones below, for instance the median. This test was invented in the 1940s. According to the professor, nobody uses this test nowadays—I suppose its power doesn’t increase fast enough with respect to sample size?—but it’s still of some historical interest.