Knox Space-Time Interaction Test

Sriram Pemmaraju

Jan 22, 2020

Suppose n events $e_1, e_2, \ldots, e_n$ have occurred and for each event e_i we have a spatial location ℓ_i and a time stamp t_i. For example, the events could be all known influenza cases in the state of Iowa in the 2018-19 influenza season. Then, the location of each case might be a zip code and its time stamp might be a date. We want to test if there is *space-time interaction* among these events. This is just a fancy way of asking whether pairs of events that are close in space also tend to be close in time and whether this tendency is greater than what can be explained by randomness. The Knox test, proposed in 1964 by G. Knox ("The detection of space-time interactions", *Applied Statistics* 13, 25-29, 1964) is a widely used statistical test for space-time interaction among a set of events. There are other tests for space-time interactions that have a similar flavor including the Mantel test, Jacquez test, k-nearest neighbor test, etc.

The Knox statistic

Fix thresholds $D > 0$ and $T > 0$. For any pair of events e_i and e_j, $1 \leq i < j \leq n$, define

$$a_{ij}^{\ell} = \begin{cases} 1 \text{ if } \|\ell_i - \ell_j\| \leq D, \\ 0 \text{ otherwise.} \end{cases}$$

$$a_{ij}^{t} = \begin{cases} 1 \text{ if } \|t_i - t_j\| \leq T, \\ 0 \text{ otherwise.} \end{cases}$$

Define

$$y = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_{ij}^{\ell} \cdot a_{ij}^{t}.$$

Here $\| \cdot \|$ denotes "distance" either in space or in time. Therefore, a_{ij}^{ℓ} simply indicates if cases e_i and e_j are at distance at most D in space from each other. Similarly, a_{ij}^{t} indicates if cases e_i and e_j are at distance at most T in time from each other. This means that the product $a_{ij}^{\ell} \cdot a_{ij}^{t}$ in the above sum is 1 iff cases e_i and e_j are at distance at most D in space from each other *and* distance at most T in time from each other. This product is 0 otherwise. As a result y is simply the *count* of all the event pairs (e_i, e_j), $1 \leq i < j \leq n$ that occur within D space units and within T time units from each other. Note that y is some integer value in the range 0 through $n(n-1)/2$. One possible setting of D and T for our influenza example could be $D = 10$ miles and $T = 3$ days.

Distribution of Knox statistic, assuming no space-time interaction

To determine if y is large relative to the count we would obtain if there were *no* association between the locations of events and time stamps of events, we proceed as follows. Let π be a permutation of $(1, 2, \ldots, n)$ chosen uniformly at random from all $n!$ permutations. From the given events $e_1, e_2, \ldots, e_n$ we create n random events $e'_1, e'_2, \ldots, e'_n$ by using the following rules: (i) each e'_i has the same location as e_i, and (ii) the time stamp of e'_i is the time stamp of $e_{\pi(i)}$. In plain English, what we are doing is constructing a set of n random events $e'_1, e'_2, \ldots, e'_n$ from $e_1, e_2, \ldots, e_n$ by leaving the locations as they are, but permuting the time stamps according to the random permutation π. Note that the set of locations and the set of time stamps stay as they are in the given data; we are simply breaking up the given association between the locations and time stamps and replacing this with a random association. Now let Y denote the count of all the random event pairs (e'_i, e'_j), $1 \leq i < j \leq n$ that occur within D space units and within T time units from each other.

The Knox test

Note that Y is a discrete random variable because it can take on different (integer) values with specific probabilities. The Knox test aims to compare the observed count y that we get from the given set of events with the distribution of Y. Suppose we have the *null hypothesis* H of "No space-time interaction". The probability $\text{Prob}(Y \geq y)$ is defined as the *p value* of the test. If y occurs at the extreme right of the distribution of Y, then $\text{Prob}(Y \geq y)$ will be very small and we can reject the null hypothesis of no space-time interaction. In plain English, if y occurs at the extreme right of the distribution of Y it means that y is very unlikely to be observed under the null hypothesis. Thus, in this case we can reject the null hypothesis.

Implementing the Knox test

My final remark is about computing $\text{Prob}(Y \geq y)$. The number of permutations π of $(1, 2, \ldots, n)$ is $n!$, which is enormous even for relatively small n. As a result, we don't know how to efficiently compute $\text{Prob}(Y \geq y)$ exactly and so we estimate this distribution by using the Monte Carlo method. We simply generate a large number of permutations π at random and for each π we compute Y. This gives us an estimate of the distribution of Y and we then calculate $\text{Prob}(Y \geq y)$ from this estimated distribution.
