University of California, Los Angeles
Department of Statistics

Statistics C173/C273 **Instructor: Nicolas Christou**

Homework 3

EXERCISE 1

Consider the county adjacency file downloaded from
`https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html`. You can read the data in R using:

```
a <- read.table("county_adjacency.txt", sep="\t", fill=FALSE, strip.white=TRUE)[,c(1,3)]
```

Answer the following questions:

a. Construct the adjacency matrix for the counties of the State of Iowa, $w_{ii} = 0, w_{ij} = 1$, etc., as we discussed in class.

b. Find the coordinates of the seat of each county of the State of Iowa and compute the Euclidean distance matrix.

EXERCISE 2

Show that the Moran's I statistic is not affected if the data are multiplied by a constant λ. For example, using $Z(s_i), i = 1, \ldots, n$ or $\lambda Z(s_i), i = 1, \ldots, n$ will give the same I statistic.

EXERCISE 3

Consider the 2×3 regular lattice with observations $Z(s_1) = 5, Z(s_2) = -3, Z(s_3) = -6, Z(s_4) = 2, Z(s_5) = 4, Z(s_6) = -2$ as in the following layout:

	Column 1	Column 2	Column 3
Row 1	$Z(s_1) = 5$	$Z(s_2) = -3$	$Z(s_3) = -6$
Row 2	$Z(s_4) = 2$	$Z(s_5) = 4$	$Z(s_6) = -2$

For the adjacency matrix assume that the lattices are connected as in the "rook" definition, i.e. there is connectivity east or west of the lattice and north or south of the lattice. For example, lattice 2 is adjacent to lattices 1 to west direction, 3 to east direction, and 5 to south direction.

Answer the following questions:

a. Find the mean and variance of the Moran's I statistic using the permutation test by enumerating all the 6! permutations of the lattice. Construct the histogram of the 6! values of I.

b. Under randomization, the mean of I is given by $E(I) = -\frac{1}{n-1}$ and
$EI^2 = \frac{n(n^2-3n+3)S_1 - nS_2 + 3S_0^2 - b[(n^2-n)S_1 - 2nS_2 + 6S_0^2]}{(n-3)(n-2)(n-1)S_0^2}$.
Note:
$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$,
$S_1 = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$,
$S_2 = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} + w_{ji} \right)^2$,
$b = n \frac{\sum_{i=1}^n [Z(s_i) - \bar{Z}]^4}{\left(\sum_{i=1}^n [Z(s_i) - \bar{Z}]^2 \right)^2}$.

Compute the empirical $p-$value using the permutation test under the hypothesis of no spatial autocorrelation (use the data from (a)) and compare it with the $p-$value based on normal approximation using the formulas in (b).

EXERCISE 4

Answer the following questions:

a. Suppose $X_1, \ldots, X_n$, the number of events that occur in lattices $i = 1, \ldots, n$, are i.i.d. Poisson(λ). Let $S = \sum_{i=1}^{n} X_i$. Show that $(X_1 = x_1, \ldots, X_n = x_n)$ conditioned on $S = N$ follows the multinomial distribution with parameters S and $(\frac{1}{n}, \ldots, \frac{1}{n})$.
Hint 1: Find the joint pmf of $X_i's$. Given that $S = N$, $\sum_i X_i = N$. Use this result in the joint pmf of the $X_i's$.
Hint 2: Continue by expressing the conditional pmf as the ratio of the joint and the marginal pmf's.
Note on the multinomial probability distribution:
A sequence of n independent experiments is performed and each experiment can result in one of r possible outcomes with probabilities $p_1, p_2, \ldots, p_r$ with $\sum_{i=1}^{r} p_i = 1$. Let X_i be the number of the n experiments that result in outcome i, $i = 1, 2, \ldots, r$. Then, $P(X_1 = x_1, X_2 = x_2, \ldots, X_r = x_r) = \frac{n!}{n_1!n_2!\cdots n_r!} p_1^{x_1} p_2^{x_2} \cdots p_r^{x_r}$.

b. Refer to question (a). Suppose we know that $E[X(X-1)] = \lambda^2$ and $\mathrm{var}[X(X-1)] = 2\lambda^2 + 4\lambda^3$. Please explain how you would verify these two results. Let $T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i(X_i - 1)$. Show that T_1 is unbiased estimator for λ^2 and compute its variance.

c. Refer to question (a). Let $T_2 = E[T_1|S]$. Show that $T_2 = \frac{S(S-1)}{n^2}$. Note that $E[X_i|S] = S\frac{1}{n}$ and $\mathrm{var}(X_i|S) = S\frac{1}{n}\left(1 - \frac{1}{n}\right)$. Is T_2 unbiased estimator of λ^2? Find $\mathrm{var}(T_2)$.