Mathematical Statistics - Section 1 - NYU Spring 2019 - Homework 9

Test of independence based on correlation Let Y, Z be two real-valued random variables and let P be their joint distribution, i.e. the distribution of the random vector (Y, Z).

We recall that if Y, Z are *independent*, then they are *uncorrelated* in the sense that

$$\rho := \frac{\mathrm{Cov}(Y, Z)}{\sqrt{\mathrm{Var}(Y)\mathrm{Var}(Z)}}$$

is equal to 0. We want to design a "test of independence" based on this fact.

Let $X_1 = (Y_1, Z_1), \ldots, X_n = (Y_n, Z_n)$ be an observation with n data points. *Reminder*: it means that the $X_i's$ are independent and have the same distribution P.

We write $\bar{Y}_n := \frac{1}{n}\sum_{i=1}^n Y_i$ and $\bar{Z}_n := \frac{1}{n}\sum_{i=1}^n Z_i$. We also denote by $\overline{YZ}_n$ the quantity

$$\overline{YZ}_n := \frac{1}{n}\sum_{i=1}^n Y_i Z_i.$$

1. **Give** an estimator for the covariance $\mathrm{Cov}(Y, Z)$, and also for $\mathrm{Var}(Y)$ and $\mathrm{Var}(Z)$. *We recalled this in class.*

2. **Give** an estimator $\hat{\rho}$ for the correlation of Y, Z. *Simply put the answers to the previous question together.*

3. **State** asymptotic normality results for the following quantities: $\bar{Y}_n, \bar{Z}_n$ and $\overline{YZ}_n$.

   An application of CLT. Make sure that your answer includes the means and variances involved.

   <u>In the following we believe and assume that Y and Z are independent.</u>

4. **Compute** the covariance matrix of the vector $\begin{pmatrix} \bar{Y}_n \\ \bar{Z}_n \\ \overline{YZ}_n \end{pmatrix}$.

   Keep in mind: we have assumed that Y, Z are independent.

5. The next step is to get an asymptotic result for $\hat{\rho}$. This would lead to some involved computations. Let's admit the following thing[1]: we obtain

$$\hat{\rho} = 0 + \frac{1}{\sqrt{n}}E,$$

[1] This is not the true result, but the details are not relevant

where E is some centered random variable, with $\mathrm{Var}(E) \leq 1$. For any $c > 0$, **give an upper bound** on

$$\mathbb{P}\left[|\hat{\rho}| \geq c\right].$$

Use an inequality that we have not used in a long time.

6. **Define** a test of independence based on $\hat{\rho}$.

7. **Comment** on type I and type II errors.