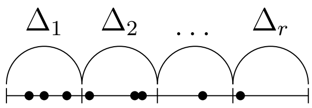
MSAI Statistics Home Assignment 4

Problem 1. (6 points) Let's prove that Pearson χ^2 test statistic follows $\chi^2(r-1)$ distribution.



Suppose that we have a sample $X = (x_1, x_2, \dots, x_n) \in \mathbb{R}$ of real numbers. Let's split the real line $\mathbb{R}$ into r intervals Δ_i and let's count how many samples reside in each interval:

$$\nu_i = \sum_k \mathbb{I}\mathrm{nd}\{\mathrm{x_k} \in \Delta_\mathrm{i}\}$$

Under null hypothesis we are able to estimate the probabilities of a sample appearing in each interval under the null hypothesis:

$$p_i = \int_{\Delta_i} p(x)\mathrm{d}x$$

Then the Pearson χ^2 test statistic is:

$$T(X) = \sum_{i=1}^{r} \frac{(\nu_i - np_i)^2}{np_i} = \sum_{i=1}^{r} \xi_i^2$$

- (1 point) Find ξ_i
- (1 point) Show that $\xi = (\xi_1, \dots, \xi_r)$ is degenerate, i.e. $\exists b : \xi^\top b = 0$
- (1 point) Use CLT to prove that ξ_i has normal distribution and find its parameters
- (3 points) Find $\mathbb{E}[\xi_i \xi_j]$ for $i = j$ and $i \neq j$

This almost concludes the proof. The rest of the proof is as follows:

Now we know the elements of $B = \mathbb{E}[\xi\xi^\top]$. We can use the multivariate CLT to have that $\frac{1}{\sqrt{n}}\xi \to \eta \sim \mathcal{N}(0, B)$.

Next, we can show that $B = E - bb^\top$. With a bit of linear algebra with can also prove that B has $r-1$ eigenvalues equal to 1 and 1 eigenvalue equal to 0. Therefore there exists orthogonal matrix U such that $UBU^\top = \mathrm{diag}(\underbrace{1, \dots, 1}_{\mathrm{r-1}}, 0)$.

Finally, if we consider random variable $\hat{\eta} = U\eta$, we will see that $\hat{\eta} \sim \mathcal{N}(0, UBU^\top)$. The covariance matrix is diagonal, therefore elements of $\hat{\eta}$ are independent and follow the univariate normal distribution: $\hat{\eta}_i \sim \mathcal{N}(0, 1)$. Therefore, $T(X) = \sum_{i=1}^{r} \xi_i^2 \to \sum_{i=1}^{r-1} \hat{\eta}_i^2 \sim \chi^2(r-1)$.

Problem 2. (2 points) Suppose you have the following sample:

interval	0	1	2	3	4	5	6	7	8	9
counts	74	92	83	79	80	73	77	75	76	91

Use the Pearson χ^2 test to test hypothesis at $\alpha = 0.05$ level

$$H_0 : X \sim U[0, 9]$$
$$H_1 : X \not\sim U[0, 9]$$

	X	Y
1	-1.75	-0.29
2	-0.33	0.09
3	-1.26	1.70
4	0.32	-1.09
5	1.53	-0.44
6	0.35	-0.29
7	-0.96	0.25
8	-0.06	-0.54
9	0.42	-1.38
10	-1.08	0.32

Table 1: Data for Problem 3.

Problem 3. (2 points) Suppose you have the following two independent samples, $X \sim \mathcal{N}(\mu_x, \sigma^2)$ and $Y \sim \mathcal{N}(\mu_y, \sigma^2)$ (see Table 1). Use Student's two-sample t-test to test the following hypothesis at $\alpha = 0.05$ level (you may assume equal variances):

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y$$

Problem 4. (2 bonus points) In the Table 2 below, X and Y are the reaction times to light and sound signal of the test subjects.

	X	Y
1	176	168
2	163	215
3	152	172
4	155	200
5	156	191
6	178	197
7	160	183
8	164	174
9	169	176
10	155	155
11	122	115
12	144	163

Table 2: Data for Problem 4.

Use Wilcoxon signed rank test to test the following hypothesis at $\alpha = 0.05$ level:

$$H_0 : \mu_x = \mu_y$$
$$H_1 : \mu_x \neq \mu_y$$

For the ease of computation, instead of Wilcoxon statistc W, use standardized test statistic

$$T = \frac{W - \mathbb{E}[W]}{\sqrt{\mathbb{V}\mathrm{ar(W)}}} = \frac{W}{\sqrt{\frac{1}{6}n(n+1)(2n+1)}} \sim \mathcal{N}(0, 1)$$

Problem 5. (4 bonus points) Let $X_1, \ldots, X_n \sim \mathcal{N}(\theta, 1)$. Consider the following test

$$H_0 : \theta = \theta_0 = 0$$
$$H_1 : \theta = \theta_1 = 1$$

Let the rejection region be $R = \{X : T(X) > c\}$, where $T(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$.

- (2 bonus points) Find c so that test has size α
- (2 bonus points) Find test power W

Problem 6. (4 bonus points) Let $X_1, \dots, X_n \sim \mathcal{N}(\theta, \sigma^2)$, where σ^2 is known. Consider the following test

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

- (2 bonus points) Compute likelihood ratio $\Lambda(X)$
- (1 bonus point) Compute $\lambda(X) = 2 \log \Lambda(X)$, find its distribution
- (1 bonus point) Specify rejection region $R = \{X : \lambda(X) > c\}$ using test size α. Introduce change of variables such that the test statistic in new variables follows standard normal distribution. Specify the rejection region R in new variables.