Question 1: Chapter 10 # 6

There is a theory that people can postpone their death until after an important event. To test the theory, Phillips and King (1988) collected data on deaths around the Jewish holiday Passover. Of 1919 deaths, 922 died the week before the holiday and 997 died the week after. Think of this as a binomial and test the null hypothesis that $\theta = 1/2$. Report and interpret the p-value. Also construct a confidence interval for θ.

Test

Let:

$$H_0 : \theta = \frac{1}{2} \quad \text{versus} \quad H_1 : \theta \neq \frac{1}{2}.$$

Let $X \sim Bin(p, n)$. In this case, we sampled 1919 samples so $X \sim Bin(p, 1919)$.

The MLE for p is $\hat{p} = X/n$. We have it that $\hat{p} = 992/1919$

The SE for this estimator is

$$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \approx 0.011$$

The Wald statistic for this estimator is:

$$w = \frac{\hat{\theta} - \theta_0}{\hat{se}} \approx -1.71$$

The p-value for this is (where Φ is the cdf of of the normal distribution):

$$p = 2(1 - \Phi(|W|) \approx 0.087$$

The confidence interval for this statistic is

$$C = \left(\hat{\theta} - z_{0.05} \cdot \hat{se}, \hat{\theta} + z_{0.05} \cdot \hat{se}\right) = (0.458, 0.503)$$

Conclusion

The p-value for this test is 0.087 is inconclusive as 0.05 is usually the cutoff point. It is unclear if the hypothesis is true that people can postpone their deaths.

Code

```
import math
import scipy.stats as st

n = 1919
X = 922

p_hat = X / n
print(p_hat)

se = math.sqrt(p_hat * (1 - p_hat) / n)
print("SE: ", se)

Wald = (p_hat - 0.5) / se
print("Wald: ", Wald)

p_value = 2 * st.norm.sf(abs(Wald))
print("p-value: ", p_value)

confidence_interval = (p_hat - 1.96 * se, p_hat + 1.96 * se)
print("Confidence interval: ", confidence_interval)
```

```
0.48045857217300675
SE:  0.011405138868050746
Wald:  -1.7133879782678245
p-value:  0.08664118646589904
Confidence interval:  (0.4581044999916273, 0.5028126443543862)
```

Question 2: Chapter 10 # 10

Here are the number of elderly Jewish and Chinese women who died just before and after the Chinese Harvest Moon Festival.

Week	Chinese	Jewish
-2	55	141
-1	33	145
1	70	139
2	49	161

Compare the two mortality patterns. (Phillips and Smith (1990))

2.1 Chinese

For this question, we can use Pearson's χ^2 test for this multinomial data where $X = (X_{\text{before}}, X_{\text{after}}) \sim Multinomial(n, p)$ for the Chinese and the Jewish deaths.

We are testing against $p = (1/2, 1/2)$, as we expect that if there is no correlation between the Chinese new year and the death data for the Chinese, then half of all the deaths should be before the new year and half should be after.

$$H_0 : p = (1/2, 1/2) \quad H_1 : p \neq (1/2, 1/2)$$

$X_1 = 88$ and $X_2 = 119$ and $n = 207$

The statistic is thus:

$$T_c = \sum_{j=1}^{2} \frac{(X_j - np_{0j})^2}{np_{0j}} = \tag{1}$$

$$= (88 - 102.5)^2/102.5 + (119 - 102.5)^2/102.5 \tag{2}$$

$$= 4.643 \tag{3}$$

We now need to test this against a χ^2 distribution, as $T_c \to \chi^2_1$.

$$p = 1 - \Phi_1(T_c)$$

Where Φ_1 is the CDF for a χ^2 distribution with 1 degree of freedom. Computing this, we get a p-value of 0.031

2.2 Jewish

This test is the same as above, but with the different data for the Jewish data.

$X_1 = 286$ and $X_2 = 300$ and $n = 586$

The statistic is thus:

$$T_j = \sum_{j=1}^{2} \frac{(X_j - np_{0j})^2}{np_{0j}} = \tag{4}$$

$$= (286 - 293)^2/293.0 + (300 - 293.0)^2/293.0 \tag{5}$$

$$= 0.334 \tag{6}$$

We now need to test this against a χ^2 distribution, as $T_c \rightarrow \chi_1^2$.

$$p = 1 - \Phi_1(T_c)$$

Where Φ_1 is the CDF for a χ^2 distribution with 1 degree of freedom.

Computing this, we get a p-value of 0.563

2.3 Conclusion

The results seem to prove that there is indeed a correlation between the Chinese new year and the death data for the Chinese, while there is not a correlation for the Jewish death data. The p-value for the Chinese Pearson Test is 0.03 which falls below the 0.05 threshold that is typically used in hypothesis testing, which provides strong evidence against the null hypothesis that the number of deaths before and after the Chinese new year would be the same.

On the other hand, for the Jewish data, preforming the same test gave us a p-value of 0.563 which is not strong evidence at all to reject the null hypothesis.

Both tests overall provide strong evidence to the idea that there is some correlation between the Chinese new year and the Chinese death data as there is a strong trend of more Chinese people dying after the new year as opposed to before that exists in the Chinese death data, but does not have strong evidence in the Jewish death data.

Question 3:

Let Let $Y_k = \sum_{i=1}^{k} Z_i^2$ be the sum of k independent squared standard Gaussians (i.e., $Z \sim N(0,1)$). We call such a random variable a χ^2 distribution with k degrees of freedom.

3.1 Expected Value

$$\mathbb{E}(Y_k) = \mathbb{E}\left(\sum_{i=1}^{k} Z_i^2\right) \tag{7}$$

$$= \sum_{i=1}^{k} \mathbb{E}\left(Z_i^2\right) \tag{8}$$

$$= \sum_{i=1}^{k} Var(Z_i) \tag{9}$$

$$= \sum_{i=1}^{k} 1 \qquad \because Z_i \sim N(0,1) \tag{10}$$

$$= k \tag{11}$$

$$\therefore \mathbb{E}(Y_k) = k \tag{12}$$

3.2 Variance

$$Var(Y_k) = Var\left(\sum_{i=1}^{k} Z_i^2\right) \tag{13}$$

$$= \sum_{i=1}^{k} Var\left(Z_i^2\right) \tag{14}$$

$$= \sum_{i=1}^{k} Var(Z_i) + Var(Z_i) \quad \because Var\left(Z_i^2\right) = Var\left(Z_i * Z_i\right) = Var(Z_i) + Var(Z_i) \tag{15}$$

$$= \sum_{i=1}^{k} 2 \qquad \because Var(Z_i) = 1 \tag{16}$$

$$= 2k \tag{17}$$

$$\therefore Var(Y_k) = 2k \tag{18}$$

Question 4:

Let X be uniform on $[0, 1]$. Fix K a positive integer, and sample $\{x_i\}_{i=1}^n$ i.i.d. from X. For $k = 1, 2, \ldots, K$, define

$$Y_k = \left| \left\{ x_i \mid x_i \in \left[\frac{k-1}{K}, \frac{k}{K} \right] \right\} \right|$$

to be the random variable that counts the number of observations landing in the interval $\left[\frac{k-1}{K}, \frac{k}{K}\right]$.

(a) Show that the random vector $(Y_1, Y_2, \ldots, Y_K)$ defines a multinomial distribution. What are its parameters?

(b) Develop a hypothesis test framework for testing whether data in $[0, 1]$ comes from a uniform distribution, based on the observations in (a).

(c) The role of K is crucial to making (b) work. Discuss how to select a good choice of K, and what can happen if K is taken too small or too large.

(a)

X is uniform on $[0, 1]$ and for $k = 1, \ldots, K$, $\left[\frac{k-1}{K}, \frac{k}{K}\right]$ divides the interval of $[0, 1]$ into K intervals of equal size.

x_i is a random iid sample from X, which is uniform on the interval $[0, 1]$. The probability x_i is in a specific interval of $\left[\frac{k-1}{K}, \frac{k}{K}\right]$ is the same for all intervals as each interval is of equal length and covers the interval $[0, 1]$. The probability that it falls into the kth interval $\frac{1}{K}$ as there are K intervals of equal length on $[0, 1]$ and $x_i \sim Unif(0, 1)$

If we define a new random variable, Y_k to be the number of samples of x_i $i = 1, \ldots, n$ that fall within a given interval, over a number of trials n. $Y_k \sim Bernoulli(n, \frac{1}{K})$.

We define a random vector of multiple Y_ks: $(Y_1, Y_2, Y_3, \ldots Y_n)$. Each, $Y_k \in (Y_1, Y_2, Y_3, \ldots Y_n)$ is Bernoulli, so the random vector is Multinomial with parameters $(n; p_1, p_2, \ldots, p_n)$ where each p_i is such that $Y_i \sim Bernoulli(n, p_i)$.

The parameters for this Multinomial distribution are $(Y_1, Y_2, Y_3, \ldots Y_n) \sim Multinomial(n; p_1, p_2, \ldots, p_n)$. We showed above that each Y_k has the same p_k as x_i is uniform. We showed that $p_k = \frac{1}{K}$ for all $k = 1, \ldots, K$.

Thus

$$(Y_1, Y_2, Y_3, \dots Y_n) \sim Multinomial(n; \frac{1}{K}, \frac{1}{K}, \dots, \frac{1}{K})$$

(b) Hypothesis Test Framework

We want to test that:

$$H_0 : X \sim Uniform(0,1) \quad H_1 : not \;\; X \sim Uniform(0,1)$$

We would imagine that the data we would see if we took groups over the interval $[0,1]$, and counted how many random events fall within that interval to be proportional to the size of the groups and about equal for all groups.

We can develop a framework that follows this intuition as follows.

$$H_0 : p = (\frac{1}{K}, \dots, \frac{1}{K}) \quad H_1 : p \neq (\frac{1}{K}, \dots, \frac{1}{K})$$

We know that $(Y_1, Y_2, Y_3, \dots Y_n) \sim Multinomial(n; \frac{1}{K}, \frac{1}{K}, \dots, \frac{1}{K})$, where $(Y_1, Y_2, Y_3, \dots Y_n)$ are the numbers of random values that fall in to a given kth bucket. Thus, we use Pearson's χ^2 test. We can use the MLE for the vectors of probabilites, which we know for a multinomial distribution is $\hat{p} = (Y_1/n, Y_2/n, \dots, Y_n/n)$. The Pearson Statistic is:

$$T = \sum_{k=1}^{K} \frac{(Y_j - np_{0i})^2}{np_{0i}} = \sum_{k=1}^{K} \frac{(Y_j - \frac{n}{K})^2}{\frac{n}{K}}$$

Which is defined over the number of Y_ks we have which is $k = 1, \dots, K$.

The Pearson Statistic T is $T \to \chi^2_{K-1}$. So, the test for a given α confidence level is

- We fail to reject the null hypothesis (X is uniformly distributed) if $T \leq \chi^2_{K-1,\alpha}$
- We reject the null hypothesis (X is uniformly distributed) if $T > \chi^2_{K-1,\alpha}$

(c) Role of K

K

K is the number of buckets that we partition the interval $[0,1]$ into to preform our test.

Small K

Let $K = 2$. If K is 2 then we divide the region $[0, 1]$ into 2 buckets $[0, 1/2)$ and $[1/2, 1]$. Our test examines if the proportion of x_is that are sampled falls about evenly into these two buckets. But in this case when $K = 2$, that doesn't exactly conclude that X is uniform over $[0, 1]$. All we know is that $P(0 \leq x_i \leq 1/2) \approx P(1/2 \leq x_i \leq 1)$. This test fails to capture the randomness that can occur on the sub intervals, $[0, 1/2)$ and $[1/2, 1]$ that would contradict the hypothesis that X is uniform on $[0, 1]$.

For example, consider that $x_i \in [0, 1/2)$ and $x_i \in [1/2, 1]$ look Gaussian on these respective intervals, but that if we look at the number of occurrences of x_is that fall into each region, we see that they are about the same. Our test would conclude that X is uniform, but it is obviously the case that it is not.

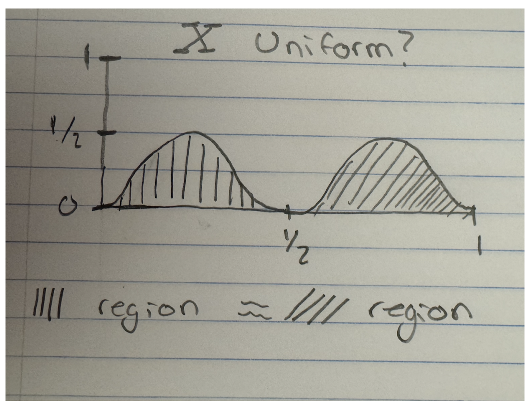

Figure 1: Our Test with K = 2

With a small K, each bin has a high number of samples in it. This makes it harder to detect non-uniform patterns as can be seen with the Gaussian example above.

Large K

As we increase K, we capture more and more of the randomness that could be hidden with a smaller K value and can be more sure that X is distributed uniformly.

However, when K is too large, the variations in X cause the test to fail more often than it should (we reject the uniform distribution hypothesis when we should not be). This is because the number of elements in a given bin is small, so it is more susceptible to the underlying variance of the random variable, as small deviations have much larger consequence.

A Good Choice

A good choice for K lies somewhere in between a large K and a small K and depends on the sample size. The larger the sample size the large the K can be. But generally, a good value for K is about a $\sqrt{n}$, which can be verified experimentally using the p-value for tests on a samples drawn from a uniform distribution.

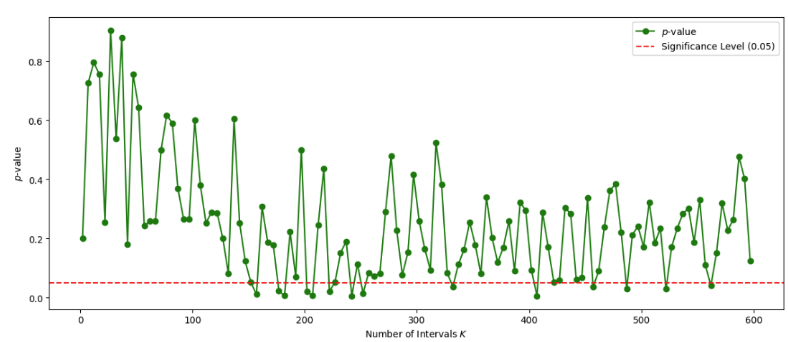

Figure 2: Good Choice for K

Here, the number of samples was 10000, and a K value that minimizes the p value for this test is somewhere between $100 - 200$