ECEN321 - Lab 4
Hypothesis Testing

Joshua Benfell - 300433229

June 6, 2020

1 Introduction

In this report a hypothesis test will be performed and analysed. The null hypothesis is that realisations of a poisson generated random variable is actually from a poisson distribution. This will be tested using a χ^2 test. Consequently the alternate hypothesis is that the poisson random variable is not from a poisson distribution.

2 Method

To test this hypothesis, $M = 100$ inter-arrivals will be generated from a Poisson random variable. This will be done by transforming a uniform random variable into a poisson random variable. To do this we take the negtaive log of the uniform random variable and divide it by the poisson parameter $\lambda = 3$. From here, the cumulative sum is taken of the inter arrival times to find the times when arrivals happen.

Now that an experimental data from a poisson random variable has been set up we can calculated the number of observed events. We take the sum of all the arrivals times that are less than time $t = 5$. This gives us the number of events that happen in less that the time period specified. This can then be plotted against the range of events that can happen $K = [0...t \times \lambda \times 2]$ which gives us an observed pdf of how many events happened in time t.

An expected distribution needs to be constructed for comparison with the χ^2 test. The poisson distribution PDF eq. (1) is used, with the above stated parameters, and scaled by the number of trials N to align it with the observed values and get a count of the expected events in time t.

$$p_{K(t)}(k) = \frac{e^{-\lambda t}(\lambda t)^k}{k!} \tag{1}$$

$$\Sigma \frac{(Observed - Expected)^2}{Expected} \tag{2}$$

To perform a χ^2 test, all the bins of the datasets need to have at least 5 events in them. We pile all values less than 5 into one bin on either side to achieve this. From here the χ^2 value is calculated eq. (2) and compared to the critical value found from a lookup table related to the confidence level (0.01) and the degrees of freedom ν. If the χ^2 value is significantly larger than the critical value, then the null hypothesis can be rejected.

3 Results

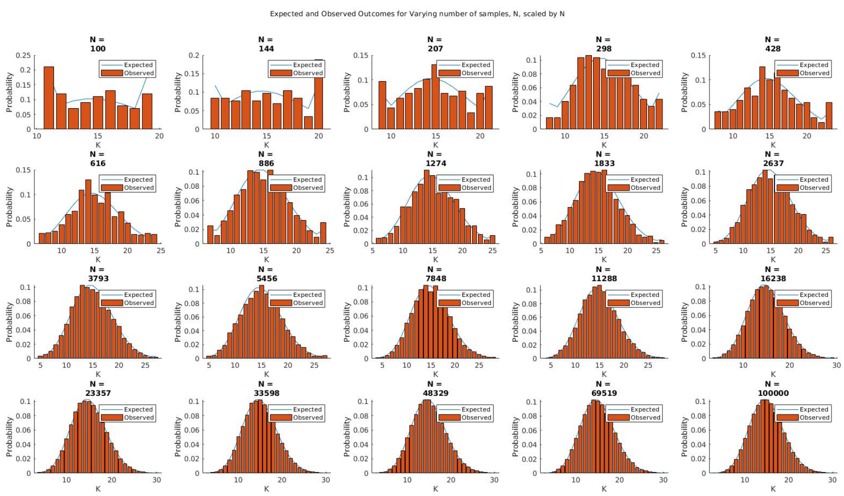

Figure 1: Plots of expected and observed outcomes for different values of N

Figure 1 contains the expected and observed datasets for our trials of various sizes N. Visually we can see that the plots all line up pretty well. At low N values, the combination of the count of events being less than 5 has a significant effect, causing the sides to spike up more than the values beside it. This is reflected in the fact that the plots look more flat than they normally would, as seen at higher values of N.

Table 1 contains all the critical and χ^2 values for each run of N trials. For almost all the values of N, the χ^2 value is less than the critical value. The value that is greater than the critical value has failed the test. On reruns

N	χ^2 **Value**	**Critical Value**
100	6.149282073	20.09023503
144	11.05540555	23.20925116
207	13.04708368	26.21696731
298	13.12529465	29.14123774
428	18.06504567	30.57791417
616	22.74178331	33.40866361
886	15.78066836	33.40866361
1274	24.72741788	36.19086913
1833	29.03443487	37.56623479
2637	33.14307315	38.93217268
3793	12.28919969	40.28936044
5456	28.08687747	40.28936044
7848	22.88582981	42.97982014
11288	26.27790647	42.97982014
16238	31.54759935	44.3141049
23357	14.80776634	45.64168267
33598	53.46831014	46.96294212
48329	34.1201107	46.96294212
69519	40.97674385	46.96294212
100000	30.55728852	46.96294212

Table 1: Table of χ^2 and critical values for different values of N

of the test, there would always be one or two cases where the χ^2 value was larger than the critical value. After N trials of $\sim$ 33000 we can see that critical value has no change, this indicates that doing any more trials than at least this amount (could be lower) you will not get any improvement on your critical value for a confidence level of 1%.

From table 1 it can be concluded that the null hypothesis of realisation of a poisson random variable belonging a poisson distribution is credible. This hypothesis can never be proven as all it takes is one case to not belong to the poisson distribution. Which it looks like we occasionally get with the χ^2 test value being higher than the critical value. So most of the time it is possible to conclude that our realisations belong to a poisson distribution.

4 Conclusion

In this report we covered a hypothesis test on the null hypothesis, realisations of a poisson random variable belong to the poisson distribution and alternate hypothesis that they don't. We tested this hypothesis by using the χ^2 test on a transformed uniform random variable. From performing this test many times for varying N realisations, it was found that this hypothesis is credible and holds true, but it occasionally may fail. It was also found that doing more than $\sim$ 33000 tests leads to no increase in the critical value and therefore no further improvement in the test.

Appendices

A MATLAB Code

```
clear

lambda = 3; %Arrival Rate
M = 100;  %Number of arrivals in time t
t = 5;  %Arbitrary arrival time.
K = 0:round(t*lambda*2); %possible values for number of
    events in time t
cl = 0.01; %confidence level

vals = round(logspace(2,5,20));
chi2Vals = zeros(length(vals));
```

```
critVals = zeros(length(vals));

figure(1)
annotation('textbox', [0 0.9 1 0.1], ...
    'String', "Expected and Observed Outcomes for Varying number of samples, N, scaled by N", ...
    'EdgeColor', 'none', ...
    'HorizontalAlignment', 'center') % Figure title
for I = 1:length(vals)
    subplot(4, 5, I)
    hold on
    N = vals(I); %Number of experiments
    Y = -log(rand(M,N))/lambda; %Inter Arrival Times; exponential random variable
    arrivalTimes = cumsum(Y);

    arrivals = sum(arrivalTimes<=t,1); %Count the number of arrivals in time t
    O_K = hist(arrivals, K); %Observed Events

    P = exp(-lambda*t) * (lambda*t).^K ./ gamma(K+1); % Poisson Dist. Gamma(k+1) = k!
    E_K = N*P; %ExpectedEvents

    % Ensure that each bin has an expected value of at least 5
    %Expected K
    [r, c, v] = find(E_K >= 5);
    E_K(c(1)) = sum(E_K(1:c(1)));
    E_K(c(end)) = sum(E_K(c(end):end));
    E_K = E_K(c(1):c(end));
    plot(K(c(1):c(end)), E_K/N);
    %Observed K
    %no find in this one to keep dimensions consistent
    O_K(c(1)) = sum(O_K(1:c(1)));
    O_K(c(end)) = sum(O_K(c(end):end));
    O_K = O_K(c(1):c(end));
    bar(K(c(1):c(end)), O_K/N);

    xlabel("K")
    ylabel("Probability")
```

```
    legend("Expected", "Observed");
    title(strcat({"N = ", N}));
    %Define degrees of freedom as num bins - 1
    nu = length(E_K) - 1;

    %Chi Square Values
    chi2 = sum((O_K-E_K).^2./E_K);
    crit = chi2inv(1-cl, nu);

    chi2Vals(I) = chi2;
    critVals(I) = crit;
end
```