How to use "Goodness of Fit Tests?"

Paweł Polak

March 24, 2016

STAT W4413: Nonparametric Statistics - Lecture 13

χ^2 test: example

Example (Gibbons and Chakraborti)

A quality control engineer has taken 50 samples of size 13 from a production process. The numbers of defectives for these samples are recorded below. Test the null hypothesis at significance level $\alpha = 0.05$. (a) Perform the χ^2 test when the null is *Binom*(0.1). (b) Perform the χ^2 test when the null is *Binom*(p) and p is not known.

Number of defects	Number of samples
0	10
1	24
2	10
3	4
4	1
5 or more	1

Part (a)

Define $p_i = P(Numberofdefects = i)$. The first step is to estimate $\hat{p}_i$.

Number of defects	Number of samples
0	10
1	24
2	10
3	4
4	1
5 or more	1

Total number of samples = 50

Therefore,

Number of defects	$\hat{p}$
0	.2
1	.48
2	.2
3	.08
4	.02
5 or more	.02

Part (a); cont'd

How to calculate π_i (the probability under null)? From Binomial distribution.

$$\pi_i = \binom{13}{i} p^i (1-p)^{13-i} = \binom{13}{i} 0.1^i (0.9)^{13-i}.$$

number of defects	$\hat{p}$ (from observations)	π (from binomial)
0	.2	0.2542
1	0.48	0.3671
2	.2	0.2448
3	.08	0.0997
4	.02	0.0277
5 or more	.02	.0065

Now we should calculate $Q = \sum_{i=0}^{5} \frac{(\hat{p}_i - \pi_i)^2}{\pi_i} = 0.0885.$

$nQ \xrightarrow{d} \chi^2(5)$. Therefore we have
$>$ kappa $< -$ qchisq(0.95,5)
$>$ kappa
[1]11.0705
$nQ = 4.425$. Therefore we accept the null hypothesis.

Part (b)

(b) Perform the χ^2 test when the null is *Binom*(p) and p is not known.
We have a composite null hypothesis. The first step is to estimate p. To do so we first calculate the likelihood:

$$P(X_1 = k_1, X_2 = k_2, \dots, X_{50} = k_{50})$$
$$= \binom{n}{k_1}\binom{n}{k_2}\dots\binom{n}{k_{50}} p^{k_1+k_2+\dots+k_{50}}(1-p)^{(13-k_1)+(13-k_2)+\dots+(13-k_{50})}.$$

$k_1 + k_2 + \dots + k_{50}$ is the total number of defects and here it is equal to

$$24 \times 1 + 10 \times 2 + 4 \times 3 + 4 \times 1 + 5 \times 1 = 65.$$

$$\frac{d}{dp} \log P(X_1 = k_1, X_2 = k_2, \dots, X_{50} = k_{50}) = 0 \Rightarrow \frac{65}{p} = \frac{650 - 65}{1-p} \Rightarrow \hat{p} = 0.1.$$

Therefore, our Q is the same as before.
However, $nQ \xrightarrow{d} \chi^2(4)$. (You lose one *df* because you estimate one parameter!)
Therefore, we have

```
> kappa <- qchisq(0.95,4)
> kappa
[1]9.487729
```

.... Still the null hypothesis is accepted since $nQ = 4.425$.

Test for Continuous CDFs

Example

When a customer arrives in a bank it takes the bank X minutes to serve him/her. One of the most well-known models for the waiting times is called M/M/1 in which we assume that the distribution of the waiting times for the customers is exponential. As a statistician you would like to check and see if this model works for the bank you are working for. For this purpose, you have collected information about the waiting time of 30 customers and you have obtained

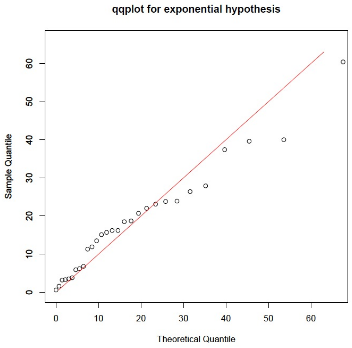
$$\begin{aligned} x \leftarrow c(&0.6, 1.6, 3.2, 3.3, 3.5, 3.8, 5.9, 6.2, \ldots \\ &6.8, 11.3, 11.9, 13.5, 15.1, 15.7, 16.2, \ldots \\ &16.2, 18.5, 18.7, 20.7, 22.0, 23.1, 23.8, \ldots \\ &23.9, 26.4, 27.9, 37.4, 39.6, 40.0, 60.4, 63.0) \end{aligned}$$

Based on these numbers you would like to judge wether the exponential model with parameter $\lambda = 1/20$ is right or not.

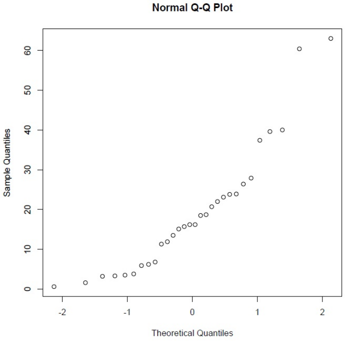
qqplot

$px < - \text{seq}(0,1,1/30)$
$y < -\text{qexp}(px, \text{ rate } = 1/20)$

qqplot(x,y)



qqnorm(x)



Case I: One-Sample, Simple Null

We know the exact null distribution:

$$H_0 : F(x) = 1 - e^{-x/20} \text{ for } x > 0 \qquad \text{versus} \qquad H_1 : F(x) \neq 1 - e^{-\frac{x}{20}}$$

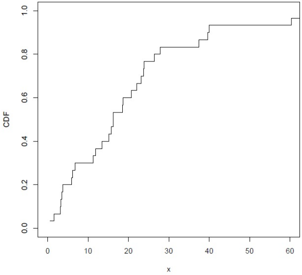
Plotting CDF:

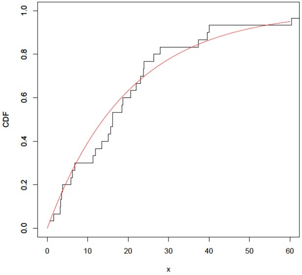
plot(sort(x), (1:30)/30, xlim = c(0,60), ylim = c(0,1), type= "s", xlab = "x", ylab= "CDF")

par(new = T)

z < − seq(0,60, .1)

plot(z, 1- exp(-z/20), type="l",xlim = c(0,60), ylim = c(0,1), xlab = "x", ylab= "CDF")





Kolmogorov-Smirnov Test

$$H_0 : F(x) = 1 - e^{-x/20} \text{ for } x > 0 \quad \text{versus} \quad H_1 : F(x) \neq 1 - e^{-\frac{x}{20}}$$

R code:

- ks.test(x,"pexp",rate=1/20)

Result:

One-sample Kolmogorov-Smirnov test data: x
D = 0.1316, p-value = 0.676
alternative hypothesis: two-sided

Our decision: *p - value* is very high. So, we accept the null hypothesis.

Case II: One-Sample, Composite Null

Checking Gaussianity:

$$H_0 : F(x) = N(\mu, \sigma^2) \quad \text{versus} \quad H_1 : F(x) \neq N(\mu, \sigma^2)$$

Note: (μ, σ^2) are NOT known.

Can we use ks.test(x,"pnorm",mean(x),sd(x))? ....

Kolmogorov-Smirnov Test for composite null

Can we use ks.test(x,"pnorm",mean(x),sd(x))?

- NO.

Let's do it anyway and see what happens.

D = 0.1538, p-value = 0.4767
alternative hypothesis: two-sided

Result: if we do the test this way, p-value is high and we should accept the "Gaussianity". This result is misleading! Why?

Kolmogorov-Smirnov Test for composite null

Correct form of KS test for composite null, aka, Lilliefros test:

- lillie.test(x)

Result:

> Lilliefors (Kolmogorov-Smirnov) normality test
> data: x
> D = 0.1538, p-value = 0.06793

Conclusion:

- Actual p-value is much lower than what we calculated before.
- Should we accept the null hypothesis?

Note

- Lilie test for nromality is in package "nortest"
- Lillie test for exponential is in package "exptest" (called ks.exp.test)

Discussion of KS test

the p-value was 0.06793 for the KS test. Should we accept the Gaussianity?

In such cases it is better to run other tests as well

Anderson-Darling test of normality

ad.test(x).

Result:

- Anderson-Darling normality test
- data: x
- A = 1.0224, p-value = 0.009251

We reject "Null".

ad test of normality is part of package "nortest".

ad test for exponential random variables is part of "exptest" package.

Wilcoxon Test

We would like to evaluate the impact of glucose treatment on the patients with Huntigotn's disease. The result for the patients that received treatment and did not receive the treatment are shown below.

Treatment	85, 89, 86, 91, 77, 93, 100, 82, 92, 86, 86
No treatment	83, 73, 65, 65, 90, 77, 78, 97, 85, 75

We would like to see whether the treatment has had any positive or negative effect. Therefore, considering $X_i \sim F$ for the "with-treatment" samples and $Y_i \sim G$ for "no-treatment" samples we would like to test:

$$H_0 : F(x) = G(x) \ \forall x \text{ vs. } H_1 : F(x) > G(x) \ \forall x \ \text{ or } \ F(x) < G(x) \ \forall x.$$

Wilcoxon Test

Test this hypothesis by Wilcoxon test. Report the p-value.

```
x<-c(85, 89, 86, 91, 77, 93, 100, 82, 92, 86, 86)
y<-c(83, 73, 65, 65, 90, 77, 78, 97, 85, 75)
wilcox.test(x,y,alternative='two.sided',mu=0)
```

```
        Wilcoxon rank sum test with continuity correction

data:  x and y
W = 4680, p-value < 2.2e-16
alternative hypothesis: true location shift is not equal to 0
```

Wilcoxon Test

Now we would like to implement permutation test with Wilcoxon statistic. Since implementing the full permutation test is time consuming we use Monte Carlo simulation with 500, 1000, 1500, 2000 Monte Carlo samples, and report the result of the permutation test with these four number of Monte Carlo samples. Then we will compare it with the result above.

Wilcoxon Test

```
z<-c(x,y)
zr<-rank(z)
p<-numeric(0)
number<-c(500,1000,1500,2000)
m<-length(y)
n<-length(x)
stat<-abs(sum(zr[(length(x)+1):length(z)])-m*(m+1)/2-m*n/2);

for(j in 1:4){
  count<-0;
    for(i in 1:number[j]){
      test<-abs(sum(sample(zr,length(y)))-m*(m+1)/2-m*n/2);

      if(test>=stat) count<-count+1;
    }
  p[j]<-count/number[j];
}
```