

BS1a Applied Statistics

Lectures 15-16

Dr Geoff Nicholls

Week 8 MT10

Monte Carlo Tests

Compare test statistic to distribution simulated under null.

Monte Carlo error (may be) small compared to asymptotic approximation.

Time consuming but accurate.

Example: we couldn't make goodness of fit tests for Bernoulli GLM models because the likelihood ratio test statistic isn't reliably $\chi^2(n-p)$ there. Fix this by simulating distribution of residual deviance under the null.

Simulation methods

We will call functions like

```
runif(n, min=0, max=1)
rnorm(n, mean=0, sd=1)
rpois(n, lambda)
```

to get rv we need. Build more complex distributions from these elements.

Simulation by inversion

Suppose X has cdf $F(x) = \Pr(X < x)$: if $U \sim U(0, 1)$ and $X = F^{-1}(U)$ then

$$X \sim F.$$

Example: from Part A probability, if for $i = 1, 2, \dots, n$ we simulate iid $U_i \sim U(0, 1)$ and set $X_i = -\log(1 - U_i)/\lambda$ then $X_i \sim \text{Exp}(\lambda)$. Here are 4 $\text{Exp}(1/10)$ rv.

```
> lambda=1/10; (X<- -log(runif(n=4))/lambda)
[1] 0.000782 6.749159 9.298347 6.831243
```

Simulation from elementary distributions

We take functions of rv to form new rv.

Example: if $X_i \sim \text{Exp}(\lambda)$ are iid, and $Y = \sum_{i=1}^n X_i$ then (as we saw in Part A) we have $Y \sim \Gamma(n, \lambda)$. A Gamma(4, 1/10) rv:

```
> Y<-sum(X)
[1] 22.87953
> #or use the R function
> rgamma(n=1,shape=4,rate=1/10)
[1] 17.93972
```

Monte Carlo Likelihood ratio test

Many tests can be Monte-Carlo'ed.

Example: Horse-kick deaths, Prussian cavalry, 1875-1894. 10 years and 20 corps.

y_i deaths $i = 1, 2, \dots, n$, $n = 200$ year-corps.

$$O_j = \sum_{i=1}^n \mathbb{I}_{y_i=j} \quad j = 0, 1, 2, \dots, k$$

Deaths	0	1	2	3	4	≥ 5
O	109	65	22	3	1	0
E	108.67	66.29	20.22	4.11	0.63	0.08

H0: y_i Poisson with a common mean?

$$\pi_j(\mu) = \mu^j \exp(-\mu) / j!, \quad j=0,1,2,3,4$$

$$\pi_5(\mu) = 1 - \sum_{j=0}^4 \pi_j(\mu).$$

MLE under H0: $\hat{\mu} = (\sum_j j O_j) / n = 0.61$

```
> O<-c(109,65,22,3,1,0)
```

```
> n<-sum(O)
```

```
> (mu<-sum((0:5)*O)/n)
```

```
[1] 0.61
```

H1: multinomial(π_j $j = 0, 1, \dots, k$). $\hat{\pi}_j = O_j/n$.

LRT statistic

$$\begin{aligned}\Lambda(y) &= 2 \sum_{j=0}^k O_j \left(-\log(\pi_j(\hat{\mu})) + \log(\hat{\pi}_j) \right) \\ &= 2 \sum_j O_j \log(O_j/n\pi(\hat{\mu}_j)) \\ &= 2 \sum_j O_j \log(O_j/E_j),\end{aligned}$$

if $E_j = n\pi_j(\hat{\mu})$ is number expected under H0.

Related Pearson statistic

$$P(Y) = \sum_j \frac{(O_j - E_j)^2}{E_j}$$

is robust for small sample sizes. Then

$$P(Y), \Lambda(Y) \xrightarrow{D} \chi^2(\dim(H1) - \dim(H0))$$

if the simpler model H0 is true.

Here, $\dim(H_0) = 1$ and $\dim(H_1) = k$
($k + 1$ parameters sum to 1, with $k = 5$)
so $P(Y), \Lambda(Y) \sim \chi^2(4)$ if approximation with
the distribution asymptotic in n is reliable.

But asymptotics unreliable (using $E_i \geq 5$) due
to cells 3,4,5.

Correct p-value (if use Pearson TS) is

$$p = \Pr(P(Y) > P(y)).$$

This equals the probability $P(Y)$ computed from
data Y simulated under H_0 exceeds the PTS
 $P(y)$ which we actually observed. Here,

$$P(y) \simeq 0.79.$$

We will estimate $\Pr(P(Y) > 0.79)$ using Monte
Carlo simulation.

Monte Carlo test:

1. Compute MLE $\hat{\mu}$ and test statistic $P(y)$.
2. What is dbn of $P(Y)$ if H_0 and $\mu = \hat{\mu}$?

(a) Simulate M iid data sets

$$Y^{(m)} \sim L(\hat{\mu}; y^{(m)}), \quad m = 1, 2, \dots, M.$$

Here $y^{(m)} = (y_1^{(m)}, \dots, y_n^{(m)})$,

$$\Pr(Y_i^{(m)} = j) = \frac{\hat{\mu}^j}{j!} e^{-\hat{\mu}}$$

iid $i = 1, \dots, n$ and $m = 1, \dots, M$. Set

$$O_j^{(m)} = \sum_{i=1}^n \mathbb{I}_{y_i^{(m)}=j} \quad j = 0, 1, 2, \dots$$

(b) Now compute $\hat{\mu}^{(m)}(y^{(m)})$, $\pi_j(\hat{\mu}^{(m)})$,

$E_j^{(m)} = n\pi_j(\hat{\mu}^{(m)})$ for $j = 1, 2, \dots, k$, and

$$P^{(m)} = \sum_j \frac{(O_j^{(m)} - E_j^{(m)})^2}{E_j^{(m)}} \quad m = 1, \dots, M.$$

3. Now the p-value = $\Pr(P(Y) > P(y))$. Estimate

$$\hat{p} = \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{P^{(m)} > P(y)}$$

with estimated variance $\hat{p}(1 - \hat{p})/M$.

Example (continued).

```
> #Step 1 (using mu=0.61 computed above)
> #pi(mu.hat) under H0 with pi Poisson dbn
> pr0<-dpois(0:4,lambda=mu)
> #5th entry is 1 - sum of the rest
> (pr0<-c(pr0,1-sum(pr0)))
[1] 0.543351 0.331444 0.101090 0.020555 0.003135
[6] 0.000425
> E<-n*pr0
> #Pearson test statistic
> (P<-sum( (O-E)^2/E ))
[1] 0.7903603

> #Monte Carlo test (auxiliary function)
> Function to count Freqs O_i
> MyCount<-function(y) {
  my<-max(y)
  O<-rep(0,my+1)
  for (j in 0:my) {O[j+1]<-O[j+1]+sum(y==j)}
  c(O,0)
}
> ym<-c(3,0,0,0,2,1,0,2,1,1); MyCount(ym)
[1] 4 3 2 1 0
```

```

> # Monte Carlo test
> M<-10000; Pm<-rep(0,M)
> for (m in 1:M) {
    #Step 2a - simulate synthetic data
    ym<-rpois(n=n,lambda=mu)
    #count the number in each cell
    Om<-MyCount(ym)

    #step 2b compute the MLE and Pearson stat
    mum<-mean(ym) #MLE

    pr0m<-dpois(0:max(ym),lambda=mum)
    pr0m<-c(pr0m,1-sum(pr0m))
    Em<-n*pr0m
    Pm[m]<-sum( (Om-Em)^2/Em )
}
> #step 3 estimate the p-value
> (p<-mean(Pm>P))
[1] 0.917
> (p.std<-sqrt(p*(1-p)/M))
[1] 0.002758822

```

The estimated p -value for the test for departure from Poisson statistics is $\hat{p} \simeq 0.917(3)$.

Conclude no evidence for departure from Poisson in Horse-Kick–Deaths Data.

Compare this with the (unreliable) value from asymptotics.

```
> # p-value for PTS  
> (1-pchisq(P,4))  
[1] 0.9397357
```

The Asymptotic approximation was out by just a little and would not in this instance have lead to a different conclusion.

Remark: we tested the H0 'Poisson common mean'. The mean-estimation added to the variance of $P(Y)$. If we were testing H0 'Poisson with common mean equal to x ' there would be no μ estimation at step 2b (and the $Y_i^{(m)}$'s would be Poisson mean x).

The general idea: generate synthetic values of the test statistic by mimicing the process that lead to the test statistic, under H0.

Exercise Suppose $\mu = 0.61$ known in PHKD example, and make simple test for fit to Poisson. Show that the lines

```
for (m in 1:M) {  
  ...  
    #step 3 compute the MLE and Pearson stat  
    mum<-mean(ym) #MLE  
  
    pr0m<-dpois(0:4,lambda=mum)  
  ...  
}
```

become

```
for (m in 1:M) {  
  ...  
    #step 3 compute the Pearson stat  
    pr0m<-dpois(0:4,lambda=mu)  
  ...  
}
```

so that we compute the Pearson test statistic using the known fixed μ -value.

Goodness of fit for continuous rv

In general: choose a function of the data.

Look for evidence that our data has a 'strange' value of this function.

Choose the function so the test is sensitive to particular kinds of departure from the null (choices must of course be made prior to seeing the data).

Kolmogorov-Smirnov one-sample test

Test for evidence that given data are not distributed according to a given distribution.

Test statistic measures difference between theory-cdf and empirical cdf of data.

$F(x) = \Pr(X \leq x)$ cdf for X .

Data y_i $i = 1, 2, \dots, n$. Sorted data $y_{(i)}$.

$F_e(x; Y)$ the empirical cdf,

$$F_e(x; Y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{Y_i \leq x}.$$

This is unbiased for $F(x)$. For the sorted data

$$F_e(y_{(i)}; y) = i/n$$

Example: Bat detection distances for 11 successful insect catches.

23 27 34 40 42 45 52 56 62 68 83

Is there any evidence for departure from exponential?

The empirical cdf is

$$F_e(x) = 0 \text{ for } x < 23,$$

$$F_e(x) = 1/11 \text{ for } 23 \leq x < 27,$$

$$F_e(x) = 2/11 \text{ for } 27 \leq x < 34,$$

and so on in steps of $1/n$ up to

$$F_e(x) = 10/11 \text{ for } 68 \leq x < 83,$$

$$\text{and } F_e(x) = 1 \text{ for } 83 \leq x.$$

X is the random variable of interest.

$F(x) = F(x; \theta)$ family of hypothesised cdfs.

H0: $\Pr(X \leq x) = F(x; \theta)$, θ unknown, MLE $\hat{\theta}$.

Under H0, $F_e(x)$ and $F(x) = F(x; \theta)$ should match. Let

$$d = \max_x |F(x) - F_e(x)|.$$

Maxima in $F_e(x) - F(x)$ at each $x = y_{(i)}$, where

$$F_e(y_{(i)}) - F(y_{(i)}) = i/n - F(y_{(i)}), i = 1, 2, \dots, n$$

Suprema in $F(x) - F_e(x)$ as $x \rightarrow y_{(i)}$, where

$$F(y_{(i)}) - F_e(y_{(i-1)}) = F(y_{(i)}) - (i-1)/n$$

with $F_e(x) = 0$ for $x < y_{(1)}$. Compute

$$d(y) = \max_{i=1, \dots, n} \left\{ \frac{i}{n} - F(y_{(i)}), F(y_{(i)}) - \frac{(i-1)}{n} \right\}.$$

One-sample KS test p -value

$$p = \Pr(d(Y) \geq d(y)|F)$$

where $Y \sim F$ is simulated data with cdf F and y is our data.

Estimate $F(x)$ as $F(x; \hat{\theta}(y))$ so our KS statistic is

$$d(y) = \max_{i=1, \dots, n} \left\{ \frac{i}{n} - F(y_{(i)}; \hat{\theta}(y)), F(y_{(i)}; \hat{\theta}(y)) - \frac{(i-1)}{n} \right\}.$$

But

$$p\text{-value} = \Pr(d(Y) \geq d(y)|F)$$

under H_0 is hard to calculate.

Use Monte Carlo. Same setup as MC-LRT.

One sample Monte Carlo KS test for GoF.

Let data $y = y_1, \dots, y_n$ and a null cdf F with unknown parameters θ be given.

1. Calculate MLEs $\hat{\theta}(y)$.
2. Calculate $d(y; \hat{\theta})$ using the sorted data.
3. Simulate and sort M iid datasets $y^{(1)}, \dots, y^{(M)}$,
 $y^{(m)} = (y_1^{(m)}, \dots, y_n^{(m)})$, $y_i^{(m)} \sim F(\cdot; \hat{\theta})$ iid.
4. For each $y^{(m)}$, compute $\hat{\theta}^{(m)} = \hat{\theta}^{(m)}(y^{(m)})$,
and

$$d^{(m)} = d(y^{(m)}; \hat{\theta}^{(m)}).$$

5. Estimate p -value

$$\hat{p} = \frac{1}{M} \sum_{m=1}^M \mathbb{I}_{d^{(m)} > d(y)}.$$

Apply this to the Bat-detection data.

$$F(x; \mu) = 1 - \exp(-x/\mu) \text{ and } \hat{\mu} = \bar{y}.$$

```
> #Step 1. data are already sorted
> y<-c(23,27,34,40,42,45,52,56,62,68,83)
> (mu<-mean(y))
[1] 48.36364
> n<-11
> #Step 2. pexp(y,lambda) = cdf for Exp(lambda)
> F<-pexp(y,rate=1/mu)
>
> #empirical cdf F_e(y_(i);y)=i/n, max F_e-F:
> (1:n)/n-F
-0.2876 -0.2460 ... 0.1542 0.1798
> #max F-F_e(y_(i-1)):
> F-(0:(n-1))/n
0.37847 0.33689 ... -0.06330 -0.08885
>
> #KS statistic d is the largest difference
> (d<-max(c((1:n)/n-F,F-(0:(n-1))/n)))
[1] 0.3784655
```

```

> #Simulate M datasets under H0
> #Compute M H0-d's.
> M<-10000
> d.sim<-rep(0,M)
> for (i in 1:M) {
+   Ym<-sort(rexp(n,1/mu)) #Step 3 (sort!)
+   mum<-mean(Ym)          #Step 4 MLE and d
+   Fm<-pexp(Ym,rate=1/mum)#H0, cdf
+   d.sim[i]<-max(c((1:n)/n-Fm,Fm-(0:(n-1))/n))
+ }
> (p<-mean(d.sim>d)) #step 5 p-hat
[1] 0.0058

```

The chances of getting such a poor fit to an exponential if the data were truly exponentially distributed are only $p \simeq 0.006$, so we should look for some other shape for this distribution.

Monte Carlo test, simple hypothesis

When null hypothesis is composite, used the MLE from the data.

Repeated this for simulated data to simulate H_0 -process generating original test statistic.

In simple hypothesis, parameter known, compute test statistic and simulated TS's using given parameter value.

The KS-test has a special simple form when the parameters of F are known.

Distribution of KS-test statistic $d(Y)$ doesn't depend on F : given $d(y)$, look up p -value in tables for KS. Monte Carlo not needed.

Recall

$$d(y) = \max_x |F(x) - F_e(x)|.$$

Suppose $a < b \Rightarrow F(a) < F(b)$.

$$\begin{aligned} d(Y) &= \max_x \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{Y_i \leq x} - F(x) \right| \\ &= \max_x \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{F(Y_i) \leq F(x)} - F(x) \right|. \end{aligned}$$

$U_i = F(Y_i) \Rightarrow U_i \sim U(0, 1)$ (since $Y_i \sim F$).

Let $z = F(x)$. Search for max over $z \in (0, 1)$:

$$d(Y) = \max_{z \in (0,1)} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{U_i \leq z} - z \right|$$

The p -value for the test of H_0 , y have cdf F :

$$p = \Pr \left(\max_{z \in (0,1)} \left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{U_i \leq z} - z \right| < d(y) \right).$$

with U_i iid uniform rv. Now $d(y)$ is fixed and $d(Y)$ doesn't depend on F .

So, estimate $\Pr(d(Y) < d|F)$ for a range of d -values from zero up to one, and make a table for the function $p = p(d)$. In future tests, get a $d(y)$ value, and then look up the p -value in this table.

This is sometimes used in GOF testing for NLM's, testing the studentised residuals of a NLM are approximately $N(0, 1)$.

The KS-test can be used to compare two data sets. Dr Lunn, (2007) lecture notes, gives the example of two collections of silver coins. The percentages of silver in the first collection are 5.9, 6.8, 6.4, 7.0, 6.6, 7.7, 7.2, 6.9, 6.2. The percentages of silver in the second collection are 5.3, 5.6, 5.5, 5.1, 6.2, 5.8, 5.8. Let $A \sim F_A$ be the percent silver in a coin in the first collection, and $B \sim F_B$ be the percent silver in a coin in the second collection. Are the two distributions the same, that is, is it the case that $F_A = F_B$? We would like to make this test without specifying a family of distributions for A and B . The two sample test is based on the distance between the two empirical cdf's.

Exercise Let $F_e^{(A)}$ and $F_e^{(B)}$ denote the empirical cdf's for the percent silver in the two collections of coins. Show that $\max_x |F_e^{(A)}(x) - F_e^{(B)}(x)| = 6/7 - 1/9 = 47/63$. Look up the p -value in a KS table (for a pair of samples of size 9 and 7) and verify that $p \simeq 0.02$, so it is very unlikely that the simple picture of a common silver standard holds.