Statistics MATH414

lecture 16

27th March 2025

Vivek T.

(+)

Finding p-values by Simulation.

(+) Finding p-values by simulation.

Recall the data

$$\{0, 0, 0, 1, 1, 1, 2, 2, 3, 4\}$$

We computed that a formula for the log likelihood ratio statistic if the data were to be fit using Poisson dist. is

$$\chi^2_L := 2\left(-n\bar{x} + n\bar{x}\log(\bar{x}) + n\lambda_0 - n\bar{x}\log\lambda_0\right)$$

② Using this formula, we calculate the statistic for the data

```
x <- c(0,0,0,1,1,1,2,2,3,4)

lr <- function(x, lambda0 = 1) {
  n <- length(x)
  m <- mean(x)
  2*(-n*m + n*m*log(m) +
     n*lambda0 - n*m*log(lambda0))
}
```

3

lr(x)

pval <- 1 - pchisq (lr(x), df = 1)

pval

0.233

③ lr(x)

pval <- 1 - pchisq(lr(x), df = 1)

pval

0.233

Computing the p-value using the $\chi^2(1)$ approximation for the test statistic.

④ Computing p-value using simulation.

4 Computing p-value using simulation.

```
data <- replicate(5000, lr(rpois(10,1),
                              lambda0 = 1 ))
```

```
sum( data >= lr(x)) / 5000
```

$\boxed{\sim 0.26}$

This is the p-value.

④ Computing p-value using simulation.

```
data <- replicate(5000, lr(rpois(10,1),
                        lambda0 = 1 ))
```

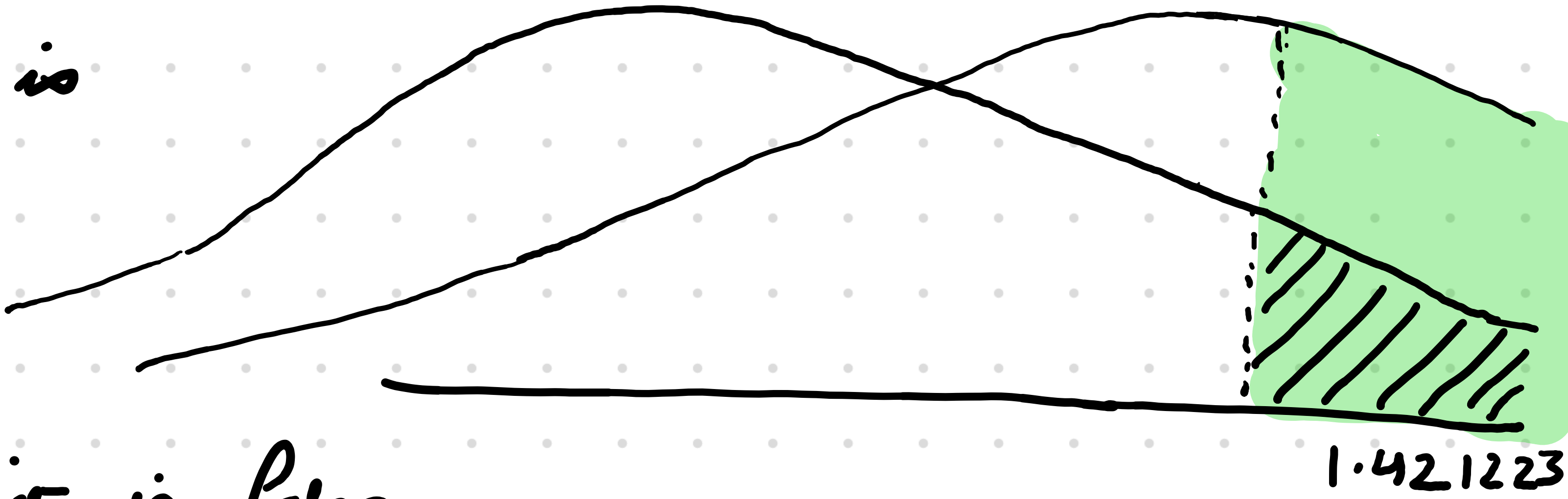
```
sum(data >= lr(x)) / 5000
```

This is the p-value.

$\boxed{\sim 0.26}$

It does not appear that we will reject the null hypothesis with this data.

5 Computing power of the test



- Type II error is the probability of not rejecting the null hypothesis when it is false.

- Power is the probability of rejecting the null hypothesis when the alternative hypothesis is true.

```
altdata <- replicate(5000, lr(rpois(10, 1.4), lambda0 =1.4))
Sum(altdata >= lr(x)) / 5000
```

Confidence Intervals

Confidence Intervals

Recall the confidence interval due to

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)}$$

and

$$\boxed{\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow{D} t_{n-1}}$$

Confidence intervals

Recall the confidence interval due to

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \xrightarrow{D} N(0,1)}$$

$$\bar{X} \pm z_{\alpha/2}\, \sigma/\sqrt{n}$$

and

$$\boxed{\frac{\bar{X}-\mu}{S/\sqrt{n}} \xrightarrow{D} t_{n-1}}$$

$$\bar{X} \pm t^{*}_{\alpha/2}\, S/\sqrt{n}$$

Confidence intervals

Wald type test

$$\boxed{\sqrt{n I(\hat{\theta})}\ (\hat{\theta} - \theta_0) \xrightarrow{D} N(0,1)}$$

Confidence Intervals

Wald type test

$$\boxed{\sqrt{n I(\hat{\theta})}\,(\hat{\theta} - \theta_0) \xrightarrow{D} N(0,1)}$$

$$\hat{\theta} \pm \frac{z^*_{\alpha/2}}{\sqrt{n I(\hat{\theta})}}$$

Confidence Intervals

Wald type test

$$\boxed{\sqrt{n I(\hat{\theta})}\,(\hat{\theta}-\theta_0) \xrightarrow{D} N(0,1)}$$

$$\hat{\theta} \pm z^*_{\alpha/2} \frac{1}{\sqrt{n I(\hat{\theta})}}$$

This is called standard error and can be computed using R.

Example

Consider the data

```
x <- c(0,0,0, 1,1,1, 2,2, 3,4)

loglik.pois <- function(theta, x.bar = 1.4,
                                 n = 10)
{
  -n * theta + n * x.bar * log(theta)
}
```

$$L(\lambda) = \prod_i \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$\ell(\lambda) = \sum_i \left(-\lambda + x_i \log \lambda\right)$$

$$= -n\lambda + n\bar{x} \log \lambda$$

Example

Consider the data

```
x <- c(0,0,0, 1,1,1, 2,2, 3,4)

loglik.pois <- function(theta, x.bar = 1.4,
                        n = 10)
{
  -n * theta + n * x.bar * log(theta)
}
```

Recall $I(\lambda) = -E\left(\frac{\partial^2 \ell(\lambda)}{\partial \lambda^2}\right)$

$= \frac{n}{\lambda}$

$$L(\lambda) = \prod_i \frac{e^{-\lambda} \lambda^{x_i}}{x_i!}$$

$$\ell(\lambda) = \sum_i \left(-\lambda + x_i \log \lambda\right)$$

$$= -n\lambda + n\bar{x} \log \lambda$$

```
m1 <- maxLik(loglik.pois, start=
                          c(lambda = 1),
                          x.bar= 1.4, n= 10)

stdEr(m1)
```

0.374

Therefore the Wald type confidence interval is

```
1.4 + c(-1,1) * qnorm(0.975) * 0.374
```

On the other hand, the log likelihood ratio gives us a confidence interval from

$$2\left(\ell(\hat{\theta}) - \ell(\theta_0)\right) \geq \text{pchisq}\left(1-\alpha,\ df = 1\right)$$

On the other hand, the log likelihood ratio gives us a confidence interval from

$$2(l(\hat{\theta}) - l(\theta_0)) \sim \chi^2(1)$$

This requires us to find two solutions to the eqn

$$2(l(\hat{\theta}) - l(\theta_0)) = \text{qchisq}(1-\alpha, df = 1).$$

$$\left.\begin{matrix} 0.025 \\ 0.975 \end{matrix}\right\}$$

We have to define the function

```
lr <- function(t)
{2*loglik.pois(coef(m1)) - 2* loglik(t)
}
```

We want to find t so that

$lr(t) =$ qchisq(0.025, df = 1) ...

This is done using Uniroot( )

This is done using uniroot()

```
uniroot(function(t) { lr(t) - qchisq(0.95,
                       df=1) },
        c(0,1)) %>% value
```

...

Inference for models with multiple parameters

Let Ω_0 and Ω_a be two sets of parameter values

$$\left.\begin{aligned} H_0 &: \underline{\theta} \in \Omega_0 \\ H_1 &: \underline{\theta} \in \Omega_a \end{aligned}\right\}$$

Inference for models with multiple parameters

Let Ω_0 and Ω_a be two sets of parameter values

$$\left.\begin{array}{l} H_0 : \underline{\theta} \in \Omega_0 \\ H_1 : \underline{\theta} \in \Omega_a \end{array}\right\}$$

$L(\hat{\Omega}_0) = \max \{ L(\theta, x) : \theta \in \Omega_0 \}$

$L(\hat{\Omega}) = \max \{ L(\theta, x) : \theta \in \Omega \}$

The likelihood ratio is

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})}$$

Theorem If $\underline{X}$ is an i.i.d. sample of size n from a model with parameters $\underline{\theta}$,

the p.d.f is smooth and its support is independent of $\underline{\theta}$ and $\underline{\theta} \in \Theta_0$.

then

$$W = -2\log(\Lambda) = 2\left(l(\hat{\theta}, x) - l(\hat{\theta}_0 ; x)\right)$$

$$\xrightarrow{D} \text{Chisq}(\dim \Omega - \dim \Omega_0) \text{ as } n \to \infty$$