Assignment 10 Solutions
(April 23, 2005)

1. *(Estimation in a Zero-Inflated Poisson Model)*

Consider n $i.i.d$ discrete observations $X_1, X_2, ..., X_n$, each with marginal probability mass function on $\{0, 1, 2, ...\}$

$$f(x|p,\lambda) = \begin{cases} p\exp(-\lambda) + (1-p) & x = 0 \\ p\frac{\exp(-\lambda)\lambda^x}{x!} & x = 1, 2, 3, ... \end{cases}$$

where $p \in (0,1)$ and $\lambda > 0$. (This marginal is a mixiture of a distribution degenerate at 0 and the $Poisson(\lambda)$ distribution. It might arise in the inspection for flaws of a mixed lot of items, some of which come from a "perfect" process and others of which come from a process that puts flaws on the items according to a Poisson distribution.)

(a) Give the likelihood equations for this problem.

(b) What are the (marginal) large sample distributions of method of moments estimators of p and λ, say $\tilde{p}$ and $\tilde{\lambda}$, when $p = 0.7$ and $\lambda = 3.0$? (Hint: What is the covariance matrix for the vector $(X_1, X_1^2)'$? What, then, does the multivariate central limit theorem say about the large sample joint distribution of $\sqrt{n}\left(\frac{1}{n}\sum X_i - E_{p,\lambda}X_1, \frac{1}{n}\sum X_i^2 - E_{p,\lambda}X_1^2\right)'$? What, then does the delta method five you for the large sample joint distribution of $\sqrt{n}\left(\tilde{p} - 0,7, \tilde{\lambda} - 3\right)'$?) It may well be useful to know that the first 4 moments of the Poisson distribution are: $\mu_1 = \lambda, \mu_2 = \lambda^2 + \lambda, \mu_3 = \lambda^3 + 3\lambda^2 + \lambda$ and $\mu_4 = \lambda^4 + 6\lambda^3 + 7\lambda^2 + \lambda$.

(c) What is the large sample joint distribution of an "MLE" of (p, λ) if $p = 0.7$ and $\lambda = 3.0$? How do the marginal distributions compare to those in b)? Below are $n = 20$ observations that I simulated from this distribution using $p = 0.7$ and $\lambda = 3.0$.

(d) Find method of moments estimators based on the data above. Then compute a "one-step Newton improvement" on $\left(\tilde{p}, \tilde{\lambda}\right)$.

(e) The MLE of (p, λ) based on these data turns out to be $\left(\hat{p}, \hat{\lambda}\right) = (.72675, 3.30242)$. Find (individual) large sample 90% confidence intervals for p and λ based on $\left(\hat{p}, \hat{\lambda}\right)$ and the observaed Fisher information matrix.

(f) Find an elliptical large sample 90% joint confidence region for (p, λ) based on $\left(\hat{p}, \hat{\lambda}\right)$ and the observed Fisher information matrix. Plot this in the (p, λ)-plane.

Solution:

(a)

The likelihood is

$$L(p,\lambda)=\prod_{i=1}^{n} f(x_i|p,\lambda)=\prod_{i=1}^{n}\left\{\left(pe^{-\lambda}+1-p\right)I(x_i=0)+p\frac{e^{-\lambda}\lambda^{x_i}}{x_i!}I(x_i\neq 0)\right\}$$

$$\Rightarrow \log L(p,\lambda)=\left(pe^{-\lambda}+1-p\right)\sum_{i=1}^{n} I(x_i=0)+\sum_{i=1}^{n}\{\log p-\lambda+x_i\log\lambda-\log x_i!\}\, I(x_i\neq 0)$$

Then the likelihood equations are:

$$\left\{\begin{array}{l}\frac{\partial}{\partial p}\log L(p,\lambda)=0\\ \frac{\partial}{\partial\lambda}\log L(p,\lambda)=0\end{array}\right. \Rightarrow \left\{\begin{array}{l}\left(e^{-\lambda}-1\right)\left(\sum_i I(x_i=0)\right)+\frac{1}{p}\sum_i I(x_i\neq 0)=0\\ \left(-pe^{-\lambda}\right)\sum_i I(x_i=0)+\sum_i\left(\frac{x_i}{\lambda}-1\right)I(x_i\neq 0)=0\end{array}\right.$$

$\mu_1=\lambda p,\qquad ,\mu_2=\left(\lambda^2+\lambda\right)p$

(b)

As
$E(x|p,\lambda)=\sum_i^{\infty} p\frac{e^{-\lambda}\lambda^x}{x!}x=p\cdot\lambda$
$E(x^2|p,\lambda)=\sum_i^{\infty} p\frac{e^{-\lambda}\lambda^x}{x!}x^2=p\cdot\left(\lambda^2+\lambda\right)=p\lambda(\lambda+1)$

i.e. $\mu_1=p\lambda,\mu_2=p\lambda(\lambda+1)$

$\Rightarrow$The MOM estimators are $\tilde{p}=\frac{\hat{\mu}_1}{\tilde{\lambda}},\tilde{\lambda}=\frac{\hat{\mu}_2}{\hat{\mu}_1}-1$

i.e. $\tilde{p}=\frac{\hat{\mu}_1^2}{\hat{\mu}_2-\hat{\mu}_1},\tilde{\lambda}=\frac{\hat{\mu}_2-\hat{\mu}_1}{\hat{\mu}_1}$

where $\hat{\mu}_1=\frac{1}{n}\sum x_i$ and $\hat{\mu}_2=\frac{1}{n}\sum x_i^2$
The variance-covariance matrix of (X_1,X_1^2) is

$$\begin{aligned}\Sigma &= p\begin{pmatrix}\lambda+(1-p)\cdot\lambda^2 & (1-p)\lambda^3+(3-p)\lambda^2+\lambda\\ (1-p)\lambda^3+(3-p)\lambda^2+\lambda & (1-p)\lambda^4+(6-2p)\lambda^3+(7-p)\lambda^2+\lambda\end{pmatrix}\\ &= \lambda p\begin{pmatrix}1+(1-p)\cdot\lambda & (1-p)\lambda^2+(3-p)\lambda+1\\ (1-p)\lambda^2+(3-p)\lambda+1 & (1-p)\lambda^3+(6-2p)\lambda^2+(7-p)\lambda+1\end{pmatrix}\end{aligned}$$

By the multivariate central limiting theorem

$$\sqrt{n}\begin{pmatrix}\hat{\mu}_1-\mu_1\\ \hat{\mu}_2-\mu_2\end{pmatrix}\xrightarrow{L} N(0,\Sigma)$$

Then by Delta method,

$$\sqrt{n}\begin{pmatrix}\tilde{p}-0.7\\ \tilde{\lambda}-0.3\end{pmatrix}\xrightarrow{L} N\left(0,l\Sigma l'\right)$$

where

$$
\begin{aligned}
l &= \begin{pmatrix} \frac{\partial p}{\partial \mu_1} & \frac{\partial p}{\partial \mu_2} \\ \frac{\partial \lambda}{\partial \mu_1} & \frac{\partial \lambda}{\partial \mu_2} \end{pmatrix} = \begin{pmatrix} \frac{\mu_1^2 - 2\mu_1\mu_2}{(\mu_2-\mu_1)^2} & -\frac{\mu_1^2}{(\mu_2-\mu_1)^2} \\ -\frac{\mu_2}{\mu_1^2} & \frac{1}{\mu_1} \end{pmatrix} \\
&= \begin{pmatrix} -\frac{2\lambda+1}{\lambda^2} & -\frac{1}{\lambda^2} \\ -\frac{\lambda+1}{\lambda p} & \frac{1}{\lambda p} \end{pmatrix}
\end{aligned}
$$

Then,

$$\sqrt{n}\begin{pmatrix} \tilde{p} - p \\ \tilde{\lambda} - \lambda \end{pmatrix} \xrightarrow{L} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, l\Sigma l'\right)$$

where,

$$
\begin{aligned}
l\Sigma l' &= \begin{pmatrix} -\frac{2\lambda+1}{\lambda^2} & -\frac{1}{\lambda^2} \\ -\frac{\lambda+1}{\lambda p} & \frac{1}{\lambda p} \end{pmatrix} \cdot \lambda p \begin{pmatrix} 1 + (1-p)\cdot\lambda & (1-p)\lambda^2 + (3-p)\lambda + 1 \\ (1-p)\lambda^2 + (3-p)\lambda + 1 & (1-p)\lambda^3 + (6-2p)\lambda^2 + (7-p)\lambda + 1 \end{pmatrix} \\
&\cdot \begin{pmatrix} -\frac{2\lambda+1}{\lambda^2} & -\frac{\lambda+1}{\lambda p} \\ -\frac{1}{\lambda^2} & \frac{1}{\lambda p} \end{pmatrix} \\
&= \begin{pmatrix} 0.365556 & -0.666667 \\ -0.666667 & 7.1428571 \end{pmatrix}
\end{aligned}
$$

(c)

Let $\theta = \begin{pmatrix} p \\ \lambda \end{pmatrix}$, $\delta_n(x) = \begin{pmatrix} \hat{p}_n \\ \hat{\lambda}_n \end{pmatrix}$ be *MLE*.

Then

$$\sqrt{n}\left(\delta_n(x) - \theta\right) \xrightarrow{L} N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, I_1^{-1}(\theta)\right)$$

where,

$$
\begin{aligned}
I_1(\theta) &= -E\left(\frac{\partial^2 \log L}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \frac{\left(e^{-\lambda}-1\right)^2}{pe^{-\lambda}+1-p} + \frac{1-e^{-\lambda}}{p} & \frac{e^{-\lambda}}{pe^{-\lambda}+1-p} \\ \frac{e^{-\lambda}}{pe^{-\lambda}+1-p} & \frac{p}{\lambda} - \frac{p(1-p)e^{-\lambda}}{pe^{-\lambda}+1-p} \end{pmatrix} \\
&= \begin{pmatrix} 4.0538844 & 0.1486843 \\ 0.1486843 & 0.2021096 \end{pmatrix}
\end{aligned}
$$

$$\Rightarrow I^{-1}(\theta) = \begin{pmatrix} 0.2535174 & -0.1865029 \\ -0.1865029 & 5.0850126 \end{pmatrix}$$

Compare with b), The *MLEs* have smaller variance than that of the MOM estimates.

(d)

Given observations:

$$0,3,2,5,3,4,0,0,4,0,0,5,3,5,5,4,2,0,1,2$$

So,

$$\begin{aligned}\tilde{p} &= \frac{\hat{\mu}_1^2}{\hat{\mu}_2 - \hat{\mu}_1} = 0.8228571 \\ \hat{\lambda} &= \frac{\hat{\mu}_2}{\hat{\mu}_1} - 1 = 2.916667\end{aligned}$$

Then "One-step Newton improvement" of $\left(\tilde{p}, \tilde{\lambda}\right)$ is

$$\begin{pmatrix} \tilde{p}_1 \\ \hat{\lambda}_1 \end{pmatrix} = \begin{pmatrix} \tilde{p} \\ \tilde{\lambda} \end{pmatrix} - l_n''^{-1} l_n' \Bigg|_{\begin{pmatrix} p \\ \lambda \end{pmatrix} = \begin{pmatrix} \tilde{p} \\ \hat{\lambda} \end{pmatrix}}$$

Where $l_n(x) = 6\ln\left(pe^{-\lambda} + 1 - p\right) + 14\ln p - 14\lambda + 48\ln\lambda$
So,

$$l_n' = \begin{pmatrix} \frac{6\left(e^{-\lambda}-1\right)}{pe^{-\lambda}+1-p} + \frac{14}{p} \\ \frac{-pe^{-\lambda}}{pe^{-\lambda}+1-p} - 14 + \frac{48}{\lambda} \end{pmatrix}$$

$$l_n'' = \begin{pmatrix} \frac{-6\left(e^{-\lambda}-1\right)^2}{\left(pe^{-\lambda}+1-p\right)^2} - \frac{14}{p^2} & -\frac{6e^{-\lambda}}{\left(pe^{-\lambda}+1-p\right)^2} \\ -\frac{6e^{-\lambda}}{\left(pe^{-\lambda}+1-p\right)^2} & \frac{6p(1-p)e^{-\lambda}}{\left(pe^{-\lambda}+1-p\right)^2} - \frac{48}{\lambda^2} \end{pmatrix}$$

so,

$$\begin{pmatrix} \tilde{p} \\ \tilde{\lambda} \end{pmatrix} = \begin{pmatrix} 0.7369791 \\ 3.3054730 \end{pmatrix}$$

```
#####one step Newton Raphson

data=c(0,3,2,5,3,4,
0,0,4,0,0,5,
3,5,5,4,2,0,
1,2)
mu1=sum(data)/20
mu2=sum(data^2)/20
lam=mu2/mu1-1
p=mu1/lam
```

```
p;lam
m=sum(data==0)
for (i in 1:1){
l=c(m*(exp(-lam)-1)/(p*exp(-lam)+(1-p))+(20-m)/p,
-m*p*exp(-lam)/(p*exp(-lam)+(1-p))-(20-m)+sum(data)/lam)
# this is under expected fisher information
q1=p*exp(-lam)+(1-p)
q2=exp(-lam)-1
e1=q2^2/q1
e2=exp(-lam)/q1
e4=(-p*exp(-lam)*(1-p))/q1
e3=e2
fish=matrix(c((1-exp(-lam))/p,0,0,p/lam),ncol=2,byrow=T)
fish=fish+matrix(c(e1,e2,e3,e4),ncol=2,byrow=T)
######
#### this is under observed fisher information
q1=p*exp(-lam)+(1-p)
q2=exp(-lam)-1
e1=6*q2^2/q1^2+14/p^2
e2=6*exp(-lam)/q1^2
e3=e2
e4=-6*p*(1-p)*exp(-lam)/q1^2+48/lam^2
fish=matrix(c(e1,e2,e3,e4),ncol=2,byrow=T)
para=c(p,lam)+solve(fish)%*%l
p=para[1]
lam=para[2]
}
para
```

(e)

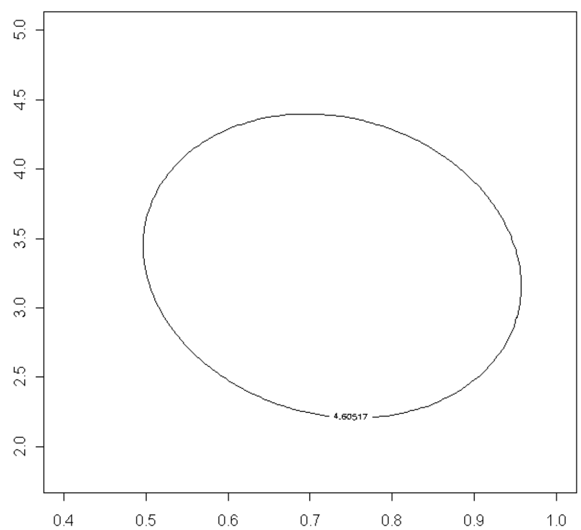
Large sample *90% confidence intervals* for p and λ based on $\left(\hat{p},\hat{\lambda}\right) = (0.72675, 3.30242)$ are

$$\hat{p} \pm Z\,(0.95)\sqrt{l_n''^{-1}\Big|_{(p,\lambda)=\left(\hat{p},\hat{\lambda}\right)}} = (2.46369, 4.14115)$$

(f)

```
###contour
p=.72675
lam=3.30242
l=c(m*(exp(-lam)-1)/(p*exp(-lam)+(1-p))+(20-m)/p,
-m*p*exp(-lam)/(p*exp(-lam)+(1-p))-(20-m)+sum(data)/lam)
#p=0.7
#lam=3
q1=p*exp(-lam)+(1-p)
```

```
q2=exp(-lam)-1
e1=6*q2^2/q1^2+14/p^2
e2=6*exp(-lam)/q1^2
e3=e2
e4=-6*p*(1-p)*exp(-lam)/q1^2+48/lam^2
fish=matrix(c(e1,e2,e3,e4),ncol=2,byrow=T)
#####upbound
c(p,lam)+qnorm(0.95)*sqrt(diag(solve(fish)))
####lowerbound
c(p,lam)-qnorm(0.95)*sqrt(diag(solve(fish)))
th1=seq(0.4,1,length=100)
th2=seq(1.8,5,length=100)
th=expand.grid(th1,th2)
ss=function(x){
l1=x[1]-p
l2=x[2]-lam
t(c(l1,l2))%*%(fish)%*%(c(l1,l2))}
sum=apply(th,1,ss)
sum=matrix(sum,100,100)
contour(th1,th2,sum,levels=qchisq(0.9,2))
```



2. Suppose that for $\alpha \in [0,1]$, X is a random variable with probability density

$$f(x|\alpha) = \alpha f_1(x) + (1-\alpha) f_0(x)$$

where $f_0(x)$ is the $N(0,1)$ density and $f_1(x)$ is the $N(1,1)$ density.

(a) Find the mean and variance of X, $E_\alpha X$ and $Var_\alpha X$.

(b) Show that the maximum likelihood estimator of α based on the single observation X, is

$$\hat{\alpha} = \begin{cases} 1 & \text{if } X > 0.5 \\ 0 & \text{otherwise} \end{cases}$$

Compute the *mean* and *variance* of this estimator. Is $\hat{\alpha}$ unbiased for α?

(c) Argue that the mean squared error of $\hat{\alpha}$ as an estimator of α is no more than $.25 + (.3085)^2$. Use this fact and compare $\hat{\alpha}$ and X in terms of mean squared error.

(d) Set up an integral giving $I(\alpha)$, the Fisher information in X concerning $\alpha \in (0, 1)$. Now consider estimation of α based on a sample $X_1, X_2, ..., X_n$ that are *iid* with density (*). Let $\hat{\alpha}$ be the MLE of α based on the n observations and let $\bar{X}_n$ be the usual sample mean.

(e) The following figure gives plots of both $1/I(\alpha)$ and $1 + \alpha - \alpha^2$. Whatdoes this figure indicate about the the large sample distributions of $\hat{\alpha}$ and X_n? Onthe basis of large sample considerations, which of these is the better estimator of α? Explain carefully.

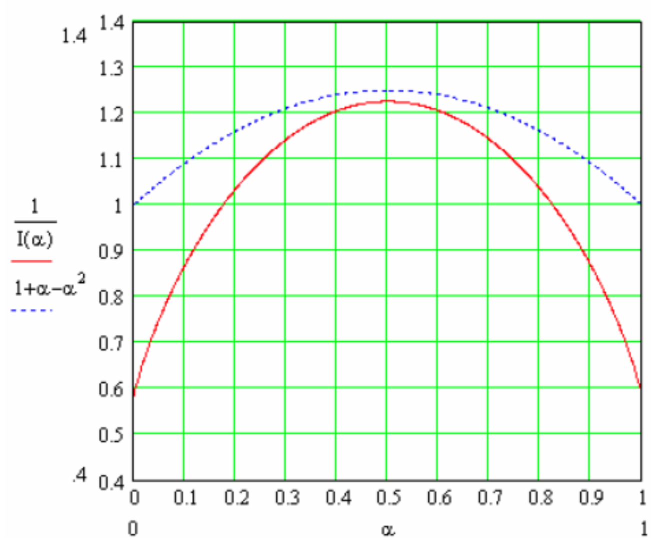


Figure1 : $1/I(\alpha)$ and $1 + \alpha + \alpha^2$

(f) A particular sample of $n = 20$ observations produces $\hat{\alpha} = .4$. What is an approximate *90% confidence interval* for α based on the "expected Fisher information" in a single observation? Explain where you are getting your limits.

Solution:

(a) $$E_\alpha X = \alpha E_1(x) + (1-\alpha) E_0(x) = \alpha \cdot 1 + (1-\alpha) \cdot 0 = \alpha$$
$$E_\alpha X^2 = \alpha E_1(x^2) + (1-\alpha) E_0(x^2) = \alpha \cdot 2 + (1-\alpha) \cdot 1 = 1 + \alpha$$
$$Var_\alpha X = E_\alpha X^2 - (E_\alpha X)^2 = (1-\alpha) - \alpha^2 = 1 + \alpha - \alpha^2$$

(b) As $f(x|\alpha) = \alpha f_1(x) + (1-\alpha) f_0(x)$ and $f_0(x) \sim N(0,1)$, $f_1(x) \sim N(1,1)$, then in order to minimize $f(x|\alpha)$

$$\begin{aligned} \text{if } f_1(x) &> f_0(x) \text{ then take } \alpha = 1 \\ \text{and if } f_1(x) &\leq f_0(x) \text{ then take } \alpha = 0 \end{aligned}$$

$$\Rightarrow \hat{\alpha}_{MLE} = I(f_1(x) > f_0(x)) = I(x > 0.5)$$

i.e.

$$\hat{\alpha}_{MLE} = \begin{cases} 1 & \text{if } x > 0.5 \\ 0 & \text{if } x \leq 0.5 \end{cases}$$

(c) $E\hat{\alpha} = P_\alpha(x > 0.5) = \alpha P_1(x > 0.5) + (1-\alpha) P_0(x > 0.5) = 0.3085 + 0.38302$
$E\hat{\alpha}^2 = E\hat{\alpha} = 0.3085 + 0.38302$
$\Rightarrow Var\hat{\alpha} = E\hat{\alpha}^2 - (E\hat{\alpha})^2 = (0.3085 + 0.38302)(0.6915 - 0.38302) \leq \left(\frac{1}{2}\right)^2 = 0.25$
Then $MSE(\hat{\alpha}) = Var(\hat{\alpha}) + Bias(\hat{\alpha})^2 = 0.25 + (\alpha - 0.3085 - 0.38302)^2 \leq 0.25 + (0.3085)^2$
on the other hand, since
$MSE(x) = Var_\alpha x = 1 + \alpha - \alpha^2 \geq 1 > 0.25 + (0.3085)^2 = MSE(\hat{\alpha})$
then MLE is better than x.

(d) $L(\alpha) = \alpha f_1 + (1-\alpha) f_0 \Rightarrow \frac{\partial \log L(\alpha)}{\partial \alpha} = \frac{f_1 - f_0}{L(\alpha)} = \frac{f_1 - f_0}{f(x|\alpha)}$
$\Rightarrow I(\alpha) = E_\alpha \left(\frac{\partial \log L(\alpha)}{\partial \alpha}\right)^2 = \int \frac{(f_1 - f_0)^2}{f(x-\alpha)} dx$

(e) Since for large n
$\bar{x}_n \xrightarrow{L} N\left(\alpha, \frac{1+\alpha-\alpha^2}{n}\right)$
and $\hat{\alpha}_n \xrightarrow{L} N\left(\alpha, \frac{1}{nI(\alpha)}\right)$,
and from Figure 1, we see that $\frac{1}{I(\alpha)} < \frac{1+\alpha-\alpha^2}{n}$
hence, $\hat{\alpha}_n$ is better than $\bar{x}_n$ as an estimator of α approximate.

(f) The *90% Confidence Interval* is

$$\hat{\alpha}_n \pm Z(0.95) \frac{1}{\sqrt{nI(\hat{\alpha}_n)}}$$

when $n = 20$, $\hat{\alpha}_n = 0.4$, and $I(\hat{\alpha}_n) = \frac{1}{1.2}$ which can be read from the Figure 1.
So *90% Confidence Interval* is:

$$\begin{aligned} &\left(0.4 \pm 1.645 \sqrt{\frac{1.20}{20}}\right) \\ = \; &(-0.0029052, 0.8029052) \end{aligned}$$