STATISTICS: Problem Set 5-Solutions

October 18, 2007

1. Let $X_1, X_2, ..X_T$ be *i.i.d.* random variables with the following density $f_{X_i}(x_i) = \lambda^{-1}\exp(-\lambda^{-1}x_i)$, $x_i > 0, \lambda > 0$, $i = 1, 2, ...T$.

(a) Work out the maximum likelihood estimator $\widehat{\lambda}$. What is the finite sample distribution of this estimator?

- The likelihood function is

$$L(\lambda) = (\prod_{i=1}^{T} f_{X_i}(X_i, \lambda)) = \lambda^{-T} e^{-\lambda^{-1}\sum_{i=1}^{T} X_i}.$$

- The log-likelihood function is

$$LogL(\lambda) = -Tln\lambda - \lambda^{-1}\sum_{i=1}^{T} X_i.$$

- The score vector is the first derivative of the log-likelihood function:

$$q(\lambda) = \frac{d}{d\lambda} LogL(\lambda) = -\frac{T}{\lambda} + \frac{\sum_{i=1}^{T} X_i}{\lambda^2}$$

- The maximum likelihood estimator $\widehat{\lambda}$ turns the score vector q to 0, therefore,

$$-\frac{T}{\widehat{\lambda}} + \frac{\sum_{i=1}^{T} X_i}{\widehat{\lambda}^2} = 0$$
$$\widehat{\lambda} = \frac{\sum_{i=1}^{T} X_i}{T} = \bar{X}.$$

The finite sample distribution of $\widehat{\lambda}$ is determined by the density function $f_{\bar{X}}$. Knowing the density function of each random variable X_i, we may use the change-of-variable method to find the density function of a random variable $X_1 + \cdots + X_T$. After that we change a variable once more to find the density function of $(X_1 + \cdots + X_T)/T = \bar{X}$. Precisely, following the same steps as in Exercise 1 (part (d)) of the Problem set 4, we find the density function

$$f_{X_1+X_2+...+X_T}(x) = \frac{\lambda^{-T}}{(T-1)!} e^{-\lambda^{-1}x} x^{T-1}$$

and

$$f_{\bar{X}}(x) = f_{X_1+X_2+...+X_T}(Tx) \cdot |T| = \frac{\lambda^{-T}}{(T-1)!} e^{-\lambda^{-1}Tx}(Tx)^{T-1}T = \frac{\lambda^{-T}T^T}{(T-1)!} x^{T-1} e^{-\lambda^{-1}Tx}, \qquad x > 0.$$

So,

$$f_{\widehat{\lambda}}(x) = \frac{\lambda^{-T}T^T}{(T-1)!} x^{T-1} e^{-\lambda^{-1}Tx}, \qquad x > 0.$$

(b) From (a) above, where you have worked out the score vector $q(\lambda)$, show that $E(q(\lambda_P)) = 0$, where λ_P is the true (or population value) of the parameter. Show also that $var(q(\lambda_P)) = \frac{T}{\lambda^2}$. Hence work out the mean and variance of $\sqrt{T}\overline{q}(\lambda_P)$, where $\overline{q}(\lambda_P)$ is defined as the average score and given by the expression $\frac{q(\lambda_P)}{T}$. Using a central limit theorem can you work out the large sample or asymptotic distribution of $\sqrt{T}\overline{q}(\lambda_P)$?

$$E(q(\lambda_p) = E\left[-\frac{T}{\lambda_p} + \frac{\sum_{i=1}^T X_i}{\lambda_p^2}\right] = -\frac{T}{\lambda_p} + \frac{\sum_{i=1}^T E(X_i)}{\lambda_p^2} = = -\frac{T}{\lambda_p} + \frac{T\lambda_p}{\lambda_p^2} = -\frac{T}{\lambda_p} + \frac{T}{\lambda_p} = 0.$$

$$var(q(\lambda_p) = var\left[-\frac{T}{\lambda_p} + \frac{\sum_{i=1}^T X_i}{\lambda_p^2}\right] = \frac{1}{\lambda_p^4} var(\sum_{i=1}^T X_i).$$

$\{X_i\}_{i=1}^T$ are i.i.d., hence $var(\sum_{i=1}^T X_i) = \sum_{i=1}^T var(X_i)$. Then

$$var(q(\lambda_p)) = \frac{1}{\lambda_p^4} T var(X_i) = \frac{T\lambda_p^2}{\lambda_p^4} = \frac{T}{\lambda_p^2}.$$

Define $\overline{q}(\lambda_p) = \frac{q(\lambda_p)}{T}$. Then

$$\begin{aligned} E(\sqrt{T}\overline{q}(\lambda_p)) &= 0 \\ var(\sqrt{T}\overline{q}(\lambda_p)) &= \frac{1}{\lambda_p^2}. \end{aligned}$$

Note,

$$q(\lambda_p) = \sum_{i=1}^T q_i(\lambda_p),$$

where $q_i(\lambda_p)$ is an *individual* score vector evaluated at λ_p:

$$q_i(\lambda_p) = \frac{d}{d\lambda}(ln f_{X_i}(x_i, \lambda))|_{\lambda=\lambda_p} = -\frac{1}{\lambda_p} + \frac{X_i}{\lambda_p^2}.$$

The individual score vectors $q_i(\lambda_p)$ are i.i.d. random variables with

$$\begin{aligned} E(q_i(\lambda_p)) &= 0 \\ var(q_i(\lambda_p)) &= \frac{var(\sum_{i=1}^T q_i(\lambda_p))}{T} = \frac{var(q(\lambda_p))}{T} = \frac{1}{\lambda_p^2}. \end{aligned}$$

We can then apply the CLT and write

$$\sqrt{T}\frac{\frac{\sum_{i=1}^T q_i(\lambda_p)}{T} - E(q_i(\lambda_p))}{\sqrt{var(q_i(\lambda_p))}} \xrightarrow[d]{} N(0,1)$$

This means

$$\sqrt{T}\overline{q}(\lambda_p) \xrightarrow[d]{} N(0, var(q_i(\lambda_p))) = N\left(0, \frac{1}{\lambda_p^2}\right)$$

So, the asymptotical distribution of $\sqrt{T}\overline{q}(\lambda_p)$ is $N(0, \frac{1}{\lambda_p^2})$.

(c) Finally, write

$$\sqrt{T}(\widehat{\lambda} - \lambda_P) \simeq \left[-\frac{Q(\lambda_P)}{T}\right]^{-1} \sqrt{T}\overline{q}(\lambda_P)$$

Using a law of large numbers, work out the probability limit of $-\frac{Q(\lambda_P)}{T}$, and hence (following all the usual steps) derive the asymptotic distribution of $\sqrt{T}(\widehat{\lambda} - \lambda_P)$.

The Taylor expansion of the score vector $q(\lambda)$ around λ_p is

$$q(\lambda) \approx q(\lambda_p) + Q(\lambda^*)(\lambda - \lambda_p)$$

λ^* belongs to the segment between λ and λ_p.

At $\lambda = \widehat{\lambda}$ this becomes

$$0 = q(\widehat{\lambda}) \approx q(\lambda_p) + Q(\lambda^*)(\widehat{\lambda} - \lambda_p)$$

λ^* belongs to the segment between $\widehat{\lambda}$ and λ_p.

Hence,

$$\begin{aligned} \widehat{\lambda} - \lambda_p &\approx [-Q(\lambda^*)]^{-1} q(\lambda_p) \\ \sqrt{T}(\widehat{\lambda} - \lambda_p) &\approx \left[\frac{-Q(\lambda^*)}{T}\right]^{-1} \frac{q(\lambda_p)}{\sqrt{T}} \end{aligned}$$

or

$$\sqrt{T}(\widehat{\lambda} - \lambda_p) \approx \left[\frac{-Q(\lambda^*)}{T}\right]^{-1} \sqrt{T}\overline{q}(\lambda_p)$$

Since $\widehat{\lambda} \underset{P}{\longrightarrow} \lambda_p$ and Q is a continuous function, $Q(\lambda^*) \underset{P}{\longrightarrow} Q(\lambda_p)$. So, as T gets large,

$$\sqrt{T}(\widehat{\lambda} - \lambda_p) \approx \left[\frac{-Q(\lambda_p)}{T}\right]^{-1} \sqrt{T}\overline{q}(\lambda_p)$$

Note, that

$$Q(\lambda_p) = \sum_{i=1}^{T} Q_i(\lambda_p),$$

where $Q_i(\lambda_p)$ is an *individual* Hessian matrix evaluated at λ_p:

$$Q_i(\lambda_p) = \frac{d^2}{d\lambda^2}(ln f_{X_i}(x_i, \lambda))|_{\lambda=\lambda_p} = \frac{1}{\lambda_p^2} - \frac{2X_i}{\lambda_p^3}.$$

Since we have only one parameter in this exercise, matrix $Q_i(\lambda_p)$ consists of one element only.

The individual hessian matrices $Q_i(\lambda_p)$ are i.i.d. with

$$E(Q_i(\lambda_p)) = -\frac{1}{\lambda_p^2}\left(= \frac{E(Q(\lambda_p))}{T}\right).$$

Then by the LLN

$$\frac{\sum_{i=1}^{T} Q_i(\lambda_p)}{T} \underset{P}{\longrightarrow} E(Q_i(\lambda_p)) = -\frac{1}{\lambda_p^2}$$

that is

$$\frac{Q(\lambda_p)}{T} \underset{P}{\longrightarrow} -\frac{1}{\lambda_p^2}$$

and then

$$-\frac{Q(\lambda_p)}{T} \underset{P}{\longrightarrow} \frac{1}{\lambda_p^2}$$

$$\left[-\frac{Q(\lambda_p)}{T}\right]^{-1} \underset{P}{\longrightarrow} \lambda_p^2$$

Combining this result with the result of the CLT in part (a) and applying the Cramer's theorem, we obtain that

$$\sqrt{T}(\widehat{\lambda} - \lambda_p) \underset{d}{\longrightarrow} \lambda_p^2 N(0, \frac{1}{\lambda_p^2}) = N(0, \lambda_p^2)$$

So, $N(0, \lambda_p^2)$ is an asymptotic distribution of $\sqrt{T}(\widehat{\lambda} - \lambda_p)$.

(d) Can you construct a chi-squared test (Wald) of the null hypothesis $H_0 : \lambda = \lambda_0$, given your result in (c)?

Since

$$\sqrt{T}(\widehat{\lambda} - \lambda_p) \underset{asy}{\sim} N(0, \lambda_p^2),$$

$$\frac{\sqrt{T}(\widehat{\lambda} - \lambda_p)}{\lambda_p} \underset{asy}{\sim} N(0, 1)$$

Hence,

$$\frac{T(\widehat{\lambda} - \lambda_p)^2}{\lambda_p^2} \underset{asy}{\sim} \chi^2_{(1)}$$

$$\frac{(\widehat{\lambda} - \lambda_p)^2}{\frac{\lambda_p^2}{T}} \underset{asy}{\sim} \chi^2_{(1)}.$$

The statistic $W = \frac{(\widehat{\lambda}-\lambda_p)^2}{\frac{\widehat{\lambda}^2}{T}}$ is called a Wald statistic.

To test the hypothesis $H_0 : \lambda = \lambda_0$ against $H_1 : \lambda \neq \lambda_0$, W is compared to the critical value of $\chi^2_{(1)}$-distribution.

If $W < c.v.$, then H_0 cannot be rejected, if $W > c.v.$, then H_0 is rejected.

2. Let $X_1, X_2, ..X_T$ be an *i.i.d.* random sample from a Poisson distribution with parameter λ. Show that

$$e^{-\overline{X}_T} \underset{P}{\longrightarrow} P(X_1 = 0)$$

where $\overline{X}_T$ is the sample mean. What theorem did you use?

We know that our X_i are identically distributed according to the Poisson (λ). Therefore, we have:

$$P(X_1 = 0) = \frac{\lambda^0 e^{-\lambda}}{0!} = e^{-\lambda}$$

Moreover, by the LLN, $\overline{X}_n \to_P E(X_i)$. Expectation of the Poisson distributed random variable is λ. Thus we know that the sample mean converges in probability to λ. Furthermore, one can notice that $f : x \longmapsto e^{-x}$

is a continuous function and thus we are allowed to use Slutsky's theorem to write:

$$plim(e^{-\overline{X}_n}) = e^{-plim\overline{X}_n}$$
$$\Longrightarrow$$
$$plim(e^{-\overline{X}_n}) = e^{-\lambda} = P(X_1 = 0)$$
$$\Longrightarrow$$
$$e^{-\overline{X}_n} \underset{P}{\longrightarrow} P(X_1 = 0)$$

3. (OPTIONAL)

An investigator asserts the model

$$y_t = \alpha z_t + \epsilon_t$$

where it is claimed that $\epsilon_t \tilde{} IN(0, \sigma_\epsilon^2)$ and $E(z_t \epsilon_s) = 0 \forall t, s$.

(a) Derive the log-likelihood function for $\theta' = (\alpha, \sigma_\epsilon^2)$, its first and second derivatives $q(\theta)$ and $H(\theta)$, and the maximum likelihood estimator $\widehat{\theta}_{ML}$ of θ.

$$\epsilon_t = y_t - \alpha z_t \sim IN(0, \sigma_\epsilon^2)$$

- The likelihood function is

$$L(\theta) = \left[\frac{1}{\sqrt{2\pi\sigma^2}}\right]^T \exp\left[-\sum_{t=1}^{T} \frac{(y_t - \alpha z_t)^2}{2\sigma_\epsilon^2}\right].$$

- The log-likelihood function is

$$LogL(\theta) = -T \ln\sqrt{2\pi} - \frac{T}{2} ln\sigma_\epsilon^2 - \sum_{i=1}^{T} \frac{(y_t - \alpha z_t)^2}{2\sigma_\epsilon^2}.$$

- The score vector is

$$q(\theta) = \begin{pmatrix} \frac{\partial}{\partial\alpha} LogL(\theta) \\ \frac{\partial}{\partial\sigma_\epsilon^2} LogL(\theta) \end{pmatrix} = \begin{pmatrix} \sum_{i=t}^{T} \frac{(y_t - \alpha z_t) z_t}{\sigma_\epsilon^2} \\ -\frac{T}{2\sigma_\epsilon^2} + \sum_{t=1}^{T} \frac{(y_t - \alpha z_t)^2}{2\sigma_\epsilon^4} \end{pmatrix}$$

- The Hessian is

$$H(\theta) = \begin{pmatrix} \frac{\partial^2}{\partial\alpha^2} LogL & \frac{\partial^2}{\partial\alpha\partial\sigma_\epsilon^2} LogL \\ \frac{\partial^2}{\partial\alpha\partial\sigma_\epsilon^2} LogL & \frac{\partial^2}{\partial(\sigma_\epsilon^2)^2} LogL \end{pmatrix} = \begin{pmatrix} -\frac{\sum_{t=1}^{T} z_t^2}{\sigma_\epsilon^2} & , -\frac{\sum_{t=1}^{T}(y_t - \alpha z_t) z_t}{\sigma_\epsilon^4} \\ -\frac{\sum_{t=1}^{T}(y_t - \alpha z_t) z_t}{\sigma_\epsilon^4} & , \frac{T}{2\sigma_\epsilon^4} - \frac{\sum_{t=1}^{T}(y_t - \alpha z_t)^2}{\sigma_\epsilon^6} \end{pmatrix}$$

- The maximum likelihood estimator $\widehat{\theta}_{ML} = argmaxlogL(\theta)$, so $q(\widehat{\theta}_{ML}) = 0$. Hence,

$$\hat{\theta}_{ML} = \begin{pmatrix} \hat{\alpha}_{ML} \\ \hat{\sigma}_{\epsilon ML}^2 \end{pmatrix} = \begin{pmatrix} \frac{\sum_{t=1}^{T} y_t z_t}{\sum_{t=1}^{T} z_t^2} \\ \frac{\sum_{t=1}^{T}(y_t - \hat{\alpha}_{ML} z_t)^2}{T} \end{pmatrix}.$$

(b) In fact suppose that the investigator has misspecified the model. Suppose that the process ϵ_t is generated as a moving average process of order 1 - i.e.

$$\epsilon_t = u_t - \lambda u_{t-1}$$

$u_t\tilde{\ } IN(0,1)$.

Assume also that z_t are determined outside the model and therefore, may be considered as fixed.

Compute $E(q(\theta_p))$ and $E(H(\theta_p))$ for the model asserted in (a) when (b) holds. θ_p is the population value of the parameter θ. Can you check if the information matrix identity holds? Why not?

If the true generating process for ϵ_t is $\epsilon_t = u_t - \lambda u_{t-1}$ and $u_t\tilde{\ } IN(0,1)$, then $\epsilon_t\tilde{\ } N(0, 1+\lambda^2)$.

Note that the score vector and the Hessian matrix evaluated at the (true) population value θ_p of the parameter θ can be written in the form:

$$q(\theta_p) = \begin{pmatrix} \sum_{i=t}^{T} \frac{\epsilon_t z_t}{\sigma_\epsilon^2} \\ -\frac{T}{2\sigma_\epsilon^2} + \sum_{t=1}^{T} \frac{\epsilon_t^2}{2\sigma_\epsilon^4} \end{pmatrix},$$

$$H(\theta_p) = \begin{pmatrix} -\frac{\sum_{t=1}^{T} z_t^2}{\sigma_\epsilon^2} & , -\frac{\sum_{t=1}^{T} \epsilon_t z_t}{\sigma_\epsilon^4} \\ -\frac{\sum_{t=1}^{T} \epsilon_t z_t}{\sigma_\epsilon^4} & , \frac{T}{2\sigma_\epsilon^4} - \frac{\sum_{t=1}^{T} \epsilon_t^2}{\sigma_\epsilon^6} \end{pmatrix}$$

Also the product $q(\theta_p)q(\theta_p)\prime$ is

$$q(\theta_p)q(\theta_p)\prime = \begin{pmatrix} \frac{(\sum_{t=1}^{T} \epsilon_t z_t)^2}{\sigma_\epsilon^4}, & -\frac{T\sum_{t=1}^{T} \epsilon_t z_t}{2\sigma_\epsilon^4} + \frac{(\sum_{t=1}^{T} \epsilon_t z_t)(\sum_{t=1}^{T} \epsilon_t^2)}{2\sigma_\epsilon^6} \\ -\frac{T\sum_{t=1}^{T} \epsilon_t z_t}{2\sigma_\epsilon^4} + \frac{(\sum_{t=1}^{T} \epsilon_t z_t)(\sum_{t=1}^{T} \epsilon_t^2)}{2\sigma_\epsilon^6}, & \frac{T^2}{4\sigma_\epsilon^4} - \frac{T}{2\sigma_\epsilon^6}\sum_{t=1}^{T} \epsilon_t^2 + \frac{(\sum_{t=1}^{T} \epsilon_t^2)^2}{4\sigma_\epsilon^8} \end{pmatrix}$$

Then, given the true distribution of ϵ_t and given that z_t are taken as fixed,

$$E(q(\theta_p)) = \begin{pmatrix} \sum_{i=t}^{T} \frac{z_t E(\epsilon_t)}{\sigma_\epsilon^2} \\ -\frac{T}{2\sigma_\epsilon^2} + \sum_{t=1}^{T} \frac{E(\epsilon_t^2)}{2\sigma_\epsilon^4} \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{T}{2\sigma_\epsilon^2} + \frac{T(1+\lambda^2)}{2\sigma_\epsilon^4} \end{pmatrix} \neq 0 \text{ unless } 1+\lambda^2 = \sigma_\epsilon^2$$

Similarly,

$$E(H(\theta_p)) = \begin{pmatrix} -\frac{\sum_{t=1}^{T} z_t^2}{\sigma_\epsilon^2}, & 0 \\ 0, & \frac{T}{2\sigma_\epsilon^4} - \frac{T(1+\lambda^2)}{\sigma_\epsilon^6} \end{pmatrix}$$

Let us check whether the information matrix identity

$$-E(H(\theta_p)) = E[q(\theta_p)q(\theta_p)\prime]$$

holds. For this purpose, first, compute the expectation of one element in the matrix $q(\theta_p)q(\theta_p)\prime$ and compare it to the "-" expectation of the corresponding element in $H(\theta_p)$.

$$E\left(\frac{(\sum_{t=1}^{T} \epsilon_t z_t)^2}{\sigma_\epsilon^4}\right) = \frac{1}{\sigma_\epsilon^4} E\left(\sum_{t=1}^{T} \epsilon_t^2 z_t^2 + \sum_{t\neq s} \epsilon_t \epsilon_s z_t z_s\right).$$

$\forall t \in 1:T$ ϵ_t is only correlated with ϵ_{t-1} and ϵ_{t+1} and

$$\begin{aligned} cov(\epsilon_t, \epsilon_{t-1}) &= cov(u_t - \lambda u_{t-1}, u_{t-1} - \lambda u_{t-2}) = cov(-\lambda u_{t-1}, u_{t-1}) = -\lambda, \\ cov(\epsilon_t, \epsilon_{t+1}) &= cov(u_t - \lambda u_{t-1}, u_{t+1} - \lambda u_t) = cov(u_t, -\lambda u_t) = -\lambda. \end{aligned}$$

Therefore, $E(\sum_{t\neq s}\epsilon_t\epsilon_s z_t z_s) = E\left(\sum_{t=1}^{T}[\epsilon_t\epsilon_{t-1}z_t z_{t-1} + \epsilon_t\epsilon_{t+1}z_t z_{t+1}]\right)$. So, we obtain:

$$\begin{aligned}
E\left(\frac{(\sum_{t=1}^{T}\epsilon_t z_t)^2}{\sigma_\epsilon^4}\right) &= \frac{1}{\sigma_\epsilon^4}\left(\sum_{t=1}^{T} z_t^2 E(\epsilon_t^2)\right) + E\left(\sum_{t=1}^{T}[\epsilon_t\epsilon_{t-1}z_t z_{t-1} + \epsilon_t\epsilon_{t+1}z_t z_{t+1}]\right) = \\
&= \frac{1}{\sigma_\epsilon^4}\left((1+\lambda^2)\sum_{t=1}^{T} z_t^2 + \sum_{t=1}^{T}[z_t z_{t-1}E(\epsilon_t\epsilon_{t-1}) + z_t z_{t+1}E(\epsilon_t\epsilon_{t+1})]\right) = \\
&= \frac{1}{\sigma_\epsilon^4}\left((1+\lambda^2)\sum_{t=1}^{T} z_t^2 + \sum_{t=1}^{T}[z_t z_{t-1}cov(\epsilon_t,\epsilon_{t-1}) + z_t z_{t+1}cov(\epsilon_t,\epsilon_{t+1})]\right) = \\
&= \frac{1}{\sigma_\epsilon^4}\left((1+\lambda^2)\sum_{t=1}^{T} z_t^2 - \lambda\sum_{t=1}^{T}[z_t z_{t-1} + z_t z_{t+1}]\right).
\end{aligned}$$

This is generally not equal to $\frac{\sum_{t=1}^{T} z_t^2}{\sigma_\epsilon^2}$ and therefore, the information matrix identity does not hold. It does not hold because the score vector and the Hessian matrix are computed using the log-likelihood function of the *misspecified* model while the expectations are calculated using the *true* distribution of ϵ_t. This confirms the importance of the distributional assumption in the ML estimation.