Statistics 531/Econ 677
Winter, 2005
Ed Ionides

SECTION A. We investigate some data from neurophysiology. An electrode implanted (painlessly) into a monkey's brain records a sequence of firing events for an individual neuron cell (neurons communicate by "firing" pulses of electrical charge). Suppose the firing times are $F_1, F_2, \ldots, F_{T+1}$, measured in milliseconds (1ms is 10^{-3}s). We take as our time series $x_t = F_{t+1} - F_t$ with $t = 1, \ldots, T$. This is the series of times intervals between firing events. The data, with $T = 415$, are plotted in Fig. 1. We wish to model x_t in order to quantify the behavior of the neuron, to later compare it with other neurons and investigate the effects of experimental treatments. The estimated ACF and PACF of x_t are shown in Fig. 2.

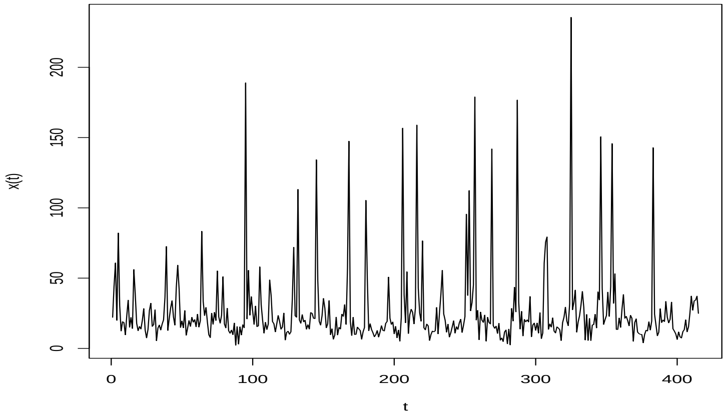

Figure 1: Time series x_t of intervals between neuron firing events

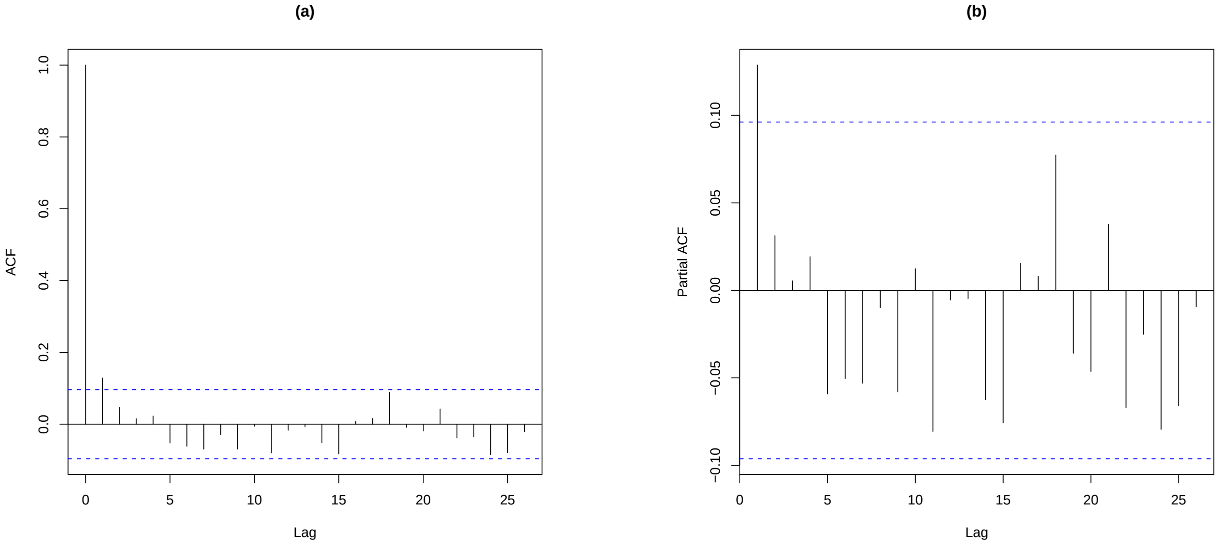

Figure 2: (a) Estimated autocorrelation function of x_t. (b) Estimated partial autocorrelation function of x_t.

A1. [2 pts] What ARMA model for x_t does Fig. 2 suggest, and why?
There is no single clear-cut answer. The PACF appears to cut off at lag 1. The ACF is also within the dashed lines after lag 1, but shows some indication of decreasing like a damped oscillation (an AR(2) property). The PACF also has some resemblance to a damped oscillation (an MA(2) property). Anything between AR(1) and ARMA(2,2) can be defended by inspecting Fig. 2!
A2. [2 pts] Another way to select a model is by comparing AIC values. A table of AIC values is shown in Table 1. What ARMA model does this suggest, and why?

$p \setminus q$	0	1	2	3
0	NA	3961.5	3962.7	3964.7
1	3961.0	3962.6	3964.6	3966.6
2	3962.6	3960.5	3959.7	3961.7
3	3964.6	3965.4	3962.6	3968.3

Table 1: AIC values from fitting ARMA(p,q) models to x_t

ARMA(2,2) has the lowest AIC, so is favored by this criterion. ARMA(2,1) and ARMA(1,0) are simpler models which also have promising AIC values.
A3. [2 pts] Find the log likelihood of an ARMA(2,1) model, and explain your calculation.
AIC $= -2\lambda + 2k$, *where* λ *is the maximized log likelihood and* k *is the number of parameters. Here,* $k = 5$ $(\phi_1, \phi_2, \theta_1, \mu, \sigma^2)$ *so*

$$\lambda = \frac{3960.5 - 10}{-2} = -1975.25$$

A4. [2 pts] Does the table of AIC values contain any evidence for or against the claim that the likelihood is correctly calculated and maximized? Explain.
The table is inconsistent — adding a parameter can only increase the maximized log-likelihood, i.e. the AIC can only increase by ≤ 2 *Compare ARMA(3,3) to ARMA(3,2). This can only come about by imperfect likelihood calculation and/or maximization.*
SECTION B. Fitting an ARMA(2,2) model gives the following *R* printout.

```
Call:
arima(x = x, order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1     ma2  intercept
      1.6009  -0.6445  -1.4982  0.5219    26.4163
s.e.  0.1886   0.1839   0.2104  0.2094     0.7954

sigma^2 estimated as 791.7:  log likelihood = -1973.88,  aic = 3959.76
```

B1. [4 pts]. Write out the fitted model, carefully stating all the model assumptions.

$$x_t = 26.4 + 1.60(x_{t-1} - 26.4) - 0.64(x_{t-2} - 26.4) + w_t - 1.50w_{t-1} + 0.5w_{t-2}$$

where w_t *is white noise with standard deviation 28.1. This calculation also assumes that* w_t *is Gaussian, i.e.* $w_t \sim N(0, 28.1^2)$

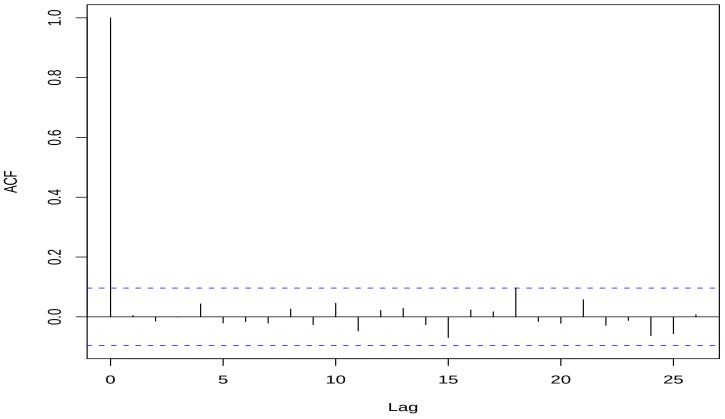

Figure 3: Estimated autocorrelation function of the residuals from fitting an ARMA(2,2) model to x_t.

B2. [3 pts] Fig. 3 shows the ACF of the residuals from fitting an ARMA(2,2) model. Comment on which modeling assumptions this figure supports, and which it does not support.

It supports the assumptions that the driving noise process is uncorrelated and has no trend. It does not reveal anything about the assumptions that the driving noise is (i) constant variance or (ii) Gaussian.

B3. [3 pts] Find the roots of the AR and the MA polynomials for the fitted ARMA(2,2) model. Is there evidence for parameter redundancy?

AR polynomial: $1 - 1.60z + 0.645z^2$ *has roots:* $\frac{1.60\pm\sqrt{1.60^2-4(0.645)}}{2(0.645)} = 1.24 \pm 0.10i$

MA polynomial: $1 - 1.50z + 0.52z^2$ *has roots:* $1.06, 1.81$.

So the AR roots ar not very close to the MA roots — there is no strong suggestion of parameter redundancy.

B4. [2 pts] Simulations from the fitted ARMA(2,2) model were computed as follows:

```
arma22<-arima(x,order=c(2,0,2))
Nt<-length(x)
sim<-rep(0,Nt)
w<-rnorm(Nt,m=0,sd=sqrt(arma22$sigma2))
for(nt in 3:Nt){
 sim[nt]<-arma22$coef["ar1"]*sim[nt-1]+arma22$coef["ar2"]*sim[nt-2]+
         arma22$coef["ma1"]*w[nt-1]+arma22$coef["ma2"]*w[nt-2]+w[nt]
}
sim<-sim+arma22$coef["intercept"]
```

Sample output is shown in Fig. 4. What does a comparison of Fig. 4 with Fig. 1 say about ARMA modeling of x_t?

Almost all well-considered and comprehensible answers were accepted here. The plots look quite different. Fig. 1 is always positive, and appears to have some regularity to the peaks. Fig. 4 has less pronounced peaks, is more symmetric about its mean, and occasionally becomes negative. The data do not resemble a Gaussian ARMA(2,2) process.

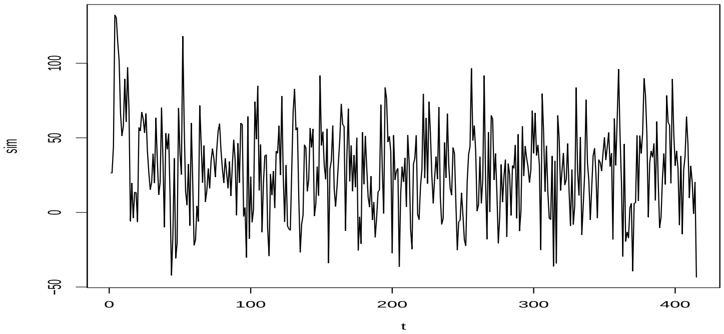

Figure 4: A simulation from the fitted ARMA(2,2) model

B5. [2 pts] Is the random process generated in B4 and plotted in Fig. 4 stationary? Answer yes or no, and explain.
No. The initial values contradict having a constant variance (the initial variance is zero). However, the process is asymptotically stationary, as in Shumway and Stoffer Problem 2.2. One could make the simulations effectively stationary by throwing away some number (say 100) values at the start of the simulation.
SECTION C. We now investigate a logarithmic transformation of the data. Below is the R printout from fitting an ARMA(2,2) model to $\log_{10} x_t$.

```
Call:
arima(x = log10(x), order = c(2, 0, 2))

Coefficients:
         ar1      ar2      ma1     ma2  intercept
      1.6975  -0.7250  -1.4647  0.4647     1.2925
s.e.  0.0740   0.0718   0.0941  0.0939     0.0021

sigma^2 estimated as 0.07637:  log likelihood = -56.79,  aic = 125.59
```

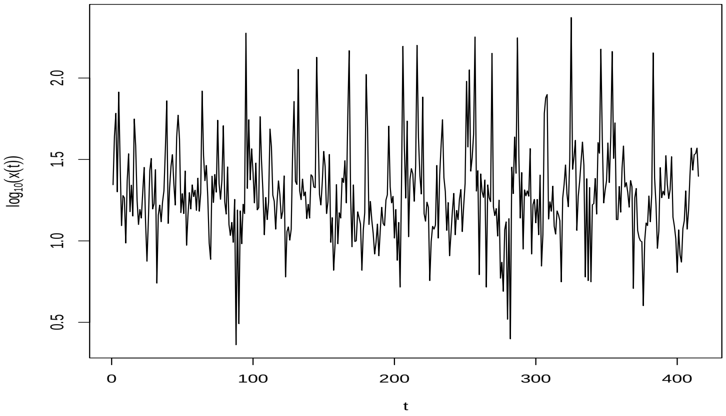

Figure 5: Time plot of $\log_{10} x_t$.

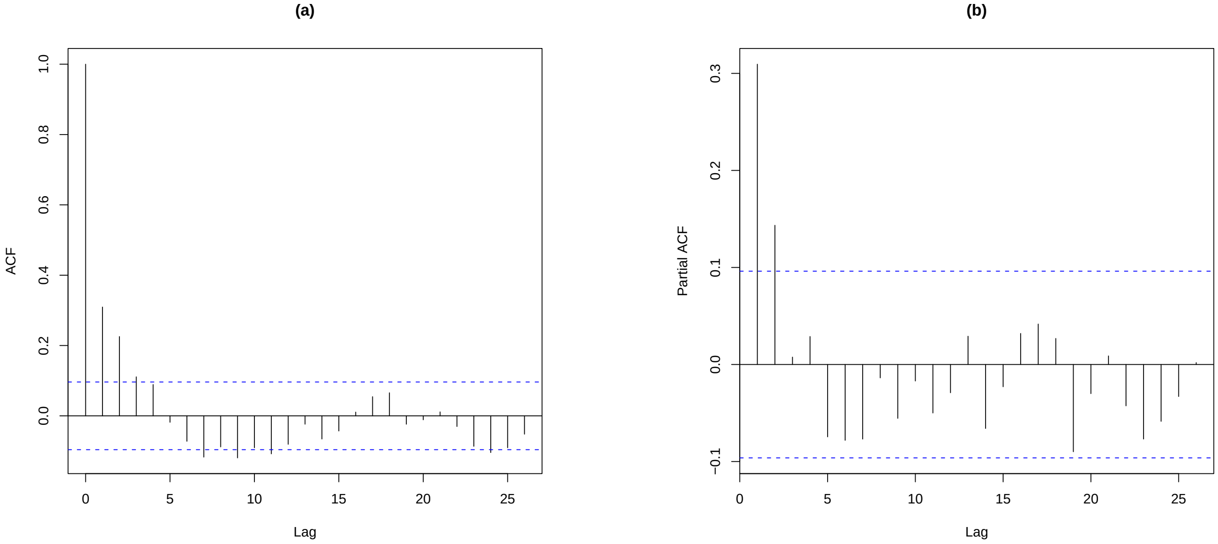

Figure 6: (a) Estimated autocorrelation function of $\log_{10} x_t$. (b) Estimated partial autocorrelation function of $\log_{10} x_t$.

C1. [2 pts] Is there any indication from Fig. 5, Fig. 6 and the fitted model printouts in Sections B and C that ARMA modeling is more successful after a log transformation? or less? Explain.

The model printouts do not tell us whether the transformation is appropriate — the AIC values are not comparable. The estimated ACF does not say anything about the transformation. The time plot looks more symmetric — symmetry is a property of Gaussian ARMA models, so this is encouraging.

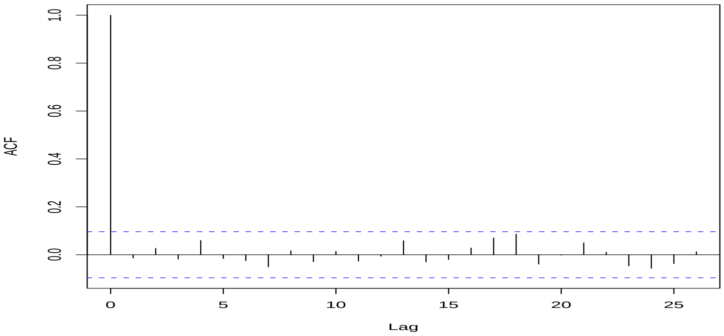

Figure 7: Estimated autocorrelation function of the residuals from fitting an ARMA(2,2) model to $\log_{10} x_t$.

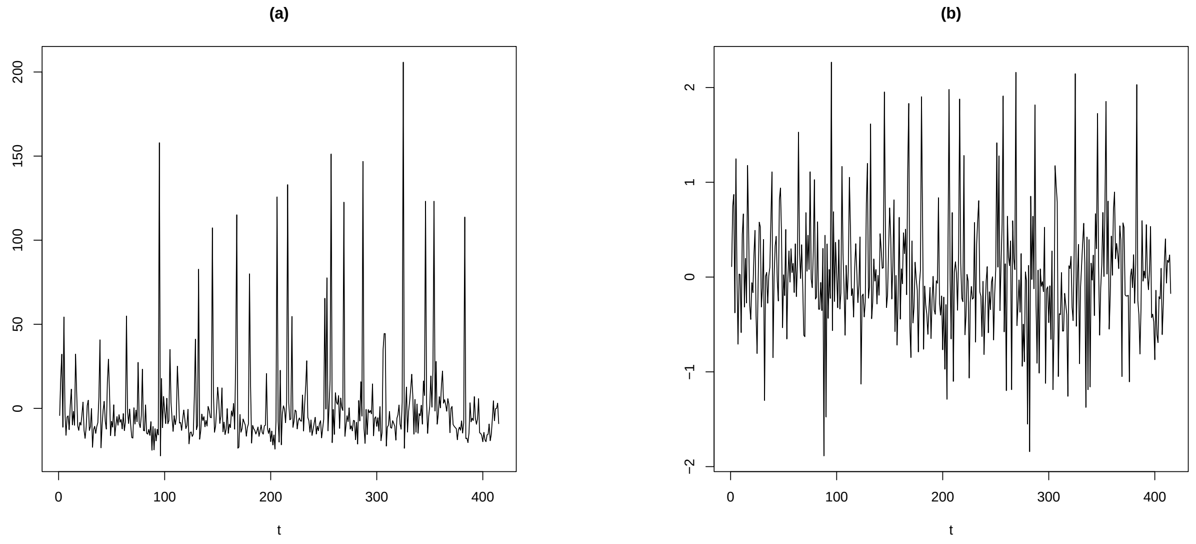

Figure 8: (a) Residuals from fitting an ARMA(2,2) model to x_t. (b) Residuals from fitting an ARMA(2,2) model to $\log_{10} x_t$.

C2. [2 pts] What do Figs. 3, 7 and 8 indicate about the success of the log transform?
Figs. 3 and 7 show that the residuals appear uncorrelated with or without the log transform. Fig. 8 shows that the residuals look much more like Gaussian white noise after the transform. Since the models are fitted using Gaussian maximum likelihood, this is a good thing. We could confirm this finding by a normal quantile plot.

C3. [3 pts] Fig. 9 shows the smoothed periodogram of $\log_{10} x_t$. Find the frequency and period corresponding to the peak in the periodogram. Your answer should include the units of these quantities. Describe briefly what this peak leads you to conclude about how the monkey's neuron behaves.
The peak at frequency approx. 0.06 cycles/firing event (or period approx. 16.7 firing events) is due tothe neuron firing in a burst (several short inter-event times) and then being less active (long inter-event times) with a characteristic period of approx. 17 firing events.

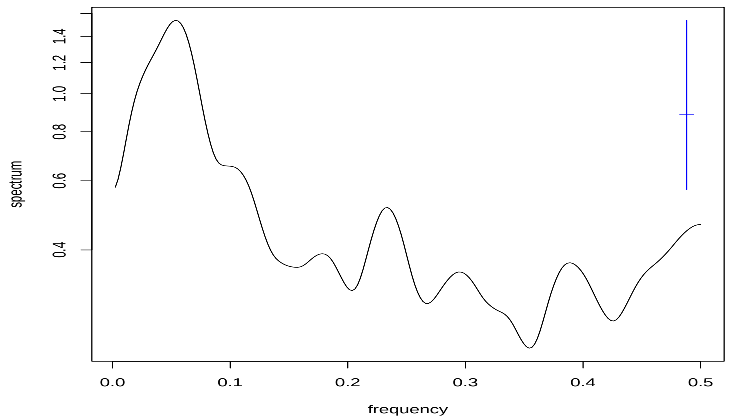

Figure 9: Smoothed periodogram of $\log_{10} x_t$.