Mathematical Formulation

Suppose the variance of a time series $Z_t$ satisfies

$$\text{var}(Z_t) = cf(\mu_t)$$

We wish to find a transformation such that, $T(\cdot)$, such that $\text{var}[T(Z_t)]$ is constant.

A first-order Taylor series of $T(Z_t)$ about $\mu_t$ is

$$T(Z_t) \approx T(\mu_t) + T'(\mu_t)(Z_t - \mu_t)$$

Now $\text{var}[T(Z_t)]$ is approximated as

$$\text{var}[T(Z_t)] \approx [T'(\mu_t)]^2 \text{var}(Z_t) = c [T'(\mu_t)]^2 f(\mu_t)$$

Therefore $T(\cdot)$ is chosen such that

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}}$$

which implies

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu)}} \, d\mu_t$$

### Box-Cox Transformation

Transforming the time series can suppress large fluctuations. The most standard transformation is the log transformation where the new series $y_t$ is given by

$$y_t = \log x_t$$

An alternative to the log transformation is the Box-Cox transformation:

$$y_t = \begin{cases} 
  (x_t^\lambda - 1)/\lambda, & \lambda \neq 0 \\
  \log x_t, & \lambda = 0 
\end{cases}$$

Many other transformations are suggested here.

### Box-Cox in R

```r
> library(MASS)
> library(forecast)
> x <- rnorm(100)^2
> ts.plot(x)
> truehist(x)
```
Box-Cox in R (II)

```r
> bc <- boxcox(x ~ 1)
> lam <- bc$x[which.max(bc$y)]
> lam
[1] 0.2222222
> truehist(BoxCox(x, lam))
```

Some Very Old Data

```r
> varve = scan("mydata/varve.dat")
> varve2 = diff(log(varve))
> ts.plot(varve2)
> acf(varve2, lwd=5)
> pacf(varve2, lwd=5)
```

Glacial Varves

Transformed Glacial Varve Series

The transformation $\nabla \log(\text{varve})$ appears appropriate although fractional differencing may be in order. Let’s take a closer look at $\nabla \log(\text{varve})$. 

```r
> varve = scan("mydata/varve.dat")
> varve2 = diff(log(varve))
> ts.plot(varve2)
> acf(varve2, lwd=5)
> pacf(varve2, lwd=5)
```
Diagnostics of ARIMA(0,1,1) on Logged Varve Data

```r
> (varve.ma = arima(log(varve), order = c(0, 1, 1)))
Call:
arima(x = log(varve), order = c(0, 1, 1))
Coefficients:
   ma1
-0.7705
s.e. 0.0341
sigma^2 estimated as 0.2353: log likelihood = -440.72, aic = 885.44
> tsdiag(varve.ma)
```

Fitting ARIMA(1,1,1) to Logged Varve Data

```r
> pacf(varve.ma$resid, lwd=5)
> (varve.arma = arima(log(varve), order = c(1, 1, 1)))
Call:
arima(x = log(varve), order = c(1, 1, 1))
Coefficients:
ar1  ma1
0.2330 -0.8858
s.e. 0.0518 0.0292
sigma^2 estimated as 0.2284: log likelihood = -431.44, aic = 868.88
> tsdiag(varve.ma)
```
Varve ARIMA(1,1,1) Diagnostics

> tsdiag(varve.arma)

Watch Out for Overfitting!

Homework 3b

Read §5.1 and §5.2 of the textbook.
Do exercise #4.5 on page 87 of the text.