STATS 207: Time Series Analysis
Autumn 2020

Lecture 3: Sample ACF; Time-Series regression

Dr. Alon Kipnis
Slides credit: David Donoho

September 21th 2020

Outline

Estimating Auto- and Crosscovariance

Classical Regression

Model Complexity & Nested Models

Sinusoidal Regression

Genera Info

- Home assignment 1 is out. It is Due two weeks later (Monday 10/5/2020).
- Slides and recordings of lectures 1 and 2 are available on Canvas.
- You will have the option to drop the final exam/assessment from the final grade.

Estimating Auto- and Crosscovariance

Recap: Autocovariance

- *Autocovariance function*

$$\gamma_x(t+h,t) \equiv \mathbb{E}\left[(x_{t+h} - \mu_{x(t+h)})(x_t - \mu_{xt})\right]$$

- *Sample autocovariance* function

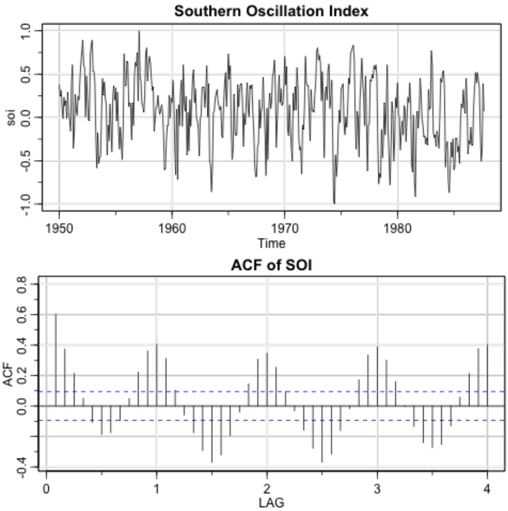
$$\hat{\gamma}_x(h) \equiv \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h} - \bar{x})(x_t - \bar{x}).$$

- (x_t) is stationary if $\mu_{xt} = \mu$ and $\gamma_x(t+h,t)$ is only a function of h, say $\gamma_x(h)$.
- Under stationarity assumption,

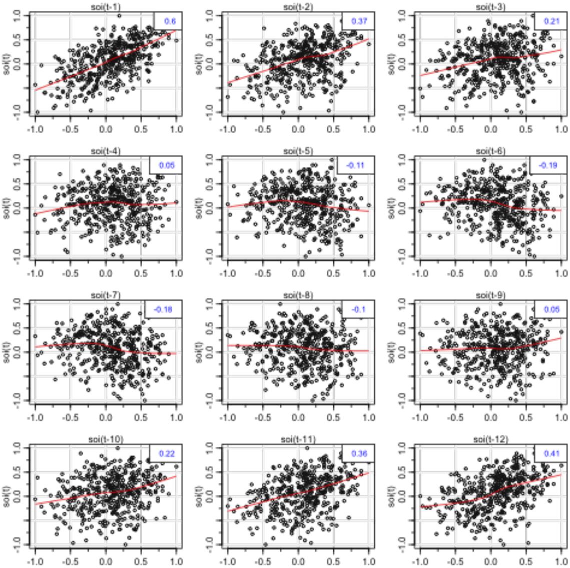
$$\hat{\gamma}_x(h) \to \gamma_x(h) \qquad (LLN).$$

- Similar result holds for the autocorrelation $\rho_x(h)$ and sample autocorrelation $\hat{\rho}_x(h)$.

Example: Southern Oscillation Index (SOI)



Southern Oscillation Index (SOI) Scatter Plot



Recap: Crosscovariance

- *Crosscovaraince function*

$$\gamma_{xy}(t+h,t) \equiv \mathbb{E}\left[(x_{t+h}-\mu_{x(t+h)})(y_t-\mu_{yt})\right]$$

- *Sample crosscovaraince* function

$$\hat{\gamma}_{xy}(h) \equiv \frac{1}{n}\sum_{t=1}^{n-h}(x_{t+h}-\bar{x})(y_t-\bar{y}).$$

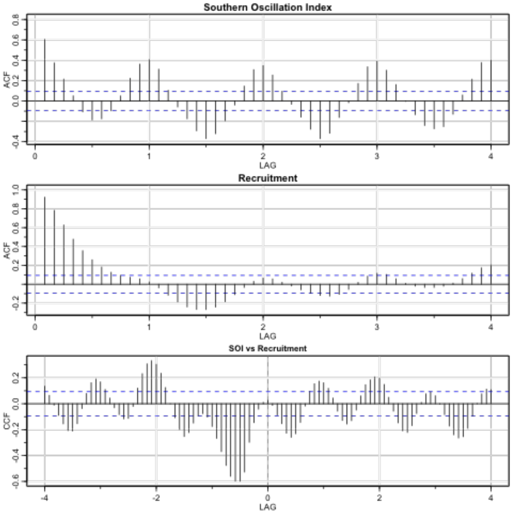
- (x_t), (y_t) are jointly stationary if both are stationary and $\gamma_{xy}(t+h,t)$ is only a function of h, say $\gamma_{xy}(h)$.
- Under joint stationarity assumption,

$$\hat{\gamma}_{xy}(h) \to \gamma_{xy}(h) \qquad (LLN).$$

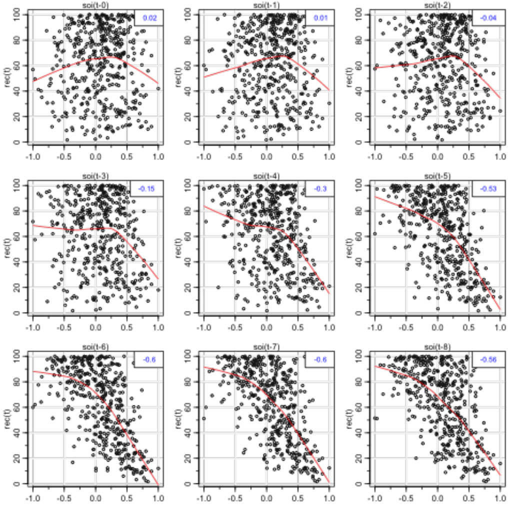
- Similar result holds for the crosscorrelation $\rho_{xy}(h)$ and sample crosscorrelation $\hat{\rho}_{xy}(h)$.

Example: CCF of SOI/Recruitment Data

Example 1.28 in [Shumway & Stoffer]



SOI/Recruitment Scatter Plot



Confidence Limits for Autocovariance

For white noise (w_t):

- Sample autocorrelation:

$$\mathrm{SE}\left[\hat{\rho}_w(h)\right] \approx \frac{1}{\sqrt{n}}.$$

- Sample cross-correlation:

$$\mathrm{SE}\left[\hat{\rho}_{wx}(h)\right] \approx \frac{1}{\sqrt{n}},$$

  where (x_t) is independent of (w_t).

(approximations assume $h \ll n$)

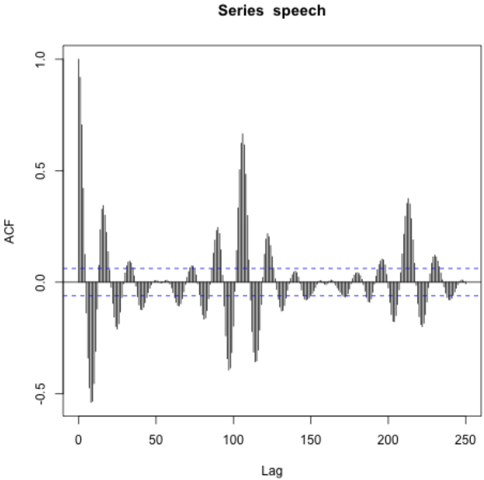
- From the CLT (Theorem A.7 and Property P1.1 in [Shumway & Stoffer]): for $h > 0$,

$$\Pr\left(|\hat{\rho}_w(h)| > 1.96/\sqrt{n}\right) \approx \Pr\left(|\mathcal{N}(0,1)| > 1.96\right) = 0.05.$$

$$\Pr\left(|\hat{\rho}_{wx}(h)| > 1.96/\sqrt{n}\right) \approx \Pr\left(|\mathcal{N}(0,1)| > 1.96\right) = 0.05.$$

ACF of Speech Data

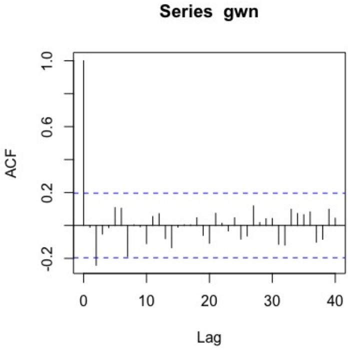
Example 1.27 in [Shumway & Stoffer]



Not a white noise!

Multiple Testing Warning

- If $|\hat{\rho}_x(h)| > 1.96n^{-1/2}$ for *some* $h > 0$, can we determine that (x_t) is not a white noise?



Classical Regression

Ordinary Least Squares (OLS) regression

- Linear regression model:

$$x_t = \beta_0 + \beta_1 z_{t1} + \ldots, \beta_q z_{tq} + w_t.$$

- $\{\beta_j\}$ unknown fixed regression coefficients.
- (w_t) is white noise
- x_t dependent variable (to be predicted)
- $\{z_{tj}\}$ independent variables (predictors)
- Matrix notation:

$$x_t = \boldsymbol{\beta}' \mathbf{z}_t + w_t$$

Example: Estimating Trend

Chicken price over time (Example 2.1):

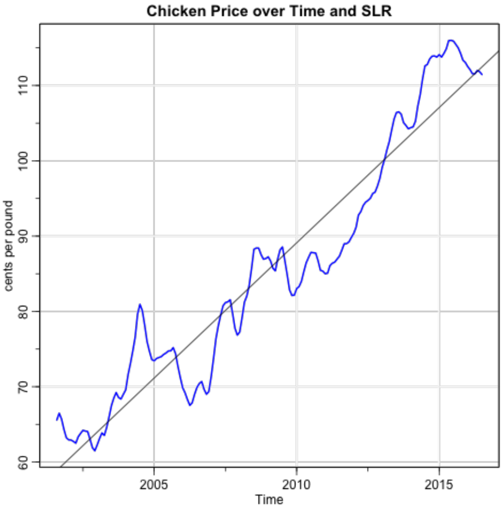
$$x_t = \beta_0 + \beta_1 z_t + w_t, \quad z_t = 2001\frac{7}{12}, 2001\frac{8}{12}, \ldots, 2016\frac{6}{12} \quad \text{(180 months)}.$$

- Regression model with $q = 1$
- OLS:

```
summary(fit <- lm(chicken~time(chicken))) # regress price on time
#Coefficients:
#                  Estimate        Std. Error
#   (Intercept)   -7.131e+03        1.624e+02
#   time(chicken)  3.592e+00        8.084e-02
```

- Interpretation: β_1 is the increment in price (cents) per month

Example 2.1



Chicken Price over Time and SLR
cents per pound
60
70
80
90
100
110
2005
2010
2015
Time

Terminal Output

```
Call: lm(formula = chicken ~ time(chicken))

Residuals:
    Min      1Q  Median      3Q     Max
-8.7411 -3.4730  0.8251  2.7738 11.5804

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -7.131e+03  1.624e+02  -43.91   <2e-16 ***
time(chicken)  3.592e+00  8.084e-02   44.43   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.696 on 178 degrees of freedom
Multiple R-squared:  0.9173,    Adjusted R-squared:  0.9168
F-statistic:  1974 on 1 and 178 DF,  p-value: < 2.2e-16
```

OLS Matrix Notation/Formulas

$$x_t = \boldsymbol{\beta}' \mathbf{z}_t + w_t$$

- $\mathbf{z}_t$ is $q + 1$-dimensional vector of predictors (features) at time t.
- $\beta = (\beta_0, \ldots, \beta_q)'$ vector of regression coefficients
- LS estimator for β:

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n} \mathbf{z}_t \mathbf{z}_t' \right)^{-1} \sum_{t=1}^{n} \mathbf{z}_t x_t .$$

- Minimized error sum of squares:

$$\mathrm{SSE} \equiv \mathrm{SSE}_q \equiv \sum_{t=1}^{n} (x_t - \hat{\boldsymbol{\beta}}' \mathbf{z}_t)^2 .$$

(as an estimator to β, $\hat{\beta}$ is unbiased and has minimum variance over the class of unbiased estimators).

OLS Formulas (cont'd)

- An unbiased estimator for σ_w^2:

$$s_w^2 \equiv \hat{\sigma}_w^2 \equiv \mathrm{MSE} \equiv \frac{\mathrm{SSE}}{n-(q+1)}$$

(aka as *adjusted* sum of squares error).

- Assume that (w_t) is white Gaussian noise. Then

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma_w^2 \cdot C), \qquad C = \left(\sum_{t=1}^{n} \mathbf{z}_t \mathbf{z}_t'\right)^{-1}.$$

Consequently,

$$t_i \equiv \frac{\hat{\beta}_i - \beta_i}{s_w \sqrt{C_{i,i}}} \text{ has a } t \text{ distribution with } n-(q+1) \text{ DoF.}$$

($\beta_i = 0$ in output of `summary(fit)`)

Competing Models

- Suppose a competing model

$$x_t = \beta_0 + \beta_1 z_{t1} + \cdots + \beta_r z_{tr} + w_t, \quad r < q,$$

equivalently:

$$H_0 \ : \ \beta_{r+1} = \ldots = \beta_q = 0.$$

- Test the reduced model ($r + 1$ coefficients) against the full model ($q + 1$ coefficients):

$$F \equiv \frac{(\mathrm{SSE}_r - \mathrm{SSE}_q)/(q - r)}{\mathrm{SSE}_q/(n - q - 1)}.$$

Under H_0, $F \sim F^{q-r}_{n-q-1}$ (F distribution with $q - r$ DoF and $n - q - 1$ DoF)

- **Coefficient of determination**:

$$R^2 \equiv \frac{\mathrm{SSE}_0 - \mathrm{SSE}_q}{\mathrm{SSE}_0}, \quad \mathrm{SSE}_0 = \sum_{t=1}^{n} (x_t - \bar{x})^2,$$

(proportion of variation accounted for by all variables compared to the sum of squares error under the model $x_t = \beta_0 + w_t$).

Terminal Output

```
Call: lm(formula = chicken ~ time(chicken))

Residuals:
    Min      1Q  Median      3Q     Max 
-8.7411 -3.4730  0.8251  2.7738 11.5804 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)   -7.131e+03  1.624e+02  -43.91   <2e-16 ***
time(chicken)  3.592e+00  8.084e-02   44.43   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.696 on 178 degrees of freedom
Multiple R-squared:  0.9173,    Adjusted R-squared:  0.9168 
F-statistic:  1974 on 1 and 178 DF,  p-value: < 2.2e-16
```

Model Complexity & Nested Models

Overfitting

- In-sample MSE

$$\mathrm{SSE}_k = \sum_{t=1}^{n} (x_t - \hat{\beta}' \mathbf{z}_t)^2$$

  always decreases in the number of features k.

- **Consequence**: If you simply pick the model with the smallest SSE, it will always be the biggest model, even if the predictor are junk! aka **overfitting**.

- D. Donoho: "Overfitting is the biggest systematic validity problem in science." (systematically claiming 'structure' that's just noise). See also J. Ioannidis.

- The situation is similar with the adjusted SSE

$$\frac{\mathrm{SSE}_k}{n - (k+1)},$$

  as the penalty in increasing k is not sever enough.

Complexity Penalization

- Ideology: we need more sever penalization than adjusted SSE.
- Definition: Akaike's Information Criterion (AIC):

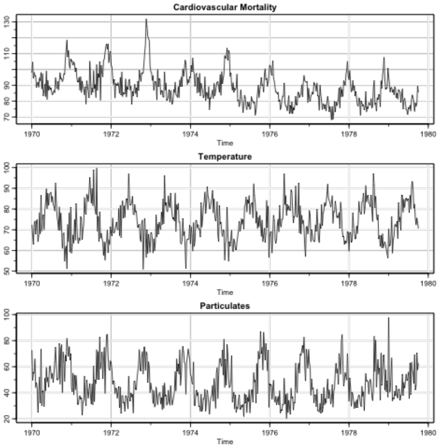
$$\mathrm{AIC}(k) \equiv \log \hat{\sigma}_k^2 + \frac{n + 2k}{n}, \qquad \hat{\sigma}_k^2 \equiv \frac{\mathrm{SSE}_k}{n},$$

  where k is the number of parameters in the model.
- Other criteria:

$$\mathrm{AIC}_c(k) \equiv \log \hat{\sigma}_k^2 + \frac{n + k}{n - k - 2}; \qquad \mathrm{BIC}(k) \equiv \log \hat{\sigma}_k^2 + \frac{k \log(n)}{n}.$$

Pollution, Temperature and Mortality (Example 2.2)

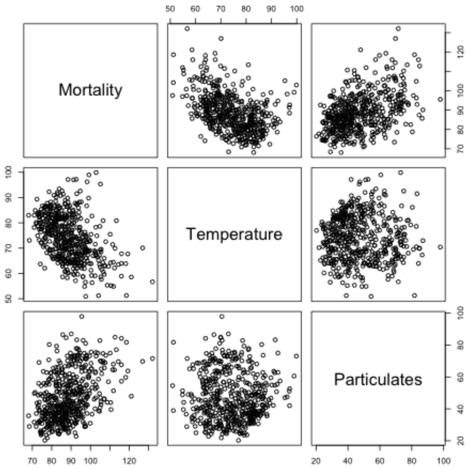
```
par(mfrow=c(3,1))
tsplot(cmort, main="Cardiovascular Mortality", ylab="")
tsplot(tempr, main="Temperature",  ylab="")
tsplot(part, main="Particulates", ylab="")
```



508 six-day smoothed averages of daily values

Pollution, Temperature and Mortality (pairs plot)

```
pairs(cbind(Mortality=cmort, Temperature=tempr, Particulates=part))
```



Pollution, Temperature and Mortality (model fitting)

- (P)ollution, (T)emperature, (M)ortality
- Several models:

$$M_t = \beta_0 + \beta_1 t + w_t$$
$$M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - \bar{T})$$
$$M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - \bar{T}) + \beta_3 (T_t - \bar{T})^2 + w_t$$
$$M_t = \beta_0 + \beta_1 t + \beta_2 (T_t - \bar{T}) + \beta_3 (T_t - \bar{T})^2 + \beta_4 P_t + w_t$$

Pollution, Temperature and Mortality (reg.stats)

```
reg.stats = function(fit){
    SSE = sum(fit$residuals^2);
    df  = fit$df.residual
    response = fit$fitted.values+fit$residuals
    R.sq = 1 - sum(fit$residuals^2)/sum((response-mean(response))^2);
             model = paste(fit$call)[2]
      MSE = SSE/df
    n = length(fit$residuals)
    k = n - df
    AIC = log(MSE) + (n+2*k)/n
    BIC = log(MSE) + k*log(n)/n
    result = data.frame(k=k,SSE=SSE, df=df, MSE=MSE,
                            R.sq=R.sq, AIC=AIC, BIC=BIC)
    result
}
```

Pollution, Temperature and Mortality (output)

```r
print(rbind(reg.stats(fit.2),
            reg.stats(fit.3),
            reg.stats(fit.4),
            reg.stats(fit.5)))
```

```
  k      SSE  df      MSE      R.sq      AIC      BIC
1 2 40019.84 506 79.09059 0.2104470 5.378468 4.395123
2 3 31413.21 505 62.20438 0.3802474 5.142236 4.167220
3 4 27984.53 504 55.52486 0.4478921 5.032579 4.065890
4 5 20508.44 503 40.77225 0.5953881 4.727687 3.769325
```

Best Model

```
fit.5
```

```
Call:
lm(formula = cmort ~ trend + temp + temp2 + part)

Coefficients:
(Intercept)        trend         temp        temp2         part
 2831.49025     -1.39590     -0.47247      0.02259      0.25535
```

- Negative trend in time.
- Negative trend for adjusted temperature.
- Quadratic effect can be seen from the pairs plot
- Pollution weights positively. Interpretation: "incremental contribution to daily deaths per unit of particulate pollution".

Sinusoidal Regression

Sinusoid in Noise

$$x_t = A\cos\left(2\pi\omega t + \phi\right) + w_t$$

- Linearization trick

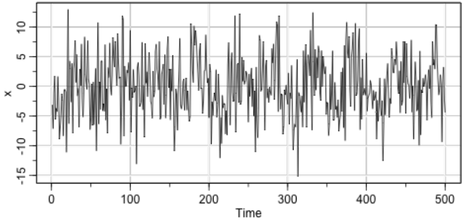
$$\beta_1 \cos(2\pi\omega t) + \beta_2 \sin(2\pi\omega t) = A\cos\left(2\pi\omega t + \phi\right)$$

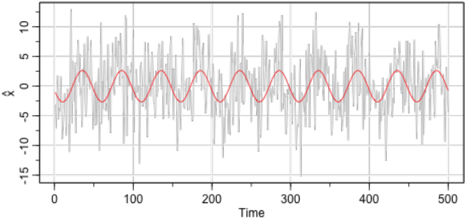
- Fit using cos and sin:

$$x_t = \beta_1 \cos\left(2\pi\omega t\right) + \beta_2 \sin\left(2\pi\omega t\right) + w_t.$$

Example 2.10: Signal Hidden in Noise

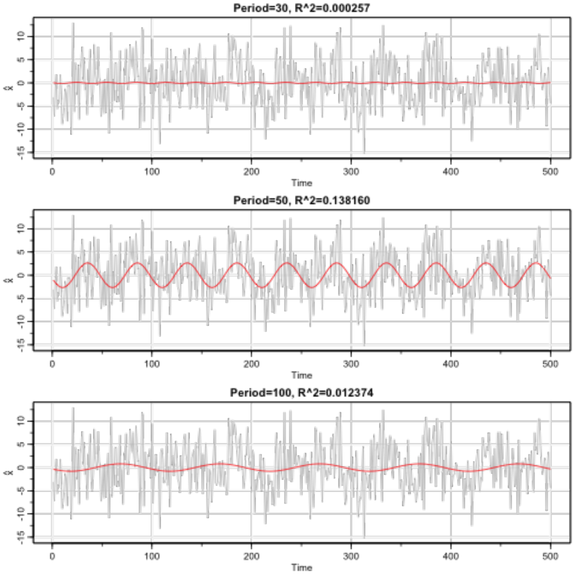
```
set.seed(1000)  # so you can reproduce these results
x = 2*cos(2*pi*1:500/50 + .6*pi) + rnorm(500,0,5)
z1 = cos(2*pi*1:500/50)
z2 = sin(2*pi*1:500/50)
summary(fit <- lm(x~0+z1+z2))  # zero to exclude the intercept
par(mfrow=c(2,1)); tsplot(x); tsplot(x, col=8, ylab=expression(hat(x)))
lines(fitted(fit), col=2)
```





How to Determine Periodicity?

- By Trial and Error:



How to Determine Periodicity?

- OLS regression coefficients

$$\hat{\beta}_1(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t \cos(2\pi jt/n), \qquad \hat{\beta}_2(j/n) = \frac{2}{n}\sum_{t=1}^{n} x_t \sin(2\pi jt/n)$$

- Measure of power in fitted model at frequency $\omega = 2\pi j/n$:

$$P(j/n) \equiv \hat{\beta}_1^2(j/n) + \hat{\beta}_2^2(j/n)$$

- R^2 at frequency j:

$$R^2 = \frac{P(j/n)}{\sum_{i=1}^{n} P(j/n)}$$

Periodogram (Section 4.3)

- Discrete Fourier Transform (aka Fast Fourier Transform):

$$d(j/n) \equiv \frac{1}{\sqrt{n}} \sum_{t=1}^{n} x_t e^{-2\pi\sqrt{-1}tj/n}, \quad j = 0, \ldots, n-1.$$

- Commutable in O(n log(n)) flops (when $n = 2^k$)
- Definition: *Periodogram*

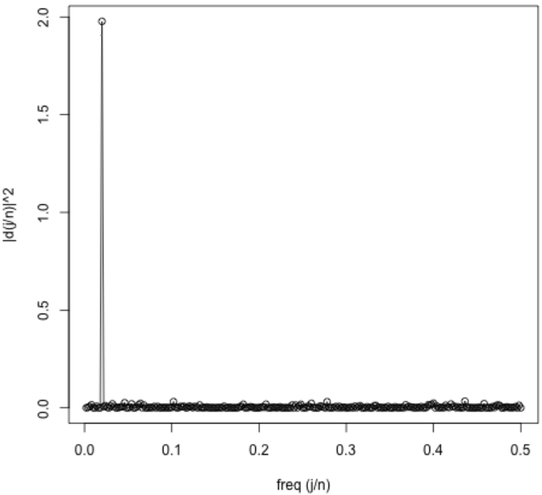
$$I(j/n) \equiv |d(j/n)|^2 .$$

- Measure of power in fitted model at frequency $\omega = 2\pi j/n$:

$$P(j/n) = \frac{4}{n} I(j/n).$$

Periodogram

```
x = 2*cos(2*pi*1:500/50 + .6*pi) + rnorm(500,0,5)
s = spectrum(x, plot=FALSE)
plot(s$freq, abs(s$spec/500)^2, ylab="|d(j/n)|^2", type="ol",
             xlab='freq (j/n)')
```



Fisher's Periodogram Test (Fisher's G-test)

- Test

$$H_0 : \text{white noise}$$
$$H_1 : \text{signal plus noise}$$

- Statistic:

$$F = \max_j \frac{I(j/n)}{\sum_{i=1}^n I(i/n)}.$$

- Reject H_0 for large values of F:

$$F \geq c_\alpha, \quad c_\alpha \approx K(\alpha)\log(n)/n.$$

```
set.seed(1000)
w = rnorm(500,0,5)
print(fisher.g.test(w))
print(fisher.g.test(2*cos(2*pi*1:500/50 + .6*pi) + w))
```

```
[1] 0.1486071
[1] 2.310001e-14
```