# MATH 8090: Stationary processes and Model-Free Estimation of Stationary Means and Covariances

Whitney Huang, Clemson University

9/9-9/11/2025

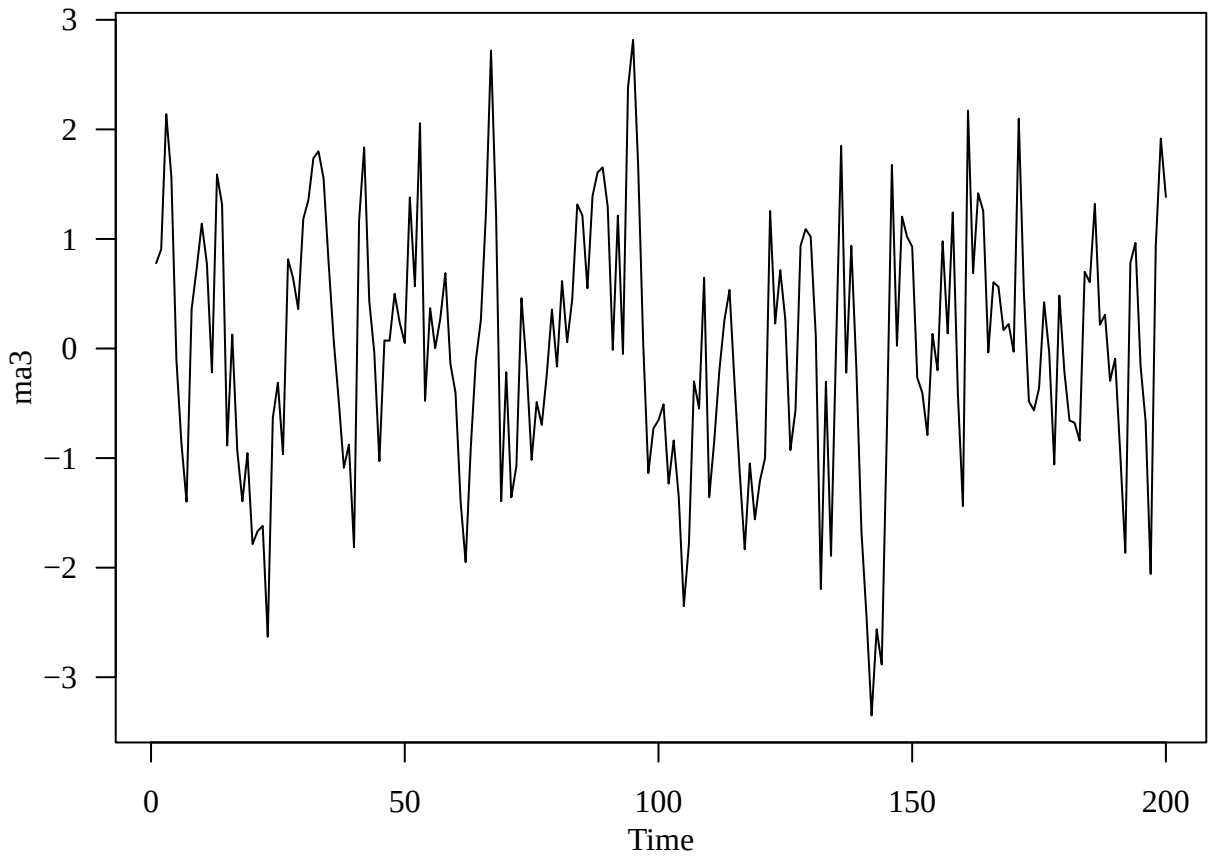
## Contents



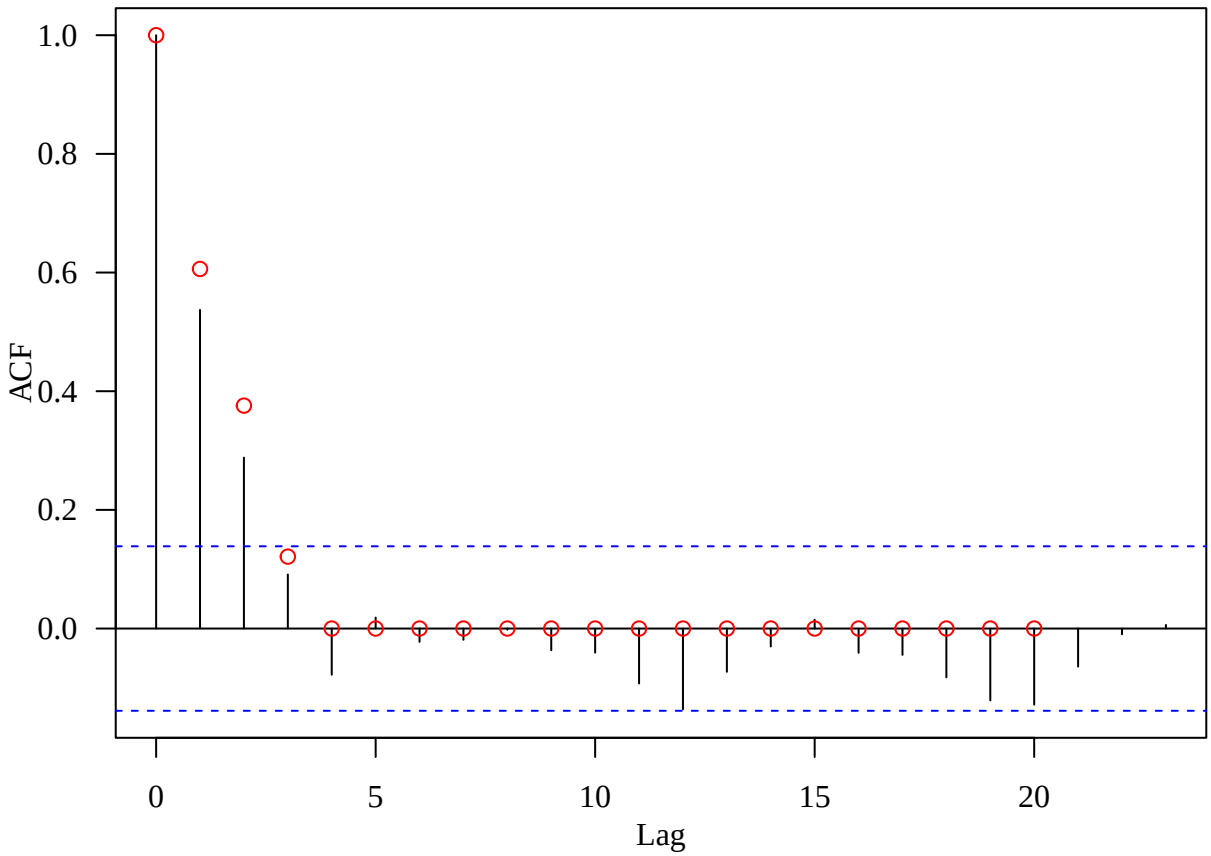
## Stationary Processes

### Examples of MA(q) processes

```
set.seed(123)
ma3 <- arima.sim(n = 200, list(ma = c(0.6, 0.5, 0.2)))
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
ts.plot(ma3)
```

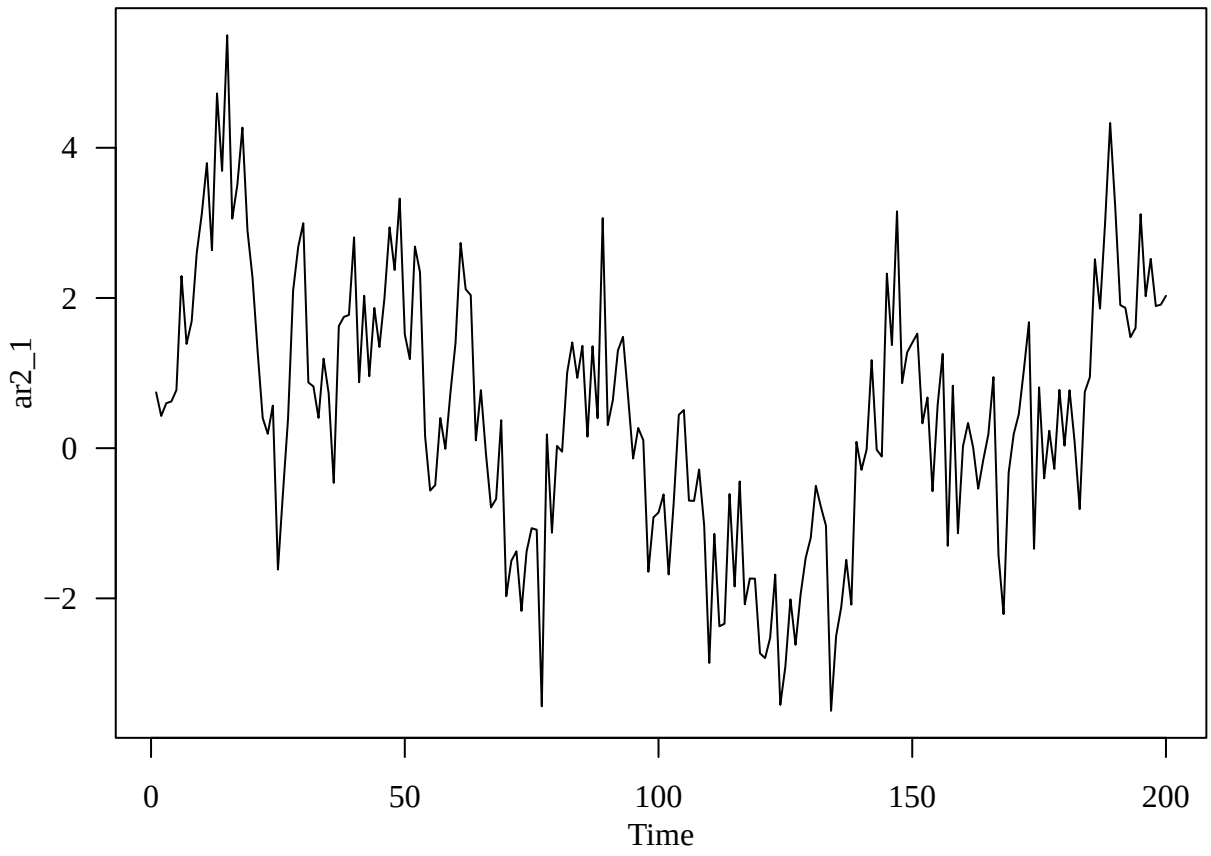



```r
acf(ma3)
acf_true <- ARMAacf(ma = c(0.6, 0.5, 0.2), lag.max = 20)
points(0:20, acf_true, col = "red")
```

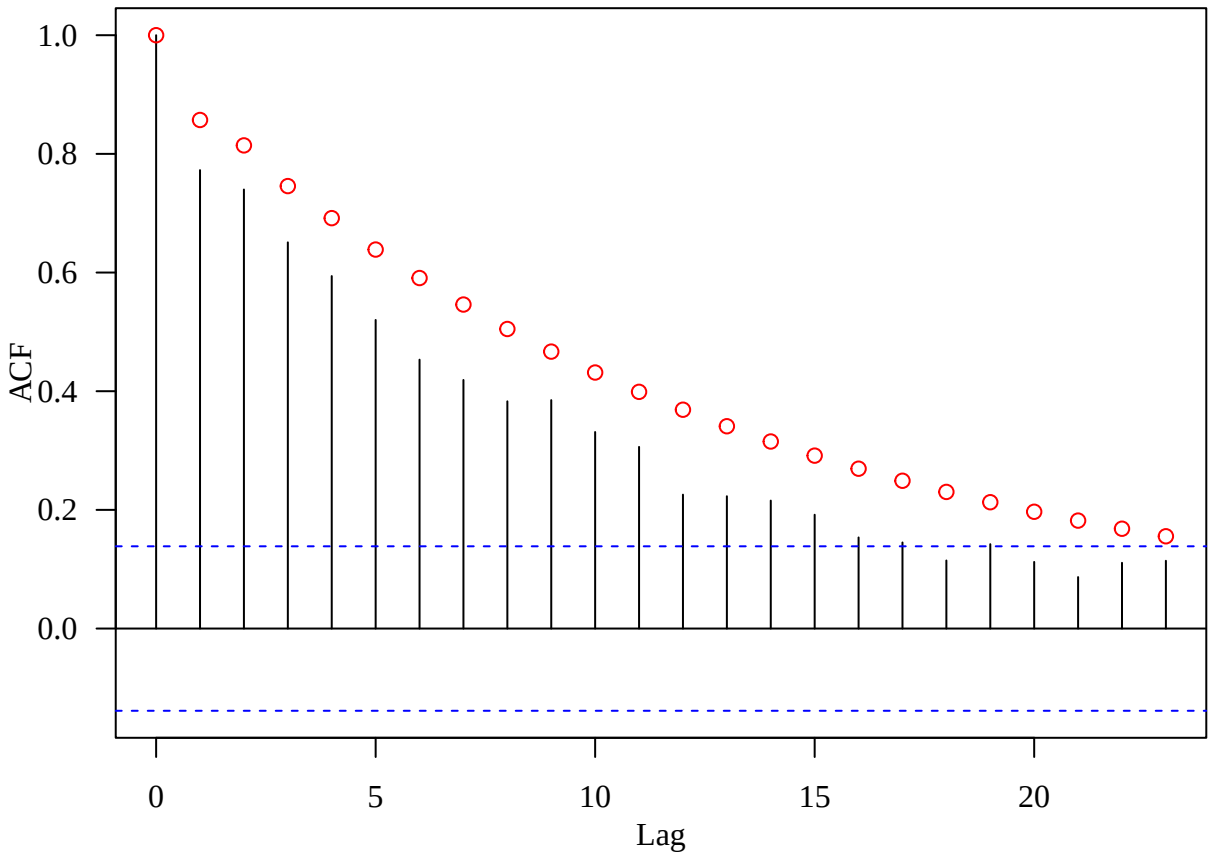



**Examples of AR(2) processes**

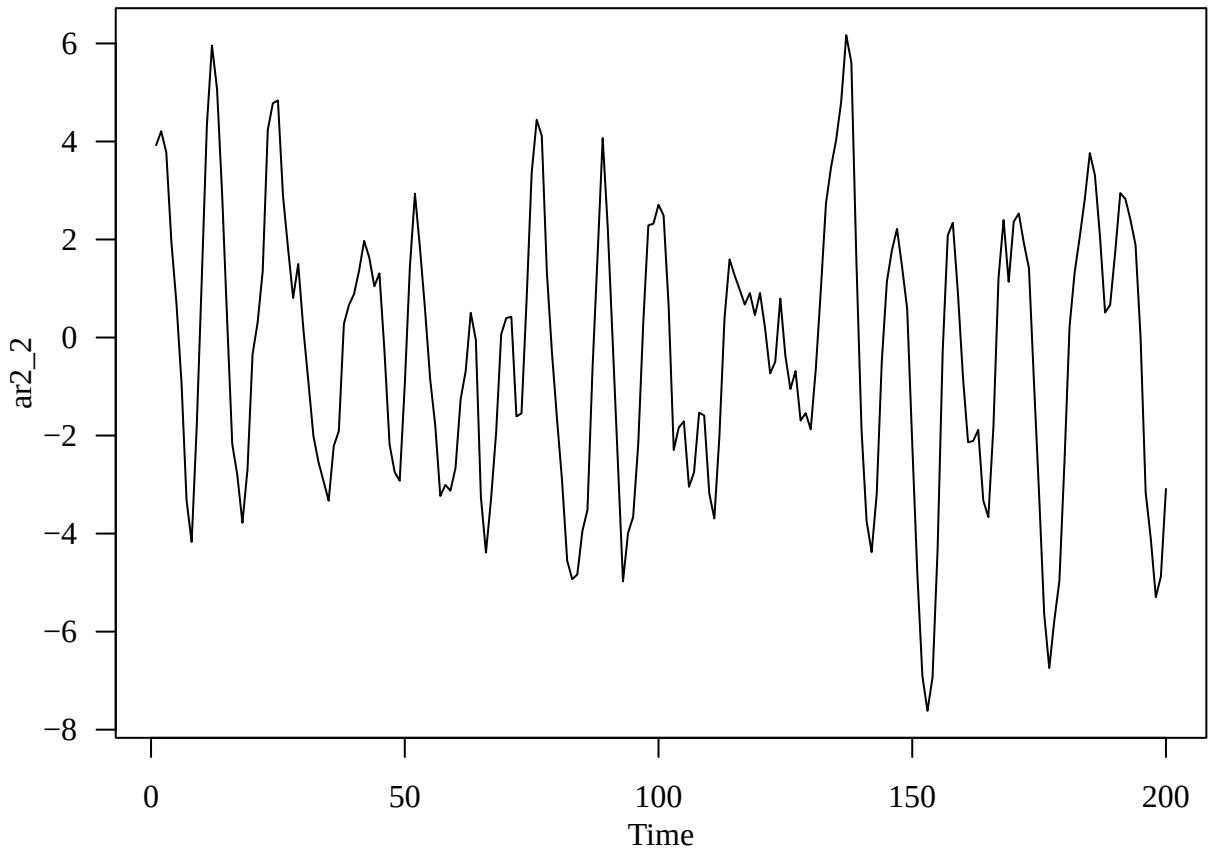
```
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
ar2_1 <- arima.sim(n = 200, list(ar = c(0.6, 0.3)))
ts.plot(ar2_1)
```



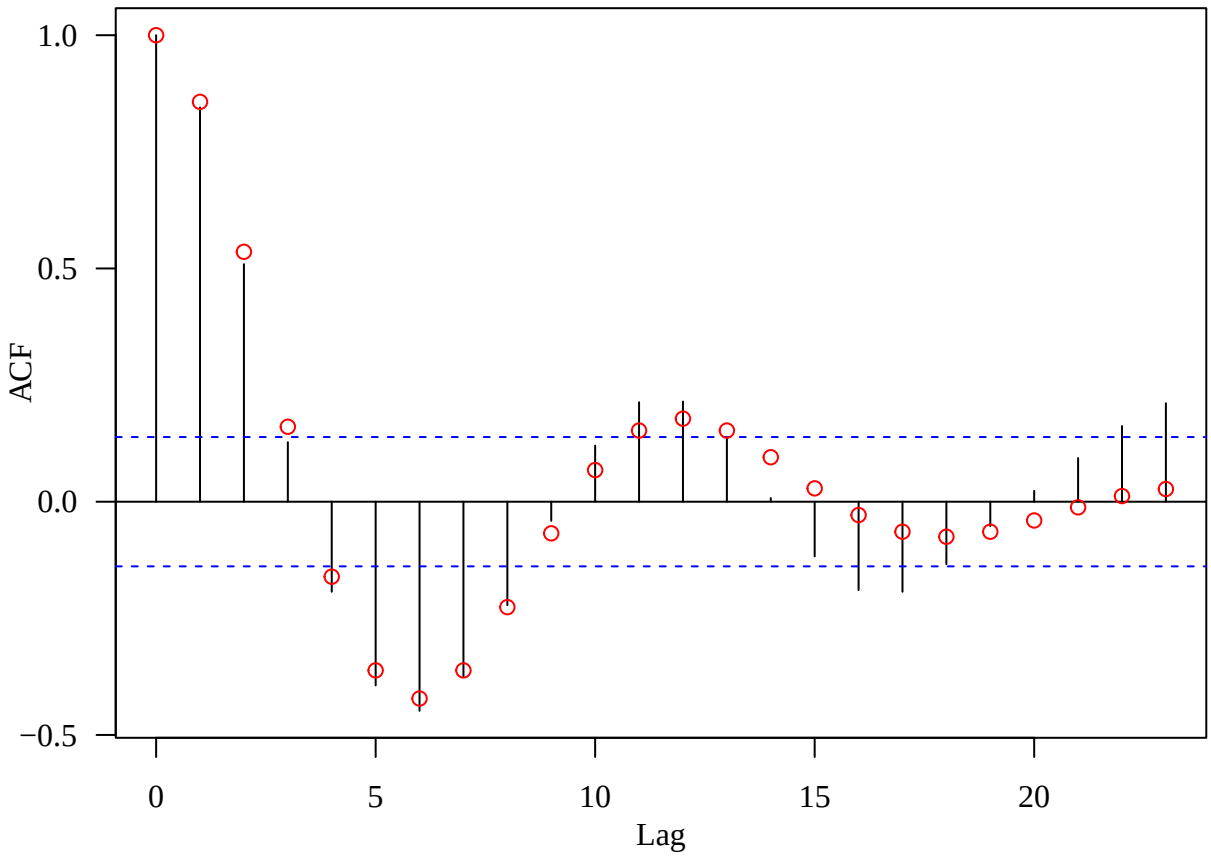

```
acf(ar2_1)
points(0:23, ARMAacf(ar = c(0.6, 0.3), lag.max = 23), col = "red")
```




```
ar2_2 <- arima.sim(n = 200, list(ar = c(1.5, -0.75)))
ts.plot(ar2_2)
```




```
acf(ar2_2)
points(0:23, ARMAacf(ar = c(1.5, -0.75), lag.max = 23), col = "red")
```




## Model-Free Estimation of Stationary Means and Covariances

### Mean Estimation

Given a stationary process $\{\eta_t\}_{t=1}^T$, the point estimator is $\bar{\eta} = \frac{1}{T}\sum_{t=1}^T \eta_t$. The varaince of this estimator is

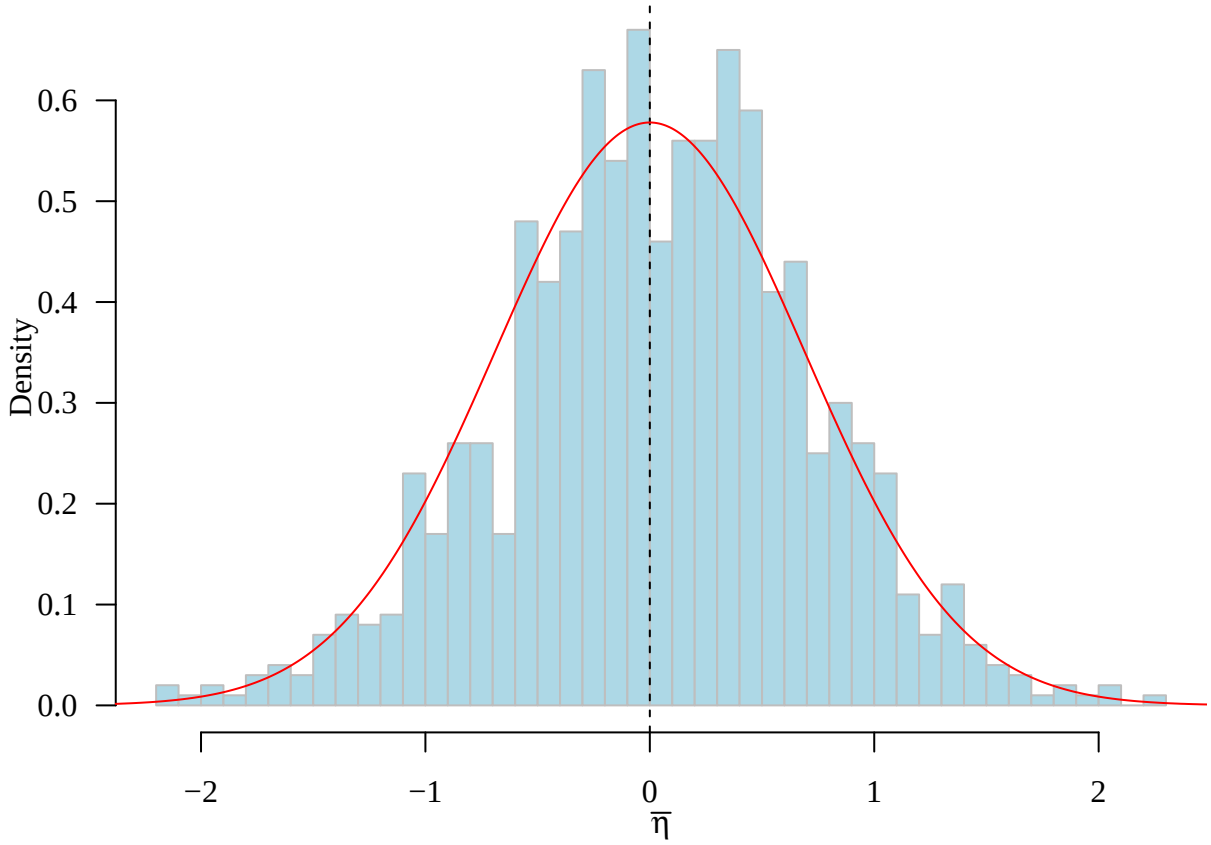
$$\nu_T = \text{Var}(\bar{\eta}) = \text{Var}\left(\frac{1}{T}\sum_{t=1}^T \eta_t\right) = \frac{1}{T}\sum_{h=-(T-1)}^{T-1}\left(1 - \frac{|h|}{T}\right)\gamma(h)$$

```
# Monte Carlo approximation
M = 1000; T = 200; phi = 0.9
set.seed(123)
sim <- replicate(M, arima.sim(n = T, list(ar = c(phi))))
eta_bar <- apply(sim, 2, mean)
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
hist(eta_bar, 40, col = "lightblue", border = "gray", las = 1, prob = T,
     main = "", xlab = expression(bar(eta)))
(nu_T_hat <- var(eta_bar))
```

```
## [1] 0.4659671
```

```
# Theoretical sampling dist
mu = 0
h <- -(T-1):(T-1)
```

```
nu_T <- (1 / T) * sum((1 - (abs(h) / T)) * (phi^(abs(h)) / (1 - phi^2)))
## Superimpose the true density curve
xg <- seq(-2.5, 2.5, 0.01)
abline(v = mu, lty = 2)
lines(xg, dnorm(xg, sd = sqrt(nu_T)), col = "red")
```




```
## Compare nu_T and nu
(nu_T <- (1 / T) * sum((1 - (abs(h) / T)) * (phi^(abs(h)) / (1 - phi^2))))
```

```
## [1] 0.4763158
```

```
(nu <- (1 / T) * (1 / (1 - phi)^2))
```

```
## [1] 0.5
```

### Autocovariance Function (ACVF)

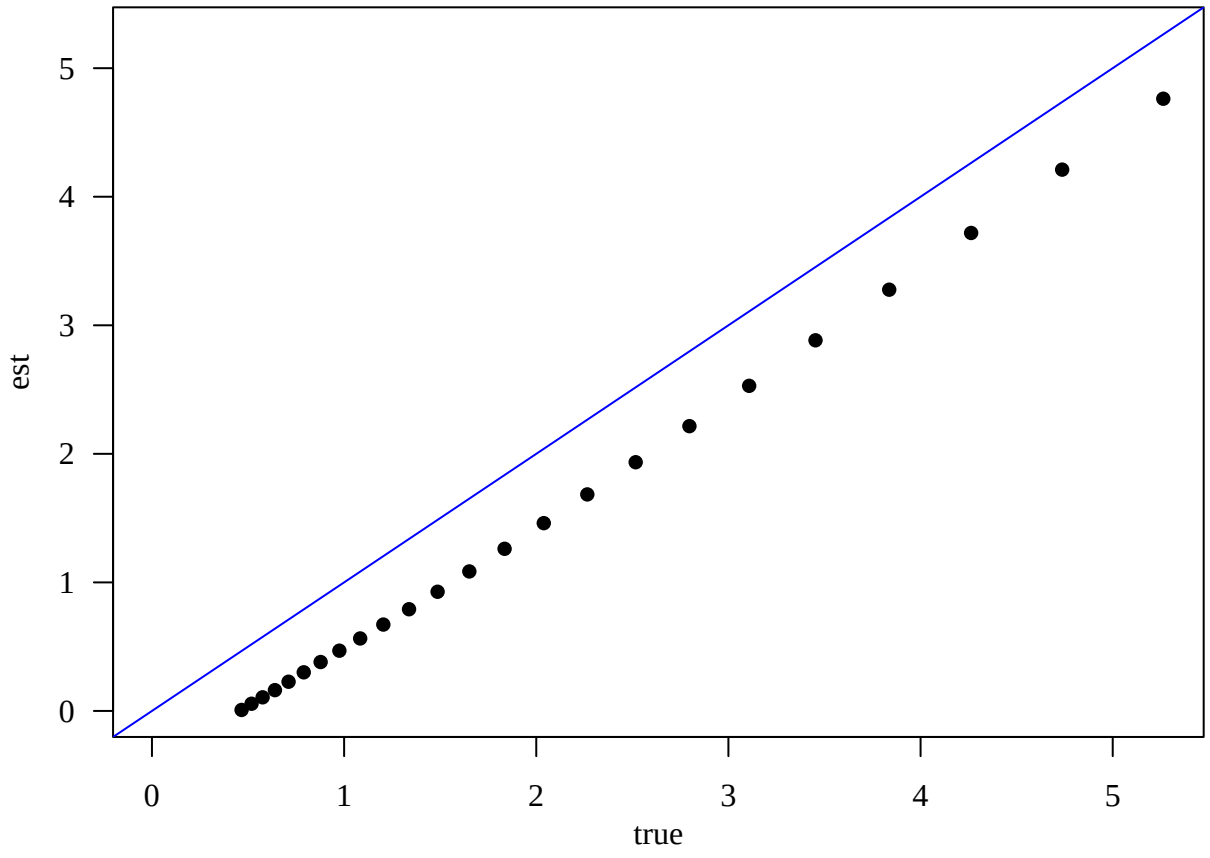
```
gamma <- apply(sim, 2, function(x) acf(x, plot = F, type = "covariance")$acf)
(est <- apply(gamma, 1, mean))
```

```
##  [1] 4.762396047 4.210822098 3.718189717 3.276955489 2.883257698 2.529033768
##  [7] 2.215391725 1.934657288 1.684410252 1.460895692 1.261443145 1.085257775
## [13] 0.927192703 0.791189571 0.671906177 0.563884323 0.468728429 0.380609541
## [19] 0.300366800 0.227258194 0.161799558 0.106047046 0.056009807 0.007812902
```

```r
(true <- ARMAacf(ar = c(0.9), lag.max = 23) * (1 / (1 - phi^2)))
```

```
##         0         1         2         3         4         5         6         7
## 5.2631579 4.7368421 4.2631579 3.8368421 3.4531579 3.1078421 2.7970579 2.5173521
##         8         9        10        11        12        13        14        15
## 2.2656169 2.0390552 1.8351497 1.6516347 1.4864712 1.3378241 1.2040417 1.0836375
##        16        17        18        19        20        21        22        23
## 0.9752738 0.8777464 0.7899718 0.7109746 0.6398771 0.5758894 0.5183005 0.4664704
```

```r
# The sample ACVF is a biased estimator
rg <- range(true, est)
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
plot(true, est, pch = 16, xlim = rg, ylim = rg, las = 1)
abline(0, 1, col = "blue")
```




```r
# ACVF is non-negative definite
h <- outer(1:24, 1:24, "-")
Sigma <- phi^abs(h) / (1 - phi^2)
eigen(Sigma)$values
```

```
##  [1] 64.7315984 26.5011371 11.7018213  6.1924112  3.7670568  2.5262589
##  [7]  1.8165664  1.3759195  1.0849122  0.8833909  0.7385536  0.6313427
## [13]  0.5500994  0.4873821  0.4382698  0.3994115  0.3684691  0.3437787
## [19]  0.3241390  0.3086742  0.2967447  0.2878873  0.2817747  0.2781898
```

**Autocorrelation Function (ACF)**

*Population ACF*

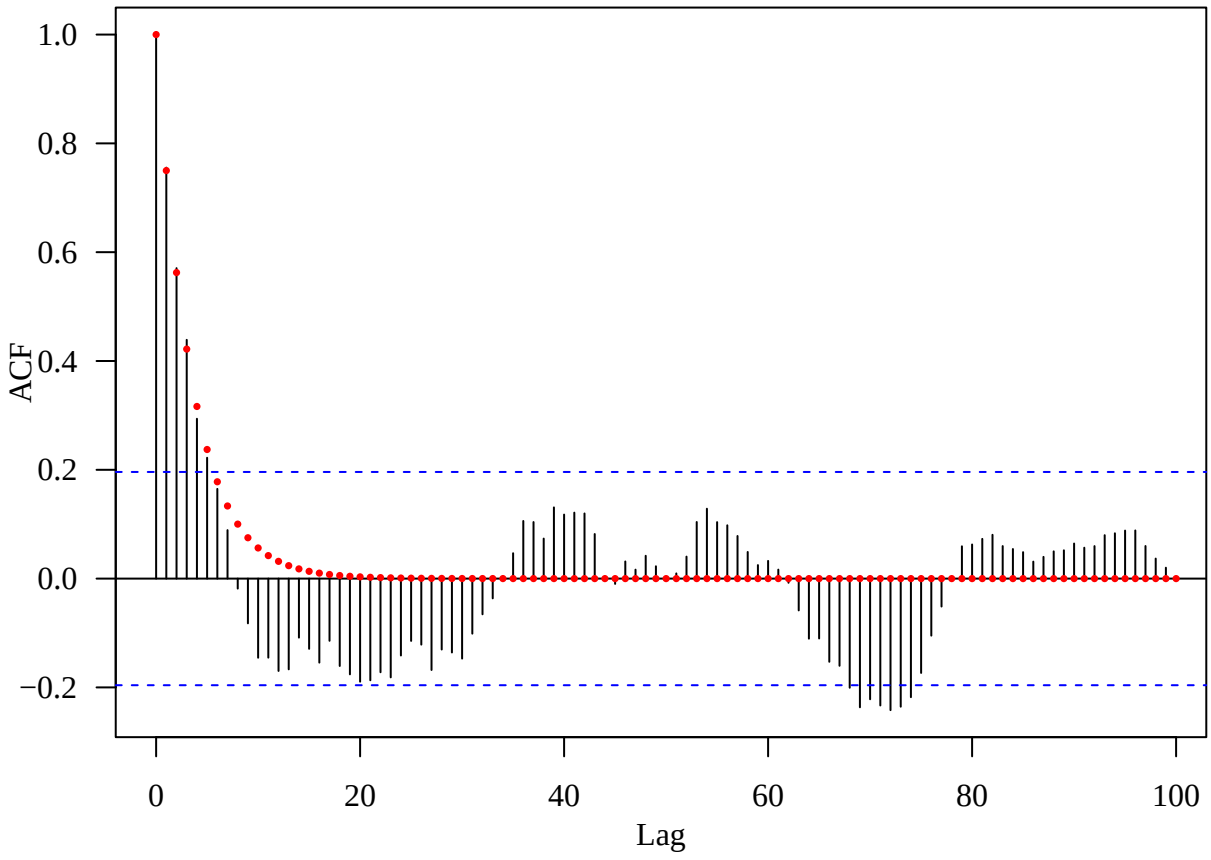
$$\rho(h) = \text{Cor}(\eta_t, \eta_{t+h}) = \frac{\text{E}\left[(\eta_t - \mu)(\eta_{t+h} - \mu)\right]}{\sqrt{\text{Var}(\eta_t)\text{Var}(\eta_{t+h})}}$$

*Sample ACF*

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma(0)}},$$

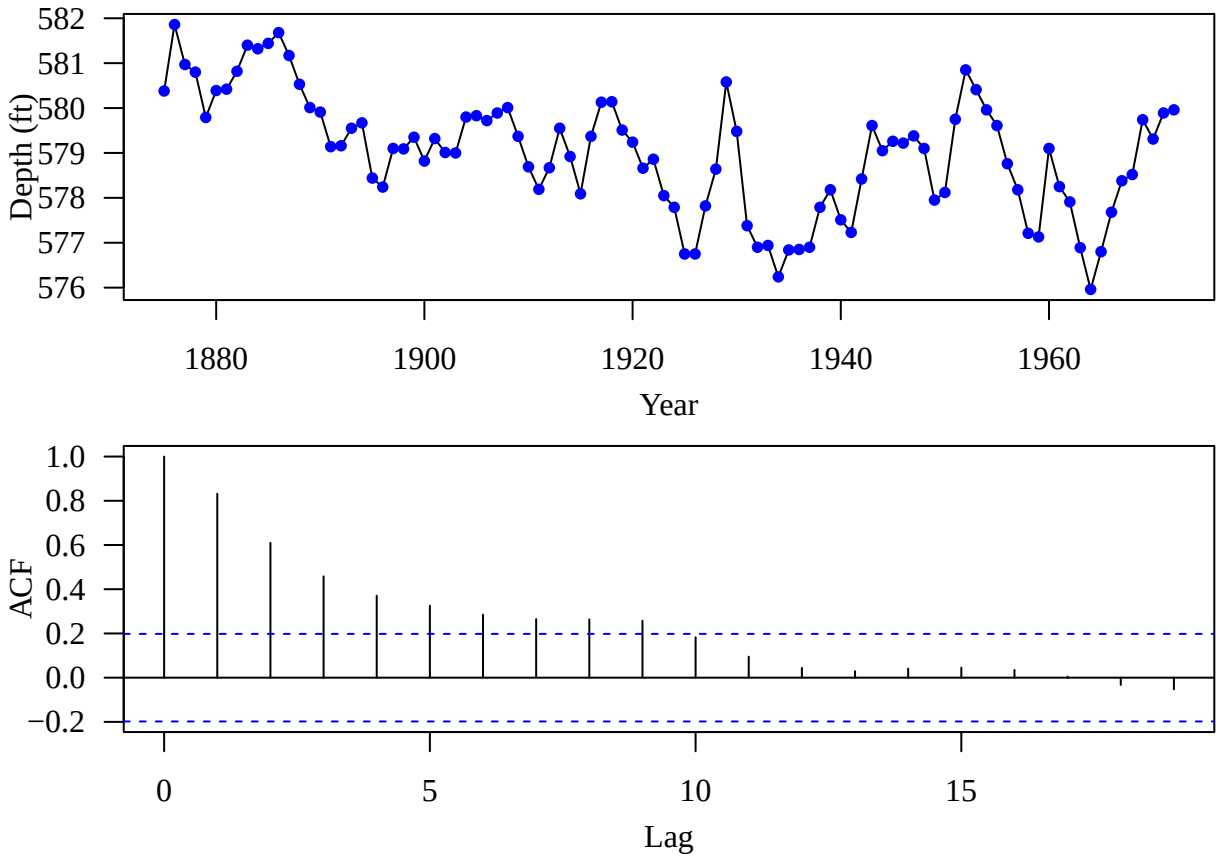
where $\hat{\gamma}(h) = \frac{1}{T}\sum_{t=1}^{T-|h|}(\eta_t - \bar{\eta})(\eta_{t+h} - \bar{\eta})$.

```
set.seed(123)
AR1 <- arima.sim(n = 100, list(ar = c(0.75)))
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
acf(AR1, lag = 100)
acf_true <- ARMAacf(ar = c(0.75), lag.max = 100)
points(0:100, acf_true, pch = 16, cex = 0.5, col = "red")
```

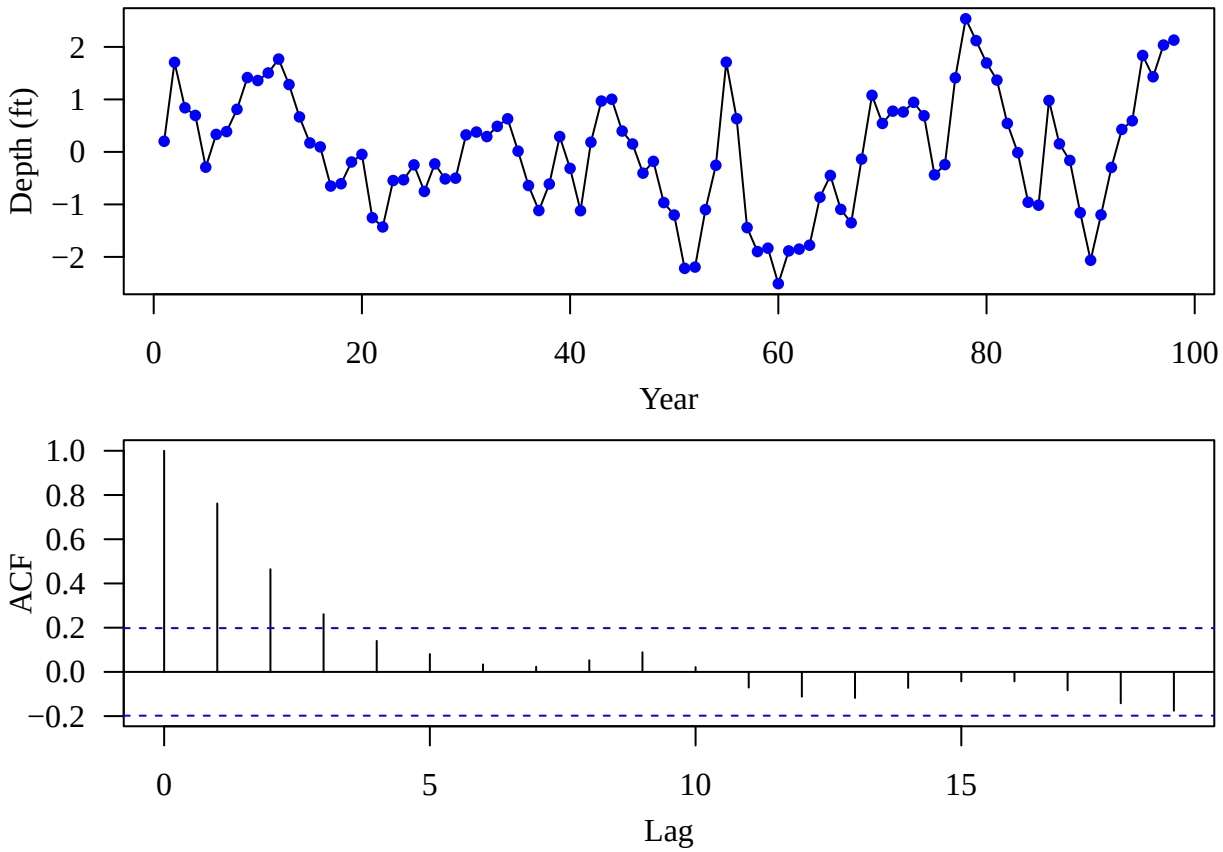



**Lake Huron Example**

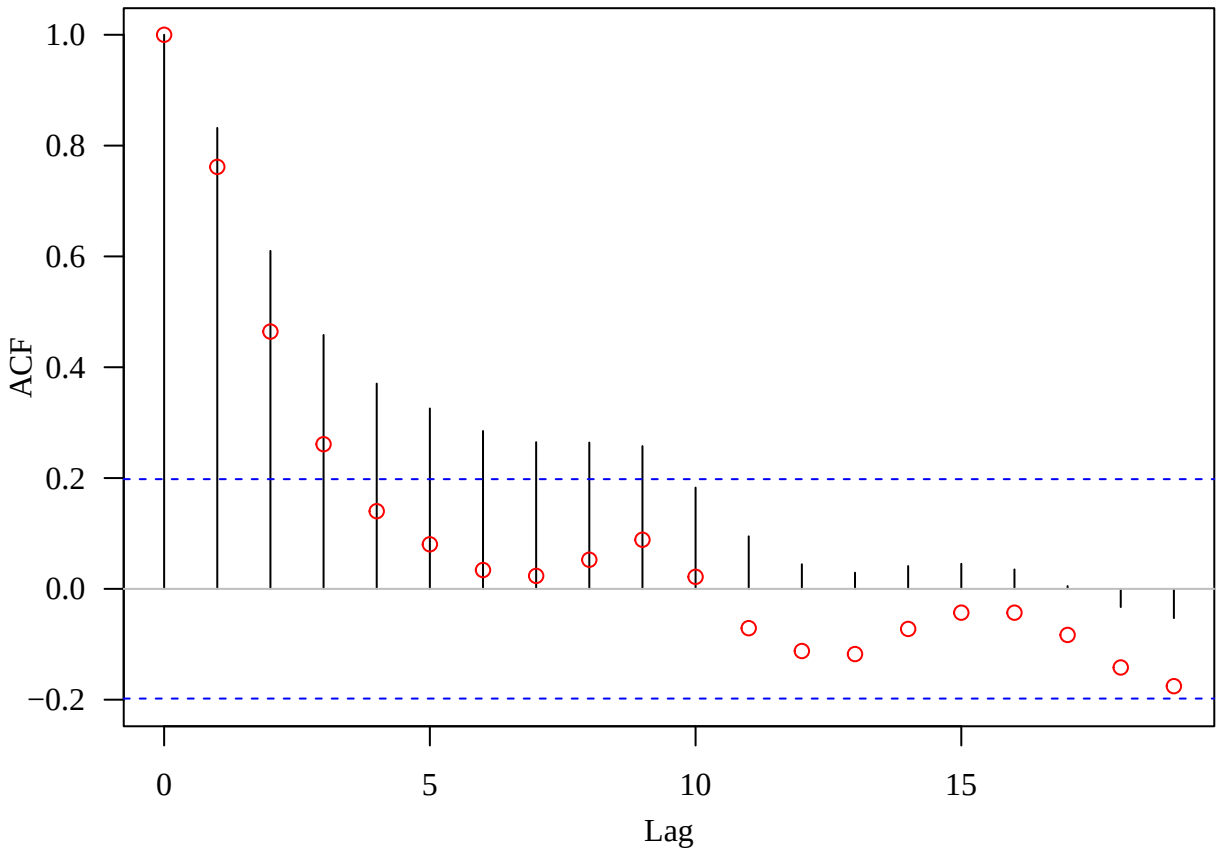
```r
data(LakeHuron)
par(mar = c(3.3, 3.5, 0.5, 0.6), mgp = c(2.2, 1, 0), mfrow = c(2, 1),
    las = 1, family = "serif")
plot(LakeHuron, ylab = "Depth (ft)", xlab = "Year")
points(LakeHuron, cex = 0.8, col = "blue", pch = 16)
acf(LakeHuron)
```




```r
# Let's remove the (linear trend)
yr <- 1875:1972
lm <- lm(LakeHuron ~ yr)
par(mar = c(3.3, 3.5, 0.5, 0.6), mgp = c(2.2, 1, 0), mfrow = c(2, 1),
    las = 1, family = "serif")
plot(lm$residuals, ylab = "Depth (ft)", xlab = "Year", type = "l")
points(lm$residuals, cex = 0.8, col = "blue", pch = 16)
acf(lm$residuals)
```




```
par(mar = c(3.3, 3.5, 0.5, 0.6), mgp = c(2.2, 1, 0), mfrow = c(1, 1),
    las = 1, family = "serif")
plot(0:19, acf(LakeHuron, plot = F)$acf, type = "h", xlab = "Lag", ylab = "ACF", ylim = c(-0.2, 1))
abline(h = 0, col = "gray")
abline(h = c(-1, 1) * qnorm(0.975) / sqrt(length(LakeHuron)) , col = "blue", lty = 2)
acf_detrend <- acf(lm$residuals, plot = F)$acf
points(0:19, acf_detrend, col = "red")
```




## Box test for temporal independence

### Box and Pierce test Box and Pierce (1970)

We wish to test:

$$H_0 : \{\eta_1, \eta_2, \cdots, \eta_T\} \text{ is an i.i.d. noise sequence}$$
$$H_1 : H_0 \text{ is false}$$

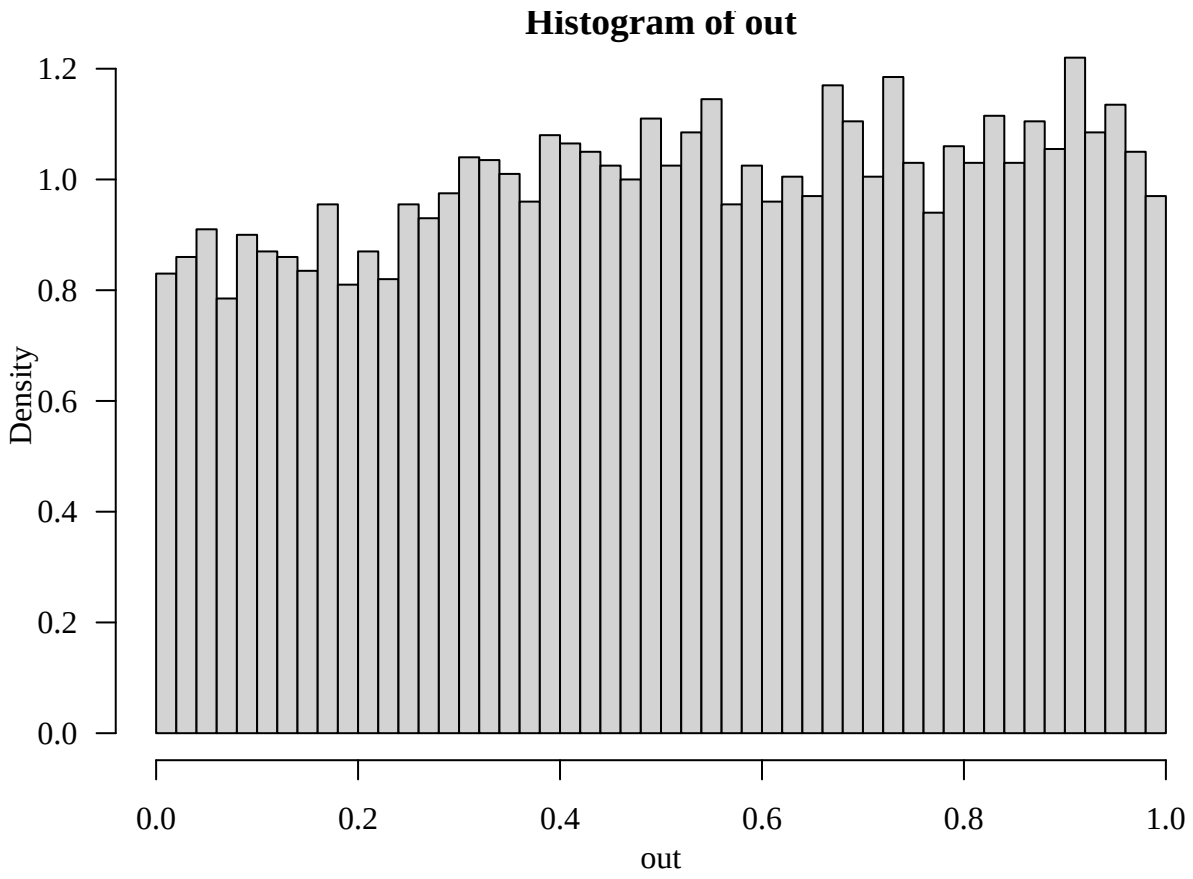
1. Under $H_0$,
$$\hat{\rho}(h) \dot{\sim} \mathrm{N}(0, \frac{1}{T}) \stackrel{d}{=} \frac{1}{\sqrt{n}}\mathrm{N}(0,1)$$
2. Hence
$$Q = T\sum_{i=1}^{k} \hat{\rho}^2(h) \dot{\sim} \chi^2_{df=k}$$
3. We reject $H_0$ if $Q > \chi^2_k(1-\alpha)$, the $1-\alpha$ quatile of the chi-squared distribution with $k$ degrees of freedom

### Ljung-Box Test

$$Q_{LB} = T(T-2)\sum_{h=1}^{k} \frac{\hat{\rho}^2(h)}{n-h} \dot{\sim} \chi^2_k.$$

The Ljung-Box test Ljung and Box (1978) can be more powerful than the Box and Pierce test

```r
out <- numeric(10000)
for (i in 1:10000) out[i] <- Box.test(rnorm(100), 5)$p.value
par(mar = c(3, 3.5, 0.5, 0.6), mgp = c(2, 1, 0), las = 1, family = "serif")
hist(out, 50, prob = T)
```




```r
Box.test(lm$residuals, 10)
```

```
##
##  Box-Pierce test
##
## data:  lm$residuals
## X-squared = 88.469, df = 10, p-value = 1.077e-14
```

```r
Box.test(lm$residuals, 10, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  lm$residuals
## X-squared = 91.776, df = 10, p-value = 2.331e-15
```

```r
library(WeightedPortTest)
Weighted.Box.test(lm$residuals, lag = 10)
```

```
##
##  Weighted Box-Pierce test (Gamma Approximation)
##
## data:  lm$residuals
## Weighted X-squared on Residuals for fitted ARMA process = 83.252, Shape
## = 3.9286, Scale = 1.4000, p-value < 2.2e-16
```

## References


Box, George EP, and David A Pierce. 1970. "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models." *Journal of the American Statistical Association* 65 (332): 1509–26.

Ljung, Greta M, and George EP Box. 1978. "On a Measure of Lack of Fit in Time Series Models." *Biometrika* 65 (2): 297–303.