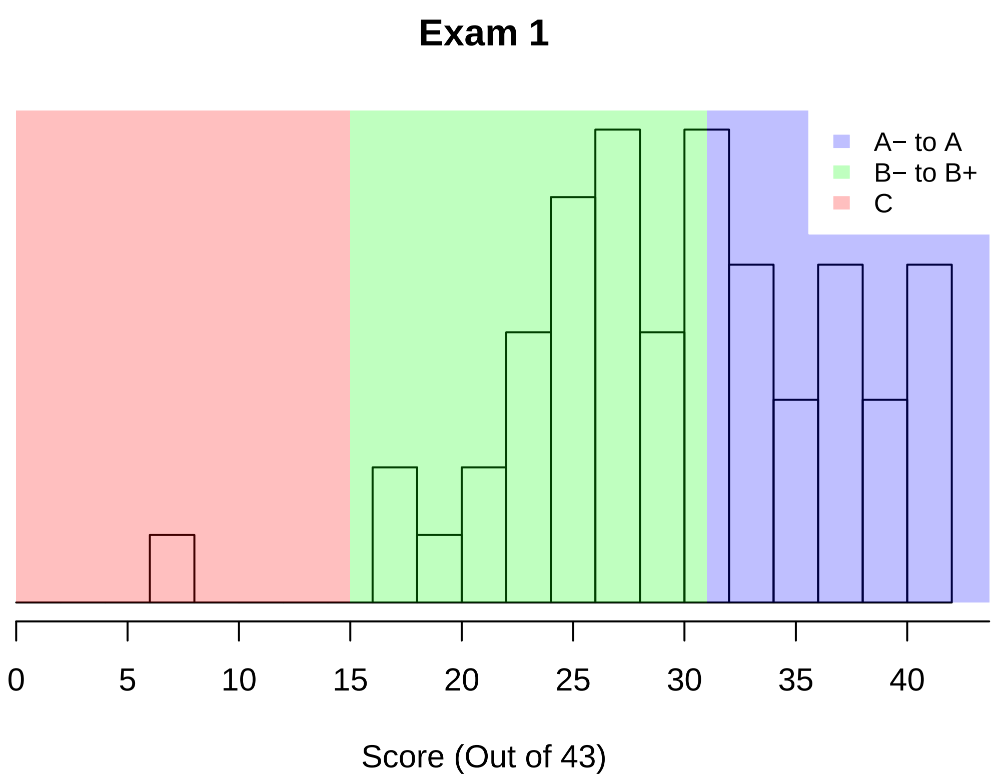
Exam 1

2/22/19

There are six questions, each of which has several parts, as well as a bonus question. Neither the questions nor the parts are necessarily in order from easiest to most difficult. Make sure you have taken a look at and attempted all of the questions in the allotted time. Stop working and immediately turn in your exam when time has been called.

Name:	______________________
Question	Maximum Possible Points
1	4
2	8
3	11
4	5
5	8
6	6
Bonus	1
Total	43

The figure below shows a histogram of scores. The mean was 30.36, the median was 31, and the standard deviation was 7.24.



1. Causality

Consider the general **AR**(1) model, $x_t = \phi_1 x_{t-1} + w_t$, where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$.

(a) Write $\rho_x(h)$ as a function of ϕ_1.

$\rho_x(h) = \phi_1^h$

(b) Describe how the sign of the following depends on ϕ_1 when $\phi_1 \geq 0$:

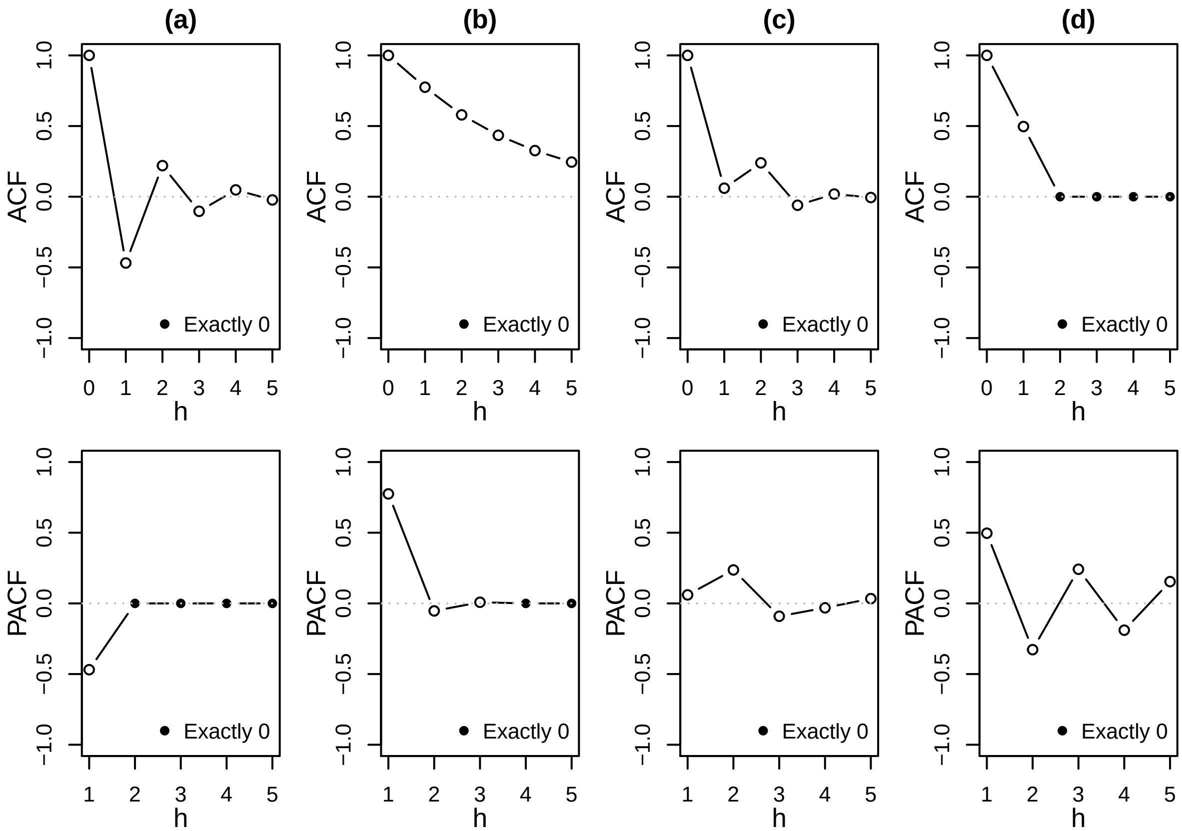
$$\rho_x(h)\log(\phi_1).$$

The function $\rho_x(h)\log(\phi_1) = \phi_1^h \log(\phi_1)$ is negative when $\phi_1 < 1$ and positive, exactly equal to 0 when $\phi_1 = 1$, and negative when $\phi_1 > 1$.

(e) The quantity derived in (b) is the derivative of $\rho_x(h)$ with respect to h when $\phi_1 > 0$. Fill in the blank with "is increasing," "is decreasing," or "could be increasing or decreasing depending on the value of ϕ_1." When an **AR**(1) model is causal with $\phi_1 > 0$, this guarantees that the **autocorrelation** function $\rho_x(h)$ is decreasing as h increases.

2. Model Order

Four pairs of true autocorrelation and partial correlation plots are given below for different **ARMA**(p, q) models with $p \leq 5$ and $q \leq 5$. They are labeled (a), (b), (c), and (d). For each pair, write “AR,” “MA” or “ARMA,” and the order, if it is possible to determine the order from the pair of plots.



(a) AR(1)
(b) AR(3)
(c) ARMA
(d) MA(1)

3. Testing

Consider the $\mathbf{AR}(1)$ model,

$$x_t = \phi_1 x_{t-1} + w_t,$$

where x_t is stationary and $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_w^2\right)$.

(a) Derive $\gamma_x(0)$ as a function of ϕ_1 and σ_w^2.

$\gamma_x(0) = \frac{\sigma_w^2}{1-\phi_1^2}$

(b) What happens to $\gamma_x(0)$ when $\phi_1 = 1$?

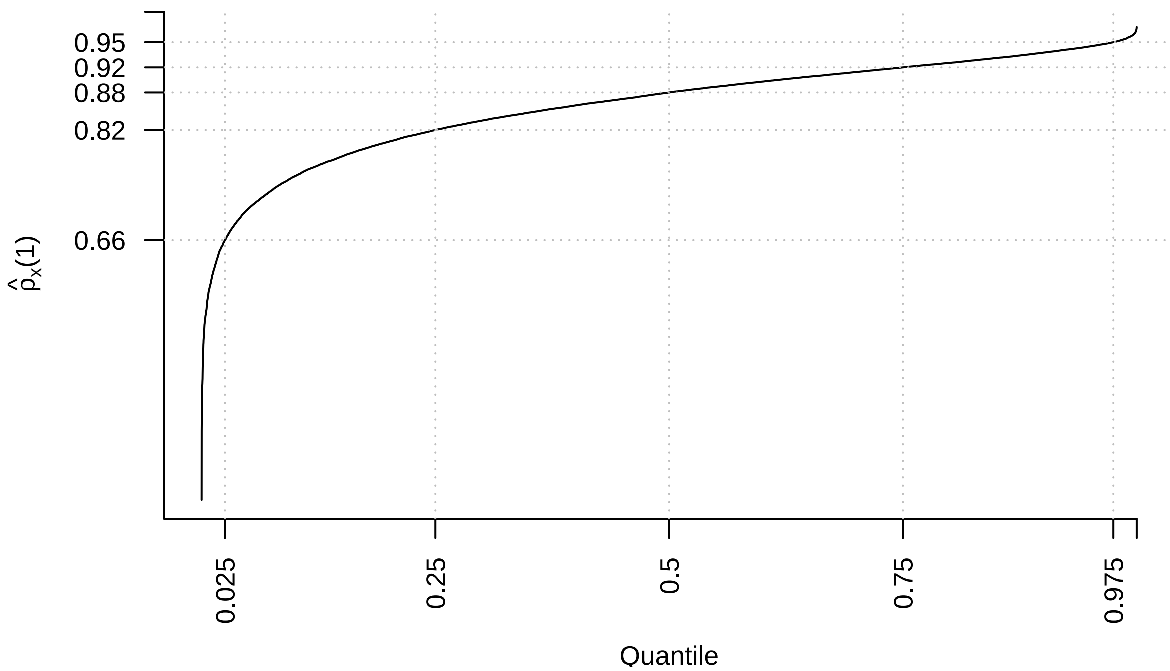
I accepted three possible solutions to this question - either "the process is not stationary" or the variance is infinite or undefined.

(c) Derive $\rho_x(1)$.

$\rho_x(1) = \phi_1$

(d) Suppose we observe a length-n time series $\boldsymbol{x}$ with $n = 50$. Suppose we wanted to test the null hypothesis that $\phi_1 = 1$, given an observed time series $\boldsymbol{x}$. This is called a **unit root test**. What sample quantity could we compute, that would give us information about ϕ_1?

Based on (c), we can compute $\hat{\rho}_x(1)$ because $\rho_x(1)$ is equal to ϕ_1.



(e) The figure above shows the inverse CDF of the distribution of $\hat{\rho}_x(1)$ when $\boldsymbol{x}$ is a a length $n = 50$ $\mathbf{AR}(1)$ time series with $\phi_1 = 1$. What are the 2.5% and 97.5% quantiles of this distribution?

The 2.5% and 97.5% quantiles of this distribution are 0.66 and 0.95, respectively.

(f) Suppose our observed time series $\boldsymbol{x}$ has a sample lag-one autocorrelation of $\hat{\rho}_x(1) = 0.86$. Using what you found in part (e), perform a level-$\alpha = 0.05$ test of the null hypothesis that the $\boldsymbol{x}$ is distributed according to an $\mathbf{AR}(1)$ model with $\phi_1 = 1$. Do we fail to reject or reject the null hypothesis that $\boldsymbol{x}$ is distributed according to an $\mathbf{AR}(1)$ model with $\phi_1 = 1$?

$\hat{\rho}_x(1) = 0.86$ is within the 95% confidence interval from (d), so we fail to reject the null hypothesis that $\boldsymbol{x}$ is distributed according to an $\mathbf{AR}(1)$ model with $\phi_1 = 1$.

A comment - the exam version of this question used the language “accept the null hypothesis.” It has come to my attention that people are discouraged from saying that they have “accepted the null hypothesis” because the phrase “accepting the null” is often conflated with the phrase “concluding that the null is true.” I used this language because (weirdly) it’s actually the language that was used in my PhD coursework and at least one canonical text on hypothesis testing, “Testing statistical hypotheses.”. When I say “accept the null hypothesis,” I mean “conclude that we don’t have enough information to decide that the null hypothesis is false.” Although this interpretation is correct, it sounds like this phrase is often discouraged so I will try to use the “fail to reject” language going forward. Sorry for the confusion, and thank you to Linxing Yao for bringing this to my attention.

(g) Based on the results of (f), in one sentence, explain whether or not you think it’s a good idea to model the observed time series $\boldsymbol{x}$ as an $\mathbf{AR}(1)$ process.

The results of (f) indicate that we cannot reject the null hypothesis that $\boldsymbol{x}$ is not stationary if we assume $\boldsymbol{x}$ is distributed according to an $\mathbf{AR}(1)$ model with $\phi_1 = 1$, so it may not be a good idea to assume to model the observed time series $\boldsymbol{x}$ as an $\mathbf{AR}(1)$ process.

4. AR Estimation

In class, we discussed how conditional least squares estimation of the parameters of an $\mathbf{AR}(p)$ model can be viewed as a regression problem. Let's suppose we assume $\mu_x = 0$, so we don't have to worry about an intercept and are only concerned with estimating $\phi_1, \dots, \phi_p$ and σ^2_w.

(a) If we're making an analogy to regression, what is the design or covariate matrix $\boldsymbol{Z}$, i.e. how are the entries z_{ij} defined in terms of elements of our time series $\boldsymbol{x}$.

$z_{ij} = x_{p+i-j}$.

(b) If we are making an analogy to regression, how many observations are we using, i.e. how many rows does $\boldsymbol{Z}$ have?

$n - p$

(c) What is the approximate distribution of $\hat{\boldsymbol{\phi}}_{CLS}$ as $n \to \infty$?

I accepted either $\hat{\boldsymbol{\phi}}_{CLS} \stackrel{d}{\to} \mathcal{N}\left(\boldsymbol{\phi}, \sigma^2_w \left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\right)$ as $n \to \infty$ or $\sqrt{n}\left(\hat{\boldsymbol{\phi}}_{CLS} - \boldsymbol{\phi}\right) \stackrel{d}{\to} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}_{\phi\phi}\right)$ as $n \to \infty$ as solutions.

5. Forecasting

In class, we derived the **forecasting equation** for computing $\hat{x}_{n+1} = \sum_{j=1}^{n} c_{nj} x_{n+1-j}$ based on $x_1, \dots, x_n$ by minimizing

$$v_n = \mathbb{E}\left[\left(x_{n+1} - \sum_{j=1}^{n} c_{nj} x_{n+1-j}\right)^2\right].$$

(a) Write down the quantity we should minimize if we want to forecast values further into the future, i.e. if we want to compute $\hat{x}_{n+k}$ for any $k > 0$.

I accepted either of the following as correct:

$$\mathbb{E}\left[\left(x_{n+k} - \sum_{i=1}^{n} c_i x_{n+1-i}\right)^2\right]$$

$$\mathbb{E}\left[\left(x_{n+k} - \sum_{i=k}^{n+k-1} c_i x_{n+k-i}\right)^2\right]$$

(b) Write out the forecasting equation that corresponds to (a). Hints: In class we talked about minimizing v_n, and we saw that the corresponding forecasting equation was $\boldsymbol{A}_n \boldsymbol{c}_n = \boldsymbol{b}_n$ where $a_{n,ij} = \gamma_x(i-j)$ and $b_{n,i} = \gamma_x(i)$. Also, for general vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ and a general matrix $\boldsymbol{Z}$, $\boldsymbol{x}'\boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$ and $\boldsymbol{x}'\boldsymbol{Z}\boldsymbol{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j z_{ij}$.

$$\begin{aligned}
\mathbb{E}\left[\left(x_{n+k} - \sum_{i=1}^{n} c_i x_{n+1-i}\right)^2\right] &= \mathbb{E}\left[x_{n+k}^2\right] - 2\sum_{i=1}^{n} c_i \mathbb{E}\left[x_{n+k}x_{n+1-i}\right] + \mathbb{E}\left[\left(\sum_{i=1}^{n} c_i x_{n+1-i}\right)^2\right] \\
&= \mathbb{E}\left[x_{n+k}^2\right] - 2\sum_{i=1}^{n} c_i \mathbb{E}\left[x_{n+k}x_{n+1-i}\right] + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \mathbb{E}\left[x_{n+1-i}x_{n+1-j}\right] \\
&= \gamma_x(0) - 2\sum_{i=1}^{n} c_i \mathbb{E}\left[x_{n+k}x_{n+k+1-i-k}\right] + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \gamma_x(i-j) \\
&= \gamma_x(0) - 2\sum_{i=1}^{n} c_i \gamma_x(1-i-k) + \sum_{i=1}^{n}\sum_{j=1}^{n} c_i c_j \gamma_x(i-j) \\
&= \gamma_x(0) - 2\boldsymbol{c}'\boldsymbol{b} + \boldsymbol{c}'\boldsymbol{A}\boldsymbol{c},
\end{aligned}$$

where $b_i = \gamma_x(1-i-k) = \gamma_x(i+k-1)$ and $a_{ij} = \gamma_x(i-j)$. The forecasting equation is obtained by taking derivatives with respect to elements of $\boldsymbol{c}$ and setting the derivatives equal to zero:

$$\boldsymbol{A}\boldsymbol{c} = \boldsymbol{b}.$$

(c) Based on your intuition, what will happen to the expected squared forecast error $\mathbb{E}\left[\left(x_{n+k} - \hat{x}_{n+k}\right)^2\right]$ as k increases? Specifically, will it be increasing or decreasing and what will it converge to?

The expected squared forecast error will be increasing as k increases, and will converge to $\gamma_x(0)$.

6. Model Selection

For this problem, we're going to fit several different **ARMA**(p, q) models to a subset of the wave data we have seen in class and discuss how to choose one. The code below outputs the number of observations in the subset.

```
www <- "http://www.maths.adelaide.edu.au/andrew.metcalfe/Data/wave.dat"
x <- ts(data = read.table (www, header=T)[, 1])
n <- length(x)
n
```

```
## [1] 396
```

The command `arima(x, order = c(p, 0, q), method = "ML")` fits an **ARMA**(p, q) model with an intercept μ_x to the data, `x` using unconditional maximum likelihood. After each fit, the corresponding value of $\log\left(\hat{\sigma}^2_{w,UW}\right)$ is returned.

```
round(log(arima(x, order = c(5, 0, 5), method = "ML")$sigma2), 2)
```

```
## [1] 9.89
```

```
round(log(arima(x, order = c(6, 0, 4), method = "ML")$sigma2), 2)
```

```
## [1] 9.89
```

```
round(log(arima(x, order = c(4, 0, 6), method = "ML")$sigma2), 2)
```

```
## [1] 9.87
```

Because the **ARMA**(p, q) estimation problems closely resemble least squares problems, we can use all of AIC, AICc, or SIC to choose a model just like we did for a linear regression model earlier this semester.

(a) Compute AIC for each model. You do not need to do out the entire calculations or simplify fractions, just right out the equations you would plug into a calculator.

Model	AIC
1, ARMA$(5, 5)$	$9.89 + (396 + 22)/396$
2, ARMA$(6, 4)$	$9.89 + (396 + 22)/396$
3, ARMA$(4, 6)$	$9.87 + (396 + 22)/396$

(b) Which model is best according to AIC?

Model 3, the ARMA$(4, 6)$ model.

(c) Which model is best according to AICc?

Model 3, the ARMA$(4, 6)$ model.

(d) Which model is best according to SIC?

Model 3, the ARMA$(4, 6)$ model.

(e) Does it matter if you use AIC, AICc, or SIC to pick from these three models? In one sentence, why or why not?

It does not matter because the total number of parameters $p + q + 1 = 11$ is the same in all three models, and we use the same number of observations $n = 396$ to fit each model.

Bonus

In one sentence, describe the most surprising thing you've learned so far in this class. You will receive full credit for your answer as long as it is not something that is false.