Practice Exercises
ASP 2010 Summer Colloquium on Forecast Verification

The following are exercises using the R statistical language.

Part 1: IID bootstrap confidence intervals

1. Load the `verification` and `boot` packages.
2. View the help file for `roc.area`.
   (a) What package provides this function?
   (b) What is the purpose of this function?
   (c) What are the required arguments (input) for this function? Are there any optional arguments?
   (d) What type of R object is returned by this function? **Hint**: it does not say in the help file, but can you infer the type of object based on the "Value" section?
   (e) What is the value (output) produced by this function?
   (f) Run the example to set up the data frame from the Mason and Graham (2002) paper.
   (g) Calculate the area under the ROC curve (AUC) for these data with and without ties as per the example.
3. Write a function to calculate the area under the ROC curve for a data frame that has `event` and `p1` as components that can be used in conjunction with the `boot` function. **Hint**: return the `A` component of the list object returned from the `roc.area` function. Remember that your function must have `d` and `i` as arguments.
4. Obtain 1000 resamples (with replacement) of the AUC for the Mason and Graham (2002) data from 2 (f) above using the `boot` function.
5. Calculate 95% CI's using the BCa method for the AUC resamples obtained in 4 above.
6. Interpret the CI's obtained in 5 above.

Part 2: Dependence

1. View the help file for the `arima.sim` function.

2. Using `arima.sim`, simulate samples of size 500 from:
   (a) a white noise process (assign it the name `z.independent`),
   (b) an AR(1) model (assign it the name `z.ar1`),
   (c) an AR(2) model (assign it the name `z.ar2`),
   (d) an MA(1) model (assign it the name `z.ma1`),
   (e) an MA(2) model (assign it the name `z.ma2`),
   (f) an ARMA(1,1) model (assign it the name `z.arma1.1`),
   (g) an ARMA(2,1) model (assign it the name `z.arma2.1`),
   (h) an ARIMA(1,1,1) model (assign it the name `z.arima1.1.1`).

3. For each simulated series from 2, inspect the time series, autocorrelation- and partial autocorrelation function graphs. What do you notice from the graphs? **Hint**: You might want to conferexample 3.1 of Gilleland (2010b, available at http://www.ral.ucar.edu/staff/ericg/Gilleland2010.pdf). The following code will produce these graphs for `z.independent`.

```
par( mfrow=c(3,1)) # make three plots on a single device.
plot( z.independent)
acf( z.independent)
pacf( z.independent)
```

4. For each series, plot their histograms. **Hint**: see the help file for `hist`.

5. For each series, plot their normal qq-plots. **Hint**: see the help file for `qqnorm`.

6. Do the series appear to be normally distributed based on the qq-plots?

7. For each series, calculate the mean and standard error of the mean. Using these values, estimate 95% normal approximation CI's for their means.

8. Find 95% BCa CI's based on the IID bootstrap (i.e., as in part 1) for the mean for each of these series. How do the intervals compare with the normal approximation intervals?

9. Write a function to calculate the mean for each series that can be used in conjunction with the `tsboot` function. Using this function with `tsboot`, resample 1000 values for the mean using block lengths of size $\lfloor\sqrt{500}\rfloor = 22$ (use the circular block bootstrap approach).

10. Calculate 95% (percentile method) CI's for the mean from each of the simulated series. How do these intervals compare with those from 7 and 8 above?

11. Suppose confidence intervals have been reported based on the data's being independent and identically distributed. It is agreed that the second assumption is reasonable, but it is discovered that the underlying data are, in fact, dependent. Time has run out, and the report needs to be submitted. There is time to slightly modify the text, but not to perform any further analyses. It is of interest to know whether the mean is statistically significantly different from a baseline mean, say zero. Using the IID results, how should the text be modified if the original conclusions were that:

    (a) At the 95% confidence level, the mean is not significantly different from zero?

    (b) At the 95% confidence level, the mean is significantly different from zero?

12. After the report from 11 has been submitted. The funding agency sends it back wanting confidence intervals that reflect the dependence in the data because of a political backlashing. In terms of the conclusions drawn from the report, is it necessary to re-do the analyses under the scenario of 11 (a)? Under the scenario of 11 (b)?

Part 3: Violation of normality

1. Simulate 100 verification sets (i.e., "observed" and "forecast" time series) of size 50 using the techniques of appendix A.3 in Gilleland (2010a, available at: `http://nldr.library.ucar.edu/collections/technotes/asset-000-000-000-846.pdf`) such that the mean of each observed series is 2 (with standard deviation of 1 for both forecast and observed series), the observed and forecast series are each correlated with each other with a correlation coefficient of 0.7, but are independent over time, and:

    (a) each forecast is unbiased (i.e., has the same mean as the observed series), and assign the collection of sets the name `z0` (e.g., as a 100×50 matrix object named `z0`);

    (b) each forecast is biased high having a mean of 4, and assign it the name `z4`;

    (c) each forecast is biased low having a mean of 1, and assign it the name `z1`.

2. Plot the observed vs. forecast simulations for some of the verification sets obtained from 1 above to make sure that they are correlated.

3. Calculate the frequency bias for each verification set from 1 above (i.e., result is three samples of size 100 of frequency bias for: (a) unbiased forecasts, (b) forecasts that are biased high, and (c) forecasts that are biased low) for the events:

   (a) $z0 > 2.5$, $z4 > 4.5$, $z1 > 2.5$;

   (b) $z0 > 3$, $z4 > 5$, $z1 > 3$.

4. For each frequency bias sample from 3 above, plot the normal qq-plots. Is the normal assumption reasonable for bias in general? Is it reasonable for any of the simulated samples?

5. Estimate the standard error of bias for each of the simulated frequency bias samples from 3 above. **Hint**: you can use the R function `sd` on each sample to obtain this estimate.

6. Calculate 95% normal approximation CI's for frequency bias for each event in 3, using the standard deviations as estimates of standard error for frequency bias in each case.

7. Find the 95% BCa bootstrap CI's for frequency bias for each event from 3. How do they compare to the normal approximation intervals from 6?