Equilibration, Uncertainty, and Bootstrapping

Or, what do my simulations actually mean?

Prof. Michael Shirts
University of Colorado, Boulder
i-CoMSE MC/MD Workshop
Oklahoma State University, July 2022

Simple uncertainty

- Average in the mean is:
  - $\langle U \rangle = \int p(x) U(x) dx$
  - But if we sample from the distribution, we can replace **integral** with a **sum over observations**
  - Monte Carlo integration
  - $\langle U \rangle = \frac{1}{N} \sum_i U_i$

How do we find the uncertainty of an estimate?

$$\langle U \rangle = \frac{1}{N} \sum_i U_i$$

- What do we MEAN by the uncertainty?
- If we did the same experiment again how different would it be?

Formulas for calculating the standard error of the mean

Sample Variance

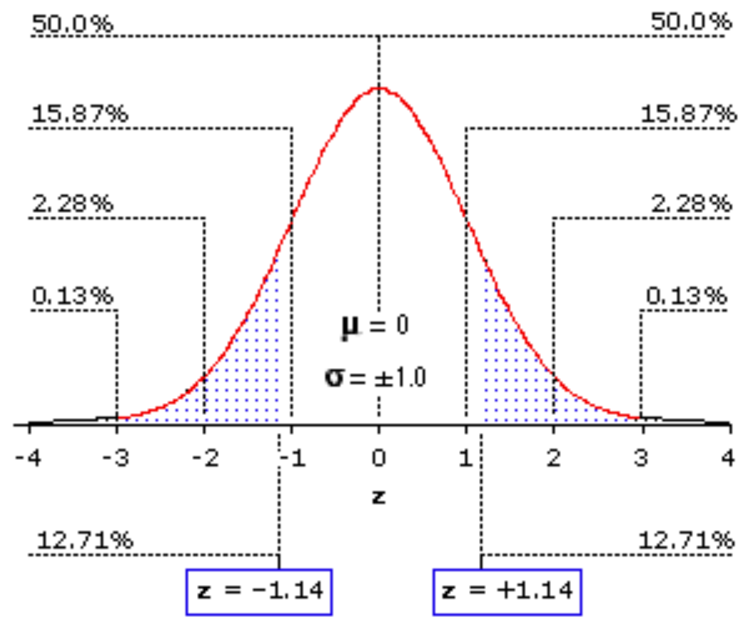
$$\delta\langle A\rangle = \sqrt{\frac{\left\langle (A - \langle A\rangle)^2 \right\rangle}{N-1}}$$

- What does this formula assume?
  - Gaussian distribution of error
  - **Stationary** time series
  - **Independent** samples

How do we REPORT the uncertainty?

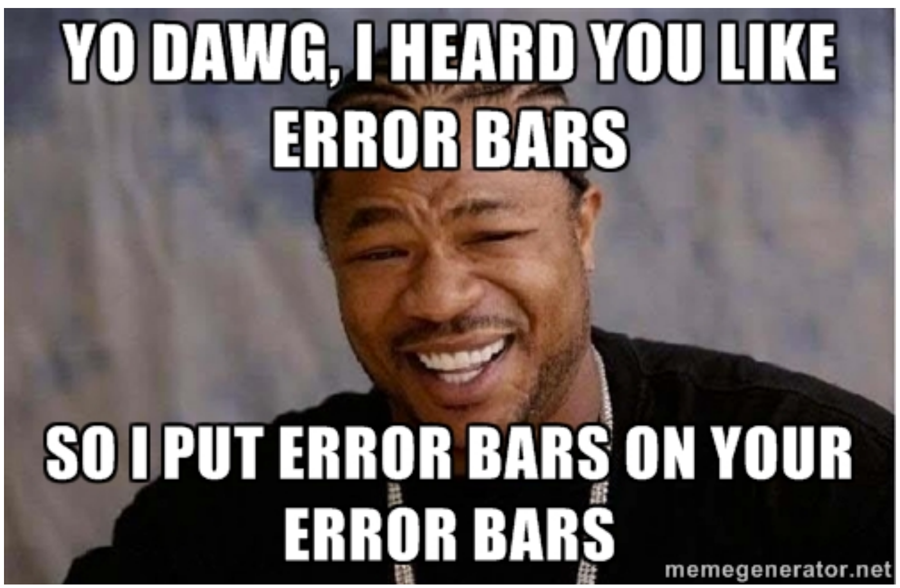
$$\langle U \rangle = \frac{1}{N} \sum_i U_i$$

- We will get a distribution of answers.
- We usually report +/- something. What is that something?
- We usually report a standard error of the mean.
- What does that mean?



What are the error bars in my error bars?

$$\langle U \rangle = \frac{1}{N} \sum_i U_i$$



- Rough rule of thumb:
- If you have 40 independent samples, then the error in the standard error is about <5%

Two main tasks

- Identifying when my simulations have become *stationary*.
- Identifying how many simulation points are *independent*.

Identifying time correlation in simulation data

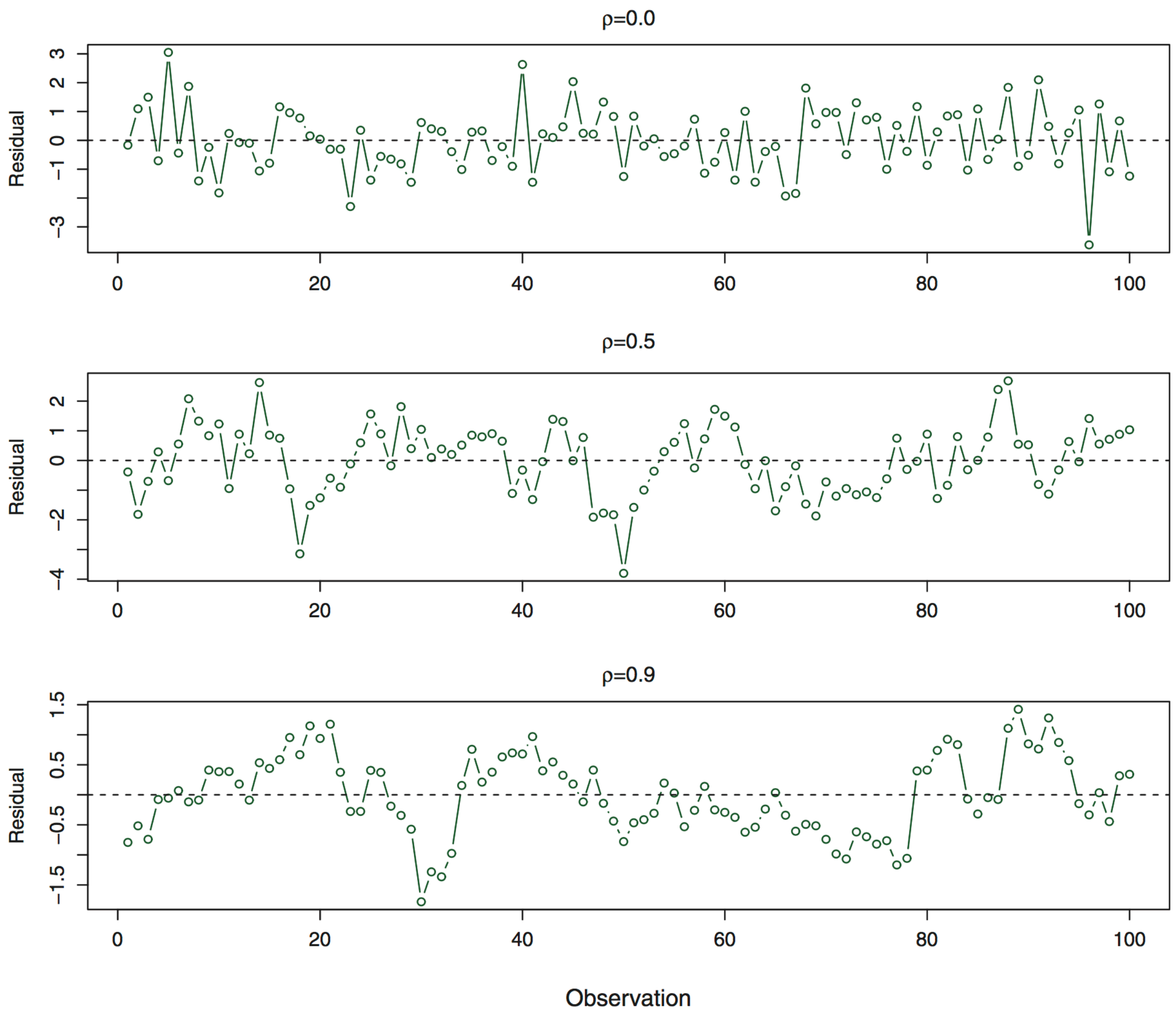


FIGURE 3.10. *Plots of residuals from simulated time series data sets generated with differing levels of correlation ρ between error terms for adjacent time points.*

What about distributions that are not normal?

- How to report confidence intervals in those distributions?
- How to report confidence intervals of the means of those distributions?

Go to notebook!

How many *independent* points do I have?

- The autocorrelation function is defined as:

$$A(\tau) = \frac{1}{Var(A(t))} \int_0^{\infty} A(t)A(t+\tau)dt$$

Where: $Var(A(t)) = \left\langle (A - \langle A \rangle)^2 \right\rangle$

- We subtract mean from A so that <A> = 0
- We divide by *Var(A(t))* to make sure that the autocorrelation functions starts at 1 when t=0.
- We usually have discrete samples, so we use:

$$A(\tau) = \frac{1}{Var(A(t))} \frac{1}{N} \sum_{t=0}^{N} A(t)A(t+\tau)$$

Only meaningful for stationary series!

Three ways to estimate when samples are independent

- Three ways to determine if they are not correlated anymore:
  - Determine when the autocorrelation function crosses to zero
  - Integrate the time under the autocorrelation function, use that as the correlation time.
  - Fit the autocorrelation to an exponential, estimate the characteristic time from $\exp(-t/\tau)$

Go to notebook!