Simulation study notes based on the following resources:

- https://onlinelibrary.wiley.com/doi/pdfdirect/10.1002/sim.8086 ((Morris paper))
- https://www.youtube.com/watch?v=sbSvLEtbCUM (Tim Morris)
- https://towardsdatascience.com/using-simulation-studies-to-motivate-modelling-decisions-be8bae2cd1c2 (article similar to Morris)
- https://www.youtube.com/watch?v=OgO1gpXSUzU (6. Monte Carlo Simulation)

What are simulation studies?

- Computer experiments that involve creating data **by pseudo-random** sampling from known probability distributions. (Source: Tim Morris paper)
- Computer experiments that use (pseudo) random sampling to understand the properties of statistical methods. (Source: Tim Morris talk)
- Simulation studies are used to obtain empirical results about the performance of statistical methods in certain scenarios.
- Example:
    1. Draw a data using a model or other procedure
    2. Analyse data using some method
    3. Repeat steps 1-2 many times
    4. Look how methods performs

Side notes:

- **What do men by "pseudo-random" ?**
    - Since a **computer is deterministic**, it can't generate real random numbers. But computers can generate sequences of pseudo-random numbers which are deterministic sequences of numbers which has the same statistical properties as sequences of real random numbers. (Source: https://www.eriksmistad.no/pseudo-random-numbers-and-sampling-from-probability-distributions/)
    - A common topic in Computer Science is the difference between true random and pseudo-random number generators. The most common of the two is pseudo-random number generators (PRNGs), which use software and a "seed" to generate a random number.
- **What do we mean by "computer is deterministic"?**
    - What deterministic means? believing that everything that happens must happen as it does and could not have happened any other way, or relating to this belief:
    - Why is a computer a deterministic machine? What we do need to understand is that computers are based on the laws of physics, and as a

result, the hardware is meant to be deterministic. Deterministic means that given the same inputs, there should be the same outputs (https://www.hoppersroppers.org/fundamentals/Hardware/2-ComputersareDeterministic.html#)
    - Deterministic algorithm is an algorithm that, given a particular input, will always produce the same output, with the underlying machine always passing through the same sequence of states (Deterministic algorithm - WikipediaWikipediahttps://en.wikipedia.org › wiki › Deterministic_algorithm)
- **Monte Carlo methods**, or **Monte Carlo experiments**, are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. The underlying concept is to use randomness to solve problems that might be deterministic in principle.
- "**Monte Carlo simulation**" means statistical techniques that use pseudo-random sampling, and has many uses that are not simulation studies.

Where/Ways to use simulation studies:

- Double checking algebra when statistical methods are derived mathematically: To check algebra (and code), or to provide reassurance that no large error has been made, where a new statistical method has been derived mathematically.
- Assessing the relevance of large-sample theory approximations in finite samples: To assess the relevance of large-sample theory approximations (eg, considering the sampling distribution of an estimator) in finite samples.
- For the absolute evaluation of a new or existing statistical method: Often a new method is checked using simulation to ensure it works in the scenarios for which it was designed.
- Comparing two or more statistical methods: For comparative evaluation of two or more statistical methods.
- Calculating a sample size or statistical power when designing a study: For calculation of sample size or power when designing a study under certain assumptions.

Why to do simulation study ?

- They are an invaluable tool for statistical research, particularly for the evaluation of new methods and for the comparison of alternative methods
- A key strength of simulation studies is the ability to understand the behavior of statistical methods because some "truth" (usually some parameter/s of interest) is known from the process of generating the data.

- It is not always possible, or may be difficult, to obtain analytic results. Simulation studies come into their own when methods make wrong assumptions or data are messy because they can assess the resilience of methods in such situations.

Key steps and decisions of simulation studies:

1. Planning (ADEMP)
    a. Aims
    b. Data-generating mechanisms
    c. Estimand/target of analysis
    d. Methods
    e. Performance measures
2. CODING AND EXECUTION
    a. Separate scripts used to analyze simulated datasets from scripts to analyze estimates datasets.
    b. Start small and build up code, including plenty of checks.
    c. Set the random number seed once per simulation repetition.
    d. Store the random number states at the start of each repetition.
    e. If running chunks of the simulation in parallel, use separate streams of random numbers
3. analysis
    a. Conduct exploratory analysis of results, particularly graphical exploration.
    b. Compute estimates of performance and Monte Carlo SEs for these estimates.
4. Reporting
    a. Describe simulation study using ADEMP structure with sufficient rationale for choices.
    b. Structure graphical and tabular presentations to place performance of competing methods side-by-side.
    c. Include Monte Carlo SE as an estimate of simulation uncertainty.
    d. Publish code to execute the simulation study including user-written routines.

"ADEMP**"** [simulation studies for the evaluation of statistical methods.]
A structured approach for planning and reporting simulation studies, which involves:

- defining aims
- data-generating mechanisms,
- estimands (what is the thing that we are interested in estimating)
- methods,
- performance measures

ADEMP in details

Aims

- We use simulation to:
  - Evaluate the performance of methods/ understand the behavior of methods, in terms of : statistical bias, variance estimation, robustness to misspecification.
  - Find scenarios where methods fail
  - When there is a lack of guidance in literature or methods are novel

Data generating mechanisms

- how random numbers are used to generate a dataset for simulation
- The choice of data-generating mechanism(s) will depend on the aims
- Data can be generated using :
  1. parametric draws from a known distribution.
     - can explore many different data-generating mechanisms (which may be completely unrealistic)
  2. by sampling with replacement from an existing dataset.
     - The true data-generating mechanism is unknown and resamples are used to study the sampling distribution.
     - resampling typically explores only one mechanism (which will be relevant for at least the study at hand)
- Tip: A few scenarios will usually be simulated, with varying sample sizes and/or effect sizes.
- involve more than one data-generating mechanism to ensure coverage of different scenarios.
- Factors that are frequently varied are sample size (several values) and true parameter values (for example, setting one or more parameters to be zero or nonzero).

Estimands

- Most simulation studies compare methods for estimating population quantities, which are usually simulated parameters in the data generating process.
- In a given setting ("target"), some performance measures are more appropriate than others:

TABLE 3 Possible targets of a simulation study and relevant performance

Statistical Task	Target	Examples of Performance Measures
Analysis		
Estimation	Estimand	Bias, empirical SE, mean-squared error, coverage
Testing	Null hypothesis	Type I error rate, power
Model selection	Model	Correct model rate, sensitivity or specificity for covariate selection
Prediction	Prediction/s	Measures of predictive accuracy, calibration, discrimination
Design		
Design a study	Selected design	Sample size, expected sample size, power/precision

Methods

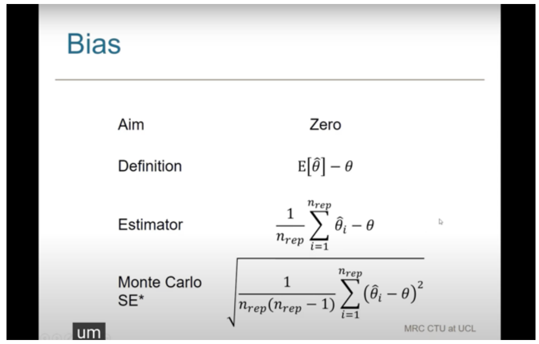
- some methods should be excluded from the simulation if they have already been shown to be flawed, since this will only add computational burden.

Performance measures

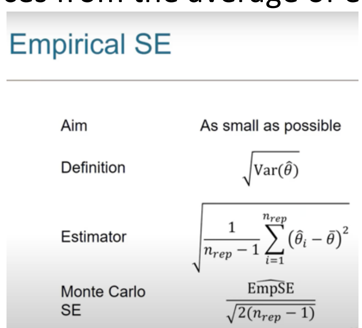
- Performance measures are quantities used to assess the performance of a method
- Statistical performance measure (properties of the methods, desirable properties of an estimator (point estimate we got from the method)
- The aims of a simulation study will typically be set out in relation to these properties, depending on what specifically we wish to learn (depend on the aims and targets (estimand) of the study)
- because simulation studies are empirical by essence, performance measures themselves are subject to error, which means analysts should be presenting estimates of uncertainty (Monte Carlo standard errors).

Examples of Performance measures

1. **Bias**: the amount by which the expectation of estimator exceeds the true value of the estimand (positive or negative)
   - aim is Zero
   - Monte Carlo SE (standard error) of bias: quantify the variability in estimating the bias due finite number of simulation (because when we run the simulation with finite sample size , what we estimate has degree on uncertainty about it.



2. **Empirical Standard Error :** repeated sampling variation of the estimator, it is the standard deviation of the estimator across all repetition.
    - may not be meaningful for biased methods. Might get high bias and low Empirical SE
    - Use the squared distances from the average of estimator



3. **Mean squared error**: kinda brings the bias and empirical SE together (combine) .
    - Sometime we accept to trade small amount of bias for a reduction in variance (bias-variance trade-off)
    - Use the squared distances from the truth (estimand)
    - If you want use this measure for comparison, you need to vary the sample size (this measure is affected by the sample size)
    - Better to report bias and empirical SE separately as the comparison is less dependant on the sample size

Mean squared error

Aim	As small as possible
Definition	$E[(\hat{\theta}) - \theta)^2]$
Estimator	$\frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} (\hat{\theta}_i - \theta)^2$

4. **average model standard error:** accuracy of the standard error (how accurate it is)
    - Root-mean squared model standard error: we compare the SE on variance scale across all repetitions
    - Gives the right description of the true empirical SE

Model standard error

Aim	Accurate representation of empirical SE
Definition	$\sqrt{E[\widehat{Var}(\hat{\theta})]}$
Estimator	$\sqrt{\frac{1}{n_{rep}} \sum_{i=1}^{n_{rep}} \widehat{Var}(\hat{\theta}_i)}$

5. **Relative error in Model SE**: provides explicit comparison of avg model SE with empirical SE, is it accurate? Ration of two measures by dividing them
6. **Coverage**: It is defined as the probability that a confidence interval contains estimand.
    - The coverage of a confidence interval is the proportion of times that the obtained confidence interval contains the true specified parameter value
    - The coverage should be approximately equal to the nominal coverage rate, e.g. 95 per cent of samples for 95 per cent confidence intervals, to properly control the type I error rate for testing a null hypothesis of no effect.
    - Over-coverage, where the coverage rates are above 95 per cent, suggests that the results are too conservative as more simulations will not find a significant result when there is a true effect thus leading to a loss of statistical power with too many type II errors.
    - under-coverage, where the coverage rates are lower than 95 per cent, is unacceptable as it indicates over-confidence in the estimates since more simulations will incorrectly detect a significant result, which leads to higher than expected type I errors.

Computational performance measure

- Speed and convergence are very important
- Sometimes we accept the weaker performance methods due their faster speed