Simulation: Discretizing SDE's & Variance Reduction

Goals:

- Describe the process by which we turn a sequence of random numbers into the path of asset prices.
- Discuss the most common variance reduction techniques.
- Introduce one of the more flexible and useful applications in simulation: Statistical Bootstrapping.

Relevant literature:

- Hirsa, Chapter 6

Review: What do we know so far about simulation?

- In the last lecture we learned how to generate random uniform numbers.
- In particular, there are two main approaches for doing this:
  - Pseudorandom number generation
  - Quasi-Random numbers / Low Discrepancy Sequences
- We also learned how to transform these uniform samples to sample of more interesting and more relevant distributions, such as:
  - Exponentially Distribution
  - Normal Distribution
- In the case of normally distributed numbers, we also learned how to turn a sequence of uncorrelated standard normals into a samples from a multivariate normal distribution.

Review: What's left?

- We still need to know what to do with these random numbers.
- We saw previously that if we knew the PDF, and knew how to sample from it directly, then we can apply these samples directly to compute our integral.
- If the PDF is unknown, or the payoff requires the entire path of the asset, we will instead need to sample from the SDE. Today we discuss a few approaches for doing this.
- We also want to know how to make sure our simulation algorithms are written as efficiently as possible, which we will also touch on today.

SDE Discretization: Overview

- To this point, we know how to generate random samples, including samples from joint normal distributions. We still need to learn how to translate those samples into a path for an arbitrary SDE.

- If we knew the solution to the SDE, that is if we knew the PDF of the instrument we were trying to price, then we wouldn't need to simulate from the SDE. However this is not common.

- SDE's are continuous and describe the evolution of an asset over infinitely small time-steps. In practice, we cannot simulate an SDE exactly but instead but apply a discretization scheme to the SDE.

- This means choosing a time grid $\Delta t = T/M$ where M is the total number of timesteps and T is the time to expiry.

SDE Discretization: Overview

- Consider the following generic SDE:

$$\begin{aligned} dS_t &= \mu(S_t)dt + \sigma(S_t)dW_t && (1) \\ S(t_0) &= S_0 && (2) \end{aligned}$$

- In integral form, this is

$$S_t = S_{t_0} + \int_{t_0}^{t} \mu(S_t)\, ds + \int_{t_0}^{t} \sigma(S_t)\, dW(s) \qquad (3)$$

- If we apply the left-point rule to each integral in (3) we get the most common discretization scheme, the Euler scheme.
- We can use the left-point rule on the ds integral in (3) as it is a deterministic integral.
- It is less obvious that we can do this to the $dW(s)$ integral.

SDE Discretization: Euler Scheme

- In order to apply the left-point rule to the ds integral in (3), we can do the following:

  $$\int_{t_0}^{T} \mu(S_t)\, ds \approx \mu(S_{t_0})(T - t_0)$$

  for $T - t_0 \approx 0$.

- Similarly, for the $dW(s)$ integral, we can use:

  $$\int_{t_0}^{T} \sigma(S_t)\, dW(s) \approx \sigma(S_{t_0})(B_T - B_0) = \sigma(S_{t_0})B_T$$

  where again we assume $T - t_0 \approx 0$

SDE Discretization: Euler Scheme

- As a result, we can define the approximation $\hat{S}_T$:

$$\hat{S}_T = S_0 + \mu(S_0)T + \sigma(S_0)B_T$$

where B_T is a standard Brownian Motion with $\mu = 0$ and $\sigma^2 = T$.

- This can be re-written as:

$$\hat{S}_{t_{j+1}} = \hat{S}_{t_j} + \mu(\hat{S}_{t_j})\Delta t + \sigma(\hat{S}_{t_j})\sqrt{\Delta t} Z_j \tag{4}$$

where Z_j is a standard normal random sample.

- The approximation $\hat{S}_T$ is called the **Euler discretization** of S_T
- If T is very small, then the distribution of $\hat{S}_T$ is very similar to the distribution of S_T
- Equation (4) defines for us how to take steps along the path of our SDE. We will be using it repeatedly in all of our simulation algorithms.

SDE Discretization: Other Methods

- We can also try to get more accuracy by performing an Ito-Taylor series expansion of (1) and (3) and trying to have our discretization schemes account for these terms as well.

- Two of the most common alternative schemes that use this approach are the **Milstein scheme** and the **Runge-Kutta scheme**. We don't have time to discuss them in detail in this course, however, they are detailed in Hirsa's book (page 229) for those interested in exploring further.

First Example: Simulating Geometric Brownian Motion

- To make sure we understand how to apply the Euler method of simulation, let's work through a few examples. To start, let's look at the Black-Scholes model.
- Recall that this is defined by the following SDE:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{5}$$

- As a result, the equation (4) becomes:

$$\hat{S}_{t_{j+1}} = \hat{S}_{t_j} + \mu \hat{S}_{t_j} \Delta t + \sigma \hat{S}_{t_j} \sqrt{\Delta t} Z_j \tag{6}$$

where μ and σ are constants.

First Example: Simulating Geometric Brownian Motion

- It just so happens that we know the solution for this SDE as well, which is:

  $$S_T = S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma dW_T\right\} \tag{7}$$

  This means that we can simulate the process directly using the solution to the SDE, and do not need to apply discretization methods. This can be done via the following formula:

  $$S_T = S_0 \exp\left\{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}Z\right\} \tag{8}$$

  where Z is a standard normal random variable.

Second Example: Simulating from the Heston Model

- In the case of simulating from a Heston model, we need to simulated a system of two SDE's.
- Unlike in the previous example, the solution to the SDE will be unknown.
- Additionally, we need to generate correlated random normals to simulate the asset and its volatility consistently with its model correlation, ρ.
- Recall that the Heston model is defined by the following system of SDE's:

$$\begin{aligned} dS_t &= rS_t\,dt + \sqrt{\nu_t}S_t\,dW_t^1 \\ d\nu_t &= \kappa(\theta - \nu_t)\,dt + \sigma\sqrt{\nu_t}\,dW_t^2 \\ Cov(dW_t^1, dW_t^2) &= \rho\,dt \end{aligned} \tag{9}$$

Second Example: Simulating from the Heston Model

- The Euler Discretization for the Heston model is:

$$\hat{S}_{t_{j+1}} = \hat{S}_{t_j} + \mu\hat{S}_{t_j}\Delta t + \sqrt{\nu_{t_j}}\hat{S}_{t_j}\sqrt{\Delta t}Z_j^1 \quad (10)$$

$$\hat{\nu}_{t_{j+1}} = \hat{\nu}_{t_j} + \kappa(\theta - \nu_{t_j})\Delta t + \sigma\sqrt{\nu_{t_j}}\sqrt{\Delta t}Z_j^2 \quad (11)$$

  where Z_j^1 and Z_j^2 are standard normal random variables with correlation ρ.

- Recall that we went through the process of generating these correlated standard normal random variables in detail in the last lecture.

Practicalities of Simulating from Heston

- One of the issues with applying simulation techniques to the Heston model is that there is some positive probability that the volatility process will become negative.

- Note that this is not possible in the continuous version of the SDE, but also becomes possible when we enter our discretization scheme.

- Obviously negative volatilities will not work well in our model and we will need to adjust accordingly. In particular, we might:
  - Treat 0 as an absorbing state for volatility
  - Truncate the value at 0 by applying a $\max$ function to the volatility process.
  - Reflect the volatility back above 0 if it goes negative.

Valuation via Simulation: Pseudocode

```
$N = 10000$, $S_0 = 100$, $T = 0.25$, $M = 252$
$dt = T/M$

for i = 1,...,N
    $\tilde{S} = S_0$
    $t = 0$

    for j=1,...,M
        $Z$ = GenerateStandardNormal()
        $dW_t$ = $\sqrt{\texttt{dt}}Z$
        $\tilde{S}$ = ApplyDwToSDE($dW_t$,t)
    end for loop

    $P(i)$ = CalcPayoffAlongPath($\tilde{S}$)
end for loop

$\tilde{\mu} = \frac{1}{N}\sum_{i=1}^{N} P(i)$
```

Variance Reduction Techniques: Overview

- Variance Reduction Techniques are an important field of study as simulation is a hugely important tool and its main drawback is efficiency.
- We could spend several lectures talking about variance reduction techniques but unfortunately only have the time to introduce them at a high level.
- They are all detailed in Hirsa's book for those interested in exploring further.
- The main techniques are:
  - Antithetic Variables
  - Control Variates
  - Importance Sampling
  - Common Random Numbers

Variance Reduction Techniuqes: Overview

- Recall from earlier lectures that the main challenge in building simulation algorithms is the rate of convergence. We can overcome this by increasing the number of samples in our estimate or alternatively, trying to reduce the variance of the samples.
- Attempting to reduce the variance of the samples is an important topic of research in simulation.
- Variance Reduction Techniques attempt to use information about the problem that we are working on (e.g. Control Variates) or information about the samples (e.g. Antithetic Variables) in order to reduce the variance.
- We do not have time to delve deeply into any of these methods, but want you to be aware of them and encourage you to research them further if they are of interest.

Variance Reduction Techniuques: Overview

- The variance reduction techniques that we discuss are:
  - Antithetic Variables
  - Control Variates
  - Importance Sampling
  - Common Random Numbers
- There are many other methods that we will not get into, but these are four of the most commonly applied, and easy to implement, techniques.

Antithetic Variables

- The antithetic variables method uses correlation between pairs of random samples in order to reduce the variance of our estimator.
- To see this, let's suppose we generate two samples X_1 and X_2 from a distribution $\phi(x)$.
- The mean of these samples, μ is equal to: $\frac{1}{2}(X_1 + X_2)$.
- The variance of μ is equal to $\frac{1}{4}\text{var}(X_1) + \frac{1}{4}\text{var}(X_2) + \frac{1}{2}\text{cov}(X_1, X_2)$.
- There are two things to note immediately:
  - If the samples are i.i.d. then the variance of μ simplifies to: $\frac{1}{4}\left(\text{var}(X_1) + \text{var}(X_2)\right)$
  - The variance is smallest when the two samples are perfectly negatively correlated.

Antithetic Variables

- Antithetic variables are generally easy to generate using standard methods.
  - One example of an antithetic variable would be to use Z and $-Z$ to generate standard normals.
  - Another example would be U and $1-U$ in the case of uniform random samples between 0 and 1.
  - In both cases the two samples are negatively correlated but still identically distributed.
- NOTE: We should only apply antithetic variables when we are estimating a mean and not when we are estimating a variance.

Control Variates

- The Control Variate method uses information about a related quantity that has known expectation in order to give us information about a harder problem.
- In practice this could mean that we are pricing an exotic option with unknown expectation, and using a European call option with known expectation as a control variate.
- Let's start with the standard estimator of our expectation, θ that is defined as:

  $$\theta = \mathbb{E}\left(h(X)\right) \tag{12}$$

  where $h(X)$ is the payoff function.
- When using a control variate, we use the estimator

  $$\tilde{\theta} = \mathbb{E}\left(h(X)\right) + c(Z - \mathbb{E}\left(Z\right)) \tag{13}$$

Control Variates

- This estimator is still unbiased, as clearly:

$$\mathbb{E}\left[\tilde{\theta}\right] = \theta \tag{14}$$

- Further, the variance of our new estimator is:

$$\begin{aligned} \text{var}(\tilde{\theta}) &= \text{var}(\mathbb{E}\left(h(X)\right) + c(Z - \mathbb{E}\left(Z\right))) \quad (15) \\ &= \text{var}(\theta) + (c^2)\text{var}(Z) + (2c)\text{cov}(\theta, Z) \quad (16) \end{aligned}$$

- We want to minimize the variance of $\tilde{\theta}$, and we get to choose c.

Control Variates

- Differentiating (16), we see that c is minimized at:

$$c^* = -\frac{\mathrm{cov}(\theta, Z)}{\mathrm{var}(Z)} \tag{17}$$

- Plugging this back into (16), we get:

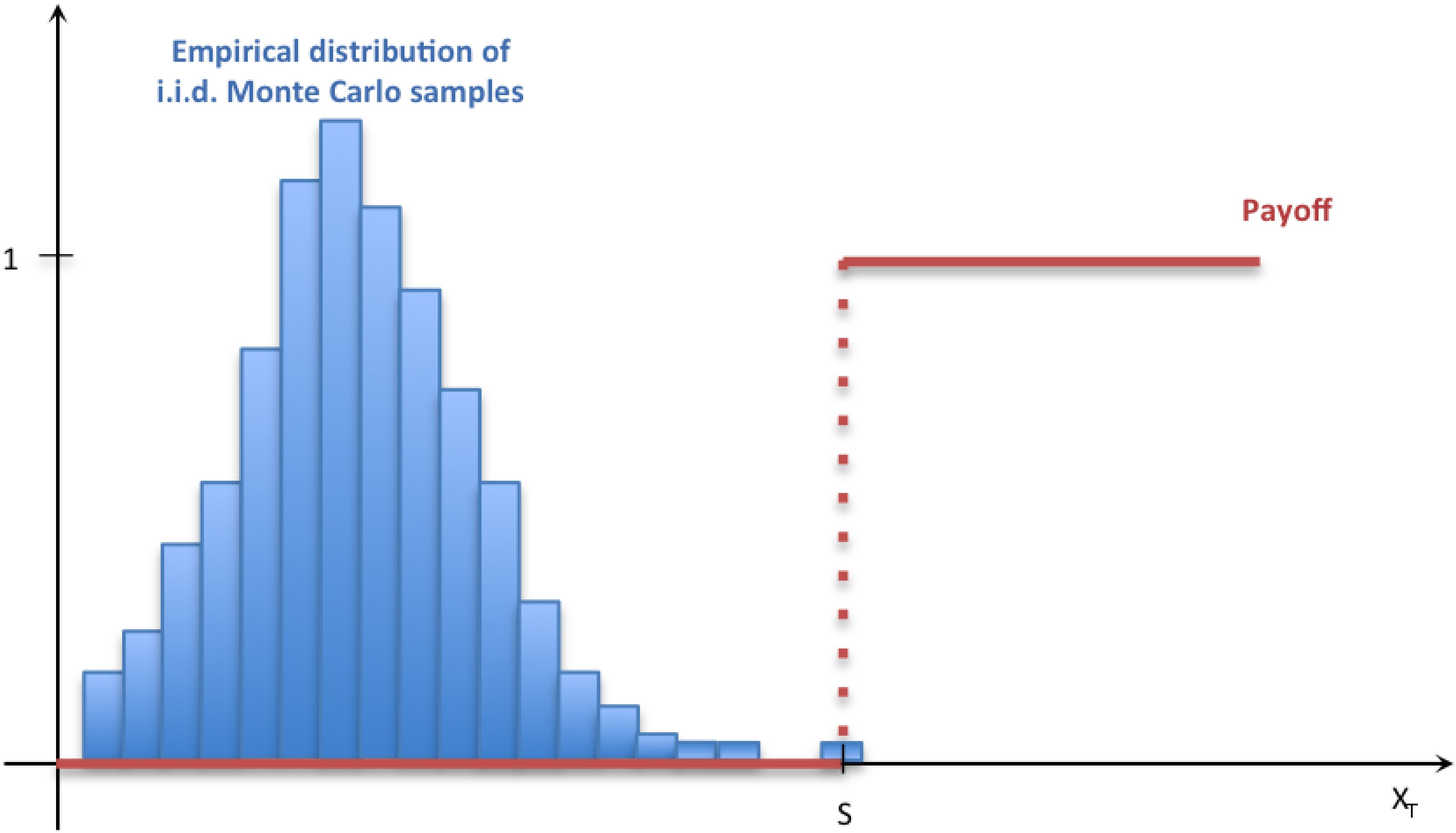
$$\mathrm{var}(\tilde{\theta}) = \mathrm{var}(\theta) - \frac{\mathrm{cov}(\theta, Z)^2}{\mathrm{var}(Z)}$$

- Therefore, the variance is lower than $\mathrm{var}(\theta)$ as long as the covariance is not equal to zero.

Importance Sampling

- Importance sampling is very closely related to Girsanov's theorem and is most commonly applied to simulating rare events.
- Essentially, when we are attempting to simulate something with very low probability, we can change measure to increase the probability of our desired event, calculate its probability, and then change measure back to adjust the probability back to our original measure.
- A canonical example of importance sampling is trying to estimate the probability of 5+ standard deviation moves in a Normal Distribution.
- It can easily be shown that the importance sampling estimator has lower variance, and is still an unbiased estimator.

Importance Sampling: Example



Importance Sampling

- Suppose we are computing the expectation of some payoff function $h(x)$ and let's define θ as our estimator: $\theta = E = \mathbb{E}_\phi[h(X)]$
- If X follows the pdf $\phi(x)$, then this can be written as:

$$\begin{aligned}\theta &= \int h(x)\phi(x)dx = \int h(x)\frac{\phi(x)}{g(x)}g(x)dx \\ &= \mathbb{E}_g\left(h(X)\frac{\phi(X)}{g(X)}\right)\end{aligned}$$

- Therefore, we can compute our estimator in a different probability measure, $g(X)$, and apply an adjusted payoff function $h(X)\frac{\phi(X)}{g(X)}$.
- That leaves us with the importance sampling estimator:

$$\tilde{\theta} = \frac{1}{N}\sum_{i=1}^{N} h(X_i)\frac{\phi(X_i)}{g(X_i)} \tag{18}$$

Common Random Numbers

- The idea behind common random numbers is to reuse the same random numbers instead of creating new samples for different parts of related calculations.
- A typical application of this technique would calculating Greeks via finite differences. In this case, we need to evaluate the price of a derivative at a few points and can do so using the same sample paths at each point.
- As an example, to calculate Δ via finite differences we would use the following formula:

$$\Delta = \frac{C(S+\Delta S) - C(S-\Delta S)}{2\Delta S} \tag{19}$$

  where ΔS is the shift amount and we are using the central difference approximation.

Common Random Numbers

- The Common Random Numbers technique suggests that we should use the same random numbers in the calculation of $C(S + \Delta S)$ and $C(S - \Delta S)$.
- Using the same random numbers in each derivative evaluation will reduce the variance of our Greek calculations.

Bootstrapping: Introduction

- Bootstrapping is a non-parametric technique for creating sample paths from an empirical distribution.
- The idea behind bootstrapping is that each realization is in actuality just a draw from some unobserved density.
- Bootstrapping provides a method for sampling from that density with minimal assumptions.
- It is also a very simple and intuitive technique, but is quite powerful nonetheless.
- In practice it involves simulating returns, and paths of asset prices by scrambling the historical data, or sampling various historical returns and creating series of them in random order.

Bootstrapping: Procedure

A bootstrapping algorithm can be summarized as follows:

- Create a matrix of historical returns for the universe of assets.
- Generate a sequence of random integers between 1 and N, where N is the number of historical days of data.
- Look up the return data for all assets that corresponds to that randomly chosen date.
- Apply these returns to the entire asset universe.
- Continue this for many time steps until the end of a path is reached. Repeat until a large number of paths are created.
- Calculate whatever payoffs, risk measures or other analytics desired from the full set of simulated paths.

Bootstrapping: Assumptions

- It can be proven that the bootstrapping algorithm will preserve the characteristics of the empirical density.
- For example, if we generate enough samples, then the mean via bootstrapping will match the mean in the historical data set. In some cases this may be desired and in other cases we may want to de-mean our input return data.
- This is true for other characteristics as well, such as variances and cross-asset correlations.
- Also note that the probability of drawing a given date is independent of the path. Phrased differently, we treat the historical data samples as i.i.d.
  - In what cases is this assumption violated?

Bootstrapping: What if there is auto-correlation in our data?

- Auto-correlation is one case that would cause the i.i.d. assumption to be violated.
- In order to overcome this we have a few options:
  - We could remove the auto-correlation from the data, generate the simulate paths, and then add the auto-correlation back. (see the Blundell-Ward Filter, for example)
  - If we know the structure of the auto-correlation, then instead of sampling individual days, we could sample indivual sections of the data. For example, if we knew the data was AR(2), then we generate samples consisting of two consecutive days.
  - We could relax the assumption that the probability of a return or historical date is independent of the previously chosen value and instead make it a function of some conditioning variables.

Bootstrapping Examples

- Re-sampling an Efficient Frontier
- Using Historical Simulation to calculate VaR
- Generating Synthetic Data
- Validating Trading Strategies
- Generating a Prior Distribution in Weighted Monte Carlo

Simulation Example: VaR of complex portfolio

- One very common example of simulation in finance is computing portfolio risk measures, such as VaR or CVaR.
- Banks, hedge funds and other financial instutions with large and complex portfolios need a means for estimating their return properties, particularly those closely related to potential losses.
- Value-at-Risk is one important, albeit somewhat flawed risk measure. It measures the max amount a portfolio will lose to a given level of certainty.
- Conditional VaR is another example, which measures the expected loss of a portfolio conditional on exceeding a given VaR threshold.
- **Important Question**: When we calculate these risk measures, should we use the risk-neutral or physical density?

VaR of complex portfolio via Gaussian Copula

- Gaussian Copulas are common techniques for creating large multivariate distributions.
- Generally speaking, copulas are a method of creating a joint distribution of many underlying variables while preserving the marginal distributions of each variable.
  - In the context of options pricing, the marginal would be each asset's risk neutral desnity.
  - In the context of VaR of other historical data based applications, the marginal would be the historical return data for each asset.
- Additionally, the Gaussian Copula specifies that the covariance structure for the joint distribution is Gaussian.

VaR of complex portfolio via Gaussian Copula: Procedure

- The procedure for using a Gaussian Copula to compute VaR is:
  - Generate many random $\mathcal{U}(0,1)$ samples for each asset
  - Transform these samples to independent standard normals.
  - Create the covariance matrix of returns from historical data by computing the asset variances and pairwise covariances.
  - If necesary, normalize the covariance matrix to make sure it is well behaved and close to full rank.
  - Using SVD or Cholesky obtain the square root of the covariance matrix.
  - Transform the independent standard normals to correlated normals using the square root of the covariance matrix. This gives us a set of simulated correlated asset returns.

VaR of complex portfolio via Gaussian Copula: Procedure

- Apply these correlated asset returns to today's prices for each asset and re-price all elements in the portfolio using the simulated prices.
- Estimate the risk measure. For example, for a 95% VaR, we would sort the portfolio return data and choose the 5th percentile.

Note: You might not have even noticed where the copula was used here. It is in how the standard normals were correlated.

In particular, we assumed that the joint distribution of asset returns was normal with a given covariance matrix.

VaR of complex portfolio via Historical Simulation (e.g. Bootstrapping)

- Historical simulation is a popular technique for computing VaR and other risk metrics.
- It is computationally very easy and makes very few assumptions about the underlying joint return distribution.
- It boils down to bootstrapping of a large number of assets.
- As it relies on bootstrapping, the return during the historical period will be used as the mean return going forward. If we don't want this historical return to persist, we should adjust the input data accordingly.
- Additionally, as we are generating all returns from the same historical day at each point, this technique will rely on the correlations embedded in the historical data.

VaR of complex portfolio via Historical Simulation (e.g. Bootstrapping): Procedure

- The procedure for Historical simulation is:
  - Collect historical return data for every asset in the portfolio for a given period.
  - Generate a sequence of uniform random integers between 1 and the numbers of days in the sample, N.
  - Translate each random number into a historical date and look up the return for each asset in the portfolio for that date.
  - Calculate the prices of all positions in the portfolio given the return for each asset. If the position is a derivative, this may involve re-calculating an option price at a new level of spot.
  - Calculate the value of the portfolio (& its return) by multiplying the positions by the simulated prices of the assets.
  - Calculate the desired risk metric. For example, for a 95% VaR

we would sort the portfolio returns and find the 5%th percentile.

- To reiterate, this is just an application of bootstrapping by another name.
- Also note that unlike with the Gaussian Copula model we didn't have to make any specifications about the returns or cross-asset correlations.

VaR of complex portfolio via Historical Simulation (e.g. Bootstrapping)

- What assumptions are we making here with respect to:
  - Mean and Variance of underlying asset returns
  - Asset correlations
  - Auto-correlation of returns
  - Conditional vs. Unconditional asset return probabilities
- If we wanted to make different assumptions, how would we do so?

VaR of complex portfolio: Comments

- When we compute a portfolio VaR, we will want to verify that our VaR number is consistent with the true distribution of portfolio returns. How can we do this?
- One way is to check how often our VaR theshold is crossed.
  - If we are computing a 95% VaR then 5% of our observations should exceed our VaR number.
  - Significant departures from this might tell us that our VaR calculation is biased.
  - Ideally, this test should be done both in and out-of-sample.
    - ∗ The in-sample test is effectively just a means of making sure your algorithm is set up correctly.
    - ∗ The out-of-sample test is a form of back-test and involves computing VaR every day and comparing the previous VaR to the subsequent return.