Simulation and Bootstrapping

FRM Part I

1 A story: play the game ten thousand times

- A game too complex for paper: play it many times
- Count the wins: that fraction estimates the chance
- A computer plays the game with random numbers
- The result is an estimate, never an exact answer

IN THE LESSON

- Imagine a card game with such complicated rules that no one can work out your chance of winning on paper.
- What could you do?
- You could play it ten thousand times and count how often you win.
- If you win three thousand times, your best estimate is thirty per cent.
- That is the whole idea of Monte Carlo simulation.
- The name comes from the famous casino, and the method was developed in the nineteen forties by scientists working on nuclear physics problems.
- The trick is that a computer can play the game for you, quickly and cheaply, using random numbers.
- The price is that the answer is only an estimate.

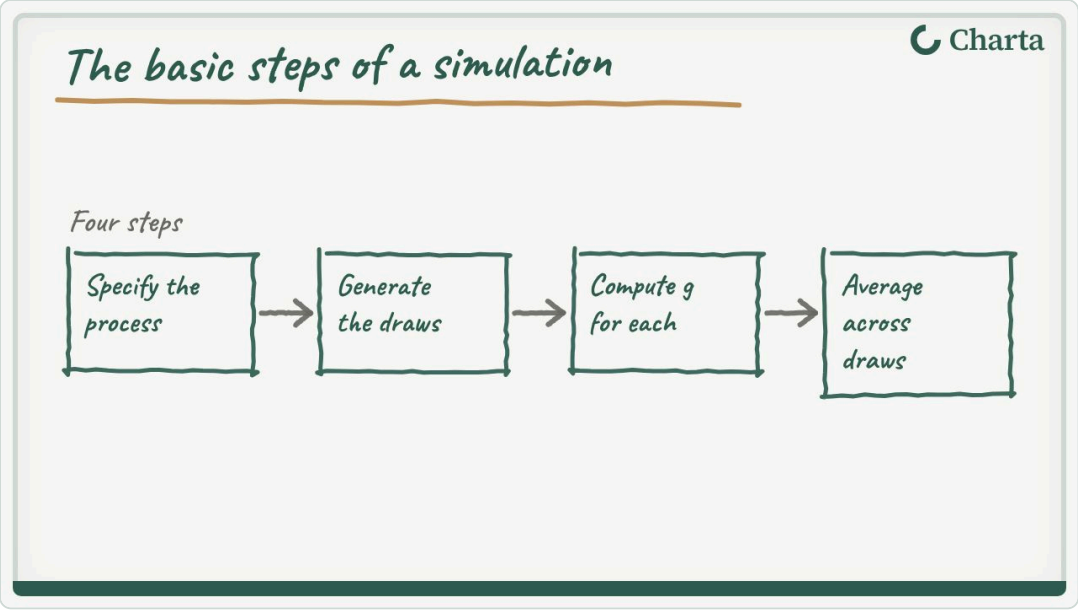
2 Monte Carlo: averaging random draws

- Goal: estimate an expectation E[g(X)]
- Method: average g over many random draws
- Second moment: mean of x^2, not (mean of x)^2
- Probability: fraction of draws meeting a condition
- The answer is an estimate with sampling error

IN THE LESSON

- Here is the idea in finance terms.
- We want an expected value of some quantity that depends on a random input.
- Call the quantity of interest the function g of the random variable.
- It might be an option payoff, a portfolio loss or a power of a return.
- The law of large numbers says that the average of g over many independent random draws gets close to the true expected value.
- So the second moment is estimated by the average of the squared draws, not by the square of the average draw.
- A probability is just the fraction of draws that meet a condition.
- And whatever we compute is an estimate, with sampling error.

3 The basic steps of a simulation



IN THE LESSON

- There are four steps, and they are the same whatever the problem.
- First, specify the data generating process and its parameters.
- That means saying how the random inputs behave: for example, that a stock return is normal with a given mean and volatility.
- Second, generate a large number of random draws or paths from that process.
- Third, for each draw, compute the quantity of interest, such as the payoff, the loss or the value whose moment you want.
- Fourth, average across all the draws.
- If the goal is a probability, the average is the fraction of draws meeting the condition.

4 The estimator and its standard error

$$E[g(X)] \approx (1 \div N) \times \Sigma\, g(x_i)$$
$$SE = s \div \sqrt{N}$$

Quantity of interest	Payoff, loss or moment per draw
Number of draws	Number of random draws, N
Spread of the draws	Sample standard deviation of g
Standard error	Sampling error of the estimate

IN THE LESSON

- In symbols, the expected value of g is estimated by the sample mean of g over the draws.
- That is one divided by the number of draws, times the sum of g at each draw.
- The estimate has sampling error, and we measure it with the standard error.
- The standard error equals s divided by the square root of the number of draws, where s is the sample standard deviation of the simulated values.
- Notice what this says.
- A more spread out quantity is harder to pin down.
- And the number of draws enters through a square root, which is slow.
- Quadrupling the draws only halves the error.
- Everything later in this lesson responds to that fact.

5 Example 1: first and second moments

Illustration: five draws of a standard normal: 0.5, −1, 1.5, 0, −0.5

Sample mean	$0.5 \div 5$	**0.1**
Mean of the squares	$3.75 \div 5$	**0.75**
Square of the mean	0.1^2	**0.01**
Variance estimate	$0.75 - 0.01$	**0.74**

IN THE LESSON

- Let's try it with five made up draws from a standard normal distribution: point five, minus one, one point five, zero and minus point five.
- The sample mean is the sum, which is point five, divided by five: point one.
- That estimates the first moment, whose true value is zero.
- Now the second moment.
- Square each draw: point two five, one, two point two five, zero and point two five.
- They add to three point seven five, and dividing by five gives point seven five.
- Compare that with the square of the mean, point one squared, which is only point zero one.
- Using the square of the mean would be a mistake.
- The variance estimate is point seven five minus point zero one, equal to point seven four.

6 Example 2: a probability and its error

Illustration: 200 simulated losses, 8 beyond the threshold

Fraction beyond threshold	$8 \div 200$	**4.0%**
Standard error	$\sqrt{(0.04 \times 0.96 \div 200)}$	**1.39%**
Rough range, 2 errors	$4.0\% \pm 2 \times 1.39\%$	**1.2% to 6.8%**
Standard error, 800 draws	$\sqrt{(0.04 \times 0.96 \div 800)}$	**0.69%**

IN THE LESSON

- Now a probability.
- Suppose we simulate two hundred portfolio losses and eight of them exceed a threshold we care about.
- The estimated probability is eight over two hundred, which is four per cent.
- How reliable is that?
- For a fraction, the standard deviation of each draw comes from the fraction itself, so the standard error is the square root of point zero four times point nine six, divided by two hundred.
- That is one point three nine per cent.
- A range of two standard errors either side runs from one point two to six point eight per cent.
- Now suppose we ran eight hundred draws and saw the same fraction.
- The standard error becomes point six nine per cent, half the size, for four times the work.

7 Example 3: pricing an option by simulation

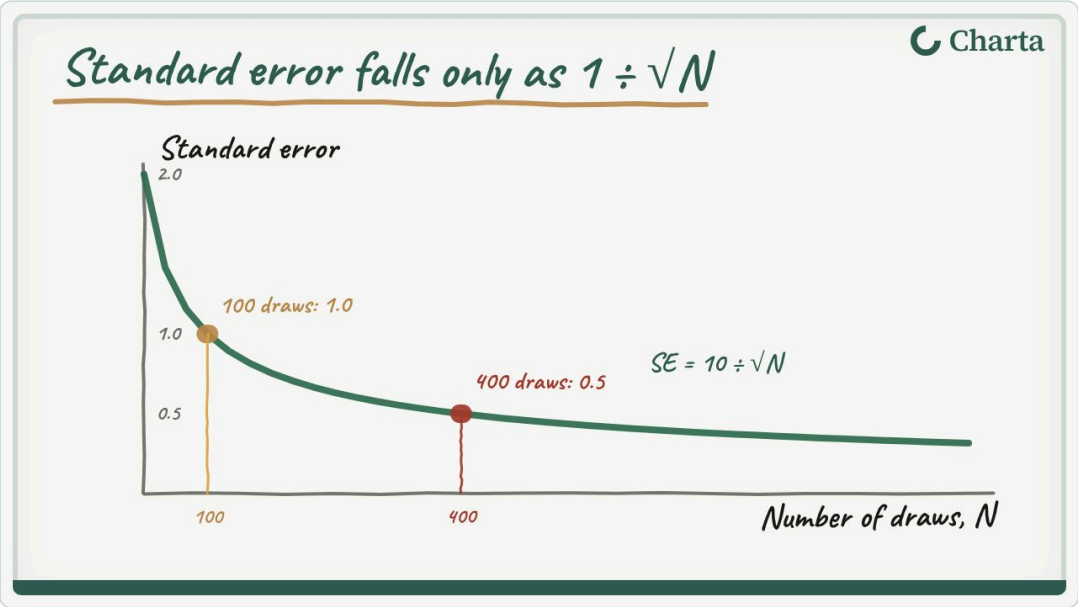
Illustration: S0 100, strike 100, r 5%, vol 20%, 1 year. Draws: −1, −0.5, 0.5, 1

Year-end prices, drift 0.03	100·e^(0.03 + 0.2z)	**84,93,114,126**
Payoffs	0, 0, 13.88, 25.86	**Sum 39.74**
Average payoff	39.74 ÷ 4	**9.94**
Discount factor	e^(−0.05 × 1)	**0.9512**
Price estimate	9.94 × 0.9512	**9.45**

IN THE LESSON

- For derivatives, there is a special rule: simulate the price under the risk neutral drift, then discount the average payoff.
- Take a call option, with a strike of one hundred, on a stock now worth one hundred.
- The interest rate is five per cent, volatility twenty per cent, and one year to expiry.
- Use four made up normal draws, minus one, minus point five, point five and one.
- The year end prices are about eighty four, ninety three, one hundred fourteen and one hundred twenty six.
- Payoffs are zero, zero, thirteen point eight eight and twenty five point eight six.
- Their sum is thirty nine point seven four, so the average payoff is about nine point nine four.
- Discounting at the five per cent rate multiplies by point nine five one two, giving an estimate of nine point four five.

8 Standard error falls only as 1 ÷ √N



IN THE LESSON

- Let's picture the standard error.
- Suppose the spread of the simulated values, s, is ten.
- Along the bottom is the number of draws, and up the side is the standard error.
- With one hundred draws, the error is ten divided by the square root of one hundred, which is one.
- Now go to four hundred draws, four times as many.
- The error is ten over twenty, or point five.
- Four times the work has bought only half the error.
- The curve drops steeply at first, then flattens.
- To get from one down to point one, you would need ten thousand draws, one hundred times the work for ten times the accuracy.
- That is why accuracy in simulation is costly.

9 Ways to reduce Monte Carlo sampling error

- Standard error = s ÷ √N: cut s, or raise N
- More draws: halving the error needs 4× as many
- Antithetic and control variates: use correlation
- Stratified and importance sampling: steer the draws
- Common random numbers; quasi-random (no usual SE)

IN THE LESSON

- Monte Carlo sampling error is simply the standard error, s over the square root of the number of draws.
- So there are two ways to shrink it.
- The direct way is more replications, but halving the error needs four times as many draws.
- The cleverer way is to cut the variance of each draw, the quantity s itself.
- The main techniques are antithetic variates, control variates, stratified sampling and importance sampling.
- Stratified sampling forces draws to cover every part of the range; importance sampling steers them toward the rare region that matters.
- Common random numbers help when you compare two similar valuations, because shared draws cancel the noise in the difference.
- Low discrepancy sequences cover the space evenly, but they are not independent, so the usual error formula does not apply.

10 Example 4: does the technique pay?

Efficiency = variance per draw × time per draw. Lower is better

Plain method	4 × 1	**4**
Technique A	2 × 2	**4**
Technique B	1 × 1.5	**1.5**
Verdict	B pays, A does not	**B wins**

IN THE LESSON

- Variance reduction is not free, so always check the cost.
- A fair test multiplies the variance of one draw by the computing time of one draw.
- Lower is better, because for a fixed time budget the final variance is proportional to this product.
- Suppose the plain method has a variance of four and takes one unit of time per draw.
- The product is four.
- Technique A halves the variance to two, but each draw takes twice as long.
- The product is two times two, equal to four.
- No gain at all, despite the impressive sounding variance cut.
- Technique B brings the variance down to one and takes one and a half units of time.
- The product is one point five, so it does pay, by a factor of nearly three.

11 Antithetic variates: use the mirror image

- Pair each draw with its mirror image
- Average the two payoffs of the pair
- Monotonic payoff: pair is negatively correlated
- Even payoff, such as z^2: correlation +1, no gain

IN THE LESSON

- The first technique is antithetic variates.
- Each time we draw a random input, we also use its mirror image, the same draw with the sign flipped.
- We compute the payoff for both and average the pair.
- Why does this help?
- If the payoff is monotonic, meaning it moves in one direction as the input rises, then a high draw gives a high value and its mirror gives a low one.
- The two values are negatively correlated, so when we average them, the noise cancels.
- The case where it fails is an even payoff, such as the square of the draw.
- The draw and its mirror image give the same value, the correlation is plus one, and we gain nothing.
- We have done the work twice for one piece of information.

12 Variance of an antithetic pair

Var(pair average) = $\sigma^2 \times (1 + \rho) \div 2$
Two independent draws: $\sigma^2 \div 2$

Variance of one value	σ^2: variance of a single value
Correlation	ρ: correlation within the pair
Variance of the average	Variance of the pair average

IN THE LESSON

- Here is the formula.
- The variance of the pair average equals the variance of one value, sigma squared, times one plus rho, all divided by two.
- Rho is the correlation between the two values in the pair.
- Compare it with independent draws.
- Two independent values averaged would have a variance of sigma squared over two.
- So the pair does better than independent draws whenever rho is negative, because one plus rho is then below one.
- If rho were minus one, the variance would vanish entirely.
- If rho is zero, the pair is no better than independent draws.
- If rho is plus one, the variance is the full sigma squared.
- A common mistake is to confuse the variance of the pair average with the variance of a single value.

13 How the correlation changes the variance

Pair-average variance when $\sigma^2 = 4$: $4 \times (1 + \rho) \div 2$

Correlation within the pair	Variance of the average
$\rho = -1$	0
$\rho = -0.6$	0.8
$\rho = 0$ (independent)	2.0
$\rho = +1$ (even payoff)	4.0, no gain

IN THE LESSON

- Let's see the numbers with sigma squared equal to four.
- If the pair were perfectly negatively correlated, the variance of the average would be four times zero over two, which is zero.
- For a correlation of minus point six, it is four times point four over two, equal to point eight.
- For a correlation of zero, which is the case of independent draws, the variance is four times one over two, equal to two.
- And for a correlation of plus one, as with an even payoff, it is four.
- Point eight is well below the independent benchmark of two, a cut of sixty per cent.
- The more negative the correlation, the bigger the gain.

14 Control variates: borrow a known answer

- Control: related quantity with a known expected value
- Example: analytic price of a similar option
- Simulate target and control on the same draws
- Subtract the control's sampling error from the estimate
- Works through correlation: stronger link, bigger gain

IN THE LESSON

- The second technique is the control variate.
- A control is a related quantity whose true expected value we already know.
- A typical example is the analytic price of a similar option, such as a simpler option that does have a formula.
- We simulate both the target and the control using the same random draws.
- Then we look at how far the simulated control landed from its known mean.
- That gap is pure sampling noise.
- Since the target moves with the control, we assume it carries some of the same noise, and we subtract a suitable multiple of the control's error from the target's estimate.
- The more closely the two are correlated, whether positively or negatively, the more noise we can strip out.

15 The control variate adjustment

Adjusted = Y − b × (C − E[C])
b = Cov(Y, C) ÷ Var(C)
New variance = old variance × (1 − ρ²)

Raw estimate	Y: the raw simulated estimate
Control variate	C: control with known mean E[C]
Slope	Best slope: Cov(Y,C) ÷ Var(C)
Correlation	ρ: correlation of Y with C
Variance reduction	Times 1 − ρ²; the cut is ρ²

IN THE LESSON

- The adjusted estimate equals the raw estimate, minus a slope times the bracket of the simulated control minus its known mean.
- The known mean is what the control should have averaged, so the bracket is the control's error.
- The best slope is the covariance of the estimate and the control, divided by the variance of the control.
- With that slope, the variance falls by the factor one minus rho squared, where rho is the correlation of the estimate with the control.
- Two traps.
- First, you subtract the control's error, not the control itself.
- Second, rho squared is the share of variance removed; the variance factor is one minus rho squared, so do not confuse the reduction with rho itself.

16 Example 5: a control variate at work

Illustration: Var(Y) 9, Var(C) 4, Cov 5.4. Y = 12, C = 5.5, E[C] = 5.0

Correlation	5.4 ÷ (3 × 2)	**0.9**
Best slope b	5.4 ÷ 4	**1.35**
Adjusted estimate	12 − 1.35 × (5.5 − 5.0)	**11.325**
Variance factor	$1 - 0.9^2$	**0.19**
New variance	9 × 0.19	**1.71**

IN THE LESSON

- Suppose our raw estimate has a variance of nine per draw, and the control has a variance of four, so standard deviations of three and two.
- Their covariance is five point four.
- The correlation is five point four divided by three times two, which is point nine.
- The best slope is the covariance over the control's variance: five point four over four, equal to one point three five.
- Say the raw estimate came out at twelve, while the control averaged five point five against its known mean of five.
- The control overshot by point five, so we subtract one point three five times point five.
- The adjusted estimate is eleven point three two five.
- The variance factor is one minus point nine squared, point one nine, so the variance drops from nine to one point seven one.

17 Bootstrapping: reuse the data you have

- Resample the observed data with replacement
- Draw as many as the data has; compute the statistic
- Repeat thousands of times; spread = standard error
- No assumed distribution: fat tails and skew carry over
- Several assets: resample whole dates, keep dependence

IN THE LESSON

- Monte Carlo needs a model of how the data are generated.
- Bootstrapping does not.
- It resamples the observed data with replacement.
- If you have a given number of observations, you draw that many from those same data points, so some appear more than once and others not at all.
- You compute your statistic on this new sample, then repeat the whole thing thousands of times.
- The spread of the results is the standard error, or you can read a confidence interval from their percentiles.
- The great advantage is that no distribution is assumed, so fat tails, skewness and other real features carry over automatically.
- With several assets, resample whole dates, so that each draw keeps that day's returns together and the dependence across assets is preserved.

18 Example 6: one bootstrap resample

Illustration: five returns (%): 4, −2, 6, 0, 2

Original mean	(4 − 2 + 6 + 0 + 2) ÷ 5	**2.0**
Resample of five	6, 6, 4, 2, −2	**Sum 16**
Resample mean	16 ÷ 5	**3.2**
Repeat many times	one new mean each time	**Many means**

IN THE LESSON

- Here is a tiny case with five yearly returns, in per cent: four, minus two, six, zero and two.
- Their sum is ten and the mean is two.
- To bootstrap, we draw five positions at random, with replacement.
- Say we pick the third observation twice, then the first, then the fifth, then the second.
- That gives six, six, four, two and minus two.
- Notice that the six appears twice and the zero never appears.
- This is allowed, and it is exactly what makes each resample different.
- The sum is sixteen and the resample mean is three point two.
- That is one replicate.
- Repeating this again and again shows how much our estimate would wobble.

19 Example 7: bootstrap standard error

Illustration: six bootstrap means: 1.0, 1.4, 0.8, 1.2, 0.6, 1.0

Average of the means	6.0 ÷ 6	**1.0**
Sum of squared deviations	0+.16+.04+.04+.16+0	**0.40**
Variance of the means	0.40 ÷ 5	**0.08**
Bootstrap standard error	√0.08	**0.283**

IN THE LESSON

- Now suppose we ran six resamples and recorded these means: one, one point four, point eight, one point two, point six and one.
- A real study uses thousands.
- The average of the means is six divided by six, which is one.
- The deviations from that average are zero, point four, minus point two, point two, minus point four and zero.
- Their squares add to point four zero.
- Dividing by five, the number of replicates less one, gives a variance of point zero eight.
- The bootstrap standard error is the square root, point two eight three.
- Notice what we did not do: we did not divide by the square root of the number of replications.
- The spread of the replicates is already the standard error.

20 Monte Carlo versus bootstrapping

Both give estimates with sampling error; the source of draws differs

	Monte Carlo | Bootstrap
Where draws come from	Assumed model | Actual data
Fat tails and skew	Only if modeled | Carry over
Several assets	Modeled | Whole dates kept
Standard error	s ÷ √N | sd of replicates
Key assumption	Model right | iid data

IN THE LESSON

- Let's compare the two.
- In Monte Carlo the draws come from an assumed model, so you must specify the distribution and its parameters.
- In bootstrapping the draws come from the actual data, so nothing is assumed.
- That is how the bootstrap improves on Monte Carlo: fat tails and skewness carry over automatically, where a simple normal model would miss them.
- With several assets, Monte Carlo needs you to model the correlations, while the bootstrap keeps them by resampling whole dates.
- The standard error also differs.
- In Monte Carlo the standard error of a mean is s over the square root of the draws.
- In the bootstrap, it is the standard deviation of the replicate estimates.
- Both rest on assumptions: Monte Carlo that the model is right, the bootstrap that the data are independent and representative.

21 Example 8: what a resample leaves out

Illustration: a sample of 100 observations, resampled with replacement

Miss it on one draw	1 − 1 ÷ 100	**0.99**
Miss it on all 100 draws	0.99^100	**36.6%**
Chosen at least once	1 − 36.6%	**63.4%**
Limit for a huge sample	1 ÷ e	**36.8%**

IN THE LESSON

- A well known feature of the bootstrap is that each resample leaves some observations out.
- Suppose the sample has one hundred observations.
- On any one draw, the chance of not picking a particular observation is one minus one over one hundred, which is point nine nine.
- A resample makes one hundred draws, so the chance of missing it every time is point nine nine to the power of one hundred.
- That is about thirty six point six per cent.
- So the chance it appears at least once is sixty three point four per cent.
- For a very large sample, the miss probability approaches one over Euler's number, about thirty six point eight per cent.
- That is the source of the rule of thumb that about thirty seven per cent of observations are left out of any resample.

22 How pseudo-random numbers are made

- Deterministic algorithm whose output looks random
- Passes tests for uniformity and independence
- Same seed gives the same stream: reproducible
- The stream repeats after its period: want a long one
- Parallel runs: a distinct seed each; never reuse draws

IN THE LESSON

- Computers cannot flip coins, so they use a deterministic algorithm to produce numbers that look random.
- We call them pseudo random.
- A good sequence passes statistical tests for uniformity and independence, yet it is fully reproducible.
- Start from a seed and the same stream comes out every time.
- That is useful for checking and debugging results.
- But it also has a consequence: the stream eventually repeats once its period is used up, so a long period matters.
- In parallel computing, each process should have its own distinct seed.
- A frequent error is to assume that identical seeds give independent runs, when they give identical numbers.
- Another is to reuse draws.
- Other distributions are produced by transforming the uniform numbers.

23 The linear congruential generator

Next = (a × x + c) mod m
u = x ÷ m z = N^{-1}(u)

Seed	Seed: the first state
Multiplier and increment	Multiplier a and increment c
Modulus	Modulus m: the state wraps at m
Uniform number	Uniform number, 0 to 1
Normal draw	Normal draw by inverse transform

IN THE LESSON

- A common example is the linear congruential generator.
- Start with a seed, which is the first state.
- Each new state is the multiplier times the current state, plus an increment, and then you take the remainder after dividing by the modulus.
- The multiplier, the increment and the modulus are fixed constants.
- To turn a state into a uniform number between zero and one, divide it by the modulus.
- To get a normal draw, use the inverse transform: apply the inverse of the normal cumulative distribution function to the uniform number.
- Box Muller is another route, which turns pairs of uniform numbers into normal draws.
- In every case the quality of the normals depends on the quality of the uniforms underneath.

24 Example 9: a toy generator

a = 5, c = 3, m = 16, seed 7: restart, same five states

Draw 1	(5 × 7 + 3) mod 16	**6, u 0.375**
Draw 2	(5 × 6 + 3) mod 16	**1, u 0.0625**
Draw 3	(5 × 1 + 3) mod 16	**8, u 0.5**
Draw 4	(5 × 8 + 3) mod 16	**11, u 0.6875**
Draw 5	(5 × 11 + 3) mod 16	**10, u 0.625**

IN THE LESSON

- Let's run a toy generator with a multiplier of five, an increment of three and a modulus of sixteen, starting from the seed seven.
- Real generators use enormous moduli.
- First state: five times seven plus three is thirty eight, and thirty eight leaves a remainder of six when divided by sixteen.
- The uniform number is six over sixteen, point three seven five.
- Second: five times six plus three is thirty three, remainder one, uniform point zero six two five.
- Third: eight, uniform point five.
- Fourth: five times eight plus three is forty three, remainder eleven, uniform point six eight seven five.
- Fifth: ten, uniform point six two five.
- Run it again from seven and you get exactly the same numbers.

25 Example 10: from uniform to normal

Standard normal: $z = N^{-1}(u)$, where u is a cumulative probability

u = 0.5	median of the normal	**z = 0**
u = 0.975	97.5% lies below	**z = 1.96**
u = 0.025	2.5% lies below	**z = −1.96**
Restart with the same seed	same u, same z	**Identical**

IN THE LESSON

- Now turn uniform numbers into normal draws using the inverse transform.
- The idea is that the uniform number is a cumulative probability, and we ask which normal value has that probability below it.
- A uniform number of point five is the median, and the standard normal value is zero.
- A uniform number of point nine seven five gives one point nine six, because ninety seven and a half per cent of the standard normal distribution lies below one point nine six.
- A uniform number of point zero two five gives minus one point nine six, by symmetry.
- And if we restart with the same seed, we get the same uniforms and the same normals.

26 When bootstrapping does not work well

- Serial dependence: use a block bootstrap instead
- Structural breaks or regimes: resamples mix them
- Tails: no resample beats the worst loss seen
- Very small samples: may not represent the population
- A bootstrap can only reuse what is in the sample

IN THE LESSON

- The standard bootstrap resamples single observations independently, so it works well only when the data are independent, identically distributed and representative.
- It struggles in four situations.
- First, serially dependent data, such as autocorrelation or volatility clustering.
- Resampling individual points destroys the dependence, and the usual fix is a block bootstrap, which resamples runs of consecutive observations.
- Second, structural breaks or regime changes, which mix different distributions together.
- Third, tail and extreme value questions.
- A resample cannot contain a loss worse than the worst one observed, and a single crash day will be missing from about thirty seven per cent of resamples of one hundred observations.
- Fourth, very small samples, which may not be representative of the population.
- A bootstrap can only reuse what is in the sample.

27 Example 11: a block bootstrap

Illustration: 10 daily returns in time order, blocks of 3 consecutive days

Observations in the series		**10**
Block length		**3**
Overlapping blocks available	10 − 3 + 1	**8**
Blocks drawn per resample	4 × 3 = 12, trim to 10	**4**

IN THE LESSON

- Here is how the block bootstrap repairs the damage to dependence.
- Suppose we have ten daily returns in time order, and we choose a block length of three.
- Instead of picking single days, we pick blocks of three consecutive days, so that a run of turbulent days stays together.
- How many overlapping blocks can we cut from the series?
- Ten minus three plus one, which is eight.
- Each resample needs ten days, and each block supplies three, so we need ten divided by three, rounded up, which is four blocks.
- Four blocks give twelve days, and we trim the last two to leave ten.
- The block length is a choice: too short loses the dependence, too long leaves few distinct blocks.

28 Drawbacks of simulation in finance

- Costly: halving the error needs four times the runs
- Estimates, not exact: they differ across seeds and runs
- Specific to the model and parameters assumed
- Bad model: precise answer to the wrong question
- Hard to debug and explain; report a standard error

IN THE LESSON

- Simulation is powerful, but it has real drawbacks.
- It is computationally expensive, because the standard error falls only with the square root of the number of replications.
- Halving it needs four times the runs, and nested or high dimensional problems multiply the cost.
- Results are estimates, not exact values, and they differ across seeds and replications, so a standard error must be reported and differences between runs compared with it.
- Results are specific to the model and the parameters assumed.
- A poor data generating process, such as normal returns when the data are fat tailed, gives a precise answer to the wrong question, and more replications cannot repair it.
- Simulation can also be hard to debug and to explain.

29 Sampling error versus model error

Two kinds of error: more draws reduce only the first

	Sampling | Model
Cause	Few draws | Wrong process
Shrinks with more draws?	Yes | No
Size falls like	1 ÷ √N | does not fall
The cure	Draws | Better model
Example	Few paths | Normal vs fat

IN THE LESSON

- The most important distinction here is between two kinds of error.
- Sampling error comes from using a finite number of draws.
- It shrinks as one over the square root of the number of draws, so more draws help, and so do the variance reduction tricks.
- Model error comes from a wrong data generating process, for example assuming normal returns when the real ones have fat tails.
- It does not shrink as you add draws.
- So when someone says the results are not accurate enough, ask which error is the problem.
- More draws fix the first one only.
- A better model is the only cure for the second.

30 Example 12: the cost of precision

Illustration: spread of the draws s = 10

Draws for error 0.05	(10 ÷ 0.05)²	**40,000**
Draws for error 0.025	(10 ÷ 0.025)²	**160,000**
Halving the error costs	160,000 ÷ 40,000	**4 times**
Nested 1,000 × 1,000	outer × inner paths	**1,000,000**

IN THE LESSON

- Let's price the cost.
- Suppose the spread of the simulated values is ten, and we want a standard error of point zero five.
- The number of draws needed comes from rearranging the formula: ten divided by point zero five, squared.
- That is two hundred squared, or forty thousand draws.
- Now halve the target to point zero two five.
- Ten divided by point zero two five is four hundred, squared is one hundred sixty thousand draws, four times as many.
- Finally, nested simulation, such as simulating many future scenarios and then valuing a derivative inside each one.
- With one thousand outer scenarios and one thousand inner paths, we need one million runs.

31 Four mistakes to avoid

- Second moment: average squares, not square the mean
- No error bar: report s ÷ √N; draws can't fix a model
- Bootstrap: blocks for dependence; breaks, tails no
- Same seed = same numbers: use a distinct seed per run

IN THE LESSON

- Mistake one: squaring the mean when you want the second moment.
- You must average the squared draws, not square the average.
- Mistake two: reporting a simulated number with no standard error, or thinking more draws is always the answer.
- Halving the error needs four times the draws, and no number of draws repairs a poor model.
- Mistake three: using the bootstrap where it fails.
- Use blocks for dependent data, and a different method for structural breaks and tail questions.
- Mistake four: believing that the same seed gives independent runs.
- It gives the same numbers.
- Use a distinct seed for each parallel process.

32 Recap: the main ideas

- Monte Carlo: specify, draw, compute, average
- Antithetic: $\sigma^2(1 + \rho) \div 2$. Control: $Y - b(C - E[C])$
- Bootstrap: resample data, no assumed model
- Fails for dependence, breaks, tails, small samples
- Pseudo-random: seed, period; model error stays

IN THE LESSON

- Let's pull it together.
- Monte Carlo simulation specifies a process, generates draws, computes a quantity for each, and averages.
- The result is an estimate whose standard error is s over the square root of the draws, so halving it costs four times the work.
- We can reduce error with antithetic variates, which use negative correlation within a pair, and control variates, which subtract the error of a quantity with a known mean.
- Stratified sampling, importance sampling and common random numbers help too, but check the cost.
- Bootstrapping resamples real data with replacement and needs no assumed model, but fails with dependence, breaks, tails and tiny samples.
- Pseudo random numbers are deterministic and seed driven.
- And simulation never fixes a bad model.

33 Try three quick questions

1. s = 12 and N = 144. Standard error? Draws to halve it?
   Answer: 12 ÷ 12 = 1. Four times the draws: 576.
2. $\sigma^2 = 10$, $\rho = -0.5$. Variance of the antithetic pair average?
   Answer: 10 × (1 − 0.5) ÷ 2 = 2.5, versus 5 for independent draws.
3. Why can a bootstrap not give a loss beyond the worst observed?
   Answer: It only resamples the data it has, so tails are missing.

IN THE LESSON

- Now try three quick questions yourself.
- Pause after each one.
- Question one.
- A simulation has a sample standard deviation of twelve and uses one hundred forty four draws.
- What is the standard error, and how many draws would halve it?
- Twelve over the square root of one hundred forty four is one.
- Halving needs four times the draws: five hundred seventy six.
- Question two.
- The variance of a single value is ten and the pair correlation is minus point five.
- What is the variance of the antithetic pair average?
- Ten times point five over two is two point five, against five for independent draws.
- Question three.
- Why can a bootstrap not estimate a loss worse than the worst one observed?
- Because it only resamples the data it has.