









Statistics of Financial Markets

Seminar 4: Probability



Daniel Traian PELE

Bucharest University of Economic Studies
IDA Institute Digital Assets
Blockchain Research Center
AI4EFin Artificial Intelligence for Energy Finance
Romanian Academy, Institute for Economic Forecasting
MSCA Digital Finance

Acronyms used in this material

- **AI**: Artificial Intelligence
- **AR**: AutoRegressive (model)
- **BET**: Bucharest Exchange Trading (index)
- **CDF**: Cumulative Distribution Function
- **EODHD**: EOD Historical Data (market-data provider)
- **GARCH**: Generalised AutoRegressive Conditional Heteroskedasticity
- **GBM**: Geometric Brownian Motion
- **LCG**: Linear Congruential Generator
- **PCG64**: Permuted Congruential Generator, 64-bit (NumPy default)
- **PMF**: Probability Mass Function
- **RANDU**: RANDU (IBM random number generator of the 1960s)

Today's Question and Seminar Structure

- Question of the seminar
  - **how do we describe, condition and simulate random returns and prices?**
- Exercises: in the order of the Exercise Map (next slide)
- Outputs to produce
  - the answers to the A exercises; the table or figure each B exercise asks for
  - a short argued answer to C1; the errors found in the AI answer of C2 and the corrected answer
  - the interpretation sentences each exercise asks for
- Nothing is handed in: the seminar is for practice; the solutions of the [Proposed] exercises are discussed in class
- Notebook for today: open the seminar notebook in Google Colab; each task names its section
- Contact: Prof. Daniel Traian Pele, danpele@ase.ro

Exercise Map

Task	Question	Type	Model
A1, A2	moments of a discrete return; is an uncorrelated pair independent?	Solved, Proposed	A1
A3, A4	portfolio volatility; a two-regime model and the law of total variance	Solved, Proposed	A3
A5, A6	a binomial tree; random walk, martingale and AR(1)	Solved, Proposed	A5
A7, A8	inverse transform by hand; GBM and the Wiener process	Solved, Proposed	A7
B1, B2	are returns independent? conditional mean and variance	Solved, Proposed	B1
B3, B4	a loss probability by Monte Carlo; testing a random number generator	Solved, Proposed	B3
B5, B6	GBM against the BET; does variance grow linearly with the horizon?	Solved, Proposed	B5
C1, C2	how often is there a bad year? what is wrong in an AI answer?	Proposed	B3, B5

- **[Solved]**: full solution in the slides and in the notebook, a model to follow; **[Proposed]**: you solve it, following the model
- **A**: calculations on paper; **B**: real data, in the notebook; **C**: open questions and the critique of an AI answer

Data Used

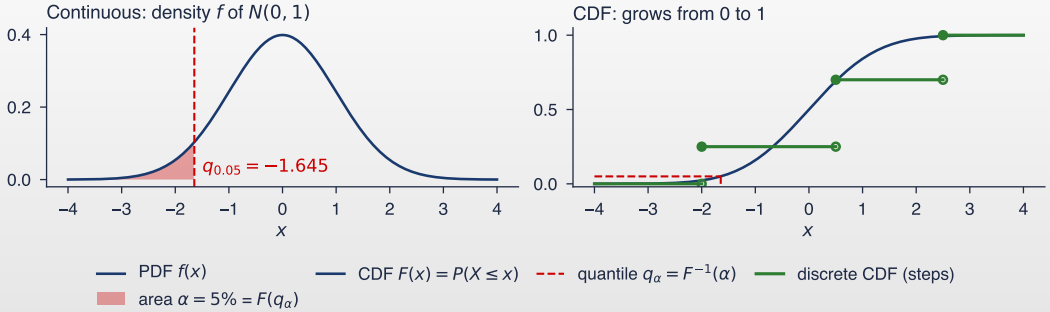
Series	Source	Price used	Period
S&P 500, DAX, BET	EODHD	close	2000–2026
Bitcoin	EODHD	close, 7 days a week	2014–2026

- Daily log returns in %: $r_t = 100\,(\ln P_t - \ln P_{t-1})$
- Weekends and repeated holiday closes are dropped (except Bitcoin); each series on its own calendar
- Simulations use NumPy's default generator with fixed seeds, so every number can be reproduced

Prerequisites for Today (1/21): Distributions

- A **random variable** X: a number determined by chance (tomorrow's return)
  - discrete X: **PMF** (probability mass function) $p(x_i) = P(X = x_i)$, the probability of each value x_i
  - continuous X: **PDF** (probability density function) f; the probability of an interval is the area under f
- **CDF** (cumulative distribution function): $F(x) = P(X \leq x)$
  - the probability that X does not exceed x; it grows from 0 to 1
- **Quantile** of order α: $q_\alpha = F^{-1}(\alpha) = \inf\{x : F(x) \geq \alpha\}$
  - the smallest x with $F(x) \geq \alpha$; the median is $q_{0.5}$

Prerequisites for Today (2/21): PDF, CDF and Quantile on a Chart



- Left: the area under the PDF between a and b is the probability of that interval

$$P(a < X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)$$

- Right: a continuous CDF is smooth; a discrete one is a staircase, with a jump $p(x_i)$ at each x_i
  - quantile q_α: from α across to the CDF, then down; for $N(0,1)$, $q_{0.05} = -1.645$

SFM_ch4_seminar

Prerequisites for Today (3/21): Mean and Variance

- **Expectation**: $E[X] = \sum_i x_i p(x_i)$ (discrete) or $\int x f(x)\,dx$ (continuous)
  - the probability-weighted average of the values
- **Variance**: $\mathrm{Var}(X) = E[X^2] - (E[X])^2$
  - the mean squared distance from the mean; standard deviation $\sigma = \sqrt{\mathrm{Var}(X)}$, the volatility of a return
- **Kurtosis**: $E[(X-\mu)^4]/\sigma^4$, with $\mu = E[X]$
  - it equals 3 for any Normal distribution; above 3: heavier tails (excess kurtosis = kurtosis -3)
  - for $N(0, \sigma^2)$: $E[X^4] = 3\sigma^4$

Prerequisites for Today (4/21): Sample Moments

- From observed returns $r_1, \ldots, r_n$ we estimate the moments

$$\bar{r} = \frac{1}{n}\sum_{t=1}^{n} r_t, \qquad \hat{\sigma}^2 = \frac{1}{n-1}\sum_{t=1}^{n}(r_t - \bar{r})^2$$

  - $\bar{r}$: the sample mean, an estimate of μ; $\hat{\sigma}$: the sample standard deviation; the hat marks an estimate
- **Sample kurtosis** and excess kurtosis

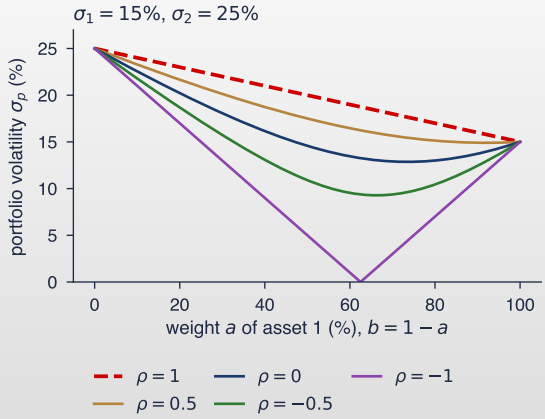
$$\hat{K} = \frac{\frac{1}{n}\sum_t (r_t - \bar{r})^4}{\left(\frac{1}{n}\sum_t (r_t - \bar{r})^2\right)^2}, \qquad \text{excess kurtosis} = \hat{K} - 3$$

  - close to 0 for Normal data; a large positive excess: more extreme returns than a Normal law gives
- $\mathrm{Corr}(r_t^2, r_{t-1}^2)$: the sample correlation of the pairs (r_t^2, r_{t-1}^2), $t = 2, \ldots, n$
  - near 0 for i.i.d. (independent, identically distributed) returns; positive when large moves follow large moves (volatility clustering)

Prerequisites for Today (5/21): Covariance and Portfolios

- **Covariance**: $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$
  - positive when X and Y tend to move in the same direction
- **Correlation**: $\rho = \mathrm{Corr}(X, Y) = \mathrm{Cov}(X, Y)/(\sigma_X \sigma_Y) \in [-1, 1]$
  - σ_X, σ_Y: the standard deviations; ρ measures linear dependence only
- **Variance of a sum**: $\mathrm{Var}(aX + bY) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\,\mathrm{Cov}(X, Y)$
  - a, b: the portfolio weights; the covariance term makes diversification work when $\rho < 1$

Prerequisites for Today (6/21): Diversification on a Chart



- Portfolio with weights a and $b = 1 - a$

$$\sigma_p = \sqrt{a^2\sigma_1^2 + b^2\sigma_2^2 + 2ab\rho\sigma_1\sigma_2}$$

  - σ_1, σ_2: the volatilities of the two assets; ρ: their correlation
- $\rho = 1$ (dashed): $\sigma_p = a\sigma_1 + b\sigma_2$, the weighted average
  - every $\rho < 1$ gives a curve below this line: the lower ρ, the larger the gain
- Minimum-variance weight

$$a^* = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}$$

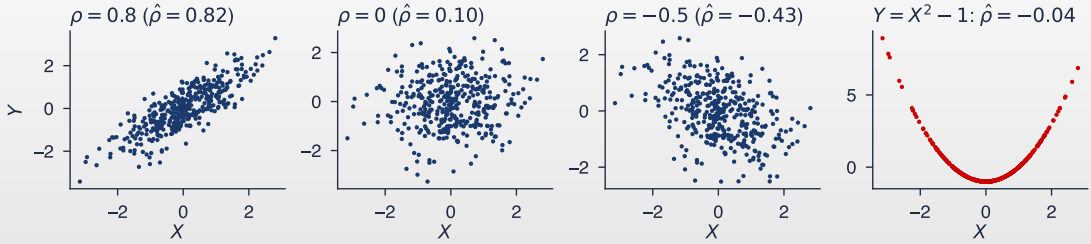
  - the lowest point of each curve; $\rho = -1$: $\sigma_p = 0$ at $a^* = \sigma_2/(\sigma_1 + \sigma_2)$

SFM_ch4_seminar

Prerequisites for Today (7/21): Independence

- **Independent**: $P(X = x, Y = y) = P(X = x)P(Y = y)$ for every cell of the joint table
  - knowing X does not change the distribution of Y
  - independent $\Rightarrow$ uncorrelated, but not conversely
- **Sample correlation** of n pairs (x_i, y_i): $\hat{\rho} = \dfrac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$
  - $\bar{x}$, $\bar{y}$: the sample means; $\hat{\rho}$ estimates ρ and serves as a test of zero correlation
  - if the data are i.i.d. (independent and identically distributed), it lies in the band $\pm 1.96/\sqrt{n}$ with probability 95%
  - a value outside the band: evidence against independence

Prerequisites for Today (8/21): Correlation on a Chart



- Four simulated samples of 400 pairs (X, Y), $X \sim N(0,1)$
  - first three: $Y = \rho X + \sqrt{1-\rho^2}\, Z$, $Z \sim N(0,1)$ independent of X, so $\mathrm{Corr}(X, Y) = \rho$
  - $\hat{\rho}$ differs slightly from ρ: sampling error
- Right: $Y = X^2 - 1$ is fully determined by X, yet $\hat{\rho} \approx 0$
  - $\mathrm{Cov}(X, X^2) = E[X^3] = 0$ for a symmetric X with mean 0: correlation sees only straight-line dependence

SFM_ch4_seminar

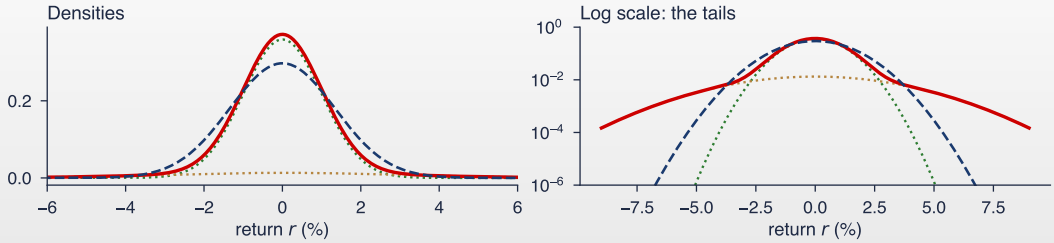
Prerequisites for Today (9/21): Conditioning

- $P(A \mid B) = P(A \cap B)/P(B)$
  - the probability of A once we know that B occurred; $A \cap B$: both occur
- $E[Y \mid X]$: the mean of Y once X is known
  - the best forecast of Y from X
  - **tower property**: $E[E[Y \mid X]] = E[Y]$, the average of the forecasts is the overall mean
- **Law of total variance**: $\mathrm{Var}(Y) = E[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(E[Y \mid X])$
  - the average variance inside the groups defined by X, plus the variance of the group means

Prerequisites for Today (10/21): Conditioning for Returns

- $\mathcal{F}_{t-1}$: the information available up to yesterday (past prices, news)
  - $\mu_t = E[r_t \mid \mathcal{F}_{t-1}]$: the conditional mean of today's return
  - $\sigma_t^2 = \text{Var}(r_t \mid \mathcal{F}_{t-1})$: the conditional variance
  - a σ_t that changes in time is what GARCH models (Chapter 9)
- A changing σ_t makes returns a mixture of Normal distributions, with heavy tails (next slide)

Prerequisites for Today (11/21): a Mixture of Two Normal Regimes



- Calm day with probability π: $N(0, \sigma_1^2)$; turbulent with probability $1-\pi$: $N(0, \sigma_2^2)$

$$f(r) = \pi\,\varphi_{\sigma_1}(r) + (1-\pi)\,\varphi_{\sigma_2}(r), \qquad \mathrm{Var}(r) = \pi\sigma_1^2 + (1-\pi)\sigma_2^2$$

  - φ_σ: the density of $N(0, \sigma^2)$; here $\pi = 0.9$, $\sigma_1 = 1$, $\sigma_2 = 3$: kurtosis 8.3, although each regime is Normal
- Log scale (right): far from 0 the mixture density is many times the Normal one with the same variance: heavy tails

SFM_ch4_seminar

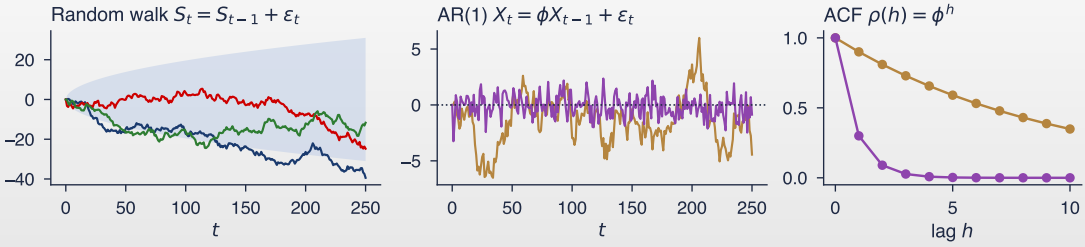
Prerequisites for Today (12/21): Processes

- **White noise** ε_t: mean 0, variance σ^2, uncorrelated over time
  - **i.i.d.** noise: in addition independent and identically distributed
- **Random walk**: $S_t = S_{t-1} + \varepsilon_t$
  - $E[S_t] = S_0$ (the starting value), $\mathrm{Var}(S_t) = t\sigma^2$: the variance grows with time
- **Martingale**: $E[X_{t+1} \mid \mathcal{F}_t] = X_t$
  - a fair game: the best forecast of tomorrow is today's value
  - $\mathcal{F}_t$: the information available at time t
- **Weakly stationary** process: $E[X_t]$ and $\mathrm{Cov}(X_t, X_{t+h})$ do not depend on t
  - h: the lag between two dates; a random walk is not stationary, its increments are

Prerequisites for Today (13/21): AR(1) and the Binomial Tree

- **AR(1)**: $X_t = c + \phi X_{t-1} + \varepsilon_t$, stationary if $|\phi| < 1$
  - c: a constant; ϕ: the persistence, the share of yesterday's value carried into today
  - mean $c/(1-\phi)$, variance $\sigma^2/(1-\phi^2)$; autocorrelation at lag h: $\rho(h) = \phi^h$
- **Binomial tree**: each step the price is multiplied by u (probability p) or by d (probability $1-p$)
  - after n steps $S_n = S_0 u^K d^{n-K}$, with $K \sim B(n,p)$ the binomial number of up moves
  - $E[S_n] = S_0(pu + (1-p)d)^n$; fair game for $p^* = (1-d)/(u-d)$, which solves $p^* u + (1-p^*)d = 1$

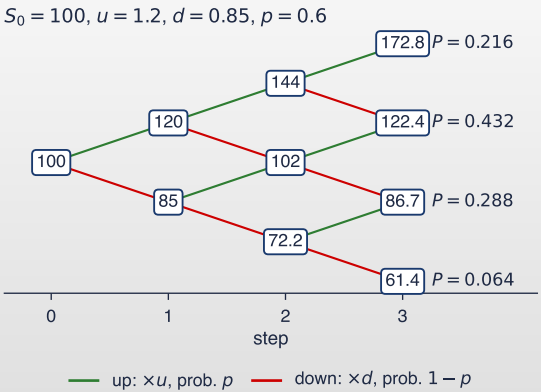
Prerequisites for Today (14/21): Random Walk and AR(1) on a Chart



- Left: three random walks with $\sigma = 1$; the band $\pm 1.96\sqrt{t}\,\sigma$ widens because $\mathrm{Var}(S_t) = t\sigma^2$
- Middle: AR(1) paths ($c = 0$) keep returning to their mean 0; with $\phi = 0.9$ they stay away from it longer
- Right: $\rho(h) = \phi^h$ dies out geometrically; the random walk is the limit case $\phi = 1$, which never returns

SFM_ch4_seminar

Prerequisites for Today (15/21): the Binomial Tree on a Chart



- Number of up moves in n steps: $K \sim B(n, p)$

$$P(K = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n$$

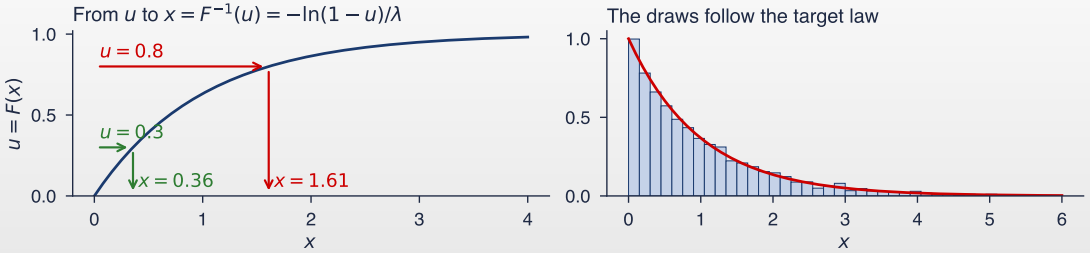
  - $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$: the number of paths with k up moves
- The tree recombines: up then down gives the same price as down then up, $S_0 ud$
- Chart: $n = 3$; the final prices with their probabilities, which add up to 1
  - $E[S_3] = S_0(pu + (1-p)d)^3 = 100 \times 1.06^3 = 119.1$

SFM_ch4_seminar

Prerequisites for Today (16/21): Random Numbers and Simulation

- **Inverse transform**: if $U \sim U(0,1)$ (uniform on $[0,1]$), then $F^{-1}(U)$ has CDF F
  - exponential with rate λ: $-\ln(1-U)/\lambda$; Normal: $\mu + \sigma\Phi^{-1}(U)$, with Φ the $N(0,1)$ CDF
- **LCG** (linear congruential generator): $x_{k+1} = (ax_k + c) \bmod M$, $u_k = x_k/M$
  - a, c, M: integers; $\bmod M$: the remainder of the division by M; period: the number of values before repetition
- **Bootstrap**: draw with replacement from the observed returns
  - the inverse transform applied to the empirical CDF $F_n(x) = \frac{1}{n}\#\{t : r_t \leq x\}$, the share of observations $\leq x$

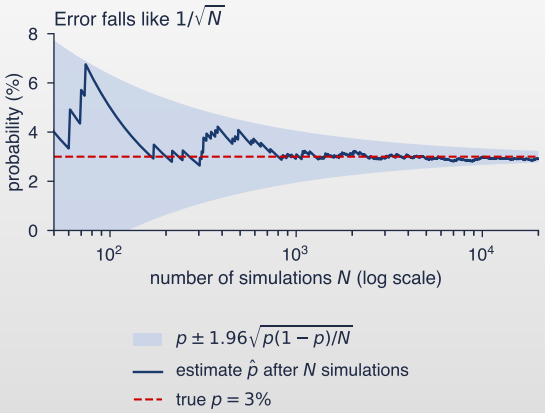
Prerequisites for Today (17/21): the Inverse Transform on a Chart



- Left: draw u on the vertical axis, read $x = F^{-1}(u)$ on the horizontal axis
  - exponential law with rate λ: solving $u = 1 - e^{-\lambda x}$ gives $x = -\ln(1-u)/\lambda$; here $\lambda = 1$
- Right: 5 000 such draws have the target density $\lambda e^{-\lambda x}$; a larger u always gives a larger x

SFM_ch4_seminar

Prerequisites for Today (18/21): Monte Carlo Error and the Binomial Test



- Monte Carlo estimate of a probability p

$$\hat{p} = \frac{1}{N}\sum_{j=1}^{N} 1\{A_j\}, \qquad \mathrm{SE}(\hat{p}) = \sqrt{\frac{p(1-p)}{N}}$$

  - $1\{A_j\} = 1$ if the event occurs in simulation j, 0 otherwise
- Chart: $p = 3\%$; the band $\pm 1.96\,\mathrm{SE}$ narrows like $1/\sqrt{N}$
  - 100 times more simulations: a 10 times smaller error
- **Binomial test**: an event observed x times in m independent periods
  - if the model probability p is right, $X \sim B(m, p)$; two-sided p-value $2\min\{P(X \le x), P(X \ge x)\}$
  - a small p-value (below 5%): the data contradict the model

SFM_ch4_seminar

Prerequisites for Today (19/21): the Wiener Process and GBM

- **Wiener process**: $W(0) = 0$, independent increments, $W(t) - W(s) \sim N(0, t - s)$
  - the variance of an increment equals the elapsed time; $\mathrm{Cov}(W(s), W(t)) = \min(s, t)$
- **GBM** (geometric Brownian motion)

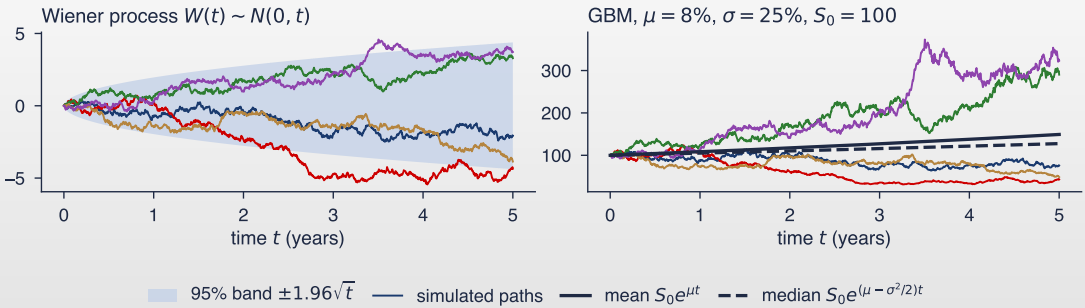
$$S_T = S_0 \exp\left((\mu - \sigma^2/2)T + \sigma W(T)\right)$$

  - μ: annual drift; σ: annual volatility; T: horizon in years; daily log returns i.i.d. Normal
  - mean $S_0 e^{\mu T}$, median $S_0 e^{(\mu - \sigma^2/2)T}$
- The log price is Normal, so probabilities and quantiles of S_T follow from Φ

$$\ln S_T \sim N\big(\ln S_0 + (\mu - \sigma^2/2)T,\ \sigma^2 T\big), \qquad P(S_T < x) = \Phi\Big(\frac{\ln(x/S_0) - (\mu - \sigma^2/2)T}{\sigma\sqrt{T}}\Big)$$

  - quantile of order α: $S_0 \exp\left((\mu - \sigma^2/2)T + z_\alpha \sigma\sqrt{T}\right)$, with $z_\alpha = \Phi^{-1}(\alpha)$

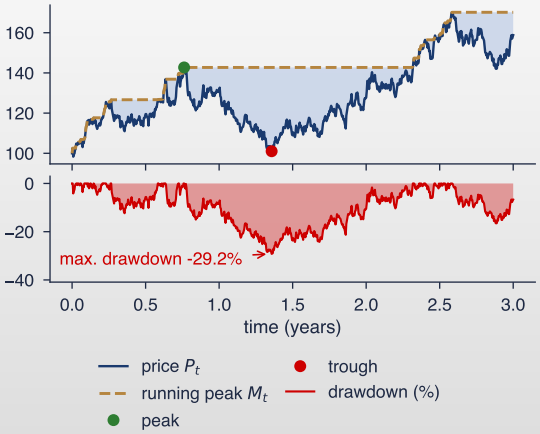
Prerequisites for Today (20/21): Wiener Paths and GBM Paths



- Left: five Wiener paths; $W(t) \sim N(0, t)$, so about 95% of the paths stay within $\pm 1.96\sqrt{t}$
- Right: the same paths turned into GBM prices, $S_0 = 100$, $\mu = 8\%$, $\sigma = 25\%$
  - the mean $S_0 e^{\mu t}$ lies above the median $S_0 e^{(\mu - \sigma^2/2)t}$: a few very high paths pull the mean up

SFM_ch4_seminar

Prerequisites for Today (21/21): Drawdown



- **Running peak**: the highest price up to day t

$$M_t = \max_{s \le t} P_s$$

- **Drawdown**: the relative fall from the running peak

$$DD_t = \frac{P_t}{M_t} - 1 \le 0$$

  - $DD_t = 0$ at every new peak

- **Maximum drawdown**: $\mathrm{MDD} = \min_t DD_t$
  - the largest fall from a peak to a later trough
  - chart: a simulated GBM path, $\mathrm{MDD} = -29.2\%$, from the green to the red point

A1: Moments of a Discrete Return [Solved]

Task

- A stock's daily return X (in %) takes the values -3, 0 and 2 with probabilities 0.2, 0.3 and 0.5.
- 1. Compute $E[X]$, $E[X^2]$, $\mathrm{Var}(X)$ and the standard deviation.
- 2. Write the CDF $F(x)$ for every x.
- 3. Compute $P(X < 0)$ and the median.
- Report: four moments, the CDF and two numbers.

Solution

- 1. $E[X] = -0.6 + 0 + 1 = 0.40$; $E[X^2] = 1.8 + 2 = 3.80$; $\mathrm{Var} = 3.80 - 0.40^2 = 3.64$; s.d. 1.908
- 2. $F(x) = 0$ for $x < -3$; 0.2 on $[-3, 0)$; 0.5 on $[0, 2)$; 1 for $x \geq 2$
- 3. $P(X < 0) = 0.2$; median $= \inf\{x : F(x) \geq 0.5\} = 0$
- The mean is positive although a loss is possible: risk and expected return are different questions.

Back Exercise list

A2: Uncorrelated but Dependent [Proposed]

Solution discussed in class

Task

- X = market move (-1 fall, 0 flat, $+1$ rise); $Y = 1$ if the day is "large". Joint probabilities: $P(-1, 1) = 0.25$, $P(0, 0) = 0.5$, $P(1, 1) = 0.25$, all other cells 0. Model: A1.
- 1. Compute the marginal distributions of X and Y and their means.
- 2. Compute $\mathrm{Cov}(X, Y)$ and ρ.
- 3. Decide whether X and Y are independent, using one cell of the table.
- 4. Compute $\mathrm{Var}(X + Y)$.
- Report: two marginals, ρ, a verdict and a variance.

A3: Volatility of a Two-Asset Portfolio [Solved]

Task

- Two assets with annual volatilities $\sigma_1 = 20\%$ and $\sigma_2 = 30\%$; weights 0.6 and 0.4.
- 1. Compute the portfolio volatility for $\rho = 0.3$.
- 2. Repeat for $\rho = 0$, $\rho = -0.5$ and $\rho = 1$.
- 3. Explain when the portfolio is less risky than the weighted average of the two volatilities.
- Report: four volatilities and one sentence.

Solution

- 1. $\sigma_p = \sqrt{374.4} = 19.35\%$
  - $\sigma_p^2 = 0.36 \times 400 + 0.16 \times 900 + 2 \times 0.24 \times 0.3 \times 600$
- 2. $\rho = 0$: 16.97%; $\rho = -0.5$: 12.00%; $\rho = 1$: 24.00%
- 3. Only $\rho = 1$ gives the weighted average $0.6 \times 20 + 0.4 \times 30 = 24\%$; every $\rho < 1$ diversifies.

Back Exercise list

A4: Two Regimes and the Law of Total Variance [Proposed]

Solution discussed in class

Task

- A day is calm with probability 0.8, $r \mid \text{calm} \sim N(0, 0.8^2)$, and turbulent otherwise, $r \mid \text{turbulent} \sim N(0, 2^2)$ (in %). Model: A3.
- 1. Use the law of total variance to compute $\text{Var}(r)$ and the standard deviation.
- 2. Compute $E[r^4]$ with the fourth moment of $N(0, \sigma^2)$ (Prerequisites for Today (3/21)) in each regime, then the kurtosis $E[r^4]/\text{Var}(r)^2$.
- 3. Explain why the mixture has heavy tails although each regime is Normal.
- Report: a variance, a kurtosis and one sentence.

Back Exercise list

A5: A Three-Step Binomial Tree [Solved]

Task

- $S_0 = 100$, $u = 1.1$, $d = 0.9$, $p = 0.55$, three independent steps.
- 1. List the possible values of S_3 and their probabilities.
- 2. Compute $E[S_3]$ in two ways: from the table and from $S_0(pu + (1-p)d)^3$.
- 3. Compute $P(S_3 < 100)$.
- 4. Find the p^* that makes S_t a martingale.
- Report: a table, two expectations, a probability, p^*.

Solution

- 1. $K = 0, 1, 2, 3$ up moves: $72.9, 89.1, 108.9, 133.1$ with $0.0911, 0.3341, 0.4084, 0.1664$ $\left(\binom{3}{k} p^k (1-p)^{3-k}\right)$
- 2. $E[S_3] = 100 \times 1.01^3 = 103.03$ both ways
- 3. $P(K \leq 1) = 0.4253$
- 4. $p^* = (1 - 0.9)/(1.1 - 0.9) = 0.50$: with $p = 0.55 > p^*$ the price drifts up

Back Exercise list

A6: Random Walk, Martingale and AR(1) [Proposed]

Solution discussed in class

Task

- $S_t = S_{t-1} + \varepsilon_t$, ε_t i.i.d. with mean 0 and variance σ^2; $X_t = 0.2 + 0.6X_{t-1} + \varepsilon_t$ with $\sigma = 1$. Model: A5.
- 1. Compute $E[S_t]$ and $\mathrm{Var}(S_t)$, and show that S_t is a martingale.
- 2. Show that $S_t^2 - t\sigma^2$ is also a martingale.
- 3. Compute the mean, the variance and $\rho(1)$, $\rho(2)$, $\rho(5)$ of the stationary X_t.
- 4. Compute $E[X_{t+1} \mid X_t = 2]$ and $\mathrm{Var}(X_{t+1} \mid X_t = 2)$.
- Report: two derivations, five moments, a forecast.

Back Exercise list

A7: The Inverse Transform Step by Step [Solved]

Task

- You have the uniform numbers $u = 0.3$ and $u = 0.9$ (and $u = 0.15, 0.45, 0.8$ for question 2).
- 1. Turn $u = 0.3$ and $u = 0.9$ into exponential draws with rate $\lambda = 0.5$.
- 2. Turn $u = 0.15, 0.45, 0.8$ into draws of the return X of A1.
- 3. Turn $u = 0.01$ into a draw of $N(0.05, 1.2^2)$.
- Report: six simulated values.

Solution

- 1. $x = -\ln(1-u)/0.5$: 0.713 and 4.605
- 2. $F^{-1}(u) = \inf\{x : F(x) \geq u\}$ with $F = 0.2, 0.5, 1$: -3, 0 and 2
- 3. $0.05 + 1.2 \times (-2.326) = -2.742$: the 1% quantile
- A large u gives a large draw: the inverse transform keeps the order of the uniforms.

Back Exercise list

A8: GBM and the Wiener Process [Proposed]

Solution discussed in class

Task

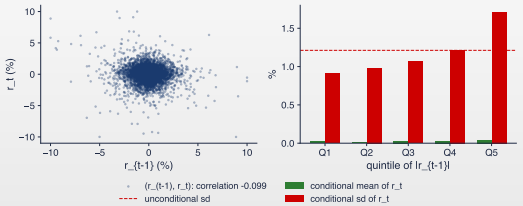
- An index follows GBM with $\mu = 8\%$, $\sigma = 20\%$ a year, $S_0 = 100$; W is a Wiener process. Model: A7.
- 1. Compute the mean and the median of S_1 and of S_5.
- 2. Compute $P(S_T < 100)$ for $T = 1$ and $T = 5$.
- 3. Compute the 5% quantile of S_1 and of S_5.
- 4. Compute $\mathrm{Var}(W(2) - W(0.5))$ and $\mathrm{Cov}(W(1), W(3))$.
- Report: eight GBM numbers and two Wiener moments.

Back Exercise list

B1: Are S&P 500 Returns Independent? [Solved]

- **Question**: are the daily returns of the S&P 500 independent, and what does yesterday tell us about today?
- **Inputs**: S&P 500 closes, 2000–2026, daily log returns in %
- **Tasks**
  1. Compute $\mathrm{Corr}(r_t, r_{t-1})$ and $\mathrm{Corr}(r_t^2, r_{t-1}^2)$ and compare them with the band $\pm 1.96/\sqrt{n}$.
  2. Split the days into quintiles of $|r_{t-1}|$ (five equal groups, Q1 the calmest, Q5 the most turbulent) and compute the mean and the standard deviation of r_t in each quintile.
  3. Check the law of total variance with the quintiles as groups.
  4. Interpretation: which is predictable from yesterday, the mean or the variance of today's return?
- **Report**: two correlations, a table of five quintiles, one decomposition, one sentence
- Notebook: section B1

B1: Solution [Solved]



- $n = 6\,714$, band ± 0.024; $\mathrm{Corr}(r_t, r_{t-1}) = -0.099$; $\mathrm{Corr}(r_t^2, r_{t-1}^2) = 0.313$
- Conditional means Q1–Q5: $0.021, 0.006, 0.029, 0.027, 0.042$ (%)
- Conditional s.d.: $0.91, 0.98, 1.07, 1.21, 1.71$; Q5/Q1 $= 1.88$
- Total variance $1.4691 = 1.4690$ (within) $+0.00013$ (between, 0.01%)
- Interpretation: the variance is predictable, the mean almost not
  - the negative lag-1 correlation comes from crisis days with large reversals
  - the squares are correlated, so returns are not independent

SFM_ch4_seminar

B2: The Same for DAX, BET and Bitcoin [Proposed]

- **Question**: do the DAX, the BET and Bitcoin show the same pattern as the S&P 500?
- **Inputs**: daily log returns, 2000–2026 (Bitcoin from 2014); model: B1
- **Tasks**
  1. Compute both lag-1 correlations and the band for each series.
  2. Compute the conditional standard deviation in Q1 and Q5 and their ratio.
  3. Compute the share of the variance explained by the quintile means.
  4. Interpretation: why is the BET the only series with a clearly positive $\mathrm{Corr}(r_t, r_{t-1})$?
- **Report**: one table and one sentence
- Notebook: section B2

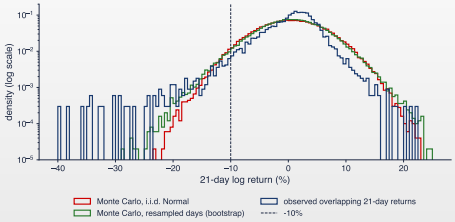
Back Exercise list

B3: A Monthly Loss Probability by Monte Carlo [Solved]

- **Question**: how likely is a 21-day loss of more than 10% for the S&P 500, and does the answer depend on the model?
- **Inputs**: S&P 500 daily log returns, 2000–2026; 100 000 simulated months
- **Tasks**
  1. Compute the exact probability under i.i.d. Normal daily returns with the sample mean and standard deviation.
  2. Estimate it by Monte Carlo under the same model and give the standard error.
  3. Estimate it by resampling 21 observed days with replacement (the empirical inverse transform).
  4. Compute the share of overlapping 21-day windows in the data with a loss beyond 10%.
  5. Interpretation: which differences are simulation error and which are model error?
- **Report**: four probabilities, two standard errors, one sentence
- Notebook: section B3

Back Exercise list

B3: Solution [Solved]



- Normal, exact: 2.922%; Monte Carlo Normal: 2.917% (s.e. 0.053%)
  - bootstrap: 3.184% (s.e. 0.056%)
- Data: 2.67% of 6 694 overlapping windows (179 windows, from a few episodes)
  - non-overlapping months: 3.45% of 319
- Interpretation
  - Monte Carlo and exact differ by less than one s.e.: simulation error
  - bootstrap and Normal differ by about 3.5 s.e.: model error (heavier daily tails)
  - the data are noisy because overlapping windows are dependent

SFM_ch4_seminar

B4: Testing a Random Number Generator [Proposed]

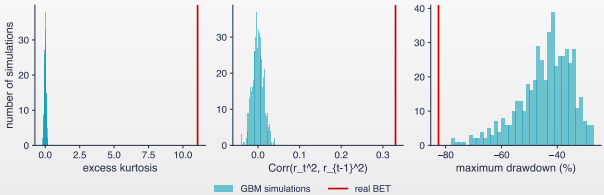
- **Question**: can a generator pass simple tests and still be unusable?
- **Inputs**: RANDU, $x_{k+1} = 65539\, x_k \bmod 2^{31}$, seed 1, 100 000 numbers; NumPy PCG64 with the same seed; model: B3
- **Tasks**
  1. Compute the mean, the variance and the lag-1 correlation of both sequences and compare with $1/2$, $1/12$ and 0.
  2. Count the distinct values of $9u_k - 6u_{k+1} + u_{k+2}$ for both generators.
  3. Find the period of the LCGs $(a, c, M) = (5, 1, 16)$, $(5, 0, 16)$ and $(4, 1, 16)$ started at 7.
  4. Interpretation: why do one-dimensional tests not detect the flaw of RANDU?
- **Report**: a table of six statistics, two counts, three periods, one sentence
- Notebook: section B4

B5: Is the BET a Geometric Brownian Motion? [Solved]

- **Question**: which features of the BET can a GBM with the same drift and volatility reproduce?
- **Inputs**: BET closes, 2000–2026; 500 GBM paths of the same length
- **Tasks**
  1. Calibrate GBM: annual log drift and volatility from the daily log returns.
  2. Simulate 500 paths and compute, for each, the excess kurtosis, $\mathrm{Corr}(r_t^2, r_{t-1}^2)$ and the maximum drawdown.
  3. Locate the real values in the simulated distributions (5% and 95% quantiles).
  4. Interpretation: which stylised facts does GBM miss, and which risk number is most affected?
- **Report**: two parameters, three comparisons, one sentence
- Notebook: section B5

Back Exercise list

B5: Solution [Solved]



- $n = 6\,670$ days; $\sigma = 22.2\%$, log drift 16.1% a year
- Excess kurtosis: real 11.06, GBM 90% range $[-0.10, 0.10]$
- $\mathrm{Corr}(r_t^2, r_{t-1}^2)$: real 0.330, GBM $[-0.022, 0.022]$
- Maximum drawdown: real -82.5%, GBM median -42.7%, range $[-62.8\%, -31.4\%]$
- Interpretation: GBM misses heavy tails and volatility clustering completely
  - the 2008 crash makes the real drawdown deeper than every simulated path
  - a GBM-based risk limit would have been far too loose

SFM_ch4_seminar

B6: Does Variance Grow Linearly with the Horizon? [Proposed]

- **Question**: does Var(h-day return) $= h\,$Var(1-day return) hold, as a random walk with i.i.d. increments predicts?
- **Inputs**: S&P 500, BET, Bitcoin daily log returns; non-overlapping blocks of $h = 5, 21, 63$ days; model: B5
- **Tasks**
  1. Compute the ratio Var(h-day)$/(h\,$Var(1-day)) for each series and horizon.
  2. Use the approximate standard error $\sqrt{2/(m-1)}$ of a variance ratio with m blocks to judge the distance from 1.
  3. Interpretation: what does a ratio above 1 or below 1 say about the autocorrelation of returns?
- **Report**: a 3×3 table and two sentences
- Notebook: section B6

C1: How Often Is There a Bad Year? [Proposed]

- **Question**: does GBM give the right probability of a drawdown beyond 20% (or 40%) within one calendar year?
- **Inputs**: S&P 500, DAX, BET, Bitcoin, complete calendar years 2000–2025 (Bitcoin 2015–2025); models: B3, B5
- **Tasks**
  1. Compute the maximum drawdown within each calendar year and the share of years beyond 20% and beyond 40%.
  2. Simulate 10 000 one-year paths under GBM and under the i.i.d. bootstrap of the daily returns and compute the same two probabilities.
  3. Compare data and models with a binomial test.
  4. Propose one model that could reproduce both numbers.
  5. Interpretation: what does the gap between the bootstrap and the data say about volatility clustering?
- **Report**: a table, two binomial p-values, a short project plan
- Notebook: section C1

Back Exercise list

C2: Audit an AI Answer [Proposed]

- A student asked an AI assistant to "summarise the probability facts about the S&P 500 for a risk report". The answer:
- (a) Corr(r_t, r_{t-1}) = −0.099, so daily returns are independent and tomorrow's variance cannot be predicted.
- (b) The log price is a random walk, so its variance is the same at every date and the process is stationary.
- (c) Under GBM with mu = 8% and sigma = 19%, the median value of 100 after 10 years is 100 exp(0.8) = 222.6.
- (d) A Monte Carlo probability of about 3% from 10,000 paths has a standard error of about 0.0017.
- (e) A 50/50 portfolio of S&P 500 and DAX has daily volatility sqrt(0.25 x 1.22^2 + 0.25 x 1.43^2) = 0.94%.
- **Tasks**
  1. For each statement, say whether it is correct; if not, give the correct statement and the correct number from the notebook (section C2).
- **Report**: a list of five verdicts with one line of justification each.

Points to Remember

- Zero correlation does not mean independence: check squares or absolute values too
- The conditional variance of returns moves with yesterday's news; the conditional mean hardly moves
- Mixtures of Normal regimes give heavy tails
- Monte Carlo: report the standard error; separate simulation error from model error
- GBM reproduces the path of an index roughly, but not its tails, clusters or deepest drawdowns
- An AI answer is a draft: check every definition, model and number

After the Seminar

- Lecture 4 develops each topic of today: distributions, dependence, conditioning, Monte Carlo, binomial trees, random walks, Wiener process and GBM
- C1 can grow into a team project: GARCH and regime-switching simulations, block bootstrap, more indices
- Reading: Franke, Härdle & Hafner (2019), Ch. 3–5; exercises with solutions: Borak, Härdle & López-Cabrera (2013)

References

- Borak, S., Härdle, W.K., López-Cabrera, B. (2013). *Statistics of Financial Markets: Exercises and Solutions*, 2nd ed. Springer.
- Cox, J.C., Ross, S.A., Rubinstein, M. (1979). Option pricing: a simplified approach. *Journal of Financial Economics*, 7(3), 229–263.
- Franke, J., Härdle, W.K., Hafner, C.M. (2019). *Statistics of Financial Markets: An Introduction*, 5th ed. Springer.
- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer.
- Marsaglia, G. (1968). Random numbers fall mainly in the planes. *Proceedings of the National Academy of Sciences*, 61(1), 25–28.