COMPUTATIONAL FINANCE: 422

Mathematical Preliminaries

Panos Parpas

(Slides courtesy of Daniel Kuhn)

p.parpas@imperial.ac.uk

Imperial College

London

This Lecture

- Mathematical background material
  - Functions
  - Differential calculus
  - Optimization
- Basic probability theory
  - Random variables
  - Independence
  - Expectation, Variance, and Covariance
  - Normal random variables and Central Limit Theorem

Further reading:

- D.G. Luenberger: *Investment Science*, Appendix A & B
- D.J. Higham: *Financial Option Valuation*, Chapter 3

Functions

Certain functions are commonly used in finance:

- Exponential functions: $f(x) = ac^{bx}$ where a, b, and c are constants. Very often c is $e = 2.7182818...$.
- Logarithmic functions: the natural logarithm is the function denoted by $\ln(\cdot)$ which satisfies $e^{\ln(x)} = x$.
- Linear functions: a function f of several variables $x_1, x_2, \ldots, x_n$ is linear if it has the form

$$f(x_1, x_2, \ldots, x_n) = a_1x_1 + a_2x_2 + \cdots + a_nx_n \,.$$

- Inverse functions: a function f has an inverse function g if for all x we have $g(f(x)) = x$. Inverse functions are usually denoted by f^{-1}.

Differential Calculus I

We shall review some concepts that are used in the course:

- Limits: if the function f approaches the value L as x approaches x_0, we write $L = \lim_{x \to x_0} f(x)$. An example is $\lim_{x \to \infty} 1/x = 0$.
- Derivatives: the derivative of a function f at x is

$$\frac{\mathrm{d}f(x)}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

  Sometimes we write $f'(x)$ for the derivative of f at x. It is important to know these common derivatives:

  - if $f(x) = x^n$, then $f'(x) = nx^{n-1}$;
  - if $f(x) = e^{ax}$, then $f'(x) = ae^{ax}$;
  - if $f(x) = \ln(x)$, then $f'(x) = 1/x$.

Differential Calculus II

- Product rule: the derivative of the product of two functions f and g is

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x)\,.$$

- Quotient rule: the derivative of the quotient of two functions f and g is

$$(f/g)'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}\,.$$

- Chain rule: the derivative of the composition of two functions f and g is

$$[f(g)]'(x) = f'(g(x))g'(x)\,.$$

Differential Calculus III

- Higher order derivatives: higher order derivatives are formed by taking derivatives of derivatives. The second derivative of f is the derivative of f'.
- Partial derivatives: functions of several variables can be differentiated partially w.r.t. each argument. We define

$$
\begin{aligned}
&\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} \\
= \quad &\lim_{\Delta x \to 0} \frac{f(x_1, x_2, \ldots, x_i + \Delta x, \ldots, x_n) - f(x_1, x_2, \ldots, x_n)}{\Delta x}
\end{aligned}
$$

Differential Calculus IV

- Taylor approximation: a function f can be approximated in a region near a point x by using its derivatives. The following approximations are useful:

  - $f(x + \Delta x) = f(x) + f'(x)\Delta x + O(\Delta x)^2$

  - $f(x + \Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2} f''(x)(\Delta x)^2 + O(\Delta x)^3$

  where $O(\Delta x)^2$ and $O(\Delta x)^3$ denote terms of order $(\Delta x)^2$ and $(\Delta x)^3$.

Differential Calculus V

- Taylor approximation for functions of several variables: a function $f : \mathbb{R}^n \to \mathbb{R}$ can be approximated in a region near a point $(x_1, x_2, \ldots, x_n)$ by using its partial derivatives. The following approximations are useful:

$$
\begin{aligned}
& f(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) \\
= \quad & f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{n} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} \Delta x_i \\
& + \sum_{i=1}^{n} \sum_{j=1}^{n} O(\Delta x_i \Delta x_j)
\end{aligned}
$$

Differential Calculus V

- Taylor approximation for functions of several variables: a function $f : \mathbb{R}^n \to \mathbb{R}$ can be approximated in a region near a point $(x_1, x_2, \ldots, x_n)$ by using its partial derivatives. The following approximations are useful:

$$
\begin{aligned}
& f(x_1 + \Delta x_1, x_2 + \Delta x_2, \ldots, x_n + \Delta x_n) \\
= \quad & f(x_1, x_2, \ldots, x_n) + \sum_{i=1}^{n} \frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_i} \Delta x_i \\
& + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f(x_1, x_2, \ldots, x_n)}{\partial x_i \partial x_j} \Delta x_i \Delta x_j \\
& + \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} O(\Delta x_i \Delta x_j \Delta x_k)
\end{aligned}
$$

Optimization I

- Necessary conditions: a function f of a single variable x is said to have a maximum at a point x_0 if $f(x_0) \geq f(x)$ for all x. If x_0 is not a boundary point of an interval over which f is defined, then for x_0 to be a maximum it is necessary that

$$f'(x_0) = 0\,.$$

  This equation can be used to find the maximum x_0.

- Example: assume that $f(x) = -x^2 + 12x$. To find the maximum, we solve

$$f'(x_0) = -2x + 12 = 0 \quad \Rightarrow \quad x = 6\,.$$

Lagrange Multipliers I

- Constrained optimization: consider the problem of maximizing a function f of several variables $x_1, x_2, \ldots, x_n$ which are required to satisfy the constraint $g(x_1, x_2, \ldots, x_n) = 0$. Formally, this problem can be written as

$$\begin{aligned} &\underset{x}{\text{maximize}}\, f(x_1, x_2, \ldots, x_n) \\ &\text{subject to } \; g(x_1, x_2, \ldots, x_n) = 0\,. \end{aligned}$$

  We introduce a Lagrange multiplier λ and form the Lagrangian function

$$L(x_1, x_2, \ldots, x_n, \lambda) = f(x_1, x_2, \ldots, x_n) - \lambda g(x_1, x_2, \ldots, x_n)\,.$$

Lagrange Multipliers II

- To solve this constrained problem, we set the partial derivatives of the Lagrangian w.r.t. each of the variables equal to zero.

  $\Rightarrow$ This gives a system of $n+1$ equations for the $n+1$ unknowns $x_1, x_2, \ldots, x_n$ and λ.

- A problem with two constraints, for example, is solved by introducing two Lagrange multipliers λ and μ.

$$\begin{aligned} &\underset{x}{\text{maximize}}\, f(x_1, x_2, \ldots, x_n) \\ &\text{subject to } \; g(x_1, x_2, \ldots, x_n) = 0 \\ &\qquad\qquad\quad\; h(x_1, x_2, \ldots, x_n) = 0\,. \end{aligned}$$

$$L = f(x_1, x_2, \ldots, x_n) - \lambda g(x_1, x_2, \ldots, x_n) - \mu h(x_1, x_2, \ldots, x_n)\,.$$

Lagrange Multipliers III

- A problem with n variables and m constraints is assigned m Lagrange multipliers, while the Lagrange function has $n + m$ arguments. Setting all partial derivatives to zero gives $n + m$ equations for $n + m$ unknowns.
- Some problems have inequality constraints of the form $g(x_1, x_2, \ldots, x_n) \leq 0$. Two cases:
  - if $g(x_1, x_2, \ldots, x_n) < 0$ at the optimum, then the constraint is not active and can be dropped $\Rightarrow$ no Lagrange multiplier is needed;
  - if $g(x_1, x_2, \ldots, x_n) = 0$ at the optimum, then the constraint is active $\Rightarrow$ a Lagrange multiplier is introduced as before; this multiplier is nonnegative.

Random Variables

- A discrete random variable x is described by a finite number of possible values $x_1, x_2, \ldots, x_m$ which are assigned probabilities $p_1, p_2, \ldots, p_m$. Interpretation:

$$p_i = \text{prob}(x = x_i) \quad \text{for any } i = 1, 2, \ldots, m.$$

  The probabilities are nonnegative and sum to unity, that is, $\sum_{i=1}^{m} p_i = 1$.

- A continuous random variable x is described by a probability density function $p(\xi)$. The interpretation is

$$\int_a^b p(\xi)\text{d}\xi = \text{prob}(a \leq x \leq b) \quad \text{for any } a < b.$$

  The density function is nonnegative and integrates to unity, that is, $\int_{-\infty}^{+\infty} p(\xi)\text{d}\xi = 1$.

Probability Distribution

- The probability distribution of a (discrete or continuous) random variable x is the function $F(\xi)$ defined as

$$F(\xi) = \text{prob}(x \leq \xi)\,.$$

  It follows that
  - $F(-\infty) = 0$,
  - $F(+\infty) = 1$,
  - F is monotonically increasing.
- If x is a continuous random variable, then

$$F(\xi) = \int_{-\infty}^{\xi} p(\xi')\mathrm{d}\xi' \quad \Rightarrow \quad \mathrm{d}F(\xi)/\mathrm{d}\xi = p(\xi)\,.$$

Dependent Random Variables I

- Two discrete random variables x and y are described by their possible pairs of values $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$ and the corresponding probabilities $p_1, p_2, \ldots, p_n$ with the interpretation

$$p_i = \text{prob}(x = x_i \;\wedge\; y = y_i)\,.$$

- Two continuous random variables x and y are described by their joint probability density function $p(\xi, \eta)$ with the interpretation

$$\int_{a_x}^{b_x} \int_{a_y}^{b_y} p(\xi, \eta) \mathrm{d}\eta \mathrm{d}\xi = \text{prob}(a_x \leq x \leq b_x \;\wedge\; a_y \leq y \leq b_y)\,.$$

Dependent Random Variables II

- The joint probability distribution F is defined as

$$F(\xi, \eta) = \text{prob}(x \leq \xi,\, y \leq \eta)\,.$$

- From a joint distribution the distribution of any of the random variables can easily be recovered. We have
  - $F_x(\xi) = F(\xi, \infty)$;
  - $F_y(\eta) = F(\infty, \eta)$.
- In general, n random variables are defined by their joint probability distribution defined w.r.t. n variables.

Independent Random Variables

- Two discrete random variables x and y are independent if the possible joint values can be written as (x_i, y_j) for $i = 1, 2, \ldots, n_x$ and $j = 1, 2, \ldots, n_y$, while the probability p_{ij} of outcome (x_i, y_j) factors into the form

$$p_{ij} = p_{x,i}\, p_{y,j} \,.$$

- Two continuous random variables x and y are independent if the joint density function factors into the form

$$p(\xi, \eta) = p_x(\xi) p_y(\eta) \,.$$

- Example: The pair of random variables defined as the outcomes on two fair tosses of a die are independent. The probability of obtaining the pair $(3, 5)$, say, is $\frac{1}{6} \times \frac{1}{6}$.

Moments

- The expected value or expectation of a random variable x is defined as
  - $\mathrm{E}(x) = \sum_{i=1}^{n} x_i p_i$ if x is a discrete r.v.;
  - $\mathrm{E}(x) = \int_{-\infty}^{+\infty} \xi p(\xi) \mathrm{d}\xi$ if x is a continuous r.v..
- The concept of an expectation can be generalized. For any function $f : \mathbb{R} \to \mathbb{R}$, we can define
  - $\mathrm{E}[f(x)] = \sum_{i=1}^{n} f(x_i) p_i$ if x is a discrete r.v.;
  - $\mathrm{E}[f(x)] = \int_{-\infty}^{+\infty} f(\xi) p(\xi) \mathrm{d}\xi$ if x is a continuous r.v..
- The moment of order m of any random variable x is defined as $\mathrm{E}(x^m)$.

  $\Rightarrow$ The (ordinary) expectation of x is the first-order moment of x.

Variance and Standard Deviation

- The variance of a r.v. x is defined as

$$\mathrm{var}(x) = \mathrm{E}([x - \mathrm{E}(x)]^2) \,.$$

- One easily verifies the identity:

$$\mathrm{var}(x) = \mathrm{E}(x^2) - \mathrm{E}(x)^2 \,.$$

- Loosely, the expectation tells you the ‘typical’ or ‘average’ value of a r.v., while the variance gives the amount of ‘variation’ around this value.
- The standard deviation of a r.v. is defined as

$$\mathrm{std}(x) = \sqrt{\mathrm{var}(x)} \,.$$

Generalized Expectation

- The concept of an expectation can be further generalized to situations in which there are two dependent random variables x and y. For any function $f : \mathbb{R}^2 \to \mathbb{R}$, we can define

  - $\mathrm{E}[f(x,y)] = \sum_{i=1}^{n} f(x_i, y_i) p_i$ if x and y are discrete dependent random variables;

  - $\mathrm{E}[f(x,y)] = \int_{\mathbb{R}^2} f(\xi,\eta) p(\xi,\eta) \mathrm{d}\xi \mathrm{d}\eta$ if x and y are continuous dependent random variables.

- Expectations of functions of n random variables are defined analogously.

Covariances and Correlations I

- The covariance of two dependent random variables x and y is defined as

$$\mathrm{cov}(x,y) = \mathrm{E}([x-\mathrm{E}(x)][y-\mathrm{E}(y)])\,.$$

- Note that $\mathrm{cov}(x,x) = \mathrm{var}(x)$.
- The correlation of x and y is defined as

$$\varrho(x,y) = \frac{\mathrm{cov}(x,y)}{\mathrm{std}(x)\mathrm{std}(\mathrm{y})}\,.$$

- If x and y are independent, then

$$\mathrm{cov}(x,y) = \mathrm{E}[x-\mathrm{E}(x)]\mathrm{E}[y-\mathrm{E}(y)] = 0 \quad \Rightarrow \quad \varrho(x,y) = 0\,.$$

Covariances and Correlations II

- By the Cauchy-Schwartz inequality, we find

$$
\begin{aligned}
|\mathrm{cov}(x,y)| &\leq \mathrm{E}(|x-\mathrm{E}(x)|\,|y-\mathrm{E}(y)|) \\
&\leq \sqrt{\mathrm{E}([x-\mathrm{E}(x)]^2)\mathrm{E}([y-\mathrm{E}(y)]^2)} \\
&= \mathrm{std}(x)\mathrm{std}(y)\,.
\end{aligned}
$$

$\Rightarrow$ the correlation $\varrho(x,y)$ is always between -1 and $+1$.

- Two random variables x and y are said to be
  - positively correlated if $\varrho(x,y) > 0$;
  - perfectly positively correlated if $\varrho(x,y) = 1$;
  - negatively correlated if $\varrho(x,y) < 0$;
  - perfectly negatively correlated if $\varrho(x,y) = -1$;
  - uncorrelated if $\varrho(x,y) = 0$.

Covariances and Correlations III

- A random variable x is perfectly positively correlated with the random variable $y = ax + b$ for any $a, b \in \mathbb{R}$ such that $a > 0$.
- A random variable x is perfectly negatively correlated with the random variable $y = ax + b$ for any $a, b \in \mathbb{R}$ such that $a < 0$.
- Note that if x and y are independent, then they are uncorrelated. However, if x and y are uncorrelated, then they are not necessarily independent.

Covariances and Correlations IV

- Let x and y be two dependent random variables, and let α and β be real numbers. Then

$$
\begin{aligned}
\mathrm{E}(\alpha x + \beta y) &= \alpha \mathrm{E}(x) + \beta \mathrm{E}(y)\,, \\
\mathrm{var}(\alpha x + \beta y) &= \alpha^2 \mathrm{var}(x) + 2\alpha\beta\,\mathrm{cov}(x, y) + \beta^2 \mathrm{var}(y)\,.
\end{aligned}
$$

- Let $x_1, x_2, \ldots, x_n$ be n dependent random variables. The covariance matrix of these random variables is defined as the $n \times n$-matrix V with entries

$$V_{ij} = \mathrm{cov}(x_i, x_j) \quad \text{for } i, j = 1, \ldots, n.$$

- if $\alpha_1, \alpha_2, \ldots, \alpha_n$ are n real numbers, then

$$\mathrm{E}\left(\sum_{i=1}^{n} \alpha_i x_i\right) = \sum_{i=1}^{n} \alpha_i \mathrm{E}(x_i) \quad \text{and} \quad \mathrm{var}\left(\sum_{i=1}^{n} \alpha_i x_i\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i V_{ij} \alpha_j\,.$$

Uniform Random Variables

- A continuous random variable x with density function

$$p(\xi) = \begin{cases} (\beta - \alpha)^{-1} & \text{for } \alpha \leq \xi \leq \beta, \\ 0 & \text{otherwise,} \end{cases}$$

  is said to have a uniform distribution over $[\alpha, \beta]$.

- x takes only values between α and β and is equally likely to take any such value.

- The uniform distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x < \alpha, \\ \frac{x-\alpha}{\beta-\alpha} & \text{for } \alpha \leq x \leq \beta, \\ 1 & \text{for } x > \beta. \end{cases}$$

- $\mathrm{E}(x) = (\beta + \alpha)/2$ and $\mathrm{var}(x) = (\beta - \alpha)^2/12$.

Normal Random Variables I

- A (continuous) random variable x is said to be normal or Gaussian if its probability density function is of the form

$$p(\xi) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(\xi-\mu)^2}.$$

- It follows that $\mathrm{E}(x) = \mu$ and $\mathrm{var}(x) = \sigma^2$.
- A normal r.v. is said to be standard if $\mu = 0$ and $\sigma = 1$.
- A standard normal random variable has density

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,,$$

and the standard normal distribution N is given by

$$N(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}\xi^2} \mathrm{d}\xi\,.$$

Normal Random Variables II

- There is no analytic expression for $N(x)$, but tables of its values are available.
- Let $x = (x_1, x_2, \ldots, x_n)$ be a vector of n normal random variables. We introduce the vector $\bar{x}$ whose components are the expected values of the components in x. The covariance matrix V associated with x can be written as

$$V = \mathrm{E}\left[(x - \bar{x})(x - \bar{x})^\top\right].$$

- If the n variables are jointly normal, the density of x is

$$p(x) = \frac{1}{(2\pi)^{n/2}\det(V)^{1/2}} e^{-\frac{1}{2}(x-\bar{x})V^{-1}(x-\bar{x})^\top}.$$

Normal Random Variables III

- If n jointly normal random variables are uncorrelated, then the covariance matrix V is diagonal $\Rightarrow$ the joint density function factors into a product of densities for the n separate variables.

  $\Rightarrow$ If n jointly normal random variables are uncorrelated, then they are independent.

- Summation property: if x and y are jointly normal random variables and $\alpha, \beta \in \mathbb{R}$, then $\alpha x + \beta y$ is normal.

- Generalization: if x is a vector of n jointly normal r.v.s and T is a $m \times n$-matrix, then Tx is a vector of m jointly normal r.v.s.

Normal Random Variables IV

- To express that x is a normal r.v. with expected value μ and variance σ^2 we use the shorthand notation:

$$x \sim \mathcal{N}(\mu, \sigma^2) \,.$$

- To express that x is a vector of jointly normal r.v. with expected values $\bar{x}$ and covariance matrix V we write:

$$x \sim \mathcal{N}(\bar{x}, V) \,.$$

- Some useful properties of normal r.v.s are:
  - if $x \sim \mathcal{N}(\mu, \sigma^2)$, then $(x - \mu)/\sigma \sim \mathcal{N}(0, 1)$;
  - if $y \sim \mathcal{N}(0, 1)$, then $\sigma\, y + \mu \sim \mathcal{N}(\mu, \sigma^2)$;
  - if $x_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$, $x_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ and x_1 and x_2 are independent, then $x_1 + x_2 \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$;

Central Limit Theorem I

- Let $x_1, x_2, x_2, \ldots$ be an infinite sequence of independent, identically distributed (i.i.d.) random variables, each with expected value μ and variance σ^2.
- Define $S_n = \sum_{i=1}^{n} x_i$ for $n = 1, 2, 3, \ldots$. Note that $\mathrm{E}(S_n) = n\mu$ and $\mathrm{var}(S_n) = n\sigma^2$.
- The Central Limit Theorem says that for large n the random variable $(S_n - n\mu)/(\sigma\sqrt{n})$ is approximately standard normally distributed. In mathematical terms:

$$\mathrm{prob}\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to N(x) \quad \text{as } n \to \infty \ (\forall\, x \in \mathbb{R}).$$

Central Limit Theorem II

- Real-life systems are subject to a range of external influences that can be reasonably approximated by i.i.d. random variables.

- Hence, by the C.L.T. the overall effect can be reasonably modelled by a single normal random variable with appropriate mean and variance.

- $\Rightarrow$ Because of the C.L.T. normal random variables are ubiquitous in stochastic modelling!