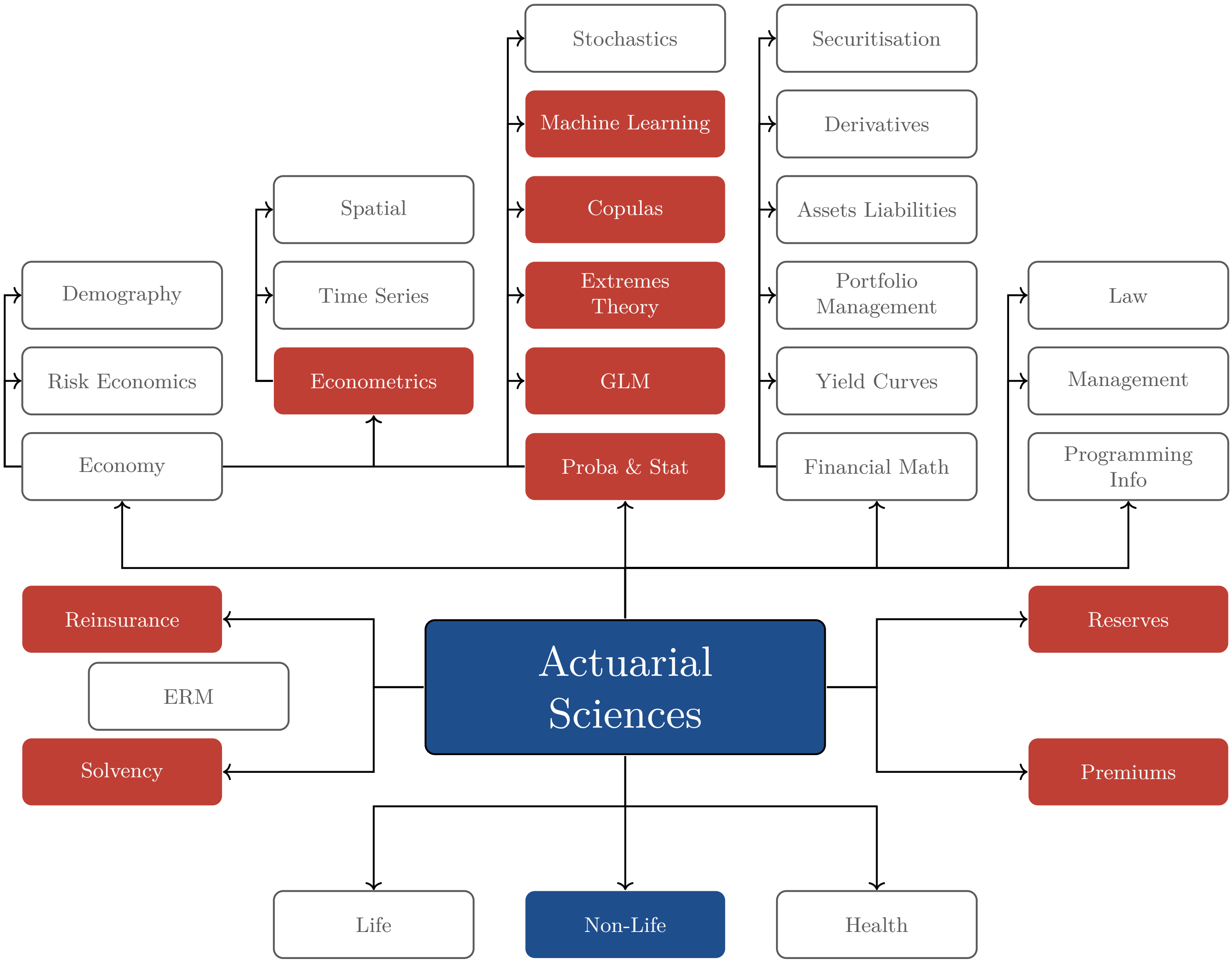
Actuarial Formula Cheat Sheet

My Actuarial Revision Sheet

Master in Insurance Economics/Econometrics
Master in Actuarial Science

October 3, 2025



Insurance Keywords

1 – Life and Non-Life Insurance

The distinction between life and non-life insurance is fundamental. An insurer cannot offer both types of insurance without holding two separate companies:

- **Life insurance**, i.e., personal insurance excluding coverage for bodily injuries.
- **Non-life insurance**, which includes property and liability insurance as well as insurance for bodily injuries.

2 – The Principles of Insurance

Insurance is assumed to :

- be based on utmost good faith,
- apply only if the insured has an insurable interest in preserving the item (property insurance),
- operate under the indemnity principle :
  - not allow enrichment from a claim settlement,
  - not even through insurance accumulation,
  - include subrogation (in Liability Insurance, if the insurer compensates the insured victim, the insured cannot then claim from the party responsible for the loss).
- not reduce the insured's efforts in prevention and protection, as a reasonable person, even if financially protected.

establish the cause of an accident in Civil Responsibility, who is not responsible if he or she does not contribute to the cause of the accident.

3 – The Insurance Policy

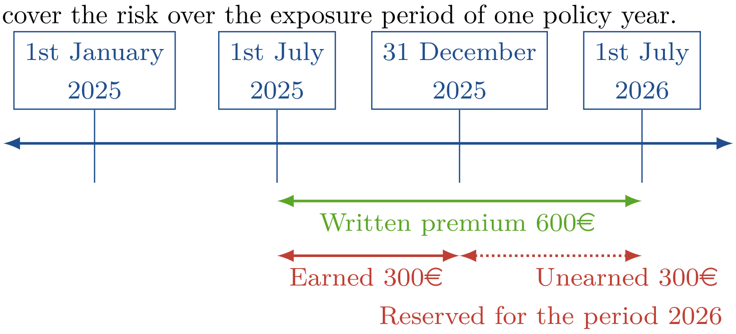
The **insurance policy** (or contract) is the contractual document that governs the relationship between the insurance company (or mutual insurance company) and the insured (policyholder). This contract defines in particular :

1. the list of covered events, including any exclusions,
2. the coverage, i.e., the assistance provided to the insured in case of a loss,
3. the obligations of the insured :
   - any preventive measures required to reduce risk,
   - time limits for reporting a claim to the insurer,
   - the amount and payment conditions of the premium (deductible, limit),
   - the conditions for cancellation of the policy (automatic renewal),
4. the obligations of the insurance company : time limits for compensation payments.

4 – The Premium and Claims

Classically, the role of the insurer is to substitute a constant C, the **contribution** or the **premium**, for a random claim S. **Pure premium** or **technical premium** aims to compensate claims without surplus or profit, overall $C_t = \mathbb{E}[S]$
The **net premium** is higher than the pure premium. It aims to cover the cost of claims and provide a safety margin.

The **gross premium** is the net premium + overhead expenses + commissions + expected profit + taxes.
For commercial reasons, the premium actually charged may differ significantly from the technical premium.
Written premium: premium charged to the insured to cover claims that may occur during the coverage period defined by the contract (generally 1 year in Property and Casualty insurance).
Earned premium: proportion of the written premium used to cover the risk over the exposure period of one policy year.



The S/P is the key indicator. For the insurer to make a profit the $S/P \ll 1$.

5 – Loss / Payment Triangle

Insurance accounting is broken down by the **accident year** of the claim. If a premium covers multiple calendar years, a proportional part will be allocated to each. Each payment and each claim reserving is assigned to the accident year. The monitoring of payments or expenses is expressed through a triangle (triangular matrix) :

$$\begin{pmatrix} C_{1,1} & C_{1,2} & \dots & & C_{1,n} \\ C_{1,1} & C_{1,2} & \dots & C_{2,n-1} & \\ \vdots & \vdots & & & \\ C_{n-1,1} & C_{n-1,2} & & & \\ C_{n,1} & & & & \end{pmatrix}$$

where $C_{i,j} = \sum_{k=1}^{j} X_{i,k}$ represents the cumulative amount of claims paid for origin year i and development year j.

6 – Solvency II and Risk Management

Solvency II is the European regulatory framework applicable to insurers and reinsurers since 2016. It is based on three interdependent pillars :

- **Pillar 1 : Quantitative Requirements**
  Determines the capital requirements :
  - **SCR** (Solvency Capital Requirement) : capital to absorb an extreme shock (99.5% over 1 year),
  - **MCR** (Minimum Capital Requirement) : absolute minimum threshold,
  - admissible assets to cover technical reserves and capital requirements.
- **Pillar 2 : Governance, Internal Control, and Risk Management**
  The core link with **ERM** (Enterprise Risk Management). The requirements cover :
  - governance : boards of directors responsible for the risk management framework;
  - an effective **internal control** system ;

– independent key functions : **actuarial**, **risk management**, **compliance**, **internal audit** ;

– **ORSA** (Own Risk and Solvency Assessment): internal assessment of risks and solvency, a central tool aligning strategy, risk appetite, and economic capital.

- **Pillar 3 : Market Discipline**
  Based on **transparency** and communication :

  – **SFCR** (Solvency and Financial Condition Report) : public, summarizes solvency and financial position,

  – **RSR** (Regular Supervisory Report) : intended for the supervisor,

  – quantitative reporting : regulatory statements (**QRTs**), regular submission of financial and prudential data.

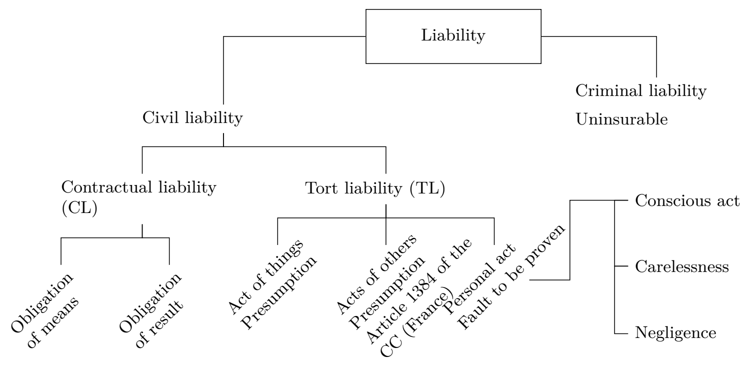
7 – Main Branches of Life and Non-Life Insurance

Life insurance covers long-term commitments, with or without a savings component :

- **Life insurance** : lump sum or annuity paid if the insured is alive at a given date.
- **Death insurance** : payment if the insured dies during the covered period.
- **Endowment insurance** : combination of life and death coverage.
- **Life annuity** : periodic payments until death.
- **Savings/retirement** : products with deferred capital or deferred annuity.
- **Unit-linked policies** : benefits dependent on the value of financial assets.
- **Group contracts** : occupational pensions, group welfare insurance.

Non-life insurance covers risks occurring in the short or medium term :

- **Automobile** : third-party liability, vehicle damage.
- **Home** : fire, theft, water damage, liability.
- **General liability** : personal liability, business liability.
- **Health and welfare** : medical reimbursements, disability, incapacity.
- **Personal accident** : capital in case of accident, disability, or death.
- **Business interruption** : financial losses related to a claim.
- **Technical risks** : construction, machinery breakdown.
- **Transport, aviation, maritime insurance** : goods in transit, specific liabilities.

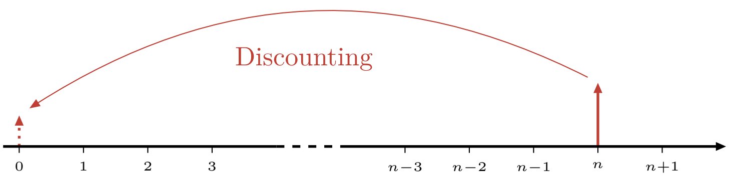


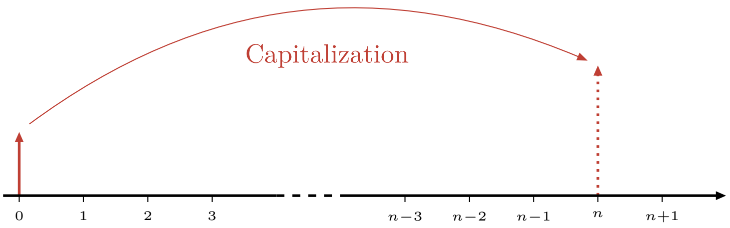
8 – Actuary

In practice, the actuary :

- prices insurance and welfare products,
- estimates technical reserves,
- projects cash flows and values long-term liabilities,
- measures economic capital (SCR, ORSA) and contributes to ERM,
- advises management on strategy, solvency, and regulatory compliance.

9 – Capitalization Discounting





10 – Interests

Discount rate d

$$d = i/(1+i)$$

Simple interest i

$$I_t = Pit = Pi\frac{k}{365}$$

Compound interest i

$$V_n = P(1+i)^n = P\left(1+\frac{p}{100}\right)^n$$

Continuous interest r

$$V_t = V_0\ e^{rt}$$

Effective rate i_e

$$i_e = \left(1+\frac{i}{m}\right)^m - 1$$

where i is the nominal rate and m the number of periods in a year.

Equivalent rate $i^{(m)}$

$$i^{(m)} = m\left((1+i)^{1/m} - 1\right)$$

Nominal rate i and periodic rate

The **nominal** or **face** rate allows calculating the interest due over one year. The **periodic** rate corresponds to the nominal rate divided by the number of periods in a year i/m. If the periodic rate is weekly, the nominal rate will be divided by 52.

11 – Valor presente y valor futuro

The present value (PV) represents the capital that must be invested today at an annual compound interest rate i to obtain future cash flows (F_k) at times t_k:

$$PV = \sum_{k=1}^{n} F_k \times \frac{1}{(1+i)^{t_k}} \tag{1}$$

When the F_k are constant

$$PV = K\frac{1-(1+i)^{-n}}{i} \tag{2}$$

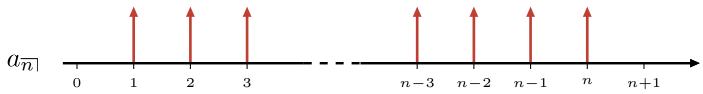
The future value (FV) represents the value of the capital at T which, with an annual compound interest rate i, capitalizes the future cash flows (F_k) at times t_k.

$$FV = V_n = \sum_{k=1}^{n} F_k \times (1+i)^{n-t_k} \tag{3}$$

More generally $FV = (1+i)^n PV$.

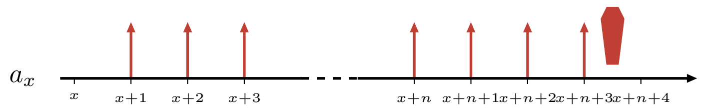
12 – Annuities

Certain annuity $a_{\overline{n}|}$ (or $a_{\overline{n}|\,i}$ if the interest rate i needs to be specified): this is the default case in financial mathematics. Its payments are, for example, guaranteed by a contract.



$$\ddot{a}_{\overline{n}|} = 1 + v + \cdots + v^{n-1} = \frac{1-v^n}{1-v} = \frac{1-v^n}{d}$$

Contingent annuity $\ddot{a}_x$: its payments are conditional on a random event, such as a life annuity of an individual aged x. In this example, payments continue until death occurs :



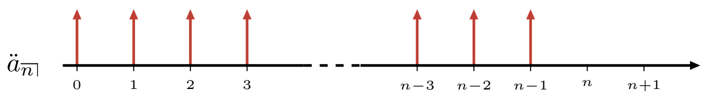
The date of death is represented here by a small coffin. This type of annuity will be extensively studied in the life actuarial section.

Annuity in arrears (immediate) $a_{\overline{n}|}$: its periodic payments are made at the end of each payment period, as with a salary paid at the end of the month. This is the default case, previously illustrated for the certain annuity.

$$\ddot{a}_{\overline{n}|} = 1 + v + \cdots + v^{n-1} = \frac{1-v^n}{1-v} = \frac{1-v^n}{d}$$

$$\mathrm{PV}^{\mathrm{due}}_{\overline{n}|} = K\ddot{a}_{\overline{n}|} = K\frac{1-v^n}{d}$$

Annuity in advance (due) $\ddot{a}_{\overline{n}|}$: its periodic payments are made at the beginning of each payment period, as with rent payments, for example.



Also denoted $\mathrm{PV}^{\mathrm{im}}$:

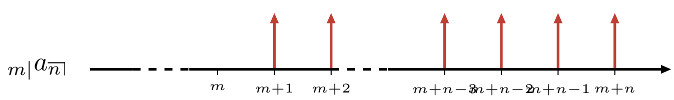
$$a_{\overline{n}|} = v + v^2 + \cdots + v^n = \frac{1-v^n}{i} = v\frac{1-v^n}{1-v}$$

$$\mathrm{PV}^{\mathrm{im}}_{\overline{n}|} = Ka_{\overline{n}|} = K\frac{1-v^n}{i}$$

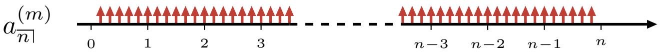
Perpetuity a or $a_{\overline{\infty}|}$:

$$a = 1/i$$

Deferred annuity ${}_{m|}a_{\overline{n}|}$: its payments do not start in the first period but after m periods, with m fixed in advance.



Periodic / monthly annuity $a^{(m)}$: the default periodicity is one year, but the unit payment can also be spread over m periods within the year.

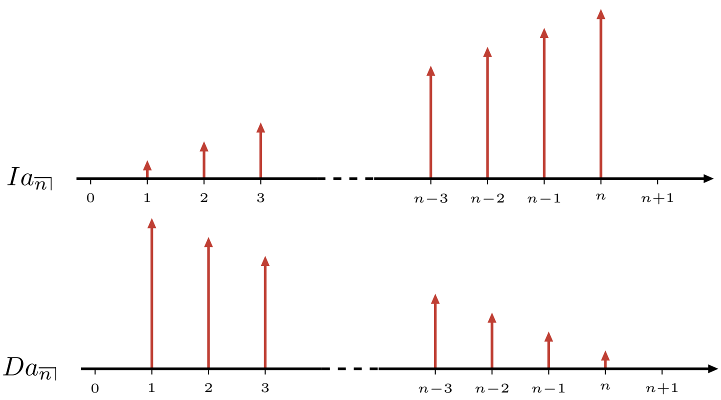


If $i^{(m)}$ represents the equivalent nominal (annual) interest rate with m periods per year, then $i^{(m)} = m\left((1+i)^{1/m} - 1\right)$.
Similarly, $d^{(m)}$ is the nominal discount rate consistent with d and m : $d^{(m)} = m\left(1-(1-d)^{1/m}\right)$.

$$\ddot{a}^{(m)}_{\overline{n}|} = \frac{1}{m}\sum_{k=0}^{mn-1} v^{\frac{k}{m}} = \frac{d}{d^{(m)}}\ddot{a}_{\overline{n}|} = \frac{1-v^n}{d^{(m)}} \approx \ddot{a}_{\overline{n}|} + \frac{m-1}{2m}\left(1-v^n\right)$$

$$a^{(m)}_{\overline{n}|} = \frac{1}{m}\sum_{k=1}^{mn} v^{\frac{k}{m}} = \frac{i}{i^{(m)}}a_{\overline{n}|} = \frac{1-v^n}{i^{(m)}} \approx a_{\overline{n}|} - \frac{m-1}{2m}\left(1-v^n\right)$$

Unit annuity a: it is used when constructing annuity formulas. For a constant annuity, the total amount paid each year is 1, regardless of m.
Dynamic annuity, increasing/decreasing Ia/Da: in its simplest form, it pays an amount that starts at 1 (n) and increases (decreases) each period arithmetically or geometrically. In the following example, the progression is arithmetic. The prefix I (increasing) is used to indicate increasing annuities and D (decreasing) for decreasing annuities.



$$(I\ddot{a})_{\overline{n}|} = 1 + 2v + \cdots + nv^{n-1} = \frac{1}{d}\left(\ddot{a}_{\overline{n}|} - nv^n\right) \tag{4}$$

with, we recall, $d = i/(1+i)$ and in arrears (immediate)

$$(Ia)_{\overline{n}|} = v + 2v^2 + \cdots + nv^n = \frac{1}{i}\left(\ddot{a}_{\overline{n}|} - nv^n\right)$$

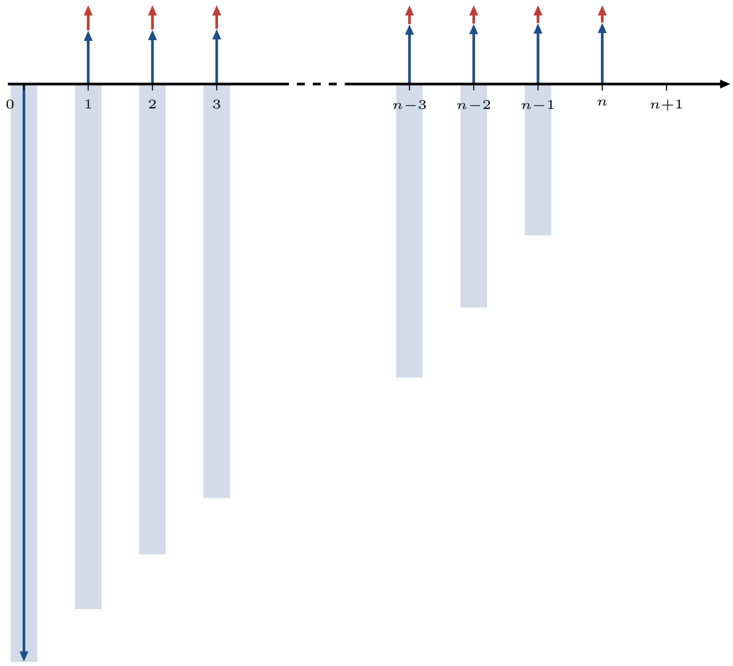
$$(D\ddot{a})_{\overline{n}|} = n + (n-1)v + \cdots + v^{n-1} = \frac{1}{d}\left(n - a_{\overline{n}|}\right)$$

and in arrears :

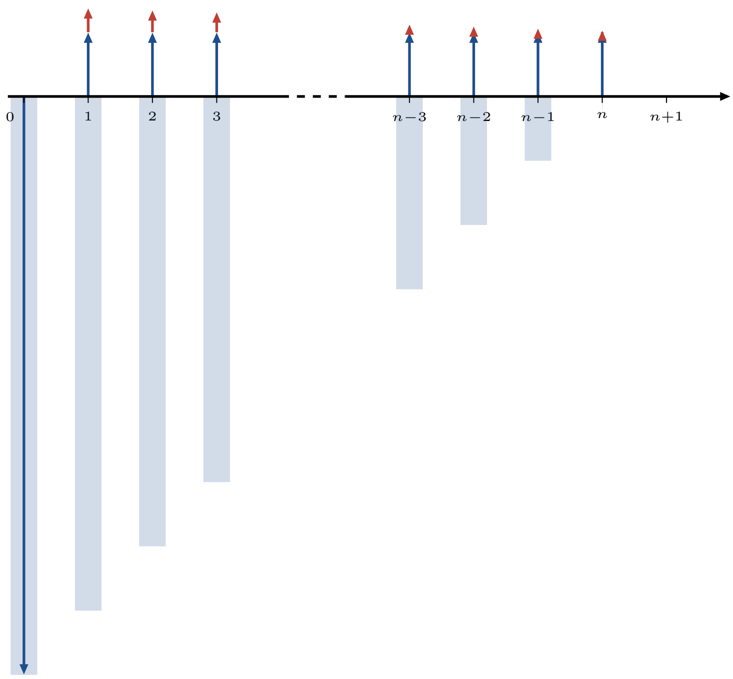
$$(Da)_{\overline{n}|} = nv + (n-1)v^2 + \cdots + v^n = \frac{1}{i}\left(n - a_{\overline{n}|}\right)$$

13 – Loans (Indivisible)

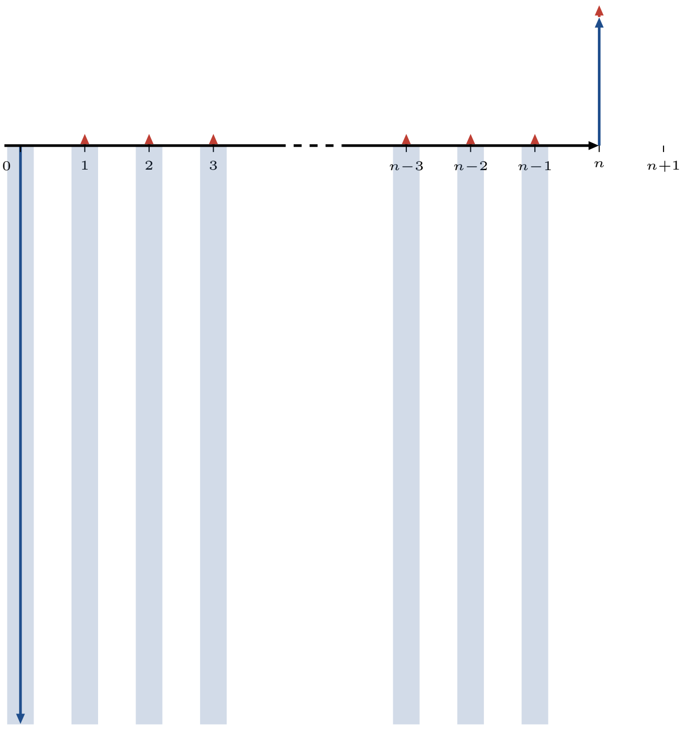
The main property of the loan is to consider separately the interest from the repayment (or amortization).
By constant repayment or constant annuity: the sum of the amortization and the interest at each period is constant.



By constant amortization.



By a bullet repayment, where the interest is constant. Only the interests are paid periodically until maturity, when the total repayment is made.



14 – Loan Amortization Schedule

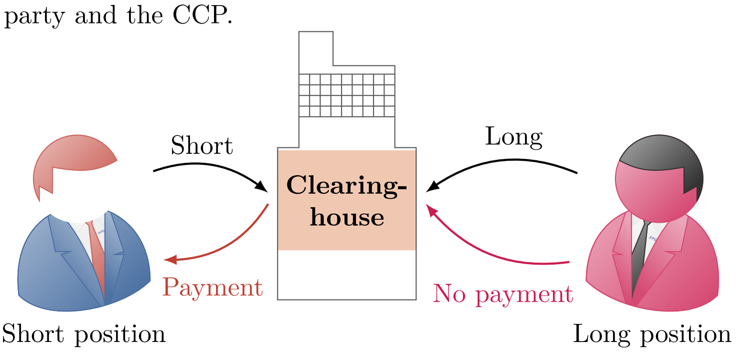
	In fine	**Constant amortizations**	**Constant annuities**
Outstanding principal S_k	$T_k = S_0$, $T_n = 0$	$S_0\left(1-\frac{k}{n}\right)$	$S_0\frac{1-v^{n-k}}{1-v^n}$
Interest U_k	$i \times S_0$	$S_0\left(1-\frac{k-1}{n}\right)i$	$K\left(1-v^{n-k+1}\right)$
Amortizations T_k	$T_k = O$, $T_n = S_0$	$\frac{S_0}{n}$	Kv^{n-k+1}
Annuity K_k	$K_k = iS_0$, $K_n = (1+i)S_0$	$\frac{S_0}{n}(1-(n-k+1)i)$	$K = S_0\frac{i}{1-v^n}$

15 – Market Functioning

The **Exchange** – a place of exchange – enables, in fact, the physical meeting between capital demanders and suppliers. The main listings concern **equities**, **bonds** (Fixed Income), and **commodities**. Listed are **securities** such as stocks or bonds, **funds** (Exchange Traded Funds that replicate equity indices, ETC or ETN that replicate more specific indices or commodities, SICAV or FCP, subscription bonds, warrant), **futures contracts**, **options**, **swaps**, and **structured products**.
The **Securities and Exchange Commission** (SEC) oversees :

- the protection of invested savings;
- the information of investors;
- the proper functioning of the markets.

Euronext (including Amsterdam, Brussels, Lisbon, and Paris) is the main stock exchange in France. Its competitors include **Deutsche Börse** (which includes Eurex, EEX) in Europe, or **ICE** (which includes NYSE (2012), NYBOT (2005), IPE (2001), LIFFE) and **CME Group** (including CBOT, NYMEX, COMEX) in the United States.
The **over-the-counter market** (**OTC**) represents a major share of volumes traded outside organized markets. Since the Pittsburgh G20 (2009), certain standardized OTC derivatives must be cleared through a central entity. These **CCPs** (Central Counterparties) thus play the role of **clearinghouses**: they replace the bilateral contract with two contracts between each party and the CCP.



16 – The Money Market

Short-term interest-bearing securities, traded on money markets, are generally at **discounted interest**. Nominal rates are then annual and calculations use **proportional rates** to adjust for durations less than one year. These securities are quoted or valued according to the discount principle and with a Euro-30/360 calendar convention.
In the American market, public debt securities are called : Treasury Bills (T-bills) : ZC < 1 year, Treasury Notes (T-notes): ZC < 10 years, Treasury Bonds (T-bonds) : coupon bonds with maturity > 10 years.
They are mainly :

- **BTF (fixed-rate Treasury bills, France) :** issued at 13, 26, 52 weeks, discounted rate, weekly auction, nominal 1 €, settlement at T+2.
- **Treasury bills > 1 year :** same rules as bonds (see next section).
- **Certificates of deposit (CDN) :** securities issued by banks at fixed/discounted rate (short term) or variable/post-discounted rate (long term), also called BMTN.
- **Eurodollars :** USD deposits outside the USA, formerly indexed on LIBOR, now declining.
- **Commercial paper :** unsecured short-term securities issued by large companies to finance their cash flow.

Price calculations of a fixed-rate Treasury bill with discounted interest
In the case of a discounted interest security according to the Euro-30/360 convention, the discount D is expressed as :

$$D = F \cdot d \cdot \frac{k}{360}$$

where F denotes the nominal value, d the annual discount rate used to value the discounted security, and k the maturity in days. If the discount rate d is known, then the price P is expressed as :

$$P = F - D = F\left(1 - d \cdot \frac{k}{360}\right)$$

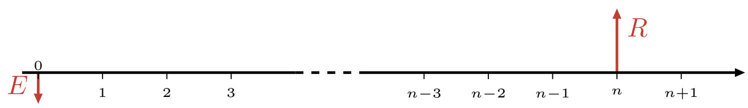
Similarly, if the price P is known, then the discount rate d is derived as :

$$d = \frac{F - P}{F} \cdot \frac{360}{k}$$

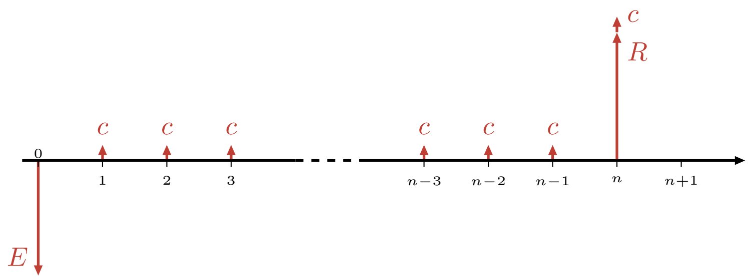
The main Futures Contracts : Federal Funds Futures (US), Three-Month SOFR Futures (US), ESTR Futures (EU), SONIA Futures (UK), Euribor Futures (EU).

17 – Bond Market

Bonds are long-term debt securities in which the issuer (central or local government, bank, borrowing company) promises the bondholder (the lender) to pay interest (**coupons**) periodically and to repay the **nominal value** (or face value, or principal) at maturity. As mentioned in the previous section, **Treasury bills with a maturity greater than one year** will be treated as bonds with maturities under 5 years because their functioning is similar.
Zero-coupon bonds : pay only the nominal value at maturity. With E the issue price and R its redemption value :



Coupon bonds : **Fixed-rate bonds** have a coupon rate that remains constant until maturity. Assuming a *bullet* repayment, with E the issue price, c the coupons, and R the redemption value, it can be illustrated as follows :



Indexed bonds (inflation-linked bonds) have coupons and sometimes also the nominal value indexed to inflation or another economic indicator, such as Treasury Assimilable Bonds indexed to inflation (OATi). The values of c vary.
Bonds with **floating rate**, **variable rate**, or **resettable rate** : have a coupon rate linked to a reference interest rate (for example, the euro short- term rate (€STR)).
Perpetual bonds have no maturity date; the principal is never repaid.
A distinction is often made between government bonds (Treasury bonds) and corporate bonds issued by private companies.
A bond is mainly defined by a **nominal value** F (Face Value), the **nominal rate** i, its duration or **maturity** n. In the default case, the bondholder lends the amount $E = F$ at issuance at time 0, receives each year a coupon $c = i \times F$, and at n, the principal or capital $R = F$ is returned. When $E = F$, the issue is said to be at par, and when $R = F$, the redemption is said to be at par.
The price of a bond is determined by the present value of the expected future cash flows (coupons and principal repayment) discounted at the market yield rate r.
The price calculation of bonds simply relies on the present value formula :

$$VP = \sum_{k=1}^{n} \frac{c}{(1+r)^k} + \frac{F}{(1+r)^n}$$

where :

- PV : price or present value of the bond,
- r : market interest rate for the relevant maturity.

For bonds with periodic coupons, the coupon is divided by the number of periods (m) per year and the formula becomes :

$$PV = \sum_{k=1}^{mn} \frac{c/m}{(1+r^{(m)})^k} + \frac{R}{(1+r^{(m)})^{mn}}$$

where c/m represents the periodic coupon payment and $r^{(m)}$ the periodic interest rate.
The bond yield is the value $r^{(m)}$, the equivalent rate of r over m periods in the year, which equates the present value PV_r with the current or market price of this bond.
The quotation of a bond is given as a percentage. Thus, a quotation of 97.9 on Euronext indicates a quoted value of $97.9/100 \times F$. It is quoted excluding **accrued coupons**, the portion of the next coupon to which the seller is entitled if the bond is sold before the payment of that coupon.

18 – Duration & Convexity

The Macaulay duration :

$$D = \sum_{t=1}^{n} t \cdot w_t, \quad \text{où} \quad w_t = \frac{PV(C_t)}{P}.$$

If the payment frequency is k per year, the duration expressed in years is obtained by dividing by k. The modified duration D^* :

$$D^* = \frac{D}{1+i}.$$

Which allows approximating the portfolio change ΔP in case of interest rate changes Δ_i

$$\Delta P \approx -P\ D^* \Delta_i$$

Similarly, the convexity

$$C = \frac{1}{P(i)} \times \frac{d^2 P(i)}{di^2},$$

which allows refining the approximation of ΔP

$$P(i + \Delta_i) \approx P(i) \left(1 - D^* \Delta_i + \frac{1}{2} C (\Delta_i)^2\right).$$

19 – CAPM

Capital Asset Pricing Model :

$$E(r_i) = r_f + \beta_i (E(r_m) - r_f)$$

- $E(r_i)$ is the expected return of asset i,
- r_f is the risk-free rate,
- $E(r_m)$ is the expected market return,
- β_i is the sensitivity coefficient of asset i with respect to market variations.

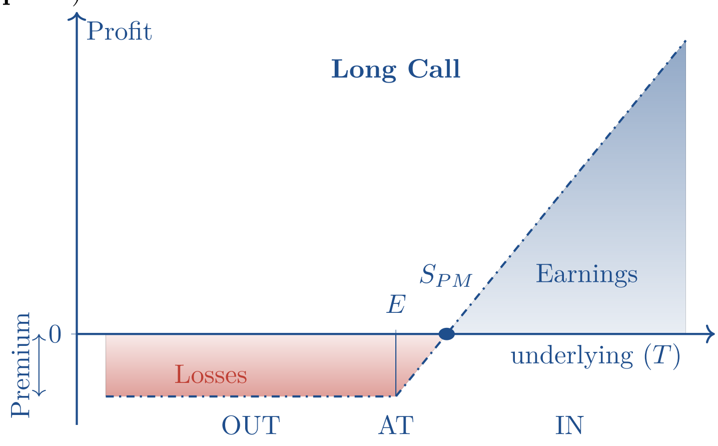
The coefficient β_i measures the volatility of asset i relative to the overall market.

20 – Derivatives Market

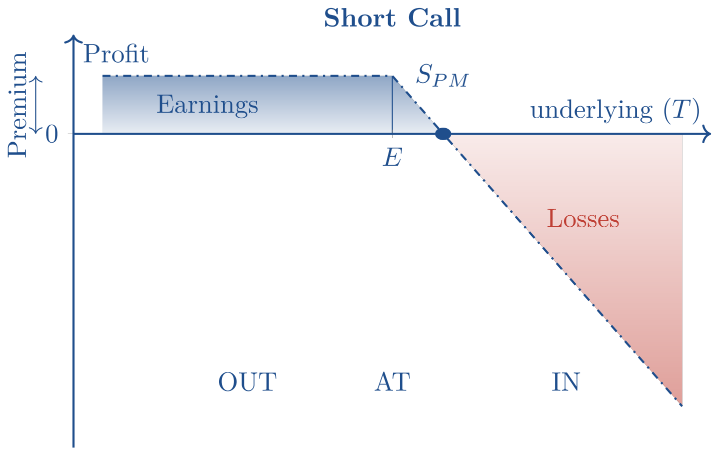
A **derivative contract** (or contingent asset) is a financial instrument whose value depends on an underlying asset or variable. Options are part of derivative contracts.
An **option** is a contract that gives the right (without obligation) to buy (call) or sell (put) an underlying asset at a fixed price (strike price) at a future date, in exchange for the payment of a premium. The buyer (long position) pays the premium; the seller (short position) receives it. **European option** (exercise possible only at maturity) and **American option** (exercise possible at any time until maturity).
Options listed on stocks are called *stock options*.

21 – Simple Strategies

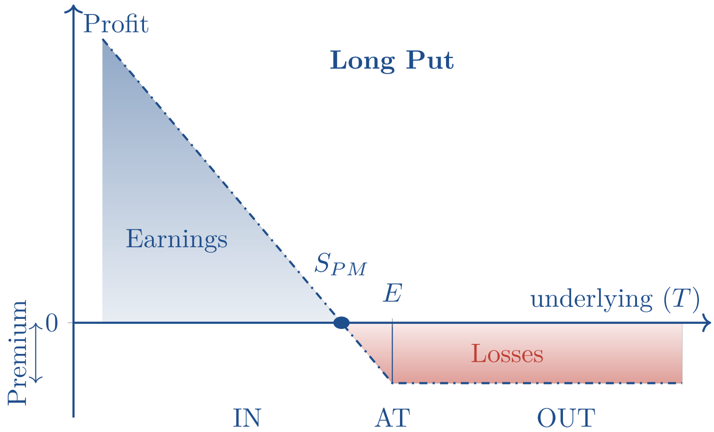
With T the maturity, K the strike price, S or S_T the underlying at maturity, the payoff is $\max(0, S_T - K) = (S_T - K)^+$. Letting C be the premium, the profit realized is $\max(0, S_T - K) - C$, with a profit if $(S_T < V_{PM} = K + C)$ (PM stands for **break-even point**).

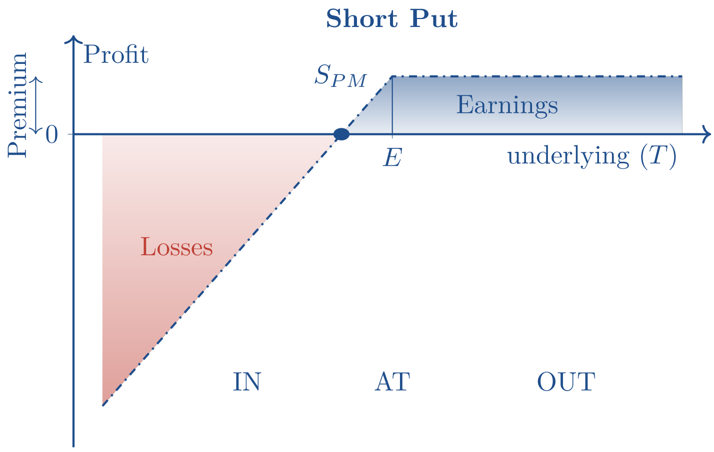


At maturity, the payoff is $\min(0, K - S_T) = -\max(0, S_T - K) = -(S_T - K)^+$ and the profit realized is $C - \max(0, S_T - K)$.



At maturity, the payoff is $\max(0, K - S_T) = (K - S_T)^+$. Letting P be the put premium, the profit realized is $\max(0, K - S_T) - P$, positive if $V_{PM} = K - P < S_T$.





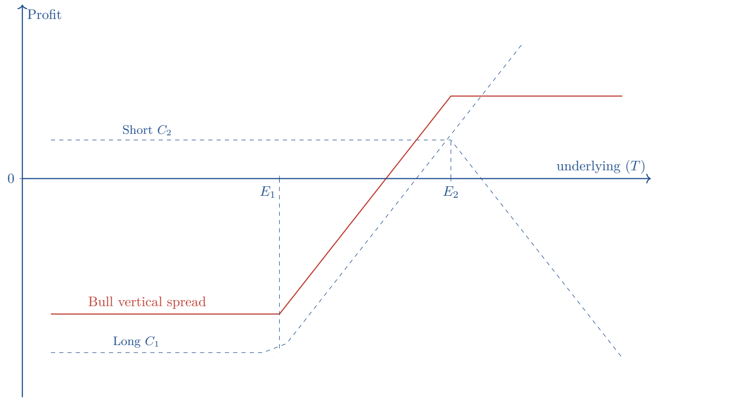
At maturity, the payoff is $\min(0, S_T - K) = -(K - S_T)^+$.

22 – Spread Strategies

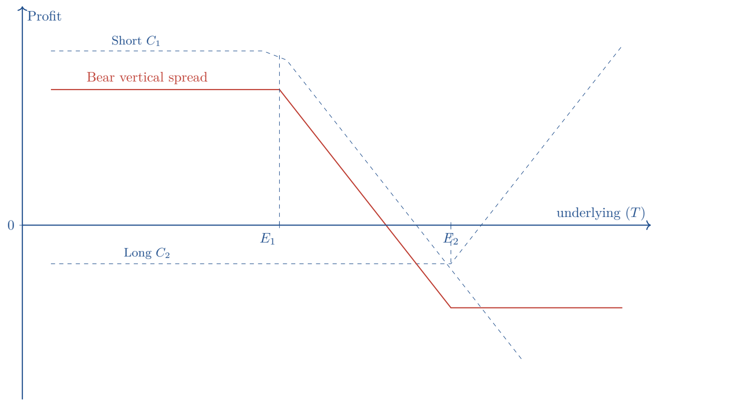
Spread strategy : uses two or more options of the same type (two call options or two put options). If the strike prices vary, it is a **vertical spread**. If the maturities change, it is a **horizontal spread**.

A vertical spread strategy involves a long position and a short position on call options on the same underlying asset, with the same maturity but different strike prices. We distinguish : **bull vertical spread** and **bear vertical spread**.

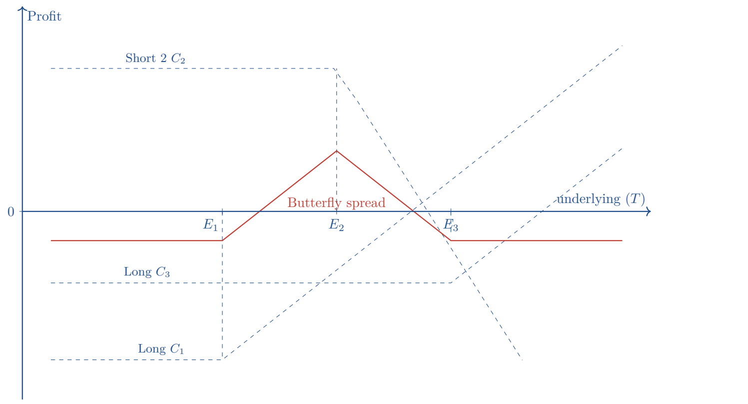
Bull vertical spread : anticipating a moderate rise in the underlying asset, the investor takes a long position on C_1 and a short position on C_2 under the condition $E_1 < E_2$. Net result at maturity :



Bear vertical spread : anticipating a moderate decline in the underlying asset, the investor sells the more expensive option and buys the cheaper one.



Butterfly spread : anticipates a small movement in the underlying asset. It is a combination of a bull vertical spread and a bear vertical spread. This strategy is suitable when large movements are considered unlikely. Requires a low initial investment.

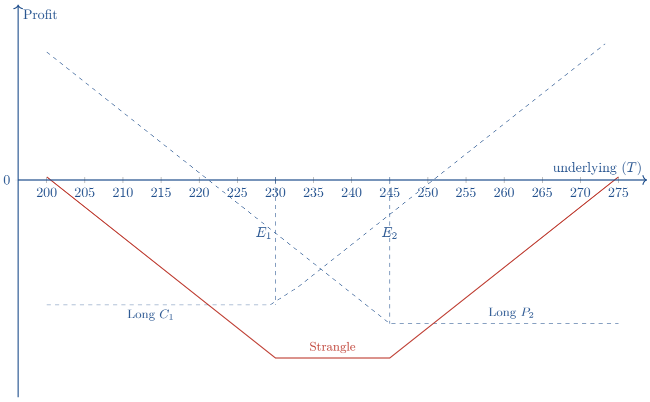


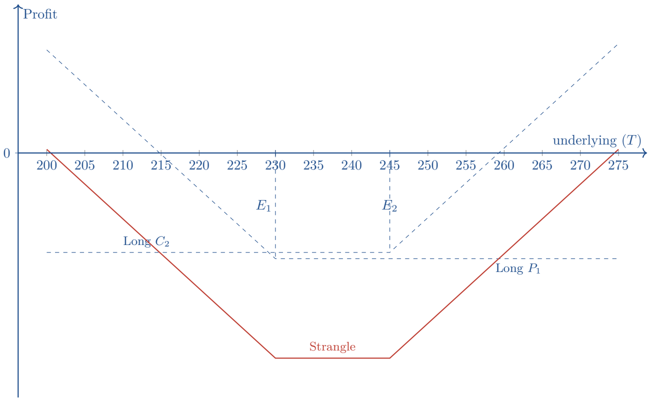
23 – Combined strategies

A **combined strategy** uses both call and put options. Notably, we distinguish between **straddles** and **strangles**.

A **straddle** combines the purchase of a call option and a put option with the same expiration date and strike price. This strategy bets on a large price movement, either upward or downward. The maximum loss occurs if the price at expiration is equal to the strike price.

A **strangle** is the purchase of a call and a put with the same expiration date but different strike prices. It assumes a very large movement in the value of the underlying asset.





24 – Absence of arbitrage opportunity

It is impossible to realize a risk-free gain from a zero initial investment. Thus, no risk-free profit is possible by exploiting price differences.

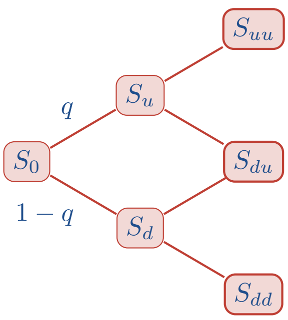
25 – Parity relation

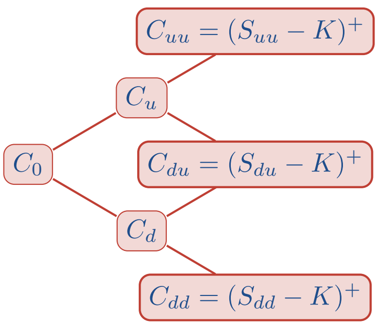
AAO implies the following relationship between the Call and the Put (stock) :

$$S_t - C_t + P_t = Ke^{-i_f \cdot \tau}$$

26 – The Cox-Ross-Rubinstein model

It is based on a discrete-time process with two possible price movements at each period : an increase (factor u) or a decrease (factor d), with $u > 1 + i_f$ and $d < 1 + i_f$. The price at $t = 1$ is then $S_1^u = S_0 u$ or $S_1^d = S_0 d$, according to a probability q or $1 - q$.





This model extends to n periods with $n + 1$ possible prices for S_T. At expiration, the value of a call option is given by $C_1^u = (S_1^u - K)^+$ and $C_1^d = (S_1^d - K)^+$.

Absence of arbitrage opportunity implies

$$d < 1 + i_f < u$$

and a risk-neutral probability

$$q = \frac{(1 + i_f) - d}{u - d}$$

Call price (with $S_1^d < K < S_1^u$) :

$$C_0 = \frac{1}{1 + i_f}\left[qC_1^u + (1 - q)C_1^d\right]$$

We can also construct a replication portfolio composed of Δ shares and B bonds, such that :

$$\begin{cases} \Delta = \frac{S_1^u - K}{S_1^u - S_1^d}, \\ B = \frac{-S_1^d}{1+i_f} \cdot \Delta \end{cases} \quad \Rightarrow \quad \Pi_0 = \Delta S_0 + B$$

Put price :

$$P_0 = \frac{1}{1 + i_f}\left[qP_1^u + (1 - q)P_1^d\right]$$

Determination of q, u, d : By calibrating the model to match the first moments of the return under the risk-neutral probability (expected value i_f, variance $\sigma^2 \delta t$), we obtain :

$$e^{i_f \delta t} = qu + (1-q)d, \qquad qu^2 + (1-q)d^2 - [qu + (1-q)d]^2 = \sigma^2 \delta t$$

With the constraint $u = \frac{1}{d}$, we obtain :

$$\begin{aligned} q &= \frac{e^{-i_f \delta t} - d}{u - d} \\ u &= e^{\sigma\sqrt{\delta t}} \\ d &= e^{-\sigma\sqrt{\delta t}} \end{aligned}$$

27 – The Black & Scholes Model

Assumptions of the model

- The risk-free rate R is constant. We define $i_f = \ln(1 + R)$, which implies $(1 + R)^t = e^{i_f t}$.
- The stock price S_t follows a geometric Brownian motion :
$$dS_t = \mu S_t dt + \sigma S_t dW_t$$
$$S_t = S_0 \exp\left(\sigma W_t + \left(\mu - \frac{1}{2}\sigma^2\right) t\right)$$
- No dividend during the option's lifetime.
- The option is "European" (exercised only at maturity).
- Frictionless market : no taxes or transaction costs.
- Short selling is allowed.

The Black-Scholes-Merton equation for valuing a derivative contract f is :

$$\frac{\partial f}{\partial t} + i_f S \frac{\partial f}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 f}{\partial S^2} = i_f f$$

At maturity, the price of a call option is $C(S, T) = \max(0, S_T - K)$, and that of a put option is $P(S, T) = \max(0, K - S_T)$.

Determinants	**call**	**put**
Underlying price	+	-
Strike price	-	+
Maturity (or time)	+ (-)	+ (-)
Volatility	+	+
Short-term interest rates	+	-
Dividend payment	-	+

The analytical solutions are :

$$\begin{aligned} C_t &= S_t \Phi(d_1) - Ke^{-i_f \tau}\Phi(d_2) \\ P_t &= Ke^{-i_f \tau}\Phi(-d_2) - S_t \Phi(-d_1) \end{aligned}$$

or :

$$d_1 = \frac{\ln(S_t/K) + (i_f + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \quad d_2 = d_1 - \sigma\sqrt{\tau}$$

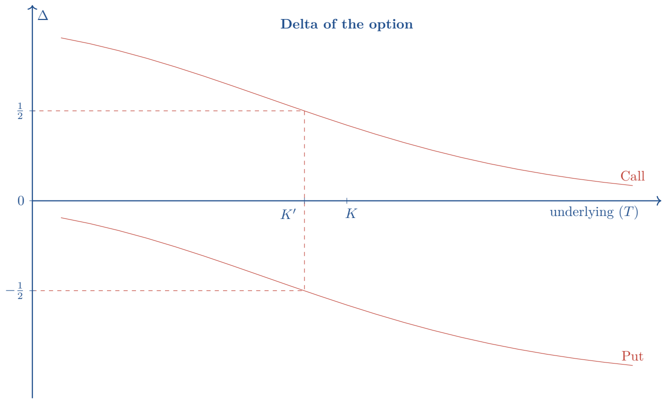
- **Delta** Δ : variation in the option price depending on the underlying.
- **Gamma** Γ : delta sensitivity.
- **Thêta** Θ : sensitivity to time.
- **Véga** $\mathcal{V}$: sensitivity to volatility.
- **Rho** ρ : interest rate sensitivity.

The **Delta** measures the impact of a change in the underlying asset :

$$\Delta_C = \frac{\partial C}{\partial S} = \Phi(d_1), \quad \Delta \in (0,1)$$
$$\Delta_P = \frac{\partial P}{\partial S} = \Phi(d_1) - 1, \quad \Delta \in (-1,0)$$

The global Delta of a portfolio Π with weights ω_i is :

$$\frac{\partial \Pi}{\partial S_t} = \sum_{i=1}^{n} \omega_i \Delta_i$$



28 – The Yield Curve

The **yield curve**, or the curve of returns, or $r_f(\tau)$, provides a graphical representation of risk-free interest rates as a function of maturity (or term). It is also called the **zero-coupon** yield curve, referring to a type of risk-free bond with no coupons (a debt composed only of two opposite cash flows, one at t_0 and the other at T). This curve also provides insight into market expectations regarding future interest rates (forward rates).

29 – The Nelson-Siegel and Svensson models

The **Nelson-Siegel** functions take the form

$$y(m) = \beta_0 + \beta_1 \frac{[1-\exp(-m/\tau)]}{m/\tau} + \beta_2 \left(\frac{[1-\exp(-m/\tau)]}{m/\tau} - \exp(-m/\tau) \right)$$

where $y(m)$ and m are as above, and β_0, β_1, β_2, and τ are parameters :

- β_0 is interpreted as the long-term level of interest rates (the coefficient is 1, it is a constant that does not decrease),
- β_1 is the short-term component, noting that :

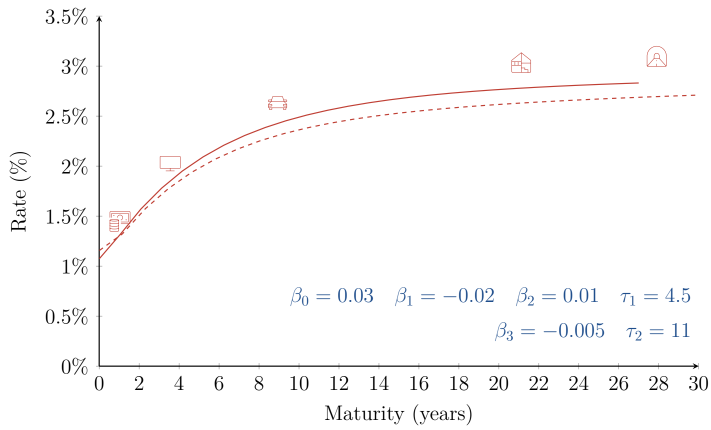
  $$\lim_{m \to 0} \frac{[1-\exp(-m/\tau)]}{m/\tau} = 1$$

  It follows that the overnight rate such as €str will equal $\beta_0 + \beta_1$ in this model.
- β_2 is the medium-term component (it starts at 0, increases, then decreases back toward zero — i.e., bell-shaped),
- τ is the scale factor on maturity; it determines where the term weighted by β_2 reaches its maximum.

Svensson (1995) adds a second bell-shaped term; this is the Nelson–Siegel–Svensson model. The additional term is :

$$+\beta_3 \left(\frac{[1-\exp(-m/\tau_2)]}{m/\tau_2} - \exp(-m/\tau_2) \right)$$

and the interpretation is the same as for β_2 and τ above; it allows for two inflection points on the yield curve.
These Nelson-Siegel and Svensson functions have the advantage of behaving well in the long term and being easy to parameterize. They are illustrated in the figure where the pictograms represent the different usual maturities for this type of property or investment. They allow for the modeling of a broad yield curve. Once adjusted, the user can then evaluate assets or define various sensitivity measures.



30 – Vasicek model

Under a risk-neutral probability $\mathbb{Q}$, the short rate (r_t) follows an Ornstein–Uhlenbeck process with constant coefficients :

$$dr_t = \kappa(\theta - r_t)\,dt + \sigma\,dW_t, \quad r_0 \in \mathbb{R}$$

or :

- $\kappa > 0$ is the speed of mean reversion,
- θ is the long-term mean level,
- $\sigma > 0$ is the volatility,
- W_t is a standard Brownian motion under $\mathbb{Q}$.

The EDS solution (application of Itô's lemma to $Y_t = r(t)e^{\kappa t}$) :

$$r_t = r_s e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) + \sigma \int_s^t e^{-\kappa(t-u)} dW_u$$

So :

$$\mathbb{E}_{\mathbb{Q}}[r_t \mid \mathcal{F}_s] = r_s e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)})$$
$$\mathrm{Var}_{\mathbb{Q}}[r_t \mid \mathcal{F}_s] = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right)$$

The process (r_t) is Gaussian; negative rates are possible.

31 – Price of a zero-coupon bond (Vasicek)

The price at time t of a zero-coupon bond maturing at T is given by :

$$ZC(t,T) = A(t,T)\, e^{-B(t,T)\, r_t}$$

où :

$$B(t,T) = \frac{1 - e^{-\kappa(T-t)}}{\kappa}$$

$$A(t,T) = \exp\left[\left(\theta - \frac{\sigma^2}{2\kappa^2}\right)(B(t,T) - (T-t)) - \frac{\sigma^2}{4\kappa} B(t,T)^2\right]$$

This formulation is possible due to the fact that $\int_t^T r_s ds$ is a Gaussian random variable conditional on $\mathcal{F}_t$.

$$ZC(t,T) = \mathbb{E}_{\mathbb{Q}}\left[\exp\left(-\int_t^T r_s\, ds\right) \Big| \mathcal{F}_t\right]$$

---

32 – Cox–Ingersoll–Ross (CIR) model

Under the risk-neutral measure $\mathbb{Q}$, the short rate (r_t) follows the dynamics :

$$dr_t = \kappa(\theta - r_t)\, dt + \sigma\sqrt{r_t}\, dW_t, \quad r_0 \geq 0$$

with :

- $\kappa > 0$: mean reversion speed,
- $\theta > 0$: long-term level,
- $\sigma > 0$: volatility,
- W_t : Brownian motion under $\mathbb{Q}$.

So :

- The square root $\sqrt{r_t}$ guarantees $r_t \geq 0$ if $2\kappa\theta \geq \sigma^2$ (Feller condition).
- The process (r_t) is a non-Gaussian diffusion process but with continuous trajectories.
- The rate is **mean-reverting** around θ.

Thus, the process (r_t) is a diffusion with explicit conditional distributions (under $\mathbb{Q}$) :

For $s < t$, the variable r_t follows a non-central χ^2 distribution :

$$r_t \mid \mathcal{F}_s \sim c \cdot \chi^2_d(\lambda)$$

with :

- $c = \dfrac{\sigma^2(1 - e^{-\kappa(t-s)})}{4\kappa}$
- $d = \dfrac{4\kappa\theta}{\sigma^2}$: degrees of freedom
- $\lambda = \dfrac{4\kappa e^{-\kappa(t-s)} r_s}{\sigma^2(1 - e^{-\kappa(t-s)})}$

and

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}[r_t \mid \mathcal{F}_s] &= r_s e^{-\kappa(t-s)} + \theta(1 - e^{-\kappa(t-s)}) \\
\mathrm{Var}_{\mathbb{Q}}[r_t \mid \mathcal{F}_s] &= \frac{\sigma^2 r_s e^{-\kappa(t-s)}(1 - e^{-\kappa(t-s)})}{\kappa} \\
&\quad + \frac{\theta\sigma^2}{2\kappa}(1 - e^{-\kappa(t-s)})^2
\end{aligned}$$

33 – Price of a zero-coupon bond (CIR)

In the CIR model, the price of a zero-coupon bond at time t with maturity T is given by :

$$ZC(t,T) = A(t,T) \cdot e^{-B(t,T)\, r_t}$$

with :

$$B(t,T) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma + \kappa)(e^{\gamma(T-t)} - 1) + 2\gamma}$$

$$A(t,T) = \left[\frac{2\gamma e^{\frac{(\kappa+\gamma)}{2}(T-t)}}{(\gamma + \kappa)(e^{\gamma(T-t)} - 1) + 2\gamma}\right]^{\frac{2\kappa\theta}{\sigma^2}}$$

or :

$$\gamma = \sqrt{\kappa^2 + 2\sigma^2}$$

---

34 – Swaption, Black model

A **swaption** is an option on an interest rate swap. It gives the right (but not the obligation) to enter into a swap at a future date T.

- **Payer swaption**: right to *pay the fixed rate* and *receive the floating rate.*
- **Receiver swaption**: right to *receive the fixed rate* and *pay the floating rate.*

Notation :

- T : swaption exercise date
- K : fixed rate (strike)
- $S(T)$: swap rate on the date T
- $A(T)$: present value of future fixed flows.
- σ : swap rate volatility

The Black (1976) model is an adaptation of the Black–Scholes model for interest rate products. Here, the swap rate $S(T)$ plays the role of the underlying asset, with a European option-type payoff.

Black's formula for a payer swaption :

$$\mathrm{SW}_{\text{payer}} = A(T)\,[S_0 N(d_1) - K N(d_2)]$$

or :

$$d_1 = \frac{\ln(S_0/K) + \frac{1}{2}\sigma^2 T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

and $N(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Formula for a receiver swaption :

$$\mathrm{SW}_{\text{receiver}} = A(T)\,[K N(-d_2) - S_0 N(-d_1)]$$

35 – Life Table Notations

Age x, y, z...

l_x is the number of people alive, relative to an initial cohort, at age x (or y, z...)

ω is the age limit of mortality tables.

$d_x = l_x - l_{x+1}$ is the number of people who die between the age x and age $x+1$.

q_x is the probability of death between the ages of x et age $x+1$.

$$q_x = d_x / l_x$$

p_x is the probability that the individual aged x survives age $x+1$.

$$p_x + q_x = 1$$

Likewise, ${}_nd_x = d_x + d_{x+1} + \cdots + d_{x+n-1} = l_x - l_{x+n}$ shows the number of people who die between the age x and age $x+n$.

${}_nq_x$ is the probability of death between the ages of x and age $x+n$.

$${}_nq_x = {}_nd_x / l_x$$

${}_np_x$ is the probability of a person of age x to survive the age $x+n$.

$${}_np_x = l_{x+n} / l_x$$

${}_{m|}q_x$, the probability that the individual of age x dies in the $m+1^{th}$ year.

$${}_{m|}q_x = \frac{d_{x+m}}{l_x} = \frac{l_{x+m} - l_{x+m+1}}{l_x}$$

e_x is the life expectancy for a person still alive at the age x. This is the number of birthdays you hope to live.

$$e_x = \sum_{t=1}^{\infty} {}_tp_x$$

36 – Coefficient or commutations

These coefficients or commutations established by actuarial functions which depend on a mortality table and a rate i ($v = 1/(1+i)$) to establish the actuarial table.

$$D_x = l_x . v^x$$

can be seen "as" the actualized number of survivors. The sums

$$N_x = \sum_{k \geq 0} D_{x+k} = \sum_{k=0}^{\omega - x} D_{x+k}$$

$$S_x = \sum_{k \geq 0} N_{x+k} = \sum_{k \geq 0} (k+1).D_{x+k}$$

will be used to simplify the calculations. Likewise

$$C_x = d_x v^{x+1}$$

can be seen "as" the number of deaths discounted to age x. The sums

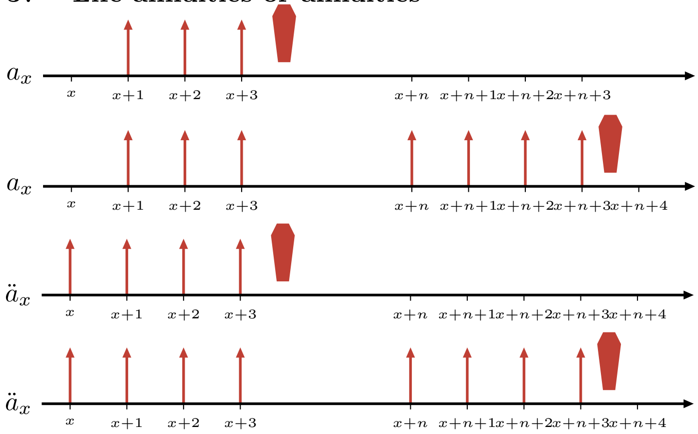
$$M_x = \sum_{k=0}^{\omega - x} C_{x+k}$$

$$R_x = \sum_{k=0}^{\omega - x} M_{x+k}$$

will be used to simplify the calculations.

The coefficients D_x N_x and S_x will be used for calculations on operations in case of life and C_x M_x and R_x for operations in case of death.

37 – Life annuities or annuities



$$a_x = \sum_{k=1}^{\infty} {}_kp_x v^k = \ddot{a}_x - 1 = \frac{N_{x+1}}{D_x}$$

$$\ddot{a}_x = \sum_{k=0}^{\infty} {}_kp_x v^k = \frac{N_x}{D_x}$$

If the periodicity corresponds to m periods per year:

$$\ddot{a}_x^{(m)} = \sum_{k=0}^{\infty} \frac{1}{m} {}_{\frac{k}{m}}p_x v^{\frac{k}{m}} \approx \ddot{a}_x - \frac{m-1}{2m}$$

Similarly, if he pays $1/m$ at the start of the m periods

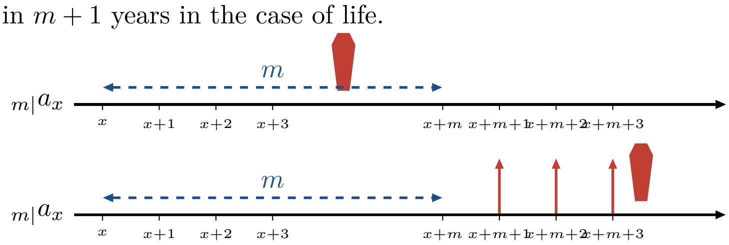
$$a_x^{(m)} \approx a_x + \frac{m-1}{2m}$$

Temporary life annuities. Whole life annuity guaranteed for n years

$$a_{x:\overline{n}|} = \sum_{k=1}^{n} {}_kp_x v^k = \frac{N_{x+1} - N_{x+n+1}}{D_x}$$

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} {}_kp_x v^k = \frac{N_x - N_{x+n}}{D_x}$$

Deferred life annuities ${}_{m|}a_x$ represent the annuities on the individual of age x deferred m years. The first payment occurs in $m+1$ years in the case of life.



38 – Death or survival benefits

Death benefits (Whole life insurance noted SP_x or A_x)

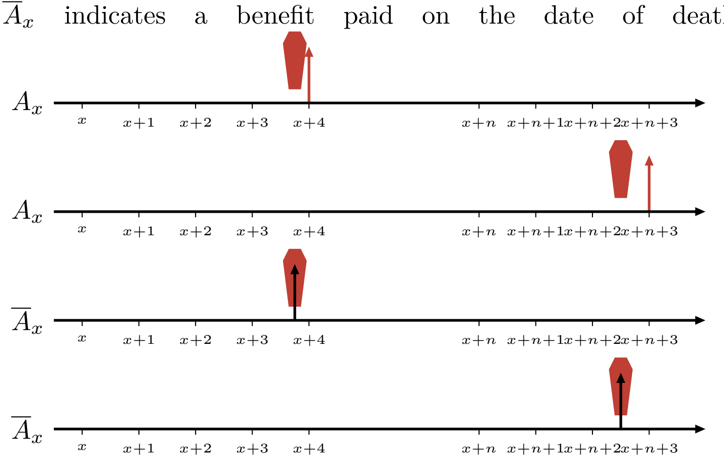
A_x indicates a death benefit at the end of the year of death (amount of 1), regardless of the date of occurrence, for an individual insured at age x at the time of subscription.

$A_{x:\overline{n}|}$ denotes a capital paid upon death if it occurs and at the latest in n years (Endowment).

$A^1_{x:\overline{n}|}$ denotes a death benefit paid if x dies within the next n years (Term insurance).

$A^{(12)}_x$ indicates a benefit payable at the end of the month of death.

$\overline{A}_x$ indicates a benefit paid on the date of death.

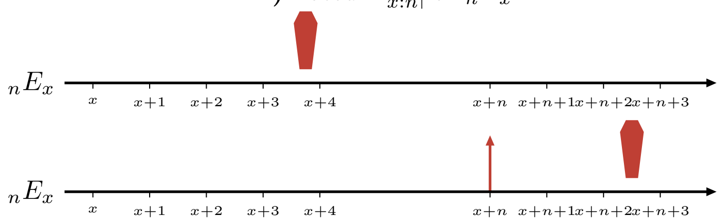


Whole life benefit

$$A_x = \sum_{k=0}^{\infty} {}_{k|}q_x \ \nu^{k+1} = \frac{M_x}{D_x}$$

$$A^1_{x:\overline{n}|} = \sum_{k=0}^{n-1} {}_{k|}q_x \ \nu^{k+1} = \frac{M_x - M_{x+n}}{D_x}$$

Deferred capital (Pure Endowment, unique capital in the event of survival) noted $A_{x:\overline{n}|}^{\ \ 1}$ or ${}_nE_x$.



$$= {}_nE_x = {}_np_x.v^n = \frac{l_{x+n}}{l_x}.v^n = \frac{D_{x+n}}{D_x}$$

Death benefit with payment of the capital in the event of survival (Endowment)

$$A_{x:\overline{n}|} = A^1_{x:\overline{n}|} + A_{x:\overline{n}|}^{\ \ 1}$$

39 – Life insurance on several individuals

a_{xyz} is an annual annuity, paid at the end of the first year and for as long as they live (x), (y) and (z).

$a_{\overline{xyz}}$ is an annual annuity, paid at the end of the first year and for as long as they live (x), (y) or (z).

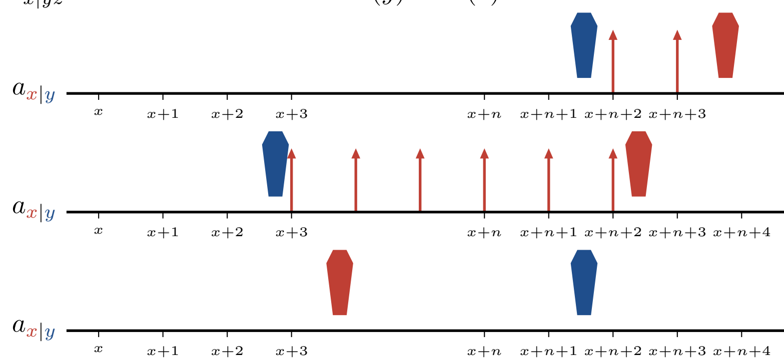
$$a_{\overline{xy}} = a_y + a_x - a_{xy}$$

A_{xyz} is an insurance that comes into effect at the end of the year of the first death of (x), (y) and (z).

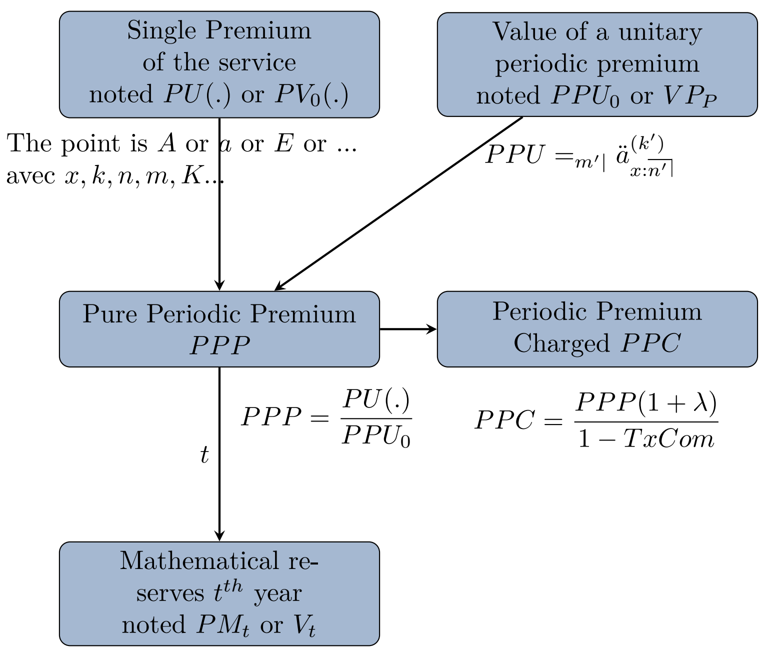
The vertical bar indicates conditionality :

$a_{x|y}$ is a survivor's annuity which benefits (x) after the death of (y).

$A_{x|yz}$ is a first-to-die insurance (y) and (z).



40 – Simplified pricing schemes for periodic premiums and reserves



$V_t = PV_t(.) - PPP \times PPU_t$

$PV_t(.)$ recalculated with $x+t, k, n, m-t, K...\ if\ t \geq m$

or $x+t, k, n-(t-m), m=0, K...\ si\ t > m$

likewise $PPU_t = {}_{(m'-t)^+|}\ \ddot{a}^{(k')}_{x+t:\overline{n'-(t-m)^+}|}$

41 – Axiomatic

A **universe** Ω, is the set of all possible outcomes that can be obtained during a random experiment.
The **random event** is an event ω_i of the universe whose outcome (the result) is not certain.
The **elementary event :**

- two distinct elementary events ω_i and ω_j are incompatible,
- the union of all the elementary events of the universe Ω corresponds to certainty.

The **sets** :

- $E = \{\omega_{i1}, \ldots, \omega_{ik}\}$ a subset of Ω (k elements).
- $\overline{E}$ the complement of E,
- $E \cap F$ the intersection of E and F,
- $E \cup F$ the union of E and F,
- $E \setminus F = E \cap \overline{F}$ E minus F,
- $\varnothing$ the impossible or empty event.

Let E be a set. We call **trib** or $\sigma-$**algebra** on E, a set $\mathcal{A}$ of parts of E which satisfies :

- $\mathcal{A} \neq \varnothing$,
- $\forall A \in \mathcal{A}, \overline{A} \in \mathcal{A}$,
- if $\forall n \in \mathbb{N}$, $A_n \in \mathcal{A}$ then $\cup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

We call **probability** $\mathbb{P}$ any application of the set of events $\mathcal{A}$ in the interval $[0, 1]$, such that :

$$\mathbb{P} : \mathcal{A} \mapsto [0, 1]$$

satisfying the following properties (or axioms) :

(P1) $A \subseteq \mathcal{A}$ then $\mathbb{P}(A) \geq 0$,

(P2) $\mathbb{P}(\Omega) = 1$,

(P3) $A, B \subseteq \mathcal{A}$, if $A \cap B = \varnothing$ then $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

The **probability space** is defined by

$$\{\Omega, \mathcal{A}, \mathbb{P}(.)\}$$

The **Poincaré equality** is written :

$$\forall A \in F, \forall B \in F, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

42 – Bayes

In probability theory, the **conditional probability** of an event A, given that another event B of non-zero probability has occurred.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

The real $\mathbb{P}(A|B)$ is read as 'probability of A, given B. Bayes' theorem allows us to write :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

43 – Random variables

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. We call **random variable** X from Ω to $\Re$ any measurable function $X : \Omega \mapsto \Re$.

$$\{X \leq x\} \equiv \{e \in \Omega \mid X(e) \leq x\} \in \mathcal{A}$$

The set of events of Ω is often not explicit.
The **distribution function** (F_X) of a real random variable characterizes its probability distribution.

$$F_X(x) = \mathbb{P}(X \leq x), x \in \Re$$

where the right-hand side represents the probability that the real random variable X takes a value less than or equal to x. The probability that X is in the interval $[a, b]$ is therefore, if $a < b$, $\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a)$
A probability law has a **probability density** f, if f is a function defined on $\mathbb{R}^+$, Lebesgue integral, such that the probability of the interval $[a, b]$ is given by

$$\mathbb{P}(a < X \leq b) = \int_a^b f(x)\mathrm{d}x \text{ for all numbers such that } a < x < b.$$

44 – Expectations

The mathematical expectation in the discrete case (discrete qualitative or quantitative variables) :

$$\mathbb{E}[X] = \sum_{j\in\mathbb{N}} x_j \mathbb{P}(x_j)$$

where $\mathbb{P}(x_j)$ is the probability associated with each event x_i.
The mathematical expectation in the continuous case :

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x.f(x)dx$$

where f denotes the density function of the random variable x, defined in our case on $\mathbb{R}$. When it comes to sum or integral, the expectation is linear, that is to say :

$$\mathbb{E}[c_0 + c_1X_1 + c_2X_2] = c_0 + c_1\mathbb{E}[X_1] + c_2\mathbb{E}[X_2]$$

$$\mathbb{E}[X] = \int x.f(x)dx = \int_0^1 F^{-1}(p)dp = \int \overline{F}(x)dx$$

45 – Convolution or law of sum

The convolution of two functions f and g, denoted $(f * g)(x)$, is defined by :

$$(f * g)(x) = \int f(t)g(x - t)\,dt$$

Convolution measures how $f(t)$ and $g(t)$ interact at different points while taking into account the shift (or translation) between If X and Y are two independent random variables with respective densities f_X and f_Y, then the density of the sum $Z = X + Y$ is given by :

$$f_Z(x) = (f_X * f_Y)(x) = \int_{-\infty}^{+\infty} f_X(t)\, f_Y(x - t)\,dt.$$

46 – Compound law or frequency/gravity model

Let N be a discrete random variable in $\mathbb{N}^+$, (X_i) a sequence of *iid* random variables with finite expectation and variance, then for $S = \sum_{i=1}^{N} X_i$:

$$\mathbb{E}(S) = \mathbb{E}(\mathbb{E}[S \mid N]) = \mathbb{E}(N.\mathbb{E}(X_1)) = \mathbb{E}(N).\mathbb{E}(X_1)$$

$$Var(S) = \mathbb{E}(Var[S \mid N]) + Var(\mathbb{E}[S \mid N])$$

47 – Fundamental theorems

Let X be a real random variable defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, and assumed to be almost surely positive or zero. The **Markov Inequality** gives :

$$\forall \alpha > 0, \mathbb{P}(X \geq \alpha) \leqslant \frac{\mathbb{E}[X]}{\alpha}.$$

The **Bienaymé-Tchebychev inequality**: For any strictly positive real number α, with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] = \sigma^2$

$$\mathbb{P}\left(|X - \mu| \geq \alpha\right) \leq \frac{\sigma^2}{\alpha^2}.$$

The **weak law of large numbers** considers a sequence $(X_i)_{i \geq n \in \mathbb{N}^*}$ of independent random variables defined on the same probability space, having the same finite expectation and variance denoted respectively $\mathbb{E}[X]$ and $\mathrm{Var}(X)$.

$$\forall \varepsilon > 0, \quad \lim_{n \to +\infty} \mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mathbb{E}[X]\right| \geq \varepsilon\right) = 0$$

Consider a sequence $(X_n)_{n \in \mathbb{N}}$ of independent random variables that follow the same probability law, integrable, i.e. $E(|X_0|) < +\infty$.

Using the notations, the **strong law of large numbers** specifies that $(Y_n)_{n \in \mathbb{N}}$ converges to $E(X)$ "almost surely".

$$\mathbb{P}\left(\lim_{n \to +\infty} Y_n = E(X)\right) = 1$$

Consider the sum $S_n = X_1 + X_2 + \cdots + X_n$.

$$Z_n \;=\; \frac{S_n - n\mu}{\sigma\sqrt{n}} \;=\; \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}},$$

the expectation and the standard deviation of Z_n are respectively 0 and 1: the variable is thus said to be centered and reduced.

The **central limit theorem** then states that the distribution of Z_n converges in law to the reduced centered normal distribution $\mathcal{N}(0, 1)$ as n tends to infinity. This means that if Φ is the distribution function of $\mathcal{N}(0, 1)$, then for any real number z :

$$\lim_{n \to \infty} \mathrm{P}(Z_n \leq z) = \Phi(z),$$

or, equivalently :

$$\lim_{n \to \infty} \mathrm{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \leq z\right) = \Phi(z)$$

48 – Multidimensional variables

A probability law is said to be **multidimensional**, or $n-$dimensional, when the law describes several (random) values of a random phenomenon. The multidimensional character thus appears during the transfer, by a random variable, of the probabilistic space $(\Omega, \mathcal{A})$ to a numerical space E^n of dimension n.

Let X be a random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, with values in $\mathbb{R}^n$ equipped with the real Borel tribe product $\mathcal{B}(\mathbb{R})^{\otimes n}$. The law of the random variable X is the probability measure $\mathbb{P}_X$ defined by :

$$\mathbb{P}_X(B) = \mathbb{P}\big(X^{-1}(B)\big) = \mathbb{P}(X \in B).$$

for everything $B \in \mathcal{B}(\mathbb{R})^{\otimes n}$.

The Cramer-Wold theorem ensures that the ($n-$dimensional) law of this random vector is entirely determined by the (one-dimensional) laws of all linear combinations of these components:

$$\sum_{i=1}^{n} a_i X_i \text{ for all } a_1, a_2, \ldots, a_n$$

49 – Marginal law

The probability distribution of the i^{th} coordinate of a random vector is called the i^{th} marginal distribution. The **marginal distribution** $\mathbb{P}_i$ of $\mathbb{P}$ is obtained by the formula :

$$\mathbb{P}_i(B) = \mathbb{P}_{X_i}(B) = \iint \mathbb{1}_{\omega_i \in B} \mathbb{P}(\mathrm{d}(\omega_1, \ldots, \omega_n)), \forall B \in \mathcal{B}(\mathbb{R}).$$

The marginal laws of an absolutely continuous law are expressed using their marginal densities.

The conditional density function X_2 given the value x_1 of X_1, can be written :

$$f_{X_2}(x_2 \mid X_1 = x_1) = \frac{f_{X_1,X_2}(x_1, x_2)}{f_{X_1}(x_1)},$$

$$f_{X_2}(x_2 \mid X_1 = x_1) f_{X_1}(x_1) = f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1 \mid X_2 = x_2) f_{X_2}$$

50 – Independence

$(X_1, X_2, \ldots, X_n)$ is a family of **independent random variables** if one of the following two conditions is met :

$$\forall (A_1, \ldots, A_n) \in \mathcal{E}_1 \times \cdots \times \mathcal{E}_n$$

$$\mathbb{P}(X_1 \in A_1 \text{ and } X_2 \in A_2 \ldots \text{ and } X_n \in A_n) \;=\; \prod_{i=1}^{n} \mathbb{P}(X_i \in A_i),$$

we have equality

$$\mathbb{E}\left[\prod_{i=1}^{n} \varphi_i(X_i)\right] \;=\; \prod_{i=1}^{n} \mathbb{E}\left[\varphi_i(X_i)\right],$$

for any sequence of functions φ_i defined on $(E_i, \mathcal{E}_i)$, with values in $\mathbb{R}$, as soon as the above expectations make sense.

$$f_X(x) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

51 – Perfect dependence in dimension 2

Let F_1, F_2 be distribution functions $\mathbb{R} \to [0,1]$.
The **Fréchet classes** $\mathcal{F}_{(F_1,F_2)}$ group together the set of distribution functions $\mathbb{R}^2 \to [0,1]$ whose marginal laws are precisely F_1, F_2.
For every $F \in \mathcal{F}(F_1, F_2)$, and for all x in $\mathbb{R}^d$

$$F^-(\boldsymbol{x}) \le F(\boldsymbol{x}) \le F^+(\boldsymbol{x})$$

où $F^+(\boldsymbol{x}) = \min\{F_1(x_1), F_2(x_2)\}$, et $F^-(\boldsymbol{x}) = \max\{0, F_1(x_1) + F_2(x_2) - 1\}$.

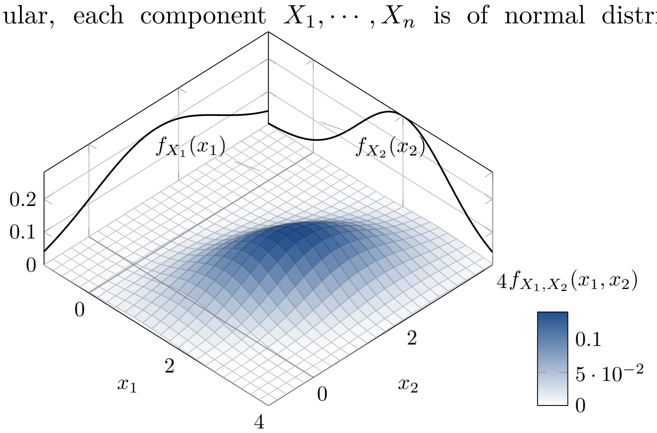
1. The pair $\boldsymbol{X} = (X_1, X_2)$ is said to be comonotonic if and only if it admits F^+ as a distribution function.
2. The pair $\boldsymbol{X} = (X_1, X_2)$ is said to be antimonotonic if and only if it admits F^- as a distribution function.

The pair $\boldsymbol{X} = (X_1, X_2)$ is said to be **comonotone** (**antimonotone**) if there exist non-decreasing (non-increasing) functions g_1 and g_2 of a random variable Z such that

$$\boldsymbol{X} = (g_1(Z), g_2(Z))$$

52 – The Gaussian vector

A vector $X = (X_1, \cdots, X_n)$ is said to be a **Gaussian vector**, with law $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, when any linear combination $\sum_{j=1}^n \alpha_j X_j$ of its components is the univariate normal law. In particular, each component $X_1, \cdots, X_n$ is of normal distribution.



- $\boldsymbol{\mu}$ of $\mathbb{R}^N$ its location,
- $\boldsymbol{\Sigma}$ positive semi-definite of $\mathcal{M}_N(\mathbb{R})$, its variance-covariance.

If $\boldsymbol{\Sigma}$ is well defined positive, therefore invertible, then

$$f_{(\boldsymbol{\mu},\boldsymbol{\Sigma})}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}\,|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}.$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

53 – Three measures of connection (correlations)

The coefficient of **Pearson linear correlation** is called the value

$$\rho_P = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

where σ_{xy} denotes the covariance between the variables x and y, and σ_x, σ_y their standard deviation. ρ takes its values in $[-1,1]$ (application of the Cauchy-Schwartz theorem).
$X \perp Y \Rightarrow \rho_P = 0$, Attention, $\rho_P = 0 \nRightarrow X \perp Y$.
Kendall's tau is defined by

$$\tau_K = \mathbb{P}((X - X')(Y - Y') > 0) - P((X - X')(Y - Y') < 0)$$

where (X, Y) (X', Y') are two independent pairs with the same joint density. This corresponds to the probability of the concordants reduced by that of the discordants :

$$\begin{aligned}
\tau_K =& \mathbb{P}\left(\operatorname{sgn}(X - X') = \mathbb{P}(\operatorname{sgn}(Y - Y')\right) - \\
& \mathbb{P}\left(\operatorname{sgn}(X - X') \neq \operatorname{sgn}(Y - Y')\right) \\
=& \mathbb{E}\left[\operatorname{sgn}(X - X') \operatorname{sgn}(Y - Y')\right] \\
=& \operatorname{Cov}(\operatorname{sgn}(X - X'), \operatorname{sgn}(Y - Y')) \\
=& 4\mathbb{P}(X < X', Y < Y') - 1
\end{aligned}$$

The correlation coefficient **Spearman's rho** of (X, Y) is defined as the Pearson correlation coefficient of the ranks of the random variables X and Y. For a sample n, the n values X_i, Y_i are converted by their ranks x_i, y_i, and ρ is calculated :

$$\rho_S = \frac{1/n \sum_i (x_i - \mathbb{E}[x])(y_i - \mathbb{E}[y])}{\sqrt{1/n \sum_i (x_i - \mathbb{E}[x])^2 \times 1/n \sum_i (y_i - \mathbb{E}[y])^2}}.$$

If we give $x_i = R(X_i)$ from 1 to N and $d_i = x_i - y_i$:

$$\rho_S = 1 - \frac{6 \sum_i d_i^2}{n(n^2 - 1)}$$

54 – Copula

A **copula** is a distribution function, denoted $\mathcal{C}$, defined on $[0,1]^d$ whose margins are uniform on $[0,1]$. A characterization is then that $\mathcal{C}(u_1, ..., u_d) = 0$ if one of the components u_i is zero, $\mathcal{C}(1, ..., 1, u_i, 1, ..., 1) = u_i$, and $\mathcal{C}$ is $d-$increasing.

Let $F^{(d)}$ be a distribution function in dimension d where the F_i are the marginal laws of F.
Sklar's theorem states that $F^{(d)}$ has a copula representation :

$$F^{(d)}(x_1, ..., x_d) = \mathcal{C}(F_1(x_1), ..., F_d(x_d))$$

If these marginal laws are all continuous, the copula $\mathcal{C}$ is then unique, and given by the relation

$$\mathcal{C}(u_1, ..., u_d) = F^{(d)}(F_1^{-1}(u_1), ..., F_d^{-1}(u_d))$$

In this case, we can then speak of the copula associated with a random vector $(X_1, ..., X_d)$. This theorem is very important since we can separate the distribution margin part from the dependence part.

The **Gaussian Copula** is a distribution on the unit cube of dimension d, $[0,1]^d$. It is constructed on the basis of a normal law of dimension d on $\mathbb{R}^d$.
Given the correlation matrix $\Sigma \in \mathbb{R}^{d \times d}$, the Gaussian copula with parameter Σ can be written :

$$\mathcal{C}_\Sigma^{Gauss}(u) = \Phi_\Sigma\left(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\right),$$

where Φ^{-1} is the inverse distribution function of the standard normal distribution and Φ_Σ is the joint distribution of a normal distribution of dimension d, with zero mean and covariance matrix equal to the correlation matrix Σ.

A copula $\mathcal{C}$ is called **Archimedean** if it admits the following representation :

$$\mathcal{C}(u_1, \ldots, u_d) = \psi^{-1}\left(\psi(u_1) + \cdots + \psi(u_d)\right)$$

where ψ is then called **generator**.
Often, copulas admit an explicit formulation of $\mathcal{C}$. A single parameter allows to accentuate the dependence of the entire copula, whatever its dimension d.

This formula provides a copula if and only if ψ is d-monotonic on $[0, \infty)$ *i.e.* the k^{th} derivative of ψ satisfies

$$(-1)^k \psi^{(k)}(x) \geq 0$$

for all $x \geq 0$ and $k = 0, 1, \dots, d-2$ and $(-1)^{d-2}\psi^{d-2}(x)$ is non-increasing and convex.

The following generators are all monotone, i.e. d-monotone for all $d \in \mathbb{N}$.

Name	Generator $\psi^{-1}(t)$,	$\psi(t)$	Setting
Ali-Mikhail-Haq	$\frac{1-\theta}{\exp(t)-\theta}$	$\log\left(\frac{1-\theta+\theta t}{t}\right)$	$\theta \in [0,1)$
Clayton	$(1+\theta t)^{-1/\theta}$	$\frac{1}{\theta}\left(t^{-\theta}-1\right)$	$\theta \in (0,\infty)$
Frank	$-\frac{1}{\theta}\exp(-t)$ $\times \log(1-(1-\exp(-\theta)))$	$-\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$	$\theta \in (0,\infty)$
Gumbel	$\exp\left(-t^{1/\theta}\right)$	$(-\log(t))^{\theta}$	$\theta \in [1,\infty)$
$\perp$	$\exp(-t)$	$-\log(t)$	
Joe	$1-(1-\exp(-t))^{1/\theta}$	$-\log\left(1-(1-t)^{\theta}\right)$	$\theta \in [1,\infty)$

55 – Brownian motion, filtration and martingales

A **filtration** $(\mathcal{F}_t)_{t\geq 0}$ is an increasing family of σ-algebras or tribe representing the information available up to time t. A process (X_t) is said to be **$\mathcal{F}_t$-adapted** if X_t is measurable with respect to $\mathcal{F}_t$ for all t.
A process $(B_t)_{t\geq 0}$ is a **standard Brownian motion** (or Wiener process) if it verifies :

- $B_0 = 0$;
- independent increments : $B_t - B_s$ independent of $\mathcal{F}_s$;
- stationary increments : $B_t - B_s \sim \mathcal{N}(0, t-s)$;
- trajectories continue almost surely.

A process (M_t) is a **martingale** (with respect to $\mathcal{F}_t$) if :

$$\mathbb{E}[|M_t|] < \infty \quad \text{and} \quad \mathbb{E}[M_t \mid \mathcal{F}_s] = M_s \quad \forall 0 \leq s < t$$

Examples : Brownian motion, stochastic integrals of the form $\int_0^t \theta_s dB_s$ (under conditions) are martingales.

Quadratic variation is denoted $\langle B\rangle_t = t, \quad \langle cB\rangle_t = c^2 t$
Covariation : for two Itô processes X, Y,

$$\langle X, Y\rangle_t := \lim_{\|\Pi\|\to 0} \sum_i (X_{t_{i+1}} - X_{t_i})(Y_{t_{i+1}} - Y_{t_i})$$

convergence in probability, where $\Pi = \{t_0 = 0 < t_1 < \dots < t_n = t\}$ is a partition of $[0, t]$.

56 – Itô's process and stochastic differential calculus

A process (X_t) is an **Itô process** if it can be written :

$$X_t = X_0 + \int_0^t \phi_s\, ds + \int_0^t \theta_s\, dB_s$$

or in differential

$$dX_t = \phi_t dt + \theta_t dB_t$$

with ϕ_t, θ_t $\mathcal{F}_t$-adapted and L^2-integrable.
Itô's formula (1D) : for $f \in C^2(\mathbb{R})$, we have :

$$df(X_t) = f'(X_t)dX_t + \frac{1}{2}f''(X_t)d\langle X\rangle_t$$

Example : if $dXt = \mu and + \sigma dBt$ then :

$$dX_t^2 = 2X_t dX_t + d\langle X\rangle_t$$

Itô's formula (multi-dimensional) :
If $X = (X^1, \dots, X^d)$ is an Itô process, $f \in C^2(\mathbb{R}^d)$:

$$df(X_t) = \sum_i \frac{\partial f}{\partial x_i}(X_t)dX_t^i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 f}{\partial x_i \partial x_j}(X_t)d\langle X^i, X^j\rangle_t$$

Integration by parts (Itô) :

$$d(X_t Y_t) = X_t dY_t + Y_t dX_t + d\langle X, Y\rangle_t$$

57 – Stochastic Differential Equations (SDE)

An SDE is a stochastic equation of the form :

$$dX_t = b(t, X_t)dt + a(t, X_t)dB_t, \quad X_0 = x$$

or :

- $b(t, x)$ is the **drift** : function $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$;
- $a(t, x)$ is the **diffusion** : function $\mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$;
- B_t is a Brownian motion ;
- X_t is the solution, adapted stochastic process.

Integral form :

$$X_t = x + \int_0^t b(s, X_s)ds + \int_0^t a(s, X_s)dB_s$$

Conditions of existence and uniqueness :

- **Lipschitz** : there exists $L > 0$ such that :
$$|b(t,x) - b(t,y)| + |a(t,x) - a(t,y)| \leq L|x-y|$$
- **Linear growth** :
$$|b(t,x)|^2 + |a(t,x)|^2 \leq C(1 + |x|^2)$$

Classic examples :

- **Geometric Brownian** : $dS_t = \mu S_t dt + \sigma S_t dB_t$
- **Ornstein–Uhlenbeck** : $dX_t = \theta(\mu - X_t)dt + \sigma dB_t$

Numerical methods : Euler–Maruyama, Milstein.

58 – Risk-neutral probability

A probability $\mathbb{Q}$ is said to be **risk neutral** if, under $\mathbb{Q}$, any asset S_t has an updated price $\frac{S_t}{B_t}$ which is a martingale where (B_t) is the numerary (e.g. $B_t = e^{rt}$).
The absence of arbitrage $\iff$ $\exists \mathbb{Q} \sim \mathbb{P}$ such that the updated prices are martingales. This is the **fundamental theorem of asset pricing**.
Application :
Under $\mathbb{Q}$, the value at time t of an asset giving a return H at date T is :

$$S_t = B_t\, \mathbb{E}^{\mathbb{Q}}\left[\frac{H}{B_T} \middle\| \mathcal{F}_t\right]$$

Note :
The measure $\mathbb{Q}$ is equivalent to $\mathbb{P}$, but reflects a "risk-free" world, useful in valuation.

59 – Simulations

The simulations allow in particular to approximate the expectation by the empirical average of the realizations $x_1, \ldots, x_n$:

$$\frac{1}{n}(x_1 + \ldots + x_n) \approx \int x dF(x) = \mathbb{E}[X]$$

Then, under the TLC, we estimate the uncertainty or confidence interval based on the normal distribution:

$$\left[\overline{x} - 1,96\frac{S_n}{\sqrt{n}}, \overline{x} + 1,96\frac{S_n}{\sqrt{n}}\right]$$

where S_n unbiasedly estimates the variance of X :

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

convergence is said to be in $\mathcal{O}(\frac{\sigma}{\sqrt{n}})$. This interval allows you to decide the number of simulations to be carried out.

60 – Pseudo-random generator on $[0,1]^d$

The computer does not know how to roll the die ($\Omega = \{⚀, ⚁, ⚂, ⚃, ⚄, ⚅\}$). It generates a pseudo-randomness, that is to say a deterministic algorithm which resembles a random event. Generators usually produce a random number on $[0,1]^d$. If the initial value (*seed*) is defined or identified, the following draws are known and replicable.

The simplest algorithm is called the method of linear congruences.

$$x_{n+1} = \Phi(x_n) = (a \times x_n + c) \mod m$$

each x_n is an integer between 0 and $m-1$. a the multiplier, c the increment, and m the modulus of the form $2^p - 1$, that is to say a Mersenne prime number (p necessarily prime) :

Marsaglia generator: $a = 69069, b = 0, m = 2^{32}$

Knuth&Lewis generator: $a = 1664525, b = 1013904223, m = 2^{32}$

Haynes Generator: $a = 6364136223846793005, b = 0, m = 2^{64}$

The Tausworth generator constitutes an 'autoregressive' extension:

$$x_n = (a_1 \times x_{n-1} + a_2 \times x_{n-2} + \cdots + a_k \times x_{n-k}) \mod m \text{ with } n \geq k$$

The generator period is $m^k - 1$, with all a_i relatively prime. If m is of the form 2^p , machine computation times are reduced.

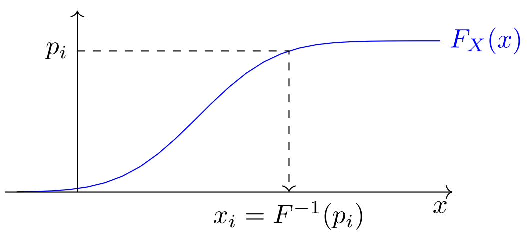
The default random generator is usually the Mersenne-Twister algorithm. It is based on a linear recurrence on a matrix F_2 (matrix whose values are in base 2, i.e. 0 or 1). Its name comes from the fact that the length of the period chosen is a Mersenne prime number.

1. its period is $2^{19937} - 1$
2. it is uniformly distributed over a large number of dimensions (623 for 32-bit numbers) ;
3. it is faster than most other generators ,
4. it is random regardless of the weight of the bit considered, and passes Diehard tests.

61 – Simulate a random variable

Simulating X of any law F_X often comes down to simulating $(p_i)_{i\in[1,n]}$ of law $\mathcal{U}ni(0,1)$.

If F_X is invertible, $x_i = F_X^{-1}(p_i)$ (or quantile function) delivers $(x_i)_{i\in[1,n]}$ a set of n simulations of law F_X.



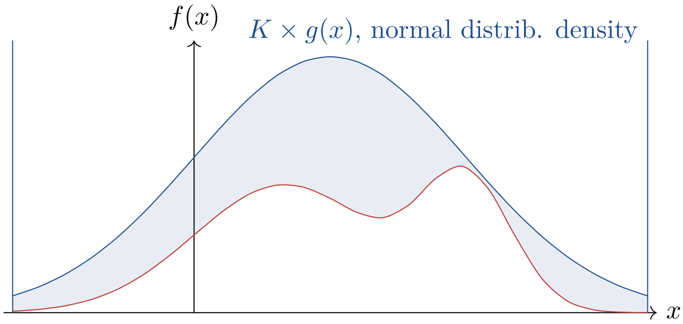
If it is a discrete variable (F^{-1} does not exist) $X_\ell = \min_\ell F(X_\ell) > p_i$, where $(X_\ell)_\ell$ is the countable set of possible values, ordered in ascending order.

In the **change of variable** method, we assume that we know how to simulate a law X, and that there exists ϕ such that $Y = \varphi(X)$ follows a law F_Y. The natural example is that of $X \sim \mathcal{N}(0,1)$ and making the change $Y = \exp(X)$ to obtain Y which follows a lognormal law.

The rejection method is used in more complex cases, for example when F^{-1} is not explicit or requires a lot of computation time. Let f be a probability density function. Assume that there exists a probability density g such that :

$$\exists K > 0 \ , \ \forall x \in \mathbb{R} \ , \ f(x) \leq Kg(x)$$

We then simulate Z according to the density law g, and $Y \sim \mathcal{U}([0; Kg(Z)])$. Then the random variable $X = \{Z | Y \leq f(Z)\}$ follows the density law f.



The performance of the algorithm depends on the number of rejections, represented by the blue area on the graph.

62 – Monte Carlo methods

Monte Carlo methods rely on the repeated simulation of random variables to approximate numerical quantities.

Convergence :

- By the **law of large numbers**, the estimator converges almost surely to the expected value.
- By the **central limit theorem**, the standard error is in $\mathcal{O}(N^{-1/2})$:

$$\sqrt{N}(\hat{\mu}_N - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

- This slow convergence justifies the use of **convergence improvement** techniques.

Variance reduction techniques :

- **Antithetical variables** : we simulate X and $-X$ (or $1-U$ if $U \sim \mathcal{U}[0,1]$), then we average the results. Reduction is efficient if f is monotone.
- **Control method** : if $\mathbb{E}[Y]$ is known, we simulate $(f(X), Y)$ and correct :

$$\hat{\mu}_{\text{corr}} = \hat{\mu} - \beta(\bar{Y} - \mathbb{E}[Y])$$

  where β optimal minimizes the variance.
- **Stratification** : we divide the simulation space into strata (subsets), and we simulate proportionally in each stratum.

- **Importance sampling** : we modify the simulation law to accentuate rare events, then we reweight :

$$\mathbb{E}[f(X)] = \mathbb{E}^Q\left[f(X)\frac{\mathrm{d}P}{\mathrm{d}Q}(X)\right]$$

used in particular to estimate the tails of the distribution (VaR, TVaR).

63 – The bootstrap

The **bootstrap** is a *resampling* method for estimating the uncertainty of an estimator without assuming a parametric form for the underlying distribution.
Let $\xi = (X_1, X_2, \ldots, X_n)$ be a sample of iid variables following an unknown distribution F. We seek to estimate a statistic $\theta = T(F)$ (e.g. mean, median, variance), via its empirical estimator $\hat{\theta} = T(\hat{F}_n)$.

1. We approximate F by the empirical distribution function :
$$\hat{F}_n(x) = \frac{1}{n}\sum_{k=1}^{n}\mathbf{1}_{\{X_k \le x\}}$$
2. We generate B bootstrap samples $\xi^{*(b)} = (X_1^{*(b)}, \ldots, X_n^{*(b)})$ by drawing **with replacement** from the initial sample.
3. For each simulated sample, we calculate the estimate $T^{*(b)} = T(\hat{F}_n^{*(b)})$.

The realizations $T^{*(1)}, \ldots, T^{*(B)}$ form an approximation of the distribution of the estimator $\hat{\theta}$.
We can deduce from this:

- an estimated bias : $\widehat{\text{bias}} = \overline{T^*} - \hat{\theta}$;
- a **confidence interval** at $(1-\alpha)$: $[q_{\alpha/2},\ q_{1-\alpha/2}]$ of the empirical quantiles of $T^{*(b)}$;
- an estimate of the **variance** : $\widehat{\text{Var}}(T^*)$.

Note: Bootstrapping is particularly useful when the distribution of T is unknown or difficult to estimate analytically.

64 – Parametric Bootstrap

The **parametric bootstrap** is based on the assumption that the data follow a parameterized family of laws $\{F_\theta\}$.
Let $\xi = (X_1, \ldots, X_n)$ be an *iid* sample according to an unknown F_θ distribution. We proceed as follows:

1. Estimate the parameter $\hat{\theta}$ from ξ (e.g. by maximum likelihood).
2. Generate B samples $\xi^{*(b)}$ of size n, simulated according to the law $F_{\hat{\theta}}$.
3. Calculate $T^{*(b)} = T(\xi^{*(b)})$ for each sample.

This method approximates the distribution of the estimator $T(\xi)$ assuming the shape of F is known. It is more efficient than the nonparametric bootstrap if the model assumption is well specified. The parametric bootstrap is faster, but inherits the biases of the model.

65 – Cross-validation

Cross-validation is a method for evaluating the predictive performance of a statistical model, used in particular in machine learning or pricing.
Principle :

- Divide the data into K blocks (or folds).
- For each $k = 1, \ldots, K$:
  - Train the model on the other $K-1$ blocks.
  - Evaluate the performance (error, log-likelihood...) on the k-th block.
- Aggregate the errors to obtain an overall estimate of out-of-sample performance.

66 – Quasi-Monte Carlo methods

Quasi-Monte Carlo methods aim to accelerate the convergence of the expectation estimator without resorting to randomness. The typical error is of the order :

$$\mathcal{O}\left(\frac{(\ln N)^s}{N}\right)$$

where N is the sample size and s the dimension of the problem. These methods rely on the use of **low-discrepancy** sequences in $[0,1]^s$. The star discrepancy, denoted $D_N^*(P)$ for a set of points $P = \{x_1, \ldots, x_N\}$, measures the maximum difference between the proportion of points contained in rectangles *anchored to the origin* and their volume. It is defined by :

$$D_N^*(P) = \sup_{u\in[0,1]^s}\left|\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}_{[0,u)}(x_i) - \lambda_s([0,u))\right|$$

with :

- $[0,u) = \prod_{j=1}^{s}[0,u_j)$ a rectangle anchored at the origin in $[0,1]^s$,
- $\mathbf{1}_{[0,u)}(x_i)$ the indicator of the membership of x_i to this rectangle,
- $\lambda_s([0,u)) = \prod_{j=1}^{s} u_j$ the volume of this rectangle.

A low discrepancy means that the points are well distributed in space, which improves the convergence of the estimate.

Van der Corput sequence (dimension 1) : Let n be an integer. We write it in base b :

$$n = \sum_{k=0}^{L-1} d_k(n)\, b^k$$

then we reverse the numbers around the decimal point to obtain :

$$g_b(n) = \sum_{k=0}^{L-1} d_k(n)\, b^{-k-1}$$

For example, for $b=5$ and $n=146$, we have $146 = (1\,0\,4\,1)_5$, so :

$$g_5(146) = \frac{1}{5^4} + \frac{0}{5^3} + \frac{4}{5^2} + \frac{1}{5} = 0{,}3616$$

Halton sequence (dimension s) : We generalize the van der Corput sequence using s distinct prime integer bases $b_1, \ldots, b_s$:

$$x(n) = \left(g_{b_1}(n), \ldots, g_{b_s}(n)\right)$$

This construction provides a sequence of points well distributed in $[0,1]^s$.

Koksma–Hlawka inequality
For a function f of finite variation $V(f)$ (in the Hardy–Krause sense) on $[0,1]^s$:

$$\left|\int_{[0,1]^s} f(u)\,du - \frac{1}{N}\sum_{i=1}^{N} f(x_i)\right| \le V(f)\, D_N$$

where D_N is the discrepancy of the sequence used.
This bound explains why quasi-Monte Carlo methods are often more efficient than Monte Carlo methods.

Distribution	Density & support	Moments & distribution function	Moment-generating function
$\mathcal{B}in(m,q)$ $(0<p<1, m\in\mathbb{N})$	$\binom{m}{x} p^x(1-p)^{m-x}$ $x=0,1,\ldots,m$	$E=mp, \mathrm{Var}=mp(1-p)$ $\gamma=\dfrac{mp(1-p)(1-2p)}{\sigma^3}$	$\left(1-p+pe^t\right)^m$
$\mathcal{B}er(q)$	$\equiv \mathrm{Binomial}(1,p)$		
$\mathcal{DU}ni(n)$ $(n>0)$	$\dfrac{1}{n}, x=0,1,\ldots n$	$\mathbb{E}=(n+1)/2$ $\mathrm{Var}=\left(n^2-1\right)/12$	$\dfrac{e^t(1-e^{nt})}{n(1-e^t)}$
$\mathcal{P}ois(\lambda)$ $(\lambda>0)$	$e^{-\lambda}\dfrac{\lambda^x}{x!}, x=0,1,\ldots$	$\mathbb{E}=\mathrm{Var}=\lambda$ $\gamma=1/\sqrt{\lambda}$ $\kappa_j=\lambda, j=1,2,\ldots$	$\exp\left[\lambda\left(e^t-1\right)\right]$
$\mathcal{NB}in(m,q)$ $(m>0, 0<p<1)$	$\binom{m+x-1}{x} p^m(1-p)^x$ $x=0,1,2,\ldots$	$\mathbb{E}=m(1-p)/p$ $\mathrm{Var}=\mathbb{E}/p$ $\gamma=\dfrac{(2-p)}{p\sigma}$	$\left(\dfrac{p}{1-(1-p)e^t}\right)^m$
$\mathcal{G}eo(q)$	$\equiv \mathcal{NB}in(1,q)$		
$\mathcal{CU}ni(a,b)$ $(a<b)$	$\dfrac{1}{b-a}; a<x<b$	$\mathbb{E}=(a+b)/2,$ $\mathrm{Var}=(b-a)^2/12,$ $\gamma=0$	$\dfrac{e^{bt}-e^{at}}{(b-a)t}$
$\mathcal{N}\left(\mu,\sigma^2\right)$ $(\sigma>0)$	$\dfrac{1}{\sigma\sqrt{2\pi}}\exp\dfrac{-(x-\mu)^2}{2\sigma^2}$	$\mathbb{E}=\mu, \mathrm{Var}=\sigma^2, \gamma=0$ $(\kappa_j=0, j\geq 3)$	$\exp\left(\mu t+\frac{1}{2}\sigma^2t^2\right)$
$\mathcal{G}am(k,\theta)$ $(k,\theta>0)$	$\dfrac{\theta^k}{\Gamma(k)}x^{k-1}e^{-\theta x}, x>0$	$\mathbb{E}=k/\theta, \mathrm{Var}=k/\theta^2,$ $\gamma=2/\sqrt{k}$	$\left(\dfrac{\theta}{\theta-t}\right)^k (t<\theta)$
$\mathcal{E}_{xp}(\lambda)$	$\equiv \mathcal{G}am(1,\lambda)$	$\mathbb{E}=1/\lambda$ $\mathrm{Var}=1/\lambda^2$	
$\chi^2(k)(k\in\mathbb{N})$	$\equiv \mathcal{G}am(k/2,1/2)$		
$\mathcal{IN}(\alpha,\beta)$ $(\alpha>0,\beta>0)$	$\dfrac{\alpha x^{-3/2}}{\sqrt{2\pi\beta}}\exp\left(\dfrac{-(\alpha-\beta x)^2}{2\beta x}\right)$	$\mathbb{E}=\alpha/\beta, \mathrm{Var}=\alpha/\beta^2,$ $\gamma=3/\sqrt{\alpha}$	$e^{\alpha(1-\sqrt{1-2t/\beta})}$ $(t\leq\beta/2)$
	$F(x)=\Phi\left(\frac{-\alpha}{\sqrt{\beta x}}+\sqrt{\beta x}\right)+e^{2\alpha}\Phi\left(\frac{-\alpha}{\sqrt{\beta x}}-\sqrt{\beta x}\right), x>0$		
$\mathcal{B}eta(\alpha,\beta)$ $(\alpha>0,\beta>0)$	$\Gamma(\alpha+\beta)\dfrac{x^{(\alpha-1}(1-x)^{\beta-1}}{\Gamma(\alpha)\Gamma(\beta)}, 0<x<1$	$\mathbb{E}=\dfrac{\alpha}{\alpha+\beta}, \mathrm{Var}=\dfrac{\mathbb{E}(1-\mathbb{E})}{\alpha+\beta+1}$	
$\mathcal{LN}orm\left(\mu,\sigma^2\right)$ $(\sigma>0)$	$\dfrac{1}{x\sigma\sqrt{2\pi}}\exp\dfrac{-(\log x-\mu)^2}{2\sigma^2}, x>0$	$\mathbb{E}=e^{\mu+\sigma^2/2}, \mathrm{Var}=e^{2\mu+2\sigma^2}-e^{2\mu+\sigma^2}$ $\gamma=c^3+3c$ où $c^2=\mathrm{Var}/\mathbb{E}^2$	
$\mathcal{P}areto(\alpha,x_{\mathrm{m}})$ $(\alpha,x_{\mathrm{m}}>0)$	$\dfrac{\alpha x_{\mathrm{m}}^\alpha}{x^{\alpha+1}}, x>x_{\mathrm{m}}$	$\mathbb{E}=\dfrac{\alpha x_{\mathrm{m}}}{\alpha-1}\quad \alpha>1, \mathrm{Var}=\dfrac{\alpha x_{\mathrm{m}}^2}{(\alpha-1)^2(\alpha-2)}\quad \alpha>2$	
$\mathcal{W}eibull(\alpha,\beta)$ $(\alpha,\beta>0)$	$\alpha\beta(\beta y)^{\alpha-1}e^{-(\beta y)^\alpha}, x>0$	$\mathbb{E}=\Gamma(1+1/\alpha)/\beta$ $\mathrm{Var}=\Gamma(1+2/\alpha)/\beta^2-\mathbb{E}^2$ $\mathbb{E}\left[Y^t\right]=\Gamma(1+t/\alpha)/\beta^t$	

67 – Concept of utility

Utility models an individual's preferences between two baskets of goods x and y in a set S, via the relation $x \succcurlyeq y$ (preferred or indifferent).

A function $U : S \to \mathbb{R}$ represents preferences if :

$$x \succcurlyeq y \iff U(x) \geq U(y)$$

Axioms necessary for the existence of a utility function :

1. **Completeness:** For all $x, y \in S$, either $x \succcurlyeq y$, or $y \succcurlyeq x$
2. **Transitivity:** If $x \succcurlyeq y$ and $y \succcurlyeq z$, then $x \succcurlyeq z$
3. **Continuity:** If $x_n \to x$ and $y_n \to y$, and $x_n \succcurlyeq y_n$ for all n, then $x \succcurlyeq y$

68 – Utility function

A function $u : \mathbb{R}_+ \to \mathbb{R}$ represents an agent's preferences in the face of uncertainty.

Expected utility criterion: The agent prefers X to Y if :

$$\mathbb{E}[u(X)] > \mathbb{E}[u(Y)]$$

He chooses X such that $\mathbb{E}[u(X)]$ is maximal.

Properties of u:

- $u' > 0$: the agent prefers more wealth (monotonicity)
- $u'' < 0$: the agent is risk averse (concavity)

Classic examples :

- *Linear* (risk neutral): $u(x) = x$
- *Logarithmic* : $u(x) = \ln(x)$
- *CRRA* (constant relative risk aversion): $u(x) = \frac{x^{1-\gamma}}{1-\gamma}$, $\gamma \neq 1$
- *CARA* (constant absolute risk aversion): $u(x) = -e^{-ax}$

69 – Risk aversion

An agent is said to be **risk-averse** (or risk-phobic) if:

$$u(\mathbb{E}[X]) > \mathbb{E}[u(X)]$$

Which is equivalent to u concave, i.e. $u''(x) < 0$

70 – Measurement of risk aversion

Absolute Aversion Index :

$$A_a(x) = -\frac{u''(x)}{u'(x)}$$

Relative Aversion Index :

$$A_r(x) = -x \cdot \frac{u''(x)}{u'(x)}$$

Jensen's inequality (concave case) :

$$u(\mathbb{E}[X]) \geq \mathbb{E}[u(X)]$$

With equality if and only if X is constant.

71 – Risk premiums

The **risk premium** π is the maximum amount an individual is willing to pay to replace a random lottery win H with its certain expectation $\mathbb{E}[H]$. It verifies:

$$\mathbb{E}[u(w + H)] = u(w + \mathbb{E}[H] - \pi)$$

π is also called *Markowitz measure*: it captures the gap between expected utility and certain utility.

Conversely, the **compensatory bonus** $\tilde{\pi}$ is the amount that must be offered to an individual so that he accepts the lottery H instead of a certain gain. It checks:

$$\mathbb{E}[u(w + H + \tilde{\pi})] = u(w + \mathbb{E}[H])$$

72 – Diversification and utility

Let's have two assets:

- A : risk
- B : certain, with $\mathbb{E}[A] = B$

A risk-averse agent prefers a combination $Z = \alpha A + (1-\alpha)B$, with $0 < \alpha < 1$, to the risky asset alone. If u is concave, then

$$\mathbb{E}[u(Z)] > \mathbb{E}[u(A)]$$

Optimal portfolio: choice of weights (w_i) maximizing expected utility:

$$\max \mathbb{E}[u(X)], \quad \text{où } X = \sum_i w_i X_i, \quad \text{s.c. } \sum w_i = 1$$

Principle: diversification reduces risk (variance) without affecting expectation.

73 – Lagrange method for constrained optimization

The Lagrange multiplier method is used to solve a constrained optimization problem.

Objective: maximize/minimize $f(\boldsymbol{x})$ under the constraint $g(\boldsymbol{x}) = c$, where $\boldsymbol{x} \in \mathbb{R}^d$ is a vector of variables.

Steps of the method:

1. **Identification** : determine the objective function $f(\boldsymbol{x})$ and the constraint $g(\boldsymbol{x}) = c$
2. **Lagrangian** :

$$\mathcal{L}(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \lambda(g(\boldsymbol{x}) - c)$$

3. **System of equations** : solve

$$\nabla_{\boldsymbol{x}} \mathcal{L} = \nabla f(\boldsymbol{x}) + \lambda \nabla g(\boldsymbol{x}) = \boldsymbol{0}, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = g(\boldsymbol{x}) - c = 0$$

4. **Resolution** of the system to obtain $\boldsymbol{x}^*, \lambda^*$
5. **Verification**: ensure that the solutions satisfy the constraint and the type of optimum (max/min)

Example (dimension 2) : maximize $f(x,y) = xy$ under the constraint $x + y = 10$

$$\mathcal{L}(x,y,\lambda) = xy + \lambda(x + y - 10)$$

We derive:

$$\frac{\partial \mathcal{L}}{\partial x} = y + \lambda = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = x + \lambda = 0, \quad \frac{\partial \mathcal{L}}{\partial \lambda} = x + y - 10 = 0$$

We solve the system:

$$\begin{cases} y + \lambda = 0 \\ x + \lambda = 0 \\ x + y = 10 \end{cases} \Rightarrow \begin{cases} \lambda = -y \\ x = -\lambda = y \\ x + y = 10 \Rightarrow 2x = 10 \end{cases} \Rightarrow \begin{cases} x^* = y^* = 5, \\ f(5,5) = 25 \end{cases}$$

Example (Optimal choice and budget constraint)
A rational agent is faced with a consumption choice (c_1, c_2) between two goods, under the constraint :

$$p_1 c_1 + p_2 c_2 = R$$

where p_1, p_2 are prices and R total revenue.
Issue : $\max_{c_1,c_2} u(c_1,c_2) \quad \text{s.t. } p_1 c_1 + p_2 c_2 = R$
Method: introduce the **Lagrangian**

$$\mathcal{L}(c_1, c_2, \lambda) = u(c_1, c_2) + \lambda(R - p_1 c_1 - p_2 c_2)$$

First order conditions (FOC) :

$$\begin{cases} \frac{\partial u}{\partial c_1} = \lambda p_1 \\ \frac{\partial u}{\partial c_2} = \lambda p_2 \\ p_1 c_1 + p_2 c_2 = R \end{cases}$$

By dividing the first two equations :

$$\frac{\partial u / \partial c_1}{\partial u / \partial c_2} = \frac{p_1}{p_2}$$

This ratio is called the **marginal rate of substitution (MRS)**: it measures the quantity of good 2 that the agent is willing to give up to obtain an additional unit of good 1, while maintaining his level of utility constant.

74 – Insurance Application (Mosin)

An agent has an initial wealth w and faces a random loss L. There exists an **insurance demand** for the insurance that pays the indemnity $0 < I(L) < L$ iff $u(w - \pi_I) \geq \mathbb{E}(u(w - L))$ and the **optimal insurance** maximizes $u(w - \pi_I)$.
In Mosin (1968) or Borch (1961) or Smith (1968), the loss model L is simply defined by s between 0 and w :

$$L = \begin{cases} 0 \text{ with prob. } 1 - p \\ s \text{ with prob. } p \end{cases}$$

The premium becomes $\pi_I = (1+\lambda)\mathbb{E}(I(L)) = (1+\lambda)pI(s)$ with λ a loading. We denote by π the case where $I(L) = L$ with $\pi = ps$. If $\lambda = 0$, then we speak of a pure or actuarially fair premium.

Co-insurance (risk sharing) : $I(l) = \alpha l$ knowing $L = l$ for $\alpha \in [0,1]$, $\pi_I(\alpha) = \alpha\pi$ and :

$$w_f = w - L + I(L) - \pi(\alpha) = w - L + \alpha L - \alpha\pi = w - (1-\alpha)L - \alpha\pi$$

$$U(\alpha) = (1-p)u(w - \alpha\pi) + pu(w - (1-\alpha)s - \alpha\pi)$$

Partial insurance $(\alpha^\star < 1)$ is optimal iff $\lambda > 0$. Total insurance $(\alpha^\star = 1)$ is optimal if the loading is zero.

Insurance with deductible (self-insurance portion): With deductible d the insurer pays an indemnity $I(l) = (l-d)_+$ knowing $L = l$.

$$\pi(d) = (1+\lambda)E\left((L-d)_+\right) = (1+\lambda)(s-d)p$$

$$w_f = w - X + (L-d)_+ - \pi(d) = w - \min(X,d) - (1+\lambda)(s-d)p$$

$$U(d) = (1-p)u(\underbrace{w + (1+\lambda)(d-s)p}_{w_f^+}) + pu(\underbrace{w - d + (1+\lambda)(d-s)p}_{w_f^-}).$$

In the deductible model, partial insurance $(d^\star > 0)$ is optimal iff the premium is not actuarially fair. Similarly, full insurance $(d^\star = 0)$ is optimal if the loading is zero.

Generalized model: The random loss risk $L > 0$ is defined on $\Re$, with distribution function F_L),

$$\pi_I = (1+\lambda)\mathbb{E}(I(L)) = (1+\lambda)\int_0^\infty I(l)dF_L(l)$$

1. Total insurance $(d^\star = 0)$ or $(\alpha^\star = 1)$ is optimal if and only if the premium is actuarially fair.
2. If $A_a(u,x)$ is decreasing, then the deductible level $d^\star$ or the coverage rate $\alpha^\star$ increases with initial wealth. For CARA preferences, $d^\star$ is independent of w or $\alpha^\star$ is constant.
3. The coverage level decreases with the loading coefficient λ when $A_a(u,x)$ is increasing or constant.
4. A more risk-averse agent chooses higher coverage.

75 – Information and insurance

Mosin with heterogeneity: Two types of individuals: H for high risk and Lo for low risk. $\theta \in [0,1]$ the proportion of individuals H. Individuals of type $i \in \{Lo, H\}$ face a risk of the same amount s occurring with a different probability p_i such that $1 > p_H > p_{Lo} > 0$.

$$L_i = \begin{cases} 0 \text{ with probability } 1 - p_i, \\ s \text{ with probability } p_i. \end{cases}$$

Market probability :

$$p_m = \theta p_H + (1-\theta)p_{Lo}.$$

Absence of adverse selection: In this model, in the presence of total information, the insurer prefers individual insurance $I_i = s$ and $\pi_i = sp_i$, $\forall i$.
The adverse selection problem : The insurer offers a non-individualized contract from the market $M = (\pi_m = p_m I, I_m(s) = I(s))$, which does not depend on i. The final fortune of an individual of type i is $W_i^m = w - \pi_m - X_i + I_m$. In the presence of a single contract, individuals of type H prefer an insurance contract such that $I_H(s) = s$ and $\pi_H = sp_m$, while individuals of type Lo prefer partial coverage with $I_L^\star < s$ and $\pi_{Lo} = I_L^\star p_m$.

Moral hazard : Insuring him reduces or interrupts his efforts now that he is insured. The efforts of

- prevention reduces the probability of disaster,
- protection reduces the amount of loss.

In the absence of any effort e to prevent or reduce risk, the final fortune w_f is simply defined by

$$\begin{cases} w_f^- = w - s - \pi(I) + I & \text{with probability } p \\ w_f^+ = w - \pi(I) & \text{with probability } 1 - p \end{cases}$$

If there is an effort e to prevent risks, we consider

$$\begin{cases} w_f^- = w - s - \pi(I) + I - e & \text{with probability } p(e) \\ w_f^+ = w - \pi(I) - e & \text{with probability } 1 - p(e) \end{cases}$$

If there is an effort e to protect against risks, we consider

$$\begin{cases} w_f^- = w - s(e) - \pi(I) + I - e & \text{with probability } p \\ w_f^+ = w - \pi(I) - e & \text{with probability } 1 - p \end{cases}$$

with

- $e \mapsto p(e)$ is strictly decreasing and strictly convex.
- $e \mapsto s(e)$ is strictly decreasing and strictly convex.
- $I \leq s$ implies $w_f^- \leq w_f^+$

76 – Definitions

Time series - is a succession of quantitative observations of a phenomena ordered in time.
There are some variations of time series :

- **Panel data -** consist of a time series for each observation of a cross section.
- **Pooled cross sections -** combines cross sections from different time periods.

Stochastic process - sequence of random variables that are indexed in time.

77 – Components of a time series

- **Trend -** is the long-term general movement of a series.
- **Seasonal variations -** are periodic oscillations that are produced in a period equal or inferior to a year, and can be easily identified on different years (usually are the result of climatology reasons).
- **Cyclical variations -** are periodic oscillations that are produced in a period greater than a year (are the result of the economic cycle).
- **Residual variations -** are movements that do not follow a recognizable periodic oscillation (are the result of eventual phenomena).

78 – Type of time series models

- **Static models -** the relation between y and x is contemporary. Conceptually :

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

- **Distributed-lag models -** the relation between y and x is not contemporary. Conceptually :

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \cdots + \beta_s x_{t-(s-1)} + u_t$$

  The long term cumulative effect in y when Δx is : $\beta_1 + \beta_2 + \cdots + \beta_s$

- **Dynamic models -** lags of the dependent variable (endogeneity). Conceptually :

$$y_t = \beta_0 + \beta_1 y_{t-1} + \cdots + \beta_s y_{t-s} + u_t$$

- Combinations of the above, like the rational distributed-lag models (distributed-lag + dynamic).

---

79 – OLS model assumptions under time series

Under this assumptions, the OLS estimator will present good properties. **Gauss-Markov assumptions** extended applied to time series :

t1. **Parameters linearity and weak dependence.**

  a. y_t must be a linear function of the β's.

  b. The stochastic $\{(x_t, y_t) : t = 1, 2, \ldots, T\}$ is stationary and weakly dependent.

t2. **No perfect collinearity.**

  - There are no independent variables that are constant : $\text{Var}(x_j) \neq 0,\ \forall j = 1, \ldots, k$
  - There is not an exact linear relation between independent variables.

t3. **Conditional mean zero and correlation zero.**

  a. There are no systematic errors : $\mathbb{E}(u \mid x_1, \ldots, x_k) = \mathbb{E}(u) = 0 \rightarrow$ **strong exogeneity** (a implies b).

  b. There are no relevant variables left out of the model : $\text{Cov}(x_j, u) = 0,\ \forall j = 1, \ldots, k \rightarrow$ **weak exogeneity.**

t4. **Homoscedasticity.** The variability of the residuals is the same for any x : $\text{Var}(u \mid x_1, \ldots, x_k) = \sigma_u^2$

t5. **No autocorrelation.** Residuals do not contain information about any other residuals :
$\text{Corr}(u_t, u_s \mid x_1, \ldots, x_k) = 0,\ \forall t \neq s$

t6. **Normality.** Residuals are independent and identically distributed (**i.i.d.**) : $u \sim \mathcal{N}(0, \sigma_u^2)$

t7. **Data size.** The number of observations available must be greater than $(k+1)$ parameters to estimate. (It is already satisfied under asymptotic situations)

80 – Asymptotic properties of OLS

Under the econometric model assumptions and the Central Limit Theorem :

- Hold t1 to t3a : OLS is **unbiased**. $\mathbb{E}(\hat{\beta}_j) = \beta_j$
- Hold t1 to t3 : OLS is **consistent**. $\text{plim}(\hat{\beta}_j) = \beta_j$ (to t3b left out t3a, weak exogeneity, biased but consistent)
- Hold t1 to t5: **asymptotic normality** of OLS (then, t6 is necessarily satisfied) : $u \underset{a}{\sim} \mathcal{N}(0, \sigma_u^2)$
- Hold t1 to t5 : **unbiased estimate** of σ_u^2. $\mathbb{E}(\hat{\sigma}_u^2) = \sigma_u^2$
- Hold t1 to t5: OLS is BLUE (Best Linear Unbiased Estimator) or **efficient**.
- Hold t1 to t6: hypothesis testing and confidence intervals can be done reliably.

---

81 – Trends and seasonality

Spurious regression - is when the relation between y and x is due to factors that affect y and have correlation with x, $\text{Corr}(x_j, u) \neq 0$. Is the **non-fulfilment of t3.**
Two time series can have the same (or contrary) trend, that should lend to a high level of correlation. This can provoke a false appearance of causality, the problem is **spurious regression.** Given the model :

$$y_t = \beta_0 + \beta_1 x_t + u_t$$

where :

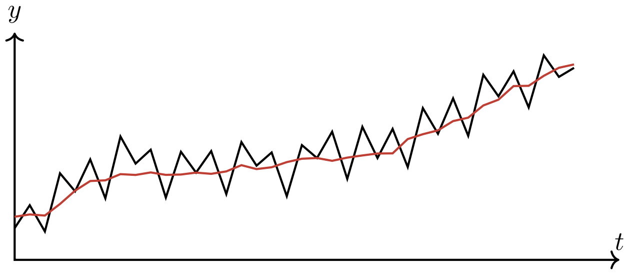
$$y_t = \alpha_0 + \alpha_1 \text{Tendance} + v_t$$
$$x_t = \gamma_0 + \gamma_1 \text{Tendance} + v_t$$

Adding a trend to the model can solve the problem :

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 \text{Tendance} + u_t$$

The trend can be linear or non-linear (quadratic, cubic, exponential, etc.)
Another way is to make use of the **Hodrick-Prescott filter** to extract the trend and the cyclical component.

82 – Seasonality
A time series with can manifest seasonality. That is, the series is subject to a seasonal variations or patterns, usually related to climatology conditions.
For example, GDP (black) is usually higher in summer and lower in winter. Seasonally adjusted series (in) for comparison.



- This problem is a **spurious regression**. Seasonal adjustment can solve it.

A simple **seasonal adjustment** could consist of creating stationary binary variables and adding them to the model. For example, for quarterly series (Qq_t are binary variables) :

$$y_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + \beta_4 x_{1t} + \cdots + \beta_k x_{kt} + u_t$$

Another way is to seasonally adjust (sa) the variables, and then, do the regression with the adjusted variables :

$$z_t = \beta_0 + \beta_1 Q2_t + \beta_2 Q3_t + \beta_3 Q4_t + v_t \rightarrow \hat{v}_t + \mathbb{E}(z_t) = \hat{z}_t^{sa}$$
$$\hat{y}_t^{sa} = \beta_0 + \beta_1 \hat{x}_{1t}^{sa} + \cdots + \beta_k \hat{x}_{kt}^{sa} + u_t$$

There are much better and complex methods to seasonally adjust a time series, like the **X-13ARIMA-SEATS**.

---

83 – Autocorrelation
The residual of any observation, u_t, is correlated with the residual of any other observation. The observations are not independent. Is the **non-fulfilment** of **t5**.

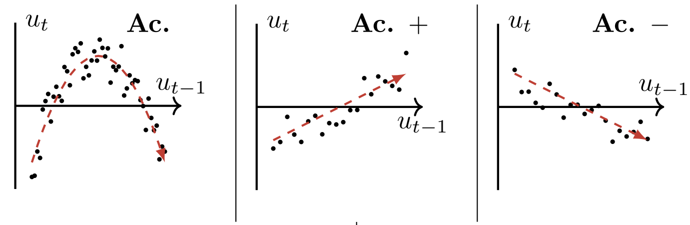
$$\text{Corr}(u_t, u_s \mid x_1, \ldots, x_k) = \text{Corr}(u_t, u_s) \neq 0, \ \forall t \neq s$$

84 – Consequences

- OLS estimators are still unbiased.
- OLS estimators are still consistent.
- OLS is **not efficient** any more, but still a LUE (Linear Unbiased Estimator).
- **Variance estimations** of the estimators are **biased**: the construction of confidence intervals and the hypothesis testing is not reliable.

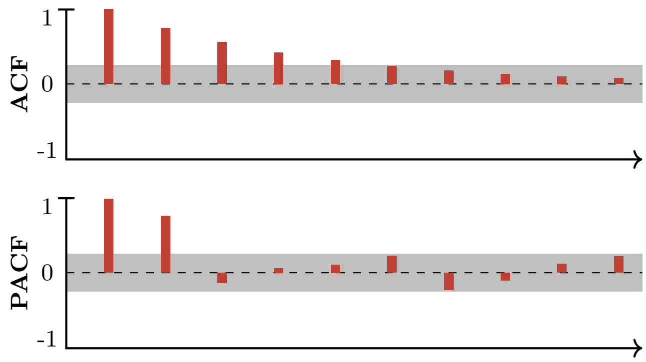
85 – Detection

- **Scatter plots** - look for scatter patterns on u_{t-1} vs. u_t.



Correlogram - autocorrelation function (ACF) and partial ACF (PACF).

- Y axis : correlation.
- X axis : lag number.
- Grey area : $\pm 1,96/T^{0,5}$



MA(q) process. <u>ACF</u> : <u>ACF</u>: only the first q coefficients are significant, the remaining are abruptly cancelled. <u>PACF</u>: attenuated exponential fast decay or sine waves.

AR(p) process. <u>ACF</u>: attenuated exponential fast decay or sine waves. <u>PACF</u>: only the first p coefficients are significant, the remaining are abruptly cancelled.

ARMA(p, q) process. <u>ACF</u> and <u>PACF</u>: the coefficients are not abruptly cancelled and present a fast decay.

If the ACF coefficients do not decay rapidly, there is a clear indicator of a lack of stationarity in mean.

- **Formal tests** - Generally, H_0 : No autocorrelation.

Supposing that u_t follows an AR(1) process :

$$u_t = \rho_1 u_{t-1} + \varepsilon_t$$

where ε_t is white noise.

AR(1) t test (exogenous regressors) :

$$t = \frac{\hat{\rho}_1}{\text{se}(\hat{\rho}_1)} \sim t_{T-k-1,\alpha/2}$$

– H_1 : Autocorrelation of order one, AR(1).

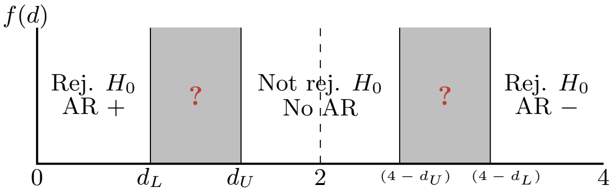
Durbin-Watson statistic (exogenous regressors and residual normality) :

$$d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \approx 2 \cdot (1 - \hat{\rho}_1)$$

Or $0 \leq d \leq 4$

– H_1 : Autocorrelation of order one, AR(1).

$d =$	0	2	4
$\rho \approx$	1	0	-1



Durbin's h (endogenous regressors) :

$$h = \hat{\rho} \cdot \sqrt{\frac{T}{1 - T \cdot v}}$$

where v is the estimated variance of the coefficient associated with the endogenous variable.

– H_1 : Autocorrelation of order one, AR(1).

Breusch-Godfrey test (endogenous regressors) : it can detect MA(q) and AR(p) processes (ε_t is w. noise) :

– MA(q) : $u_t = \varepsilon_t - m_1 u_{t-1} - \cdots - m_q u_{t-q}$

– AR(p) : $u_t = \rho_1 u_{t-1} + \cdots + \rho_p u_{t-p} + \varepsilon_t$

Under H_0 : No autocorrelation :

$$T \cdot R^2_{\hat{u}_t} \underset{a}{\sim} \chi^2_q \qquad \textbf{or} \qquad T \cdot R^2_{\hat{u}_t} \underset{a}{\sim} \chi^2_p$$

– H_1 : Autocorrelation of order q (or p).

Ljung-Box Q test :

– H_1 : Autocorrelation up to lag h.

86 – Correction • Use OLS with a variance-covariance matrix estimator that is **robust to heteroscedasticity and autocorrelation** (HAC), for example, the one proposed by **Newey-West**.

- Use **Generalized Least Squares** (GLS). Supposing $y_t = \beta_0 + \beta_1 x_t + u_t$, where $u_t = \rho u_{t-1} + \varepsilon_t$, and $|\rho| < 1$ et ε_t is white noise.
  - If ρ is **known**, use a **quasi-differentiated model** :

    $$y_t - \rho y_{t-1} = \beta_0(1-\rho) + \beta_1(x_t - \rho x_{t-1}) + u_t - \rho u_{t-1}$$
    $$y_t^* = \beta_0^* + \beta_1' x_t^* + \varepsilon_t$$

    where $\beta_1' = \beta_1$; and estimate it by OLS.
  - If ρ is **not known**, estimate it by -for example- the **Cochrane-Orcutt iterative method** (Prais-Winsten's method is also good) :
    1. Obtain $\hat{u}_t$ from the original model.
    2. Estimate $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$ and obtain $\hat{\rho}$.
    3. Create a quasi-differentiated model :

       $$y_t - \hat{\rho} y_{t-1} = \beta_0(1-\hat{\rho}) + \beta_1(x_t - \hat{\rho} x_{t-1}) + u_t - \hat{\rho} u_{t-1}$$
       $$y_t^* = \beta_0^* + \beta_1' x_t^* + \varepsilon_t$$

       where $\beta_1' = \beta_1$; and estimate it by OLS.
    4. Obtain $\hat{u}_t^* = y_t - (\hat{\beta}_0^* + \hat{\beta}_1' x_t) \neq y_t - (\hat{\beta}_0^* + \hat{\beta}_1' x_t^*)$.
    5. Repeat from step 2. The algorithm ends when the estimated parameters vary very little between iterations.
- If not solved, look for **high dependence** in the series.

87 – Exponential smoothing $f_t = \alpha y_t + (1-\alpha) f_{t-1}$ where $0 < \alpha < 1$ is the smoothing factor.

88 – Forecasts
Two types of forecasts :

- Of the mean value of y for a specific value of x.
- Of an individual value of y for a specific value of x.

Theil's U statistic - compares the forecast results with the ones of forecasting with minimal historical data.

$$U = \sqrt{\frac{\sum_{t=1}^{T-1} \left(\frac{\hat{y}_{t+1} - y_{t+1}}{y_t}\right)^2}{\sum_{t=1}^{T-1} \left(\frac{y_{t+1} - y_t}{y_t}\right)^2}}$$

- < 1 : The forecast is better than guessing.
- $= 1$: The forecast is about as good as guessing.
- > 1 : The forecast is worse than guessing.

89 – Stationarity
Stationarity allows to correctly identify relations –that stay unchanged with time– between variables.

- **Stationary process** (strict stationarity) - the joint probability distribution of the process remains unchanged when shifted h periods.
- **Non-stationary process** - for example, a series with trend, where at least the mean changes with time.
- **Covariance stationary process** - it is a weaker form of stationarity :
  - $\mathbb{E}(x_t)$ is constant. – Var(x_t) is constant.
  - For any t, $h \geq 1$, Cov(x_t, x_{t+h}) depends only of h, not of t.

90 – Weak dependence
Weak dependence replaces the random sampling assumption for time series.

- An stationary process $\{x_t\}$ is **weakly dependent** when x_t and x_{t+h} are almost independent as h increases without a limit.
- A covariance stationary process is **weakly dependent** if the correlation between x_t and x_{t+h} tends to 0 fast enough when $h \to \infty$ (they are not asymptotically correlated).

Weakly dependent processes are known as **integrated of order zero**, I(0). Some examples :

- **Moving average** - $\{x_t\}$ is a moving average of order q, MA(q) :

  $$x_t = e_t + m_1 e_{t-1} + \cdots + m_q e_{t-q}$$

  where $\{e_t : t = 0, 1, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.
- **Autoregressive process** - $\{x_t\}$ is an autoregressive process of order p, AR(p) :

  $$x_t = \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p} + e_t$$

where $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.

Stability condition : if $1 - \rho_1 z - \cdots - \rho_p z^p = 0$ for $|z| > 1$, then $\{x_t\}$ is an AR(p) stable process that is weakly dependent. For AR(1), the condition is : $|\rho_1| < 1$.

- **ARMA process -** is a combination of AR(p) and MA(q) ; $\{x_t\}$ is an ARMA(p, q) :

$$x_t = e_t + m_1 e_{t-1} + \cdots + m_q e_{t-q} + \rho_1 x_{t-1} + \cdots + \rho_p x_{t-p}$$

91 – Unit roots

A process is I(d), that is, integrated of order d, if applying differences d times makes the process stationary.
When $d \geq 1$, the process is called a **unit root process** or it is said to have an unit root.
A process have an unit root when the stability condition is not met (there are roots on the unit circle).

92 – Strong dependence

Most of the time, economic series are strongly dependent (or high persistent). Some examples of **unit root** I(1) :

- **Random walk -** an AR(1) process with $\rho_1 = 1$.

$$y_t = y_{t-1} + e_t$$

  where $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.

- **Random walk with a drift -** an AR(1) process with $\rho_1 = 1$ and a constant.

$$y_t = \beta_0 + y_{t-1} + e_t$$

  où $\{e_t : t = 1, 2, \ldots, T\}$ is an *i.i.d.* sequence with zero mean and σ_e^2 variance.

93 – Unit root tests

Test	H_0	Reject H_0
ADF	I(1)	tau < Critical value
KPSS	I(0) level	mu > Critical value
	I(0) trend	tau > Critical value
Phillips-Perron	I(1)	Z-tau < Critical value
Zivot-Andrews	I(1)	tau < Critical value

94 – From unit root to weak dependence

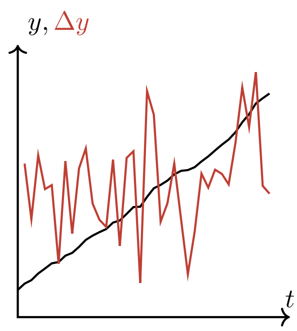
An integrated of **order one**, I(1), means that the **first difference** of the process is **weakly dependent** or I(0) (and usually, stationary). For example, let $\{y_t\}$ be a random walk :

$$\Delta y_t = y_t - y_{t-1} = e_t$$

where $\{e_t\} = \{\Delta y_t\}$ est *i.i.d.*

Note :

- The first difference of a series removes its trend.
- Logarithms of a series stabilizes the variance.



From unit root to percentage change

When an I(1) series is strictly positive, it is usually converted to logarithms before taking the first difference to obtain the (approx.) percentage change of the series :

$$\Delta \log(y_t) = \log(y_t) - \log(y_{t-1}) \approx \frac{y_t - y_{t-1}}{y_{t-1}}$$

95 – Cointegration

When **two series are I(1), but a linear combination of them is I(0)**. If the case, the regression of one series over the other is not spurious, but expresses something about the long term relation. Variables are called cointegrated if they have a common stochastic trend.

For example, $\{x_t\}$ and $\{y_t\}$ are I(1), but $y_t - \beta x_t = u_t$ where $\{u_t\}$ is I(0). (β is the cointegrating parameter).

96 – Cointegration test

Following the example above :

1. Estimate $y_t = \alpha + \beta x_t + \varepsilon_t$ and obtain $\hat{\varepsilon}_t$.
2. Perform an ADF test on $\hat{\varepsilon}_t$ with a modified distribution.

   The result of this test is equivalent to :

   - $H_0 : \beta = 0$ (no cointegration)
   - $H_1 : \beta \neq 0$ (cointegration)

   if test statistic > critical value, reject H_0.

97 – Heteroscedasticity on time series

The **assumption** affected is **t4**, which leads **OLS to be not efficient**.

Use tests like Breusch-Pagan or White's, where H_0 : No heteroscedasticity. It is **important** for the tests to work that there is **no autocorrelation**.

98 – ARCH

An autoregressive conditional heteroscedasticity (ARCH), is a model to analyse a form of dynamic heteroscedasticity, where the error variance follows an AR(p) process.

Given the model : $y_t = \beta_0 + \beta_1 z_t + u_t$ where, there is AR(1) and heteroscedasticity :

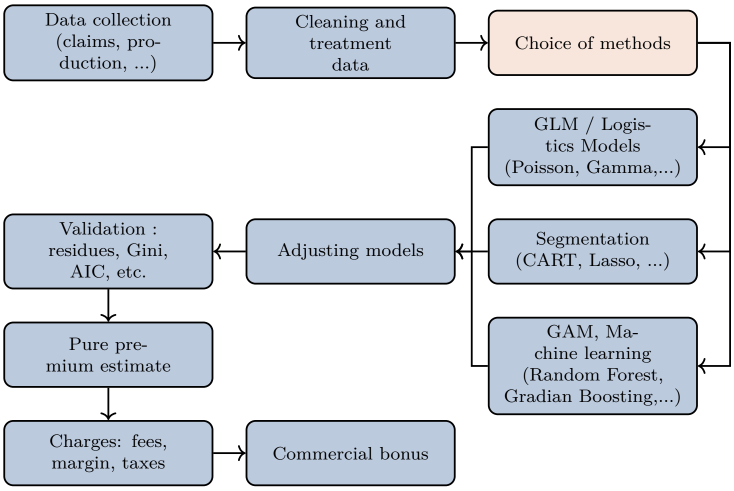
$$\mathbb{E}(u_t^2 \mid u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$$

99 – GARCH

A general autoregressive conditional heteroscedasticity (GARCH), is a model similar to ARCH, but in this case, the error variance follows an ARMA(p, q) process.

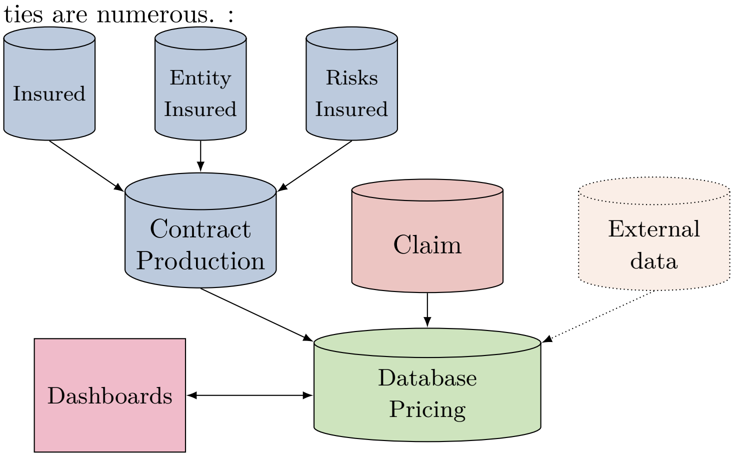
100 – Non-life insurance pricing

A general approach, but not exhaustive, because there are many possibilities :



101 – General structure of insurance data

A classic data structure in insurance. Here again, the possibilities are numerous. :



102 – Provision

The non-life actuary mainly assesses the following provisions:

- Reserves for claims reported but not settel (RBNS).
- Reserve for claims incurred but not reported (IBNR).
- Reserves for unearned premiums.
- Reserves for outstanding risks (non-life).

103 – Deterministic Chain Ladder

Let C_{ik} be the amount, cumulative up to development year k, of claims occurring in accident year i, for $1 \leq i, k \leq n$. C_{ik} may represent either the amount paid or the total estimated cost (payment already made plus reserve) of the claim. The amounts C_{ik} are known for $i+k \leq n+1$ and we seek to estimate the values of the C_{ik} for $i+k > n+1$, and in particular the ultimate values C_{in} for $2 \leq i \leq n$. These notations are illustrated in the following triangle:
The Chain Ladder method estimates the unknown amounts, C_{ik} for $i+k > n+1$, by

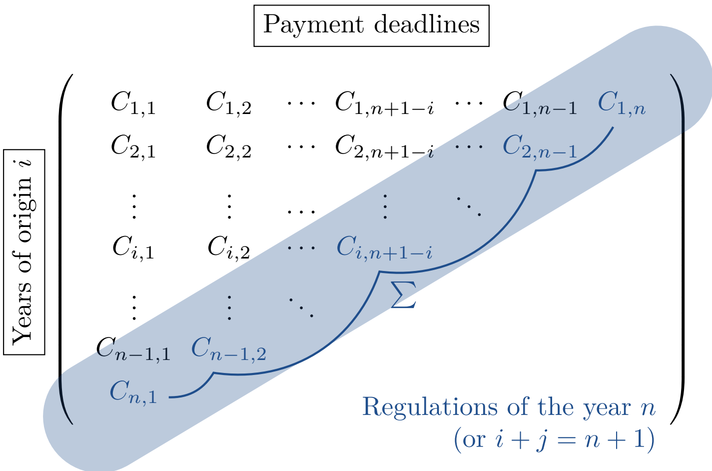
$$\hat{C}_{ik} = C_{i,n+1-i} \cdot \hat{f}_{n+1-i} \cdots \hat{f}_{k-1} \quad i+k > n+1 \tag{5}$$

or

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k+1}}{\sum_{i=1}^{n-k} C_{ik}} \quad 1 \leq k \leq n-1. \tag{6}$$

The claim reserve for the year of the accident (R_i, $2 \leq i \leq n$), is then estimated by

$$\begin{aligned} \hat{R}_i &= C_{in} - C_{i,n+1-i} \\ &= C_{i,n+1-i} \cdot \hat{f}_{n+1-i} \cdots \hat{f}_{n-1} - C_{i,n+1-i} \end{aligned}$$



104 – Mack's Method

The first two hypotheses are:

$$E\left(C_{i,k+1} \mid C_{i1}, \ldots, C_{ik}\right) = C_{ik} f_k \quad 1 \leq i \leq n, 1 \leq k \leq n-1 \tag{7}$$

$$\left\{C_{i1}, \ldots, C_{in}\right\}, \left\{C_{j1}, \ldots, C_{jn}\right\} \quad \forall i, j \quad \text{are independent} \tag{8}$$

Mack demonstrates that if we estimate the parameters of the model (7) par (6) so this stochastic model (7), combined with the assumption (8) provides exactly the same caveats as the original Chain Ladder method (5).
With the notation $f_{i,k} = \frac{C_{i,k+1}}{C_{i,k}}$, $\hat{f}_k$ is the average of the $f_{i,k}$ weighted by the $C_{i,k}$:

$$\hat{f}_k = \frac{\sum_{i=1}^{n-k} C_{i,k} \times f_{i,k}}{\sum_{i=1}^{n-k} C_{ik}}$$

The variance is written :

$$\begin{aligned} \hat{\sigma}_k^2 &= \frac{1}{n-k-1} \sum_{i=1}^{n-k} C_{ik} \left(\frac{C_{i,k+1}}{C_{ik}} - \hat{f}_k\right)^2 \\ &= \frac{1}{n-k-1} \sum_{i=1}^{n-k} \left(\frac{C_{i,k+1} - C_{i,k}\hat{f}_k}{\sqrt{C_{i,k}}}\right)^2 \end{aligned}$$

The third assumption concerns the distribution of R_i to be able to easily construct confidence intervals on the estimated reserves. If the distribution is normal, with mean the estimated value $\hat{R}_i$ and standard deviation given by the standard error $\mathrm{se}\left(\hat{R}_i\right)$. A confidence interval at 95% is then given by $\left[\hat{R}_i - 2\,\mathrm{se}\left(\hat{R}_i\right), \hat{R}_i + 2\,\mathrm{se}\left(\hat{R}_i\right)\right]$.
If the distribution is assumed to be lognormal, the bounds of a 95% confidence interval will then be given by

$$\left[\hat{R}_i \exp\left(\frac{-\sigma_i^2}{2} - 2\sigma_i\right), \hat{R}_i \exp\left(\frac{-\sigma_i^2}{2} + 2\sigma_i\right)\right]$$

105 – The collective risk model

Is the collective model the basic model in non-life actuarial science? X_i denotes the amount of the i^th claim, N denotes the number of claims and S the total amount over a year

$$S = \sum_{i=1}^{N} X_i$$

knowing that $S = 0$ when $N = 0$ and that $\{X_i\}_{i=1}^{\infty}$ is a sequence *iid* and $N \perp \{X_i\}_{i=1}^{\infty}$. The difficulty is to obtain the distribution of S, even though $\mathbb{E}[N]$ is not large in the TCL sense.

106 – The distribution of S

Either $G(x) = \mathbb{P}(S \leq x)$, $F(x) = \mathbb{P}(X_1 \leq x)$, and $p_n = \mathbb{P}(N = n)$ so that $\{p_n\}_{n=0}^{\infty}$ be the probability function for the number of claims.

$$\{S \leq x\} = \bigcup_{n=0}^{\infty} \{S \leq x \text{ and } N = n\}$$

$$\mathbb{P}(S \leq x \mid N = n) = \mathbb{P}\left(\sum_{i=1}^{n} X_i \leq x\right) = F^{n*}(x)$$

So, for $x \geq 0$

$$G(x) = \sum_{n=0}^{\infty} p_n F^{n*}(x)$$

where F^{n*} denotes the n^{th} convolution, unfortunately it does not exist in closed form for many distributions.
If $E[X] = m$

$$E[S] = E[Nm] = E[N]m$$

This result is very interesting because it indicates that the total expected amount of claims is the product of the expected number of claims and the expected amount of each claim. Similarly, using the fact that $\{X_i\}_{i=1}^{\infty}$ are independent random variables,

$$V[S \mid N = n] = V\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} V[X_i]$$

$$\begin{aligned} V[S] &= E[V(S \mid N)] + V[E(S \mid N)] \\ &= E[N]V[X_i] + V[N]m^2 \end{aligned}$$

107 – The distribution class $(a, b, 0)$

A counting distribution is said to be $(a, b, 0)$ if its probability function $\{p_n\}_{n=0}^{\infty}$ can be calculated recursively from the formula

$$p_n = \left(a + \frac{b}{n}\right) p_{n-1}$$

for $n = 1, 2, 3, \ldots$, where a and b are constants.
There are exactly three non-trivial distributions in the class $(a, b, 0)$, namely Poisson, binomial and negative binomial. Here are the values of a and b for the main distributions $(a, b, 0)$:

	a	b
$\mathcal{P}_{ois}(\lambda)$	0	λ
$\mathcal{B}_{in}(n, q)$	$-q/(1-q)$	$(n+1)q/(1-q)$
$\mathcal{NB}_{in}(k, q)$	$1-q$	$(1-q)(k-1)$
$\mathcal{G}_{eo}(q)$	$1-q$	0
Panjer's distribution	$\frac{\lambda}{\alpha+\lambda}$	$\frac{(\alpha-1)\lambda}{\alpha+\lambda}$

The geometric law is a special case of the negative binomial where k=1.

108 – Panjer's aggregation algorithm

The **Panjer algorithm** aims at estimating the distribution of a composite cost-frequency law under particular conditions.

- $(X_i)_{i=1}^{N}$ *iid* discrete defined on $\{0, h, 2h, 3h...\}$
- the law of numbers in the so-called class $(a, b, 0)$

Since we now assume that the individual claim amounts are distributed over the non-negative integers, it follows that S is also distributed over the non-negative integers. As $S = \sum_{i=1}^{N} X_i$, it follows that $S = 0$ if $N = 0$ or if $N = n$ and $\sum_{i=1}^{n} X_i = 0$. As $\sum_{i=1}^{n} X_i = 0$ only if each $X_i = 0$, it follows by independence that

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i = 0\right) = f_0^n$$

$$\begin{cases} g_0 = p_0 + \sum_{n=1}^{\infty} p_n f_0^n = P_N(f_0) \text{ if } a \neq 0, \\ g_0 = p_0 \cdot \exp(f_0 b) \text{ if } a = 0, \\ g_k = \frac{1}{1-af_0} \sum_{j=1}^{k} \left(a + \frac{bj}{k}\right) f_j g_{k-j} \end{cases}$$

gx is expressed in terms of $g0, g1, \ldots, g_{x-1}$, so that the calculation of the probability function is recursive. In all practical applications of this formula, a computer is required to perform the calculations. However, the advantage of Panjer's recursion formula over the formula for g_x is that there is no need to compute convolutions, which is much more computationally efficient. Panjer's algorithm requires the discretization of the variable X_i.

109 – Panjer and Poisson's Law

When the frequency follows a Poisson distribution, this implies that $a = 0$ and $b = \lambda$. The result is simplified :

$$\begin{cases} g_0 = e^{-\lambda(1-f_0)} \\ g_k = \frac{\lambda}{k} \sum_{j=1}^{k} j.f_j.g_{k-j} \end{cases}$$

110 – Panjer and Pollaczeck-Khinchina-Beekman

Let τ_1 be the first instant where $R_t < \kappa(= \kappa_0)$. We then set $L_1 = \kappa - R_{\tau_1}$. We restart the process with $\kappa_1 = \kappa_0 - R_{\tau_1}$ to find τ_2 and $L_2 = \kappa_1 - R_{\tau_2}$. Continuing in this way, we see that :

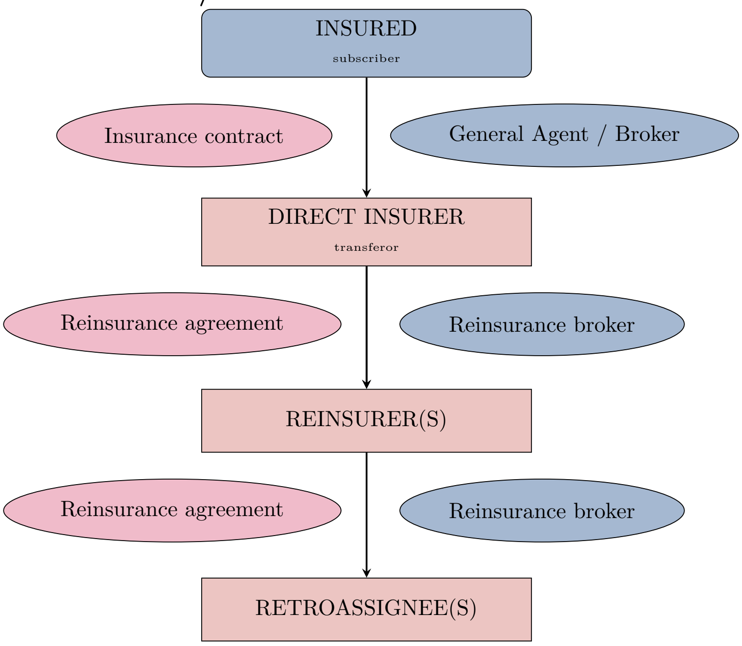
$$M = \sup_{t \geq 0}\{S_t - ct\} = \sum_{k=1}^{K} L_k$$

where $K \sim \mathcal{G}eo(q)$ with $q = 1 - \psi(0)$. Noting that the variables $(L_k)_{1 \leq k \leq K}$ are *iid* (F), we then have $\psi(k) = \mathbb{P}[M > \kappa]$ given by the Pollaczeck-Khinchine-Beekman formula.
The representation

$$\psi(\kappa) = \mathbb{P}\left[\sum_{j=1}^{K} L_j > \kappa\right]$$

allows to evaluate the probability of ruin on an infinite horizon using the Panjer algorithm.

111 – Transfer/retrocession schemes



112 – Key words of reinsurance

Cedant : client of the reinsurer, i.e., the direct insurer, who transfers (cedes) risks to the reinsurer in exchange for the payment of a **reinsurance premium**.
Cession : transfer of risks by the direct insurer to the reinsurer.
Value Exposure : limit of the amount of risk covered by a (re)insurance contract.
Proportional reinsurance : proportional participation of the reinsurer in the premiums and claims of the direct insurer.
Quota Share : type of proportional reinsurance where the reinsurer participates in a given percentage of all risks underwritten by a direct insurer in a given line of business.
Surplus Share : A type of proportional reinsurance where the reinsurer covers risks beyond the direct insurer's full retention amount. This ratio is calculated on the capacity of the risk subscribed (≈ Maximum possible loss).
Reinsurance commission : remuneration that the reinsurer grants to the insurer or brokers as compensation for the costs of acquiring and managing insurance contracts.
Excess Reinsurance : coverage by the reinsurer of claims exceeding a certain amount, against payment by the direct insurer of a specific reinsurance premium.
Retrocession : share of risks that the reinsurer cedes to other reinsurers.
Co-insurance : participation of several direct insurers in the same risk.
We then use the expression **reinsurance pool**. The main reinsurer is called **leading reinsurer**.
Reinsurance treaty : contract concluded between the direct insurer and the reinsurer on one or more of the insurer's portfolios.
Facultative reinsurance : It differs from the reinsurance treaty by underwriting risk by risk (or policy by policy) (case by case, one risk at a time).

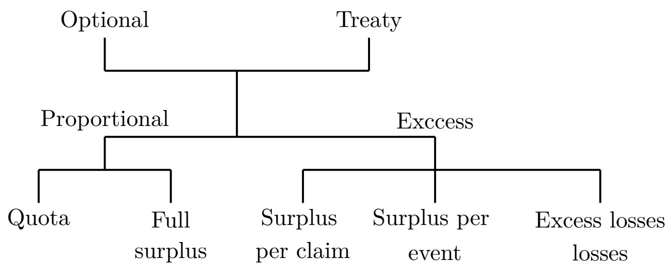
113 – The economic role of reinsurance

Insurance and reinsurance share the same purpose : the pooling of risks. Reinsurance intervenes in particular on risks :

- independent, but unitarily expensive (airplane, ship, industrial sites...),
- small amounts (breakage, car, ...) but correlated during large-scale events, resulting in expensive accumulations,
- aggregated within a portfolio of insurance policies,
- little known or new.

Reinsurance allows to increase the business issuing capacity, ensure the financial stability of the insurer, especially in the event of disasters, reduce their capital requirements, and benefit from the expertise of the reinsurer.

114 – Types of reinsurance agreements

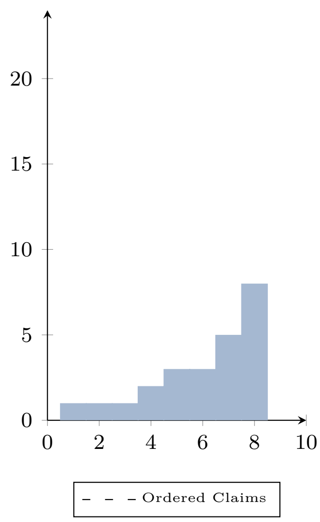


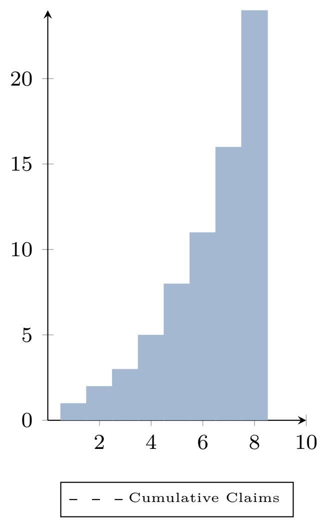
115 – Types of reinsurance through an example

Our insurer reinsures $N = 30$ insurance policies, with a total premium of 10M€ $(P = \sum_{i=1...N} P_i)$. The total capacity is 180M€ $(\sum_{i=1...30} K_i)$. S_r will be the total share of the loss covered by the insurer and P_r the total reinsurance premium. Here are the $n = 8$ policies affected by losses $(1 \leq i \leq n)$, the losses of the other policies being zero $(S_i = 0, \forall i > n)$:

Claim number	1	2	3	4	5	6	7	8
Prime (k€)	500	200	100	100	50	200	500	200
Value Exposure (M€)	8	5	3	2	3	5	8	8
Claims (M€)	1	1	1	2	3	3	5	8

The S/P is 240%.

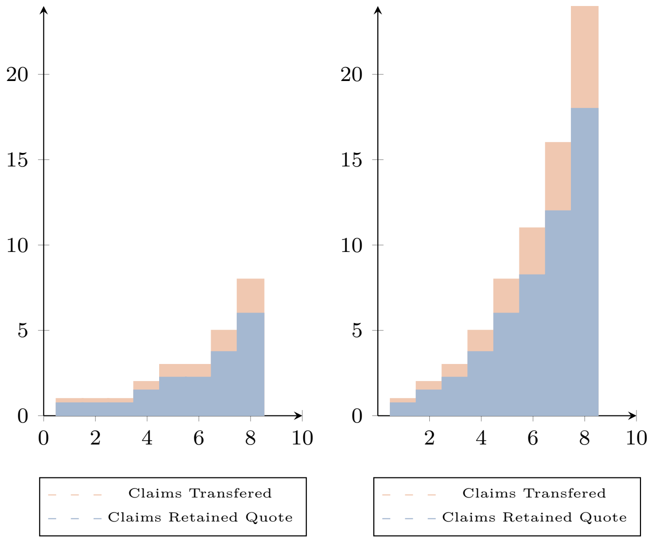




Quota:

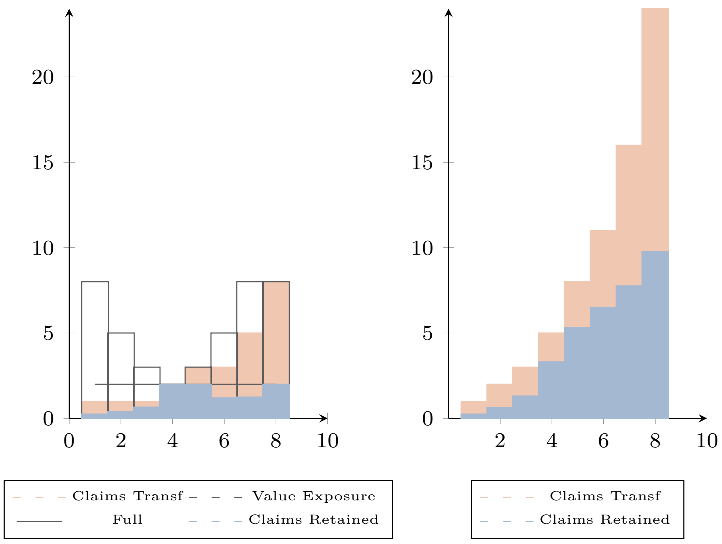
$$S_r = \alpha \sum_{i=1\ldots n} S_i \qquad P_r = \alpha \sum_{i=1\ldots N} P_i$$

where $\alpha \in [0,1]$ (25% in the figure) is the share transferred in Quota.



Full surplus, the full is noted $\boldsymbol{K}$ (2M€ in the example), α_i represents the cession rate of policy i.

$$S_r = \sum_{i=1\ldots n} \underbrace{\left(\frac{(K_i - \boldsymbol{K})_+}{K_i}\right)}_{\alpha_i} S_i \qquad P_r = \sum_{i=1\ldots N} \left(\frac{(K_i - \boldsymbol{K})_+}{K_i}\right) P_i$$

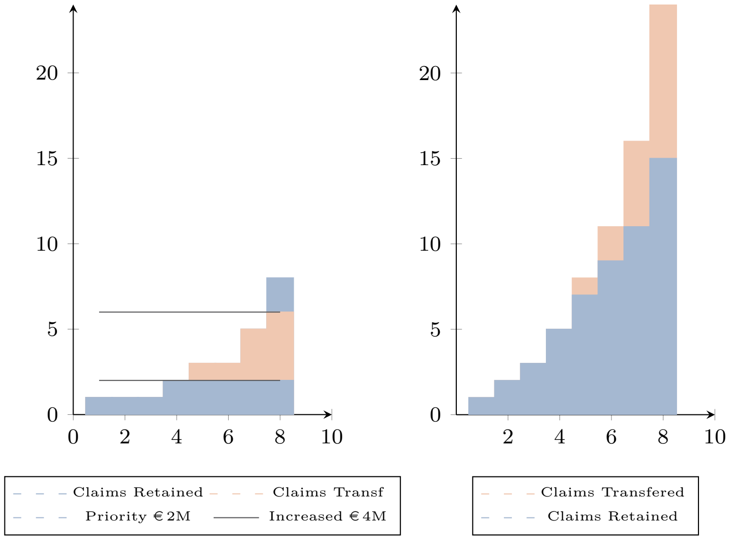


Surplus per claim

The insurer sets the priority a and the scope b (respectively 2M€ and 4M€ in the figure).

$$S_r = \sum_{i=1\ldots n} \min\left((S_i - a)^+, b\right)$$

The premium is set by the reinsurer, based on its estimate of $\mathbb{E}[S_r]$.

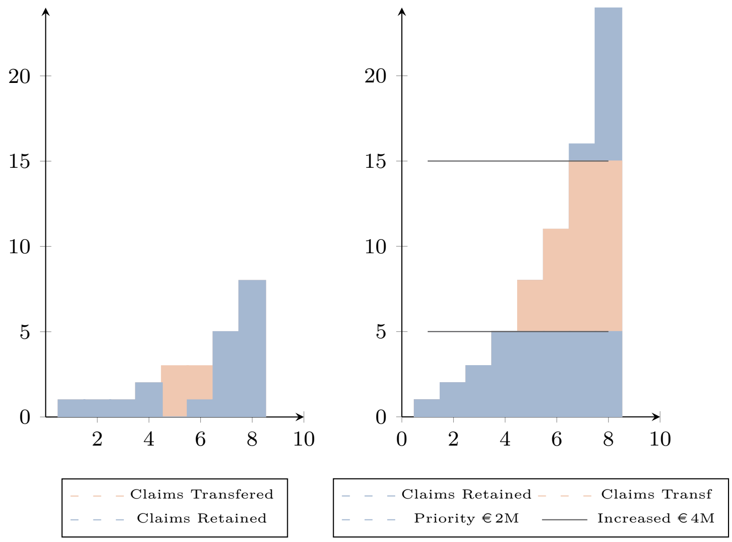


WXL-R = *Working XL per Risk*

Surplus per event

$$S_r = \sum_{\substack{Cat_j,\\ i=1\ldots N}} \min\left(\mathbb{1}_{i \in Cat_j} \times (S_i - a)^+, b\right)$$

In the illustration, claims refer to a single event, with a priority of 5M€ and a scope of 10M€.



Cat-XL = *Catastrophe XL*

Annual excess losses

Stop Loss occurs when the cumulative annual losses deteriorate. It is expressed on the basis of the ratio S/P with a priority and a scope of XL expressed in %.

$$S_r = \min\left(\left(\sum_{i=1\ldots N} S_i - aP\right)^+, bP\right)$$

116 – The main clauses in reinsurance

The deductible a^{ag} and the **aggregate limit** b^{ag} apply after the calculation of S_r.

$$S_r^{ag} = \min\left((S_r - a^{ag})^+, b^{ag}\right)$$

The objective of the **indexation clause** is to maintain the terms of the treaty over several successive financial years. The treaty limits are aligned with an economic index (salary, currency, price index, ...).

With **stabilization clause**, when the claim suffers from a long settlement, or even a very long one (at least ≥ 1 year), the treaty limits are updated in the calculation of the S_r so that the respective shares of the reinsurer and the ceding company initially planned are generally respected.

With the **interest sharing clause**, if in a transaction or court judgment a distinction has been made between compensation and interest, the interest accrued between the date of the loss and the date of actual payment of the compensation will be divided between the ceding company and the reinsurer in proportion to their respective burden resulting from the application of the treaty excluding interest.

The **guarantee reinstatement clause** only concerns *processed in excess of loss by risk or by event* which could be triggered several times during the year. The reinsurer limits its benefit to N times the scope of the XS, in return for the payment of an additional premium. Reconstitution can be done pro rata temporis (time remaining until the expiry date of the treaty) or pro rata to the absorbed capital, or both (double pro rata).
Interlocking Clause is used in event-based XS treaties, which operate by subscription exercise and not by occurrence exercise. The interlocking clause will have the effect of recalculating the treaty limits, because the same event can trigger the treaty for both n and $N-1$ subscriptions.

117 – Public reinsurance

The Caisse Centrale de Réassurance (CCR) offers, with the State guarantee, unlimited coverage for branches specific to the French market.

- exceptional risks linked to transport,
- liability insurance for operators of nuclear vessels and installations,
- the risks of natural disasters,
- the risks of attacks and acts of terrorism,
- the Public Credit Insurance Supplement (CAP).

It also manages certain Public Funds on behalf of the State, in particular the Cat Nat regime.
Also, the GAREAT is a non-profit Economic Interest Group (GIE), mandated by its members, which manages reinsurance of risks of attacks and acts of terrorism with the support of the State via the CCR.

118 – Securitization / CatBonds

Why? The financial capacity of all insurers and reinsurers combined does not cover the damage of a major earthquake in the United States. (≥ 200 B€). This amount corresponds to less than 1% of the capitalization of American financial markets.
Securitization transforms an insurance risk into a negotiable security, often into bonds called Cat-Bonds. It consists of an exchange of principal for periodic payment of coupons, in which the payment of coupons and/or the repayment of principal are conditional on the occurrence of a triggering event defined a priori. The rates on these bonds are increased based on the risk, not of default or counterparty, but of the occurrence of the event (less than 1%). The structure dedicated to this transformation is called a Special Purpose Vehicle (SPV).
The trigger can be directly linked to the results of the predecessor (Compensation), depend on a loss index, a measurable parameter (sum of excess rainfall, Richter scale, mortality rate), or a model (RMS & Equecat Storm modeling).

Criteria	Compensative	Hint	Parametric	Model
Transparency	$\ominus$	$\oplus$	$\oplus$	$\oplus$
Basis risk	$\oplus$	$\ominus$	$\ominus$	$\oplus$
Moral hazard	$\ominus$	$\oplus$	$\oplus$	$\oplus$
Universality of perils	$\oplus$	$\oplus$	$\ominus$	$\oplus$
Trigger delay	$\ominus$	$\ominus$	$\oplus$	$\oplus$

119 – Pareto's Law

Let the random variable X follow a Pareto distribution with parameters (x_{m}, k), k is the Pareto index:

$$\mathbb{P}(X > x) = \left(\frac{x}{x_{\mathrm{m}}}\right)^{-k} \text{ with } x \geq x_{\mathrm{m}}$$

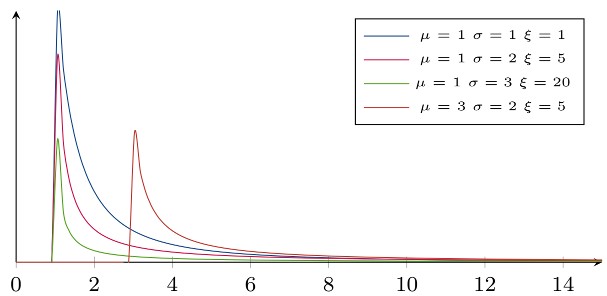
$$f_{k,x_{\mathrm{m}}}(x) = k\,\frac{x_{\mathrm{m}}^k}{x^{k+1}} \text{ for } x \geq x_{\mathrm{m}}$$

Generalized Pareto Law (GPD) has 3 parameters μ, σ and ξ.

$$F_{\xi,\mu,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi} & \text{for } \xi \neq 0, \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right) & \text{for } \xi = 0. \end{cases}$$

for $x \geq \mu$ when $\xi \geq 0$ and $\mu \leq x \leq \mu - \sigma/\xi$ when $\xi < 0$ and where $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ the scale and $\xi \in \mathbb{R}$ the form. Note that some references give the "shape parameter", as $\kappa = -\xi$.

$$f_{\xi,\mu,\sigma}(x) = \frac{1}{\sigma}\left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)} = \frac{\sigma^{\frac{1}{\xi}}}{(\sigma + \xi(x-\mu))^{\frac{1}{\xi}+1}}$$



120 – Generalized extreme value law

The distribution function of the generalized law of extremes is

$$F_{\mu,\sigma,\xi}(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for $1 + \xi(x-\mu)/\sigma > 0$, or $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ of scale and $\xi \in \mathbb{R}$ the form. For $\xi = 0$ the expression is defined by its limit at 0.

$$\begin{aligned} f_{\mu,\sigma,\xi}(x) =& \frac{1}{\sigma}\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} \\ & \times \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \end{aligned}$$

$$f(x;\mu,\sigma,0) = \frac{1}{\sigma}\exp\left(-\frac{x-\mu}{\sigma}\right)\exp\left[-\exp\left(-\frac{x-\mu}{\sigma}\right)\right]$$

121 – Gumbel's Law

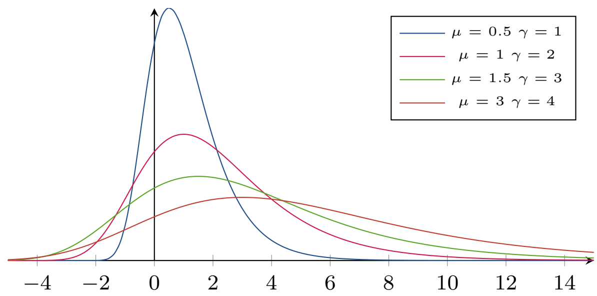
The distribution function of Gumbel's law is :

$$F_{\mu,\sigma}(x) = e^{-e^{(\mu-x)/\sigma}}.$$

For $\mu = 0$ et $\sigma = 1$, we obtain the standard Gumbel law. Gumbel's law is a special case of GEV (with $\xi = 0$).

Its density :

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma}e^{\left(\frac{x-\mu}{\sigma} - e^{-(x-\mu)/\sigma}\right)}$$



122 – Weibull's Law

The **Weibull distribution** has the distribution function defined by :

$$F_{\alpha,\mu,\sigma}(x) = 1 - e^{-((x-\mu)/\sigma)^{\alpha}}$$

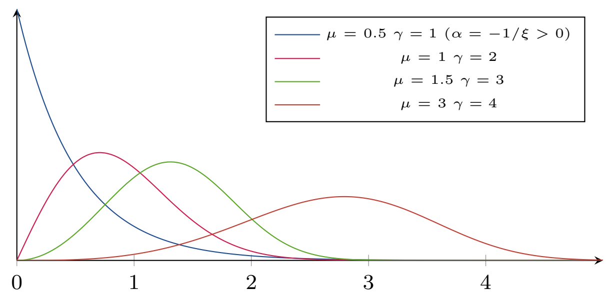
or $x > \mu$. Its probability density is :

$$f_{\alpha,\mu,\sigma}(x) = (\alpha/\sigma)((x-\mu)/\sigma)^{(\alpha-1)}e^{-((x-\mu)/\sigma)^{\alpha}}$$

or $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ of scale and $\alpha = -1/\xi > 0$ the form.

The Weibull distribution is often used in the field of lifetime analysis. It is a special case of the GEV when $\xi < 0$.

If the failure rate decreases over time then, $\alpha < 1$. If the failure rate is constant over time then, $\alpha = 1$. If the failure rate increases over time then, $\alpha > 1$. Understanding the failure rate can provide insight into the cause of failures.



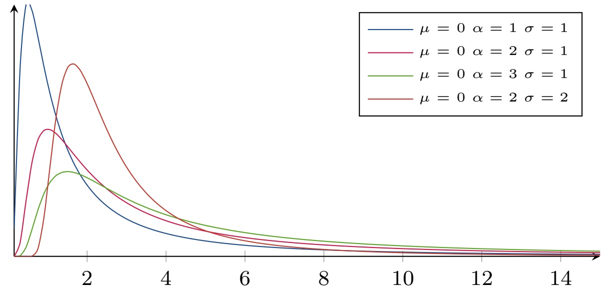
123 – Fréchet's Law

Its distribution function of **Frechet's law** is given by

$$F_{\alpha,\mu,\sigma}(x) = \mathbb{P}(X \leq x) = \begin{cases} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}} & \text{if } x > \mu \\ 0 & \text{otherwise.} \end{cases}$$

or $\mu \in \mathbb{R}$ is the location, $\sigma > 0$ the scale and $\alpha = 1/\xi > 0$ the form. This is a special case of GEV when $\xi > 0$.

$$f_{\alpha,\mu,\sigma}(x) = \frac{\alpha}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-1-\alpha} e^{-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha}}$$



124 – Link between GEV, Gumbel, Fréchet and Weibull
The shape parameter ξ governs the behavior of the distribution tail. The subfamilies defined by $\xi = 0$, $\xi > 0$ and $\xi < 0$ correspond respectively to the Gumbel, Fréchet and Weibull families :

- Gumbel or type I extreme value law
- Fréchet or type II extreme value law, if $\xi = \alpha^{-1}$ with $\alpha > 0$,
- Reversed Weibull ($\overline{F}$) or type III extreme value law, if $\xi = -\alpha^{-1}$, with $\alpha > 0$.

---

125 – General Extreme Value Theorem
Either $X_1, \dots, X_n$ *iid*, X distribution function F_X and either $M_n = \max(X_1, \dots, X_n)$.
The theory gives the exact distribution of the maximum :

$$\begin{aligned}\mathcal{P}(M_n \leq z) =& \Pr(X_1 \leq z, \dots, X_n \leq z)\\ =& \mathcal{P}(X_1 \leq z) \cdots \mathcal{P}(X_n \leq z) = (F_X(z))^n.\end{aligned}$$

If there exists a sequence of pairs of real numbers (a_n, b_n) such that $a_n > 0$ and $\lim_{n\to\infty} \mathcal{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) = F_X(x)$, or F_X is a non-degenerate distribution function, then the limit of the function F_X belongs to the family of GEV laws.

126 – Sub-exponential density
Case of powers
If $\overline{F}_X(x) = \mathbb{P}(X > x) \sim c\, x^{-\alpha}$ when $x \to \infty$ for one $\alpha > 0$ and a constant $c > 0$ then the law of X is sub-exponential.
If F_X is a continuous distribution function of expectation $\mathbb{E}[X]$ finished, we call the index of major risks by

$$D_{F_X}(p) = \frac{1}{\mathbb{E}[X]} \int_{1-p}^{1} F_X^{-1}(t)dt,\ p \in [0, 1]$$

This excess distribution decays less quickly than any exponential distribution. It is possible to consider this statistic :

$$T_n(p) = \frac{X_{(1:n)} + X_{(2:n)} + \ldots + X_{(np:n)}}{\sum_{1\leq i\leq n}(X_i)} \text{ où } \frac{1}{n} \leq p \leq 1$$

$X_{(i:n)}$ designates the i^{th} max of the X_i.

127 – Pickands-Balkema-de Haan theorem (law of excesses)
Let X be of distribution F_X, and let u be a high threshold. Then, for a large class of F_X distributions, the conditional excess distribution

$$X_u := X - u \mid X > u$$

is approximated, for u sufficiently large, by a generalized Pareto distribution (GPD) :

$$\mathbb{P}(X - u \leq y \mid X > u) \approx G_{\xi,\sigma,\mu=0}(y) := 1 - \left(1 + \frac{\xi(x-\mu)}{\sigma}\right)^{-1/\xi}$$

$y \geq 0$. In other words, for $u \to x_F := \sup\{x : F(x) < 1\}$,

$$\sup_{0\leq y<x_F-u} |\mathbb{P}(X - u \leq y \mid X > u) - G_{\xi,\sigma,\mu=0}(y)| \to 0.$$

This theorem justifies the use of the *Pareto law (generalized)* to model excesses beyond a threshold, which is precisely the framework of reinsurance treaties in *excess of loss* by risk, by event or annual accumulation.

---

128 – Reinsurance data
As reinsurance compensates for aggregations of losses or extreme losses, it often uses historical data which should be used with caution :

- updating of data (impact of monetary inflation).
- the revaluation takes into account the evolution of the risk :
  - changes in premium rates, guarantees and terms of contracts,
  - the evolution of claims costs (construction cost index, car repair cost index,...)
  - the evolution of the legal environment.
- the adjustment of the statistics to take into account the evolution of the portfolio base :
  - font profile (number, capitals, ...),
  - nature of guarantees (evolution of deductibles, exclusions...)

After these corrections, the data are said to be “as if” (in economics, we use the expression counterfactual).

129 – Burning Cost
X_i^j designates the i^e disaster of the year j as if updated, revalued and corrected, n^j the number of claims per year j, c^j the insurer's responsibility c. The pure rate by the **Burning Cost** method is given by the formula :

$$BC_{pur} = \frac{1}{s} \sum_{j=1}^{n} \frac{c^j}{a_j}$$

The Burning Cost is only an average of the crossed S/P ratios : the claims payable by the reinsurer on the premiums received by the ceding company. The Burning Cost premium is then : $P_{pure} = BC_{pur} \times a_{s+1}$.
In the case of a p XS f),

$$c^j = \sum_{i=1}^{n^j} \max\left(\left(X_i^j - f\right), p\right) \mathbb{1}_{x^j \geq f}$$

While life insurance calculates premium rates with reference to capital, non-life insurance uses the insured value as a reference, reinsurance takes the total premiums of the ceding company as a reference, called the **base**. We note a_j denotes the premium base for year j and a_{s+1}^* denotes the estimated base for the coming year and where s denotes the number of years of history.

130 – The Poisson-Pareto model

[Bonus of XS or XL] Let p and f be the scope and priority (franchise) of the XS, respectively, with the limit $l = p + f$ (p XS f).
The XS bonus corresponds to :

$$\mathbb{E}\left[S_N\right] = \mathbb{E}\left[\sum_{i=1}^{N} Y_i\right] = \mathbb{E}[N] \times \mathbb{E}[Y]$$

où

$$\mathbb{E}[Y] = l\mathbb{P}[X > l] - f \times \mathbb{P}[X \geq f] + \mathbb{E}[X \mid f \geq x \geq l]$$

If $l = \infty$ and $\alpha \neq 1$:

$$\mathbb{E}[S_N] = \lambda \frac{x_{\mathrm{m}}^{\alpha}}{\alpha - 1} f^{1-\alpha}$$

if $l = \infty$ and $\alpha = 1$ there is no solution.
If $l < \infty$ and $\alpha \neq 1$:

$$\mathbb{E}[S_N] = \lambda \frac{x_{\mathrm{m}}^{\alpha}}{\alpha - 1} \left(f^{1-\alpha} - l^{1-\alpha}\right)$$

If $l < \infty$ and $\alpha = 1$:

$$\mathbb{E}[S_N] = \lambda x_{\mathrm{m}} \ln\left(\frac{1}{f}\right)$$

131 – The Poisson-LogNormal Model

If X follows a $\mathcal{LN}orm(x_{\mathrm{m}}, \mu, \sigma)$ then $X - x_{\mathrm{m}}$ follows a $\mathcal{LN}orm(\mu, \sigma)$ It comes :

$$\mathbb{P}[X > f] = \mathbb{P}[X - x_{\mathrm{m}} > f - x_{\mathrm{m}}] = 1 - \Phi\left(\frac{\ln(f - x_{\mathrm{m}}) - \mu}{\sigma}\right)$$

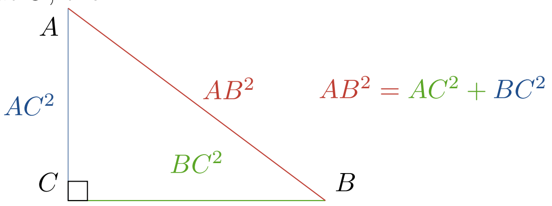
$$\begin{aligned}
&\mathbb{E}[X \mid X > f] \\
&\qquad = \mathbb{E}\left[X - x_{\mathrm{m}} \mid X - x_{\mathrm{m}} > f - x_{\mathrm{m}}\right] + x_{\mathrm{m}}\mathbb{P}[X > f] \\
&\qquad = e^{m+\sigma^2/2}\left[1 - \Phi\left(\frac{\ln(f - x_{\mathrm{m}}) - (\mu + \sigma^2)}{\sigma}\right)\right] \\
&\qquad\quad + x_{\mathrm{m}}\left(1 - \Phi\left(\frac{\ln(f - x_{\mathrm{m}}) - \mu}{\sigma}\right)\right)
\end{aligned}$$

With frankness and without limits :

$$\begin{aligned}
&\mathbb{E}[S_N] \\
&\quad = \lambda (\mathbb{E}\left[X - x_{\mathrm{m}} \mid X - x_{\mathrm{m}} > f - x_{\mathrm{m}}\right] + x_{\mathrm{m}}\mathbb{P}[X > f] - f\mathbb{P}[X > f \\
&\quad = \lambda\left(e^{m+\sigma^2/2}\left[1 - \Phi\left(\frac{\ln(f - x_{\mathrm{m}}) - (\mu + \sigma^2)}{\sigma}\right)\right]\right) \\
&\qquad + \lambda(x_{\mathrm{m}} - l)\left(1 - \Phi\left(\frac{\ln(f - x_{\mathrm{m}}) - \mu}{\sigma}\right)\right)
\end{aligned}$$

132 – Pythagoras

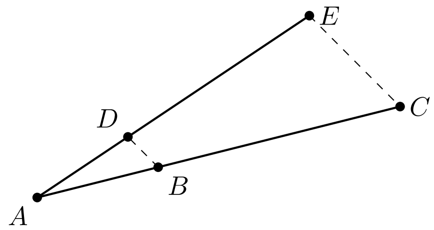
In a right triangle, the square of the hypotenuse is equal to the sum of the squares of the other two sides. If ABC is right-angled at C, then



133 – Thales

Let two lines **intersect at a point** A, and let two lines (BC) and (DE) **parallel**, intersecting the two lines at B, D and C, E, then :

$$\frac{AB}{AD} = \frac{AC}{AE} = \frac{BC}{DE}$$



134 – Quadratic equation

$$ax^2 + bx + c = 0$$

The discriminant is defined by :

$$\Delta = b^2 - 4ac$$

- If $\Delta > 0$, the equation has two distinct solutions :

$$x_1 = \frac{-b + \sqrt{\Delta}}{2a}, \quad x_2 = \frac{-b - \sqrt{\Delta}}{2a}$$

- If $\Delta = 0$, the equation has a double solution :

$$x = \frac{-b}{2a}$$

- If $\Delta < 0$, the equation has a solution in the imaginary

$$x_1 = \frac{-b + i\sqrt{\Delta}}{2a}, \quad x_2 = \frac{-b - i\sqrt{\Delta}}{2a}$$

135 – Factorial, Counting and Gamma Functions

The **factorial** function (of $\mathbb{N}$ in $\mathbb{N}$) is defined by $0! = 1$ and $n! = n \times (n-1) \times \cdots \times 2 \times 1 =$ permutations of n elements $C_n^k = \binom{k}{n} = \frac{n!}{k!(n-k)!} =$ choice of k elements among n the C_n^k are also calculated by Pascal's triangle and verify:

$$C_n^k = C_n^{n-k}, C_n^k + C_n^{k+1} = C_{n+1}^{k+1}.$$

Let E be a set of cardinal $\mathrm{Card}(E)$ and parts $\mathcal{P}(E)$:

$$\begin{aligned} \mathrm{Card}(\mathcal{P}(E)) &= 2^{\mathrm{Card}(E)} \\ \mathrm{Card}(A \times B) &= \mathrm{Card}(A) \times \mathrm{Card}(B) \\ \mathrm{Card}(A \cup B) &= \mathrm{Card}(A) + \mathrm{Card}(B) - \mathrm{Card}(A \cap B) \end{aligned}$$

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t}\, dt$$

The function Γ can be seen as the extension of the factorial : $\Gamma(n+1) = n!$.

136 – Binomial expansion

For a positive integer n,

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

137 – Sequences

Arithmetic sequences of reason r

$$\left\{ \begin{array}{ll} u_{n+1} & = u_n + r \\ u_0 & \in \mathbb{R} \end{array} \right. \Rightarrow \left\{ \begin{array}{ll} u_n & = nr + u_0 \\ \sum_{k=0}^{n} u_k & = \frac{(n+1)(2u_0+nr)}{2} \end{array} \right.$$

geometric sequences of reason q $\left\{ \begin{array}{lll} u_{n+1} & = & q \times u_n \\ u_0 & \in & \mathbb{R} \end{array} \right.$

$$\Rightarrow \left\{ \begin{array}{ll} u_n & = u_0 \times q^n \\ \sum_{k=0}^{n} u_k & = \left\{ \begin{array}{ll} (n+1)u_0 & \text{if} \\ u_0 \frac{1-q^{n+1}}{1-q} & \text{otherwise} \end{array} \right. \quad q = 1 \end{array} \right.$$

138 – Exponential and Logarithm

The exponential function e^x can be defined by the following power series expansion :

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

This series converges for all $x \in \mathbb{R}$ and allows us to define the exponential as an infinite sum.

The natural logarithm function $\ln(x)$ is defined as the antiderivative of the function $\frac{1}{x}$. In other words :

$$\frac{d}{dx} \ln(x) = \frac{1}{x}$$

with the condition $\ln(1) = 0$. This definition allows us to establish the link between the exponential and the logarithm via inversion : $e^{\ln(x)} = x$ for $x > 0$.

139 – Congruence relationship

Let $m > 0$. We say that two real numbers a and b are congruent modulo m if there exists a relative integer $k \in \mathbb{Z}$ such that :

$$a = b + km.$$

On note $a \equiv b \pmod{m}$.

In trigonometry, we often choose $m = 2\pi$ or $m = \pi$.

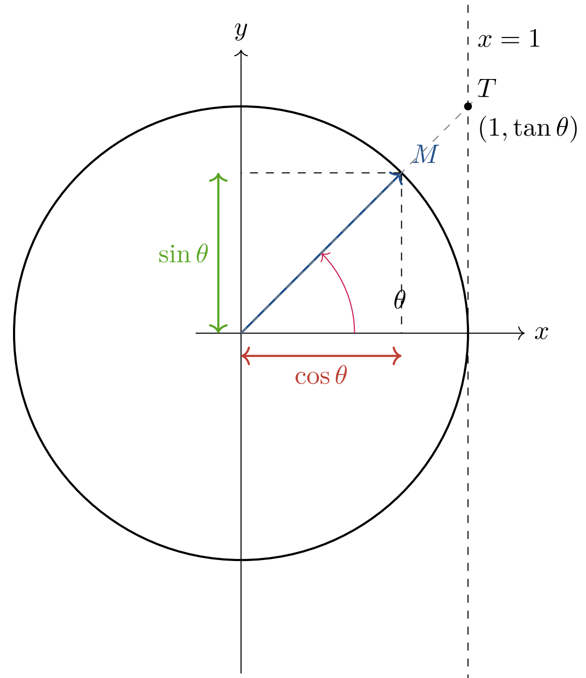
140 – Trigonometric circle — sine, cosine, tangent

$M = (\cos\theta, \sin\theta)$

If $\theta \not\equiv \frac{\pi}{2} \pmod{\pi}$, we define : $\tan\theta = \dfrac{\sin\theta}{\cos\theta}$

$\cos^2\theta + \sin^2\theta = 1$

θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0
$\sin\theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\tan\theta$	0	$\frac{\sqrt{3}}{3}$	1	$\sqrt{3}$	indéfini



$-\theta$	$\theta + \pi$	$\pi - \theta$
$\cos\theta$	$-\cos\theta$	$-\cos\theta$
$-\sin\theta$	$-\sin\theta$	$\sin\theta$
$\theta + 2\pi$	$\frac{\pi}{2} - \theta$	$\frac{\pi}{2} + \theta$
$\cos\theta$	$\sin\theta$	$-\sin\theta$
$\sin\theta$	$\cos\theta$	$\cos\theta$

$$\cos x = \cos y \iff \begin{cases} x \equiv y \pmod{2\pi} \text{ or} \\ x \equiv -y \pmod{2\pi} \end{cases}$$

$$\sin x = \sin y \iff \begin{cases} x \equiv y \pmod{2\pi} \text{ or} \\ x \equiv \pi - y \pmod{2\pi} \end{cases}$$

$$\tan x = \tan y \iff x \equiv y \pmod{\pi}$$

141 – Properties of the congruence relation

Let $m > 0$ et $a, b, c, d \in \mathbb{R}$. Then :

- **Reflexivity** : $a \equiv a \pmod{m}$.
- **Symmetry** : $a \equiv b \pmod{m} \iff b \equiv a \pmod{m}$.
- **Transitivity** : if $a \equiv b \pmod{m}$ and $b \equiv c \pmod{m}$, then $a \equiv c \pmod{m}$.
- **Additivity** : if $a \equiv b \pmod{m}$ and $c \equiv d \pmod{m}$, then $a + c \equiv b + d \pmod{m}$.

142 – Derivatives and primitives

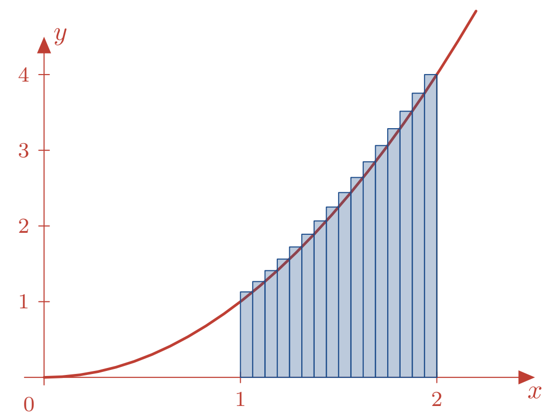
f continues in $x_{\mathrm{m}} \Leftrightarrow \lim\limits_{x \to x_{\mathrm{m}}} f(x) = f(x_{\mathrm{m}})$

f derivable in $x_{\mathrm{m}} \Leftrightarrow \exists \lim\limits_{h \to 0} \dfrac{f(x_{\mathrm{m}} + h) - f(x_{\mathrm{m}})}{h} =: f'(x_{\mathrm{m}})$

The **Riemann integral** of a function $f(x)$ on an interval $[a, b]$ is the limit, if it exists, of the sum of the areas of the rectangles approaching the area under the curve, given by :

$$\int_a^b f(x)\,dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^*)\Delta x_i,$$

- $[x_{i-1}, x_i]$ is a subdivision of $[a, b]$,
- $\Delta x_i = x_i - x_{i-1}$ is the width of the subinterval,
- $x_i^* \in [x_{i-1}, x_i]$ is an arbitrarily chosen point in each subinterval.



Example of a Riemann integral (upper)*

The **Lebesgue integral** of a function $f(x)$ on a set E is defined by measuring the area under the curve as a function of the values taken by f, given by :

$$\int_E f\,d\mu = \int_0^\infty \mu(\{x \in E : f(x) > t\})\,dt,$$

- μ is a measure (often the Lebesgue measure),
- $\{x \in E : f(x) > t\}$ represents the set of points where $f(x)$ exceeds t.

Unlike Riemann, Lebesgue groups points according to their values rather than their position.

function ($n \in \mathbb{R}$)	derivative	primitive
x	1	$\frac{x^2}{2}+C$
x^2	$2x$	$\frac{x^3}{3}+C$
$1/x$	$-1/x^2$	$\ln(x)+C$
$\sqrt{x}=x^{1/2}$	$\frac{1}{2\sqrt{x}}$	$\frac{2}{3}x^{3/2}+C$
$x^n, n \neq -1$	nx^{n-1}	$\frac{x^{n+1}}{n+1}+C$
$\ln(x)$	$1/x$	$x\ln(x)-x+C$
e^x	e^x	e^x+C
$a^x = e^{x\ln(a)}$	$\ln(a)\times a^x$	$a^x/\ln(a)+C$
$\sin(x)$	$\cos(x)$	$-\cos(x)+C$
$\cos(x)$	$-\sin(x)$	$\sin(x)+C$
$\tan(x)$	$1+\tan(x)$	$-\ln(\lvert\cos(x)\rvert)+C$
$1/\left(1+x^2\right)$	$-2x/\left(1+x^2\right)^2$	$\arctan(x)+C$

$$
\begin{aligned}
(u+v)' &= u'+v' \\
(ku)' &= ku' \\
(u\times v)' &= u'v+uv' \\
\left(\tfrac{u}{v}\right)' &= \tfrac{u'v-uv'}{v^2} \\
(u^n)' &= nu^{n-1}\times u'
\end{aligned}
\qquad
\begin{aligned}
\left(\tfrac{1}{u}\right)' &= -\tfrac{u'}{u^2} \\
(\ln(u))' &= \tfrac{u'}{u} \\
(\exp(u))' &= \exp(u)\times u' \\
(f(u))' &= f'(u)\times u' \\
(f\circ u)' &= (f'\circ u)\times u'
\end{aligned}
$$

143 – Integration by parts

Either $u(x)$ and $v(x)$ two continuously differentiable functions on the interval $[a,b]$, then

$$\int_a^b u(x)v'(x)\,dx = [u(x)v(x)]_a^b - \int_a^b u'(x)v(x)\,dx$$

where :

- $u(x)$ is a function whose derivative is known $u'(x)$,
- $v'(x)$ is a function whose primitive is known $v(x)$.

144 – Integration with change of variable

Either $f(x)$ a continuous function and $x=\phi(t)$ a change of variable, where ϕ is a differentiable function. Then :

$$\int_a^b f(x)\,dx = \int_{\phi^{-1}(a)}^{\phi^{-1}(b)} f(\phi(t))\phi'(t)\,dt$$

where :

- $x=\phi(t)$ represents the change of variable,
- $\phi'(t)$ is the derivative of $\phi(t)$,
- the limits of the integral are adjusted according to the change of variable.

145 – Taylor formula

Either $f(x)$ a function n - times differentiable at a point a. Taylor's development of $f(x)$ around a is given by :

$$
\begin{aligned}
f(x) =& f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 \\
&+ \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + \mathcal{O}_n(x)
\end{aligned}
$$

where :

- $f^{(n)}(a)$ is the n^{th} derived from f evaluated in a,
- $\mathcal{O}_n(x)$ is the remainder of the Taylor development, representing the approximation error when truncating the series after the order term n, then

$$\lim_{x\to 0}\frac{\mathcal{O}_n(x)}{x^n} \Rightarrow 0$$

146 – Intermediate Value Theorem

Either f a continuous function on a closed interval $[a,b]$ and $f(a)\neq f(b)$. The intermediate value theorem states that for any real number c between $f(a)$ and $f(b)$, there is a point $x\in[a,b]$ such as :

$$f(x)=c$$

In other words, if a function is continuous on an interval, it takes all the values between $f(a)$ and $f(b)$ at least once.

147 – Matrices and properties

Diagonal matrices : A matrix is said to be diagonal if all elements outside the main diagonal are zero. For a matrix $A\in\mathbb{R}^{n\times n}$, it is written :

$$A=\operatorname{diag}(a_1,a_2,\ldots,a_n)$$

where a_i are the diagonal elements.
Triangular matrices : A matrix is upper triangular if all elements below the diagonal are zero, that is, $A_{ij}=0$ for $i>j$. Conversely, it is lower triangular if $A_{ij}=0$ for $i<j$.

148 – Determinant of a matrix

The determinant of a square matrix $A\in\mathbb{R}^{n\times n}$ is a scalar, denoted $\det(A)$:

- If A is a square matrix $n\times n$, then A is invertible if and only if $\det(A)\neq 0$.
- The determinant of an upper or lower triangular matrix or a diagonal matrix :

$$\det(A)=\prod_{i=1}^{n}A_{ii}$$

- $$\det(AB)=\det(A)\cdot\det(B),\quad \det(\lambda B)=\lambda\det(B),$$
- $$\det(A^T)=\det(A)$$
- If a matrix A contains two identical rows or columns, then $\det(A)=0$.

Calculation of the determinant : The determinant of a matrix 2×2 is simply calculated by :

$$\det\begin{pmatrix}a & b\\ c & d\end{pmatrix}=ad-bc$$

For a matrix 3×3, it is given by :

$$\det\begin{pmatrix}a & b & c\\ d & e & f\\ g & h & i\end{pmatrix}=a(ei-fh)-b(di-fg)+c(dh-eg)$$

For higher-dimensional matrices, the determinant can be calculated by cofactors or via a reduction method (e.g., Gauss's method).

149 – Invertibility of a matrix

A matrix $A \in \mathbb{R}^{n\times n}$ is invertible if there exists a matrix A^{-1} such as :

$$AA^{-1} = A^{-1}A = I_n$$

or I_n is the identity matrix. The invertibility of a matrix is guaranteed by $\det(A) \neq 0$.

Trace : The trace of a square matrix A, noted $\mathrm{Tr}(A)$, is the sum of its diagonal elements :

$$\mathrm{Tr}(A) = \sum_{i=1}^{n} A_{ii}$$

It often represents quantities linked to the sum of the eigenvalues of a matrix.

Cholesky decomposition : Cholesky decomposition is applicable to positive definite symmetric matrices. It allows a matrix to be factorized $A \in \mathbb{R}^{n\times n}$ into a product of the form :

$$A = LL^T$$

or L is a lower triangular matrix. This decomposition is useful in numerical calculations and optimization algorithms.

150 – Gradient and Hessian matrix

For a function $f : \mathbb{R}^n \to \mathbb{R}$ of class C^2, we define :

- The **gradient** $\nabla f(x)$ as the vector of first partial derivatives :

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

- The **Hessian matrix** $\nabla^2 f(x)$ as the symmetric matrix of second derivatives :

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

151 – Implicit Function Theorem

Either $F : \mathbb{R}^2 \to \mathbb{R}$ a class function C^1, and suppose that $F(a^\star, b) = 0$ for a certain couple $(a^\star, b) \in \mathbb{R}^2$. If

$$\frac{\partial F}{\partial y}(a^\star, b) \neq 0,$$

so there is a real $h > 0$ and a single function φ, defined on a neighbourhood $(a^\star - h, a^\star + h)$, such as

$$\varphi(a^\star) = b \quad \text{et} \quad \forall x \in (a^\star - h, a^\star + h), \quad F(x, \varphi(x)) = 0$$

In addition, the implicit function φ is of class C^1 and its derivative is given by :

$$\varphi'(x) = -\left.\frac{\partial F/\partial x}{\partial F/\partial y}\right|_{y=\varphi(x)}$$

My personal notes

About the authors

Martial Phélippé-Guinvarc'h @ has been an actuary and lecturer at Le Mans University since 2012 and has also taught at EURIA since 2003. He is an actuary (EURIA, 2006), a member of the Institute of Actuaries (IA) and a member of the European Association of Agricultural Economists (EAAE).
Martial Phélippé-Guinvarc'h teaches statistics, data analysis, actuarial modeling, programming, particularly in SAS and SAS IML, financial mathematics, commodity markets, market finance and derivatives, non-life actuarial science, life actuarial science, provisioning, solvency, reinsurance, insurance economics, risk management, and asset-liability management. He is a SAS Protor and has obtained the SAS Join certificate Insurance and Economic Analytics for the MBFA Master's degree from Le Mans University.
Marcelo Moreno Porras @ is an Associate Professor at the Department of Applied Economics I and Economic History and Institutions at Universidad Rey Juan Carlos (Madrid, Spain) since 2023, specializing in econometrics, macroeconomics, and programming. He is also a Software Developer and IT specialist at a private company in Mijas (Spain). He is the author of several contributions to teaching innovation, co-author of publications on incorporating Sustainable Development Goals in economics education, and has technical expertise on multiple programming languages and statistical tools. His academic background includes a Bachelor's degree in Economics (URJC, top 1 % national ranking), a Master's in Modern Economic Analysis (URJC), and a Master's in Applied Statistics (Nebrija University).