Financial Risk Analytics I Notes

Nong Minh Hieu[1]

[1] School of Physical and Mathematical Sciences, Nanyang Technological University (NTU - Singapore)

Contents

1 Financial Modelling

1.1 Brownian Motion

Definition 1.1 (Brownian Motion).
A stochastic process $(B_t)_{t\geq 0}$ is called a Brownian motion if it satisfies the following properties:

- ***Starts at 0*** *: $B_0 = 0$ (almost surely).*
- ***Continuous sample path*** *: $t \to B_t$ is continuous (almost surely).*
- ***Independence of increments*** *: For any finite sequence of time $t_0 < t_1 < \cdots < t_n$, the increments*

$$B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

  are independent.
- *For all $0 \leq s < t$, $B_t - B_s \sim \mathcal{N}(0, t-s)$.*

Remark : Brownian motion can be seen under the lens of Random Walk when we scale the interval between steps of the walk down infinitesimally.

1.2 Geometric Brownian Motion

Definition 1.2 (Market Return).
Two definitions of market returns:

- *Standard return is defined as:*

$$\frac{\Delta S_t}{S_t} = \frac{S_{t+\Delta t} - S_t}{S_t}$$

- *Log return is defined (for $t \geq 0$) as:*

$$\Delta \log S_t = \log S_{t+\Delta t} - \log S_t = \log \frac{S_{t+\Delta t}}{S_t} = \log \left(1 + \frac{\Delta S_t}{S_t}\right)$$

Where S_t is the market price at time $t \geq 0$.

Proposition 1.1: Geometric Brownian Motion

The Geometric Brownian Motion $(S_t)_{t\geq 0}$ is a stochastic process, which is the solution to the Stochastic Differential Equation of the form

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

The solution for the above equation is given by:

$$\boxed{S_t = S_0 \exp\left(\sigma B_t + \left(\mu - \frac{\sigma^2}{2}\right) t\right), t \geq 0}$$

Proposition 1.2: Distribution of Geometric BM

At any time $T > 0$, the random variable

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

has the **log-normal distribution** with the density function:

$$\boxed{f_{S_T}(x)\frac{1}{\sigma x\sqrt{2\pi T}}\exp\left(-\frac{1}{2\sigma^2 T}\left(\log x - \log S_0 - \left(\mu - \frac{\sigma^2}{2}\right)T\right)^2\right)}$$

1.3 Distribution of market return

Overview : Even though we can model the market returns using Brownian Motion and Geometric Brownian Motion, the assumption of Gaussian market returns is often not true. We can use the following visualizations/tests to verify that assumption:

- Empirical vs. estimated Gaussian CDFs.
- QQ plot.
- The one-sample Kolmogorov-Smirnov test.

For example, the following visualization illustrates how the Gaussian return assumption can underestimate extreme events.

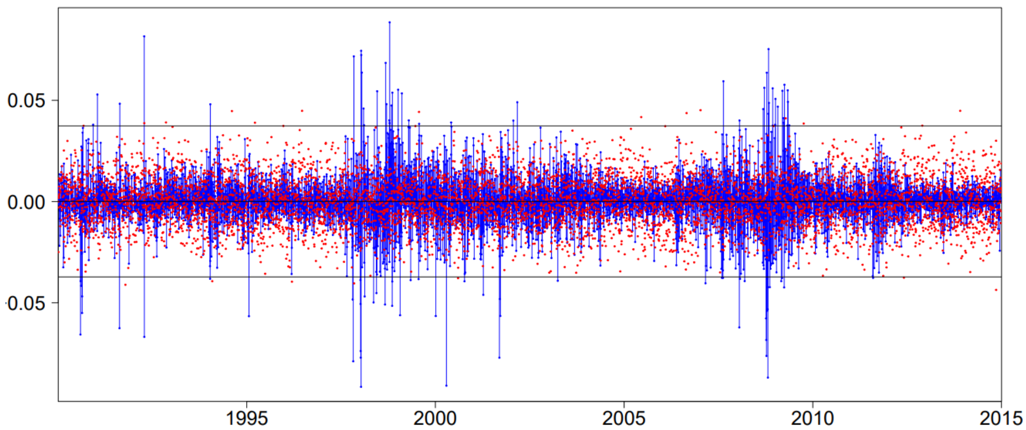

Figure 1: Market returns (blue) vs normalized Gaussian returns (red) (figure sampled from [3])

In later section, we will see that in order to alleviate this problem, **Gram-Charlier expansion** is proposed for higher-order estimate of the empirical CDF of market returns.

Definition 1.3 (Cumulants)**.**
The cumulants of a random variable X is essentially an alternative to moments. The cumulants

Return to contents

are derived from the ***cumulant generating function****:*

$$K_X(t) = \log M_X(t) = \log \mathbb{E}[e^{tX}]$$
$$\kappa_n^X = K_X^{(n)}(0) \quad (n^{th} \textit{ order cumulant})$$

Remark : We have the following remarks about the cumulant:

- $\kappa_1^X = \mathbb{E}[X]$ (Mean).
- $\kappa_2^X = Var(X) = \mathbb{E}\Big[(X - \mu_X)^2\Big]$ (Variance).
- $\kappa_3^X = \mathbb{E}\Big[(X - \mu_X)^3\Big]$ (Third central moment).
- $\kappa_4^X = \mathbb{E}\Big[(X - \mu_X)^4\Big] - 3(\kappa_2^X)^2$.
- From the fourth-order cumulant, the cumulant is no longer consistent to central moment.

Proposition 1.3: Power-series expansion of $K_X(t)$

The power-series expansion of the cumulant generating function is defined as followed:

$$K_X(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \kappa_n^X$$

The above expansion is similar to the following power-series expansion of the moment generating functions:

$$M_X(t) = \sum_{n=1}^{\infty} \frac{t^n}{n!} \mathbb{E}\Big[X^n\Big]$$

Definition 1.4 (Skewness & Excess kurtosis).

The skewness of a random variable X is defined as

$$Sk_X = \frac{\kappa_3^X}{(\kappa_2^X)^{3/2}}$$

The excess kurtosis of a random variable X is defined as

$$EK_X = \frac{\kappa_4^X}{(\kappa_2^X)^{4/2}} = \frac{\kappa_4^X}{(\kappa_2^X)^2}$$

The skewness, intuitively, measures the **degree of asymmetry** of the probability density function. The excess kurtosis measures the **peakedness** of the probability density function. An empirical distribution with skewness and excess kurtosis close to zero will be similar to a Gaussian distribution.

1.4 Gram-Charlier Expansions

To solve the problem of inaccurate approximation of the probability density function of market returns, Gram-Charlier expansion is proposed for higher order approximation.

Definition 1.5 (Hermite polynomial).

The Hermite polynomial is defined as:

$$H_n(x) = (-1)^n \frac{\varphi^{(n)}(x)}{\varphi(x)}$$

Where $\varphi(x)$ is the probability density function of the standard normal distribution. The Hermite polynomial is one of the classical orthogonal polynomials used as the orthogonal basis for the Gram-Charlier expansion [4] (defined below).

Definition 1.6 (Gram-Charlier Expansion).

Given a random variable X, the probability density function of X can be written as a series expansion as followed:

$$f_X(x) = \sum_{n=0}^{\infty} \frac{q_n}{\sqrt{\sigma}} H_n(\overline{x})\varphi(\overline{x})$$

$$Where: \begin{cases} q_0 = 1; \ q_1 = q_2 = 0 \\ \\ q_n = \frac{1}{n!}\mathbb{E}\Big[H_n(\overline{X})\Big], \quad n \geq 3 \end{cases}$$

Where $\overline{x} = \frac{x-\kappa_1^X}{\sqrt{\kappa_2^X}} = \frac{x-\mu}{\sigma}$ is the standardized observation and $\overline{X} = \frac{X-\kappa_1^X}{\sqrt{\kappa_2^X}} = \frac{X-\mu}{\sigma}$ is the standardized random variable. $\varphi(.)$ is the PDF of the standard normal distribution.

2 Superhedging risk measures

2.1 Call and Put options

Definition 2.1 (Put option).
*The **Put option** provides trader with the right but not the obligation to **sell** an asset at a strike price K in future time T. The payoff of a put option is defined as:*

$$C = (K - S_T)^+ = \begin{cases} K - S_T, & \text{when } K \geq S_T \\ \\ 0, & \text{Otherwise} \end{cases}$$

(The payoff is realized when in the future, the asset price falls below the strike price).

Definition 2.2 (Call option).
*The **Call option** provides trader with the right but not the obligation to **buy** an asset at a strike price K in future time T. The payoff of a call option is defined as:*

$$C = (S_T - K)^+ = \begin{cases} S_T - K, & \text{when } S_T \geq K \\ \\ 0, & \text{Otherwise} \end{cases}$$

(The payoff is realized when the future price rises above the strike price).

2.2 Hedging and Pricing options

Definition 2.3 (Cash settlement & Physical delivery).
Take put option issuer as the example, there are two modes of issue (deliveries):

- ***Physical delivery** : The option issuer pays the strike price of K to the option holder in exchange for one unit of asset. This usually applies for physical assets like live cattle, fuel, etc.*
- ***Cash settlement** : The option issuer fulfills the contract by transferring the amount of $(K - S_T)^+$ to the option holder.*

Definition 2.4 (Hedging and Pricing options).
Two issues of options:

- ***Option pricing** : To be fair, option holder have to pay an appropriate price upon signing the contract.*
- ***Option hedging** : Manage a given portfolio such that it contains the required payoff: $(K - S_T)^+$ for put option and $(S_T - K)^+$ for call option.*

Example : Consider the following example, a risky asset priced at time $t = 0$ at $S_0 = 4$ and taking only two possible values at time $t = 1$: $S_1 = \{5, 2\}$.

An option contract promises the payoff:

$$C = \begin{cases} 3, \text{if } S_1 = 5 \\ \\ 0, \text{if } S_1 = 2 \end{cases}$$

(**ii**) **Option hedging** : how to manage the portfolio (α, β) such that $\alpha S_1 + \beta$ matches the payoff at $t = 1$.

$$C = \begin{cases} 3 = 5\alpha + \beta \\ 0 = 2\alpha + \beta \end{cases} \implies \begin{cases} \alpha = 1 \text{ (Buy stock)} \\ \beta = -2 \text{ (Borrow from bank)} \end{cases}$$

(**i**) **Option hedging** : how to charge the option buyer.

$$\begin{aligned} V_0 &= \alpha S_0 + \beta \\ &= 1 \times 4 - 2 = 2 \end{aligned}$$

Definition 2.5 (Arbitrage-free price).
*The **arbitrage-free price** of an option contract should be the initial cost of creating the portfolio:*

$$V_0 = \alpha S_0 + \beta$$

2.3 Risk-neutral probability & Market implied probability

Definition 2.6 (Risk-neutral probability).
With the absence of arbitrage opportunities, the expected payoff ($\mathbb{E}[C]$) should equal the amount of the initial amount V_0 invested in the portfolio:

$$\mathbb{E}[C] = V_0$$

Risk-neutral probability *is the probabilities infered from the above equation.*

Definition 2.7 (Market-implied probability).
*If we match the theoretical price $\mathbb{E}[C]$ with some market price M, we derive the **Market-implied probability**.*

$$\mathbb{E}[C] = M$$

Example : Following the example from the previous section, we have:

$$\begin{aligned} \mathbb{E}[C] &= 3 \times P(S_1 = 5) + 0 \times P(S_1 = 2) \\ &= 3 \times P(S_1 = 5) \end{aligned}$$

Equating the above expected payoff to the initial value invested in the option, we have:

$$\mathbb{E}[C] = V_0 \implies 3 \times P(S_1 = 5) = 2 \implies \begin{cases} P(S_1 = 5) & = \frac{2}{3} \\ \\ P(S_1 = 2) & = \frac{1}{3} \end{cases}$$

On the other hand, the market-implied probability is:

$$\begin{cases} P(S_1 = 5) & = \frac{M}{3} \\ P(S_1 = 2) & = \frac{3-M}{3} \end{cases}$$

3 Value at Risk (VaR)

3.1 Risk Measures

Definition 3.1 (Risk measure).
*A **risk measure** is a mapping that assigns a value V_X to a given **payoff** random variable X.*

These are a few examples of risk measures:

- The **expected value premium principle** :

$$V_X = \mathbb{E}[X] + \alpha \mathbb{E}[X]$$

  For some $\alpha \geq 0$. For $\alpha = 0$, $V_X = \mathbb{E}[X]$ is called the **pure premium risk measure**.

- The **standard deviation premium principle** :

$$V_X = \mathbb{E}[X] - \alpha \sqrt{Var(X)}$$

  For some $\alpha \geq 0$.

- The **Conditional Tail Expectation (CTE)** over negative payoff X is defined as followed:

$$CTE_X = \mathbb{E}\Big[X|X < 0\Big] = \frac{\mathbb{E}[X \mathbb{1}_{\{X<0\}}]}{P(X < 0)}$$

Definition 3.2 (Coherent risk measure).
*A risk measure V is said to be **coherent** if it satisfies the following properties:*

- ***Monotonicity :*** $X \leq Y \implies V_X \leq V_Y$.
- ***Positive Homogeneity :*** $V_{\lambda X} = \lambda V_X, \lambda > 0$.
- ***Translation invariance :*** $V_{\mu + X} = \mu + V_X$.
- ***Sub-additivity :*** $V_{X+Y} \leq V_X + V_Y$.

Definition 3.3 (Distortion risk measure).
*A **distortion risk measure** is any risk measure of the form:*

$$M_X = \mathbb{E}\Big[X g_X(X)\Big]$$

*Where g_X is **non-negative, non-decreasing** function satisfying:*

- $\mathbb{E}\Big[g_X(X)\Big] = 1$.
- ***Positive homogeneity :*** $g_{\lambda X}(\lambda x) = g_X(x), \ \lambda > 0$.
- ***Translation invariance :*** $g_{X+\mu}(x + \mu) = g_X(x)$.

The distortion risk measure M_X satisfies the following properties:

- ***Positive homogeneity :*** *For* $\lambda > 0$, *we have*

$$\mathbb{E}\Big[\lambda X g_{\lambda X}(\lambda X)\Big] = \mathbb{E}\Big[\lambda X g_X(X)\Big] = \lambda\mathbb{E}\Big[X g_X(X)\Big]$$

- ***Translation invariance :*** *For* $\mu \geq 0$, *we have*

$$\mathbb{E}\Big[(X+\mu)g_{X+\mu}(X+\mu)\Big] = \mathbb{E}\Big[(X+\mu)g_X(X)\Big] = \mathbb{E}\Big[X g_X(X)\Big] + \mu\mathbb{E}\Big[g_X(X)\Big] = \mathbb{E}\Big[X g_X(X)\Big] + \mu$$

3.2 Quantile Risk Measures

3.2.1 Cumulative Distribution Function

Definition 3.4 (Cumulative Distribution Function (CDF)).
The CDF of a random variable X is defined as a function $F_X : \mathbb{R} \to [0,1]$ defined by:

$$F_X(x) = P(X \leq x), \quad x \geq 0$$

Any CDF satisfies the following properties:

- ***Non-decreasing.***
- ***Right-continuous.***
- $\lim_{x\to\infty F_X(x)=1}$.
- $\lim_{x\to-\infty F_X(x)=0}$.

(The following are fundamental results of the CDFs, the proof will not be provided).

Proposition 3.1: Continuity of CDFs

For any non-decreasing sequence $(x_n)_{n\geq 1}$ such that $x_n \to x \in \mathbb{R}$, we have:

$$\lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(X \leq x_n) = P(X < x)$$

Proposition 3.2: Discontinuity of CDFs

The CDF of a random variable X is right-continuous (not both way continuous). Hence, when there is a case where the CDF is discontinuous (figure 2) at a given point $x \in \mathbb{R}$, we can compute the probability that $X = x$:

$$P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - \lim_{y\to x^-} F_X(y)$$

Definition 3.5 (Quantile).
Given a random variable X with the CDF $F_X : \mathbb{R} \to [0,1]$ and a level $p \in (0,1)$. The p-quantile

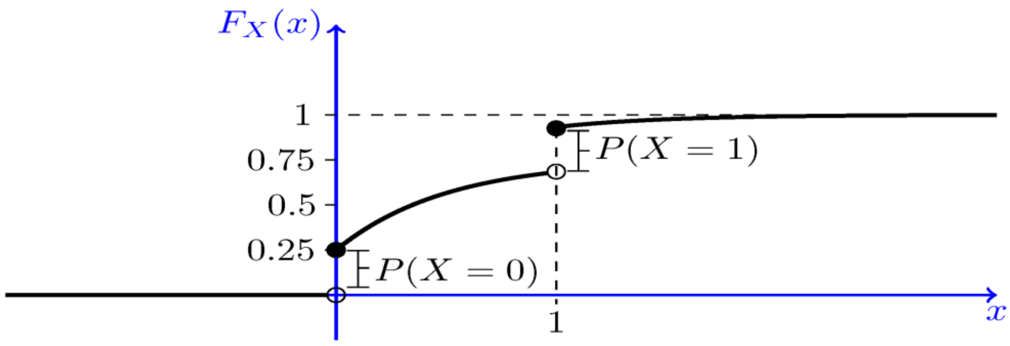

Figure 2: CDF with discontinuity (figure sampled from [3])

of X is given by:

$$q_X^p = \inf\left\{x \in \mathbb{R} : F_X(x) \geq p\right\}$$

Remark : Note that by proposition 3.2, we have: $P(X = q_X^p) = F_X(q_X^p) - \lim_{y \to (q_X^p)^-} F_X(y)$. Hence, when there is no discontinuity in q_X^p ($P(X = q_X^p) = 0$), we have :

$$\boxed{P(X = q_X^p) = 0 \implies p = \lim_{y \to (q_X^p)^-} F_X(y) = F_X(q_X^p)}$$

3.2.2 Empirial Cumulative Distribution Function

Definition 3.6 (Empirical Cumulative Distribution Function (E-CDF))**.**
*The **Empirical Cumulative Distribution Function (E-CDF)** of an N-point dataset $\{x_1, \ldots, x_N\}$ is estimated as:*

$$F_N(x) = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}_{\{x_n \leq x\}}, \quad x \geq 0$$

3.3 Value at Risk (VaR)

Definition 3.7 (Value at Risk (VaR))**.**
*The **Value at Risk (VaR)** of a random variable X at level $p \in (0,1)$ is the p-quantile of X:*

$$\boxed{V_X^p = q_X^p = \inf\left\{x \in \mathbb{R} : F_X(x) \geq p\right\}}$$

Proposition 3.3: Properties of Value at Risk

The Value at Risk (VaR) has the following properties:

- (i) The function $p \to V_X^p$ is **non-decreasing, left-continuous** and it **admits limit from the right**.
- (ii) $V_X^p \leq x \iff p \leq P(X \leq x)$.
- (iii) $V_{-X}^p = -V_X^{1-p}$.

Proof (Proposition 3.3).

Proving each property, we have:

Claim : $p \to V_X^p$ ***is non-decreasing, left-continuous and admits limit from the right***
The function $p \to V_X^p$ *(called the quantile function) is the* ***generalized inverse*** *of the Cumulative Distribution Function.*

By proposition 2.3 - [2], since $F_X(x)$ *is non-decreasing, the generalized inverse of it is non-decreasing, left-continuous and admits limits on the right.*

Claim : $V_X^p \leq x \iff p \leq P(X \leq x)$
Trivial due to the definition of Value at Risk (VaR).

Claim : $V_{-X}^p = -V_X^{1-p}$
We have:

$$\begin{aligned} F_{-X}(x) &= P(-X \leq x) = P(X \geq -x) \\ &= 1 - P(X < -x) = 1 - P(X \leq -x) \\ &= 1 - F_X(-x) \end{aligned}$$

Hence, we have:

$$p = F_{-X}(F_{-X}^{-1}(p)) = 1 - F_X(-F_{-X}^{-1}(p))$$

Which yields:

$$V_{-X}^p = F_{-X}^{-1}(p) = -F_{-X}^{-1}(1-p) = -V_X^{1-p} \tag{1}$$

$\square$.

Theorem 3.1: Coherence of Value at Risk (VaR)

Value at Risk (VaR) satisfies:

- (i) Monotonicity.
- (ii) Positive homogeneity and translation invariance.
- (iii) **But NOT** sub-additivity.

Proof (Theorem 3.1).

$\square$.

Definition 3.8 (Gaussian Value at Risk (G-VaR)).
Given $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, we have:

$$\boxed{V_X^p = \mu_X + \sigma_X q_Z^p}$$

Where the normal quantile $q_Z^p = V_Z^p$ at level p satisfies:

$$\Phi(q_Z^p) = P(Z \leq q_Z^p) = p, \quad Z \sim \mathcal{N}(0,1)$$

Meaning, we have:

$$V_X^p = \mu_X + \sigma_X \Phi^{-1}(p)$$

Lemma 3.1: $X = V_X^U, P(U \geq p) \neq P(V_X^U \geq V_X^p)$

We can write any random variable X as:

$$X = V_X^U, \; U \sim Uniform(0,1)$$

However, for any $p \in (0,1)$, we do not always have $P(U \geq p) = P(V_X^U \geq V_X^p)$. Instead, we have the following relationship:

$$\begin{aligned} P(V_X^U \geq V_X^p) &= P((V_X^U \geq V_X^p) \cap (U \geq p)) + P((V_X^U \geq V_X^p) \cap (U < p)) \\ \text{Or}: \; P(X \geq V_X^p) &= P((X \geq V_X^p) \cap (U \geq p)) + P((X \geq V_X^p) \cap (U < p)) \end{aligned}$$

This implies that the event $(V_X^U \geq V_X^p) \cap (U < p)$ can indeed have non-zero probability measure and it happens when there are discontinuities.

Proof (Lemma 3.1).
By proposition 3.3, we have $P(V_X^p \leq x) \iff p \leq P(X \leq x)$. Hence:

$$P(V_X^U \leq x) = P(U \leq P(X \leq x)) = P(X \leq x) = F_X(x)$$

To prove the second point, we have the following visual representation in figure 3. □.

Proposition 3.4: Discontinuity at V_X^p

Let V_X^p be the Value at Risk of a random variable X at level $p \in (0,1)$. Then, we have:

$$P(X = V_X^p) = 0 \iff p = F_X(V_X^p) = \lim_{y \to x^-} F_X(y)$$

In other words, if there is no discontinuity at V_X^p, then $p = F_X(V_X^p)$.

Proof (Proposition 3.4).
By proposition 3.2, we have that when $P(X = V_X^p) = 0$, we have $F_X(V_X^p) = P(X < V_X^p) = \lim_{y \to x^- F_X(y)}$. Now, we have to prove that $p = F_X(V_X^p)$.

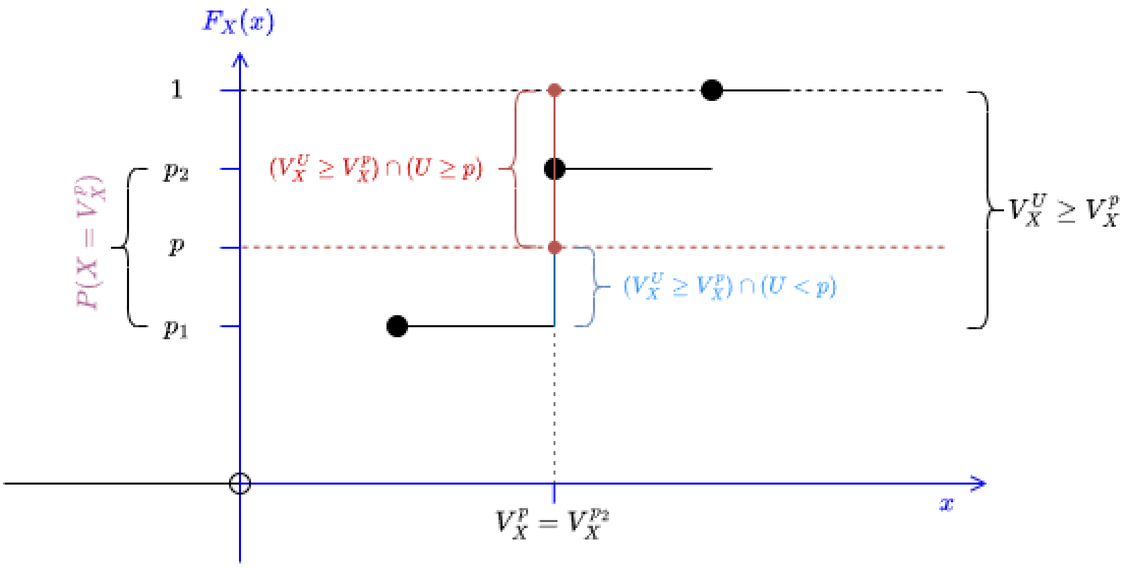

Figure 3: $P(V_X^U \geq V_X^p) = P((V_X^U \geq V_X^p) \cap (U \geq p)) + P((V_X^U \geq V_X^p) \cap (U < p))$

We have:

$$
\begin{aligned}
F_X(V_X^p) &= P(X \leq V_X^p) = 1 - P(X > V_X^p) \\
&= 1 - P(X \geq V_X^p) = 1 - P(V_X^U \geq V_X^p) \\
&= 1 - P(U \geq p) \\
&= 1 - (1 - p) = p
\end{aligned}
$$

Remark *: Note that by the visual representation in figure 3, when $P(X = V_X^p) = 0$, we have:*

$$P(V_X^U \geq V_X^p) = P(V_X^U \geq V_X^p \cap (U \geq p)) = P(U \geq p)$$

$\square$.

4 Expected Shortfall

4.1 Tail Value at Risk (TVaR)

Overview : One common shortcoming of Value at Risk is the inability to capture the behavior of the distribution beyond V_X^p (as illustrated in figure 4). Hence, one way to remedy this shortcoming is to taking the average of Value at Risk beyond the level p.

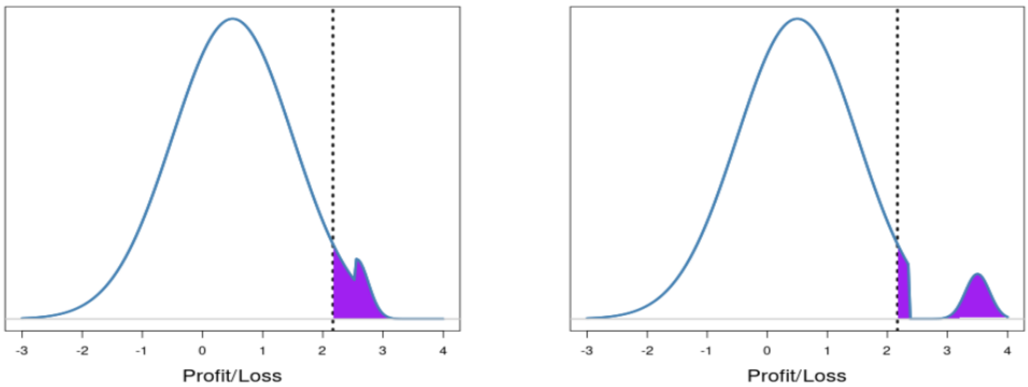

Figure 4: Two distributions having the same $V_X^{0.95} = 2.145$ (figure sampled from [3])

Definition 4.1 (Tail Value at Risk (TVaR)).
The Tail Value at Risk (TVaR) of a random variable X at the level $p \in (0,1)$ is defined by the following average:

$$\boxed{TV_X^p = \frac{1}{1-p}\int_p^1 V_X^q dq}$$

Note that $p \to V_X^p$ is a non-decreasing function (Proposition 3.3). Therefore, we always have:

$$TV_X^p = \frac{1}{1-p}\int_p^1 V_X^{\mathbf{q}} dq \geq \frac{1}{1-p}\int_p^1 V_X^{\mathbf{p}} dq = V_X^p$$

4.2 Conditional Tail Expectation (CTE)

Overview : Conditional Tail Expectation is another measure that takes into account "what happens beyond V_X^p. However, instead of taking the uniform average as Tail Value at Risk, Conditional Tail Expectation takes the conditional expectation of X conditioned on the event that $X > V_X^p$.

Definition 4.2 (Conditional Expectation).
Given a random variable X and an event A such that $P(A) > 0$. The conditional expectation of X given the event A is defined as:

$$\mathbb{E}[X|A] = \frac{1}{P(A)}\mathbb{E}[X\mathbb{1}_{\{A\}}]$$

Definition 4.3 (Conditional Tail Expectation (CTE)).
Given a random variable X such that $P(X > V_X^p) > 0$ at a level $p \in (0,1)$. The ***Conditional Tail Expectation*** *of X at level p is defined as:*

$$\boxed{CTE_X^p = \mathbb{E}\Big[X\Big|X > V_X^p\Big] = \frac{\mathbb{E}[X\mathbb{1}_{\{X>V_X^p\}}]}{P(X > V_X^p)}}$$

The Conditional Tail Expectation can be written as a Distortion Risk Measure $CTE_X^p = \mathbb{E}\Big[Xf_X(X)\Big]$ with the distortion function:

$$\boxed{f_X(x) = \frac{1}{P(X > V_X^p)}\mathbb{1}_{\{x>V_X^p\}}}$$

Proposition 4.1: $CTE_X^p - V_X^p$

For any $p \in (0,1]$, we have $CTE_X^p > \mathbb{E}[X]$ and $CTE_X^p > V_X^p$. Specifically:

$$CTE_X^p = \mathbb{E}\Big[X\Big|X > V_X^p\Big] = V_X^p + \mathbb{E}\Big[(X - V_X^p)^+|X > V_X^p\Big]$$

Proof (Proposition 4.1).
We have :

$$\begin{aligned}
\mathbb{E}\Big[X|X > V_X^p\Big] &= \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[X\mathbb{1}_{\{X>V_X^p\}}\Big] \\
&= \frac{1}{P(X > V_X^p)}\Bigg[\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X>V_X^p\}}\Big] + V_X^p\mathbb{E}\Big[\mathbb{1}_{\{X>V_X^p\}}\Big]\Bigg] \\
&= \frac{1}{P(X > V_X^p)}\Bigg[\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X>V_X^p\}}\Big] + V_X^p P(X > V_X^p)\Bigg) \\
&= V_X^p + \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X>V_X^p\}}\Big] \\
&= V_X^p + \mathbb{E}\Big[X - V_X^p|X > V_X^p\Big]
\end{aligned}$$

$\square$.

Proposition 4.2: When $CTE_X^p = TV_X^p$

When $P(X = V_X^p) = 0$, meaning there is no discontinuity at V_X^p, then we have:

$$P(X = V_X^p) = 0 \iff CTE_X^p = TV_X^p$$

Proof (Proposition 4.2).
By figure 3, we can see that when $P(X = V_X^p) = 0$, we have, with probability one, $P(V_X^U > V_X^p) =$

$P(V_X^U \geq V_X^p) = P(U \geq p)$. *Hence:*

$$\begin{aligned}
CTE_X^p &= \mathbb{E}\Big[X|X > V_X^p\Big] \\
&= \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[X\mathbb{1}_{\{X>V_X^p\}}\Big] \\
&= \frac{1}{P(U \geq p)}\mathbb{E}\Big[X\mathbb{1}_{\{U\geq p\}}\Big] \\
&= \frac{1}{P(U \geq p)}\mathbb{E}\Big[V_X^U\mathbb{1}_{\{U\geq p\}}\Big] \\
&= \frac{1}{1-p}\int_p^1 V_X^q dp = TV_X^p
\end{aligned}$$

$\square$.

Definition 4.4 (Gaussian CTE (G-CTE)).
Given $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, *we have:*

$$\boxed{CTE_X^p = \mu_X + \frac{\sigma_X}{1-p}\varphi(V_Z^p) = \mu_X + \frac{\sigma_X}{(1-p)\sqrt{2\pi}}\exp\left(-\frac{(V_Z^p)^2}{2}\right)}$$

Or we can write:

$$CTE_X^p = \mu_X + \frac{\sigma_X}{(1-p)\sqrt{2\pi}}\exp\left(-\frac{(q_Z^p)^2}{2}\right)$$

$\varphi(.)$ *is the standard normal probability density function.*

4.3 Expected Shortfall (ES)

Definition 4.5 (Expected Shortfall (ES)).
Given a random variable X, *the* ***Expected Shortfall (ES)*** *at the level* $p \in (0,1)$ *is defined by*

$$\boxed{ES_X^p = V_X^p + \frac{1}{1-p}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X\geq V_X^p\}}\Big]}$$

The Expected Shortfall (ES) can be written as a Distortion Risk Measure $ES_X^p = \mathbb{E}\Big[Xf_X(X)\Big]$ *defined as:*

$$\boxed{f_X(x) = \frac{1}{1-p}\left(\mathbb{1}_{\{X>V_X^p\}} + \mathbb{1}_{\{P(X=V_X^p)>0\}}\mathbb{1}_{\{X=V_X^p\}}\frac{1-p-P(X > V_X^p)}{P(X = V_X^p)}\right)}$$

This can be deduced from proposition 4.3 below.

Proposition 4.3: Altenative definition of ES_X^p

The Expected Shortfall (ES) at level $p \in (0,1)$ can be written as:

$$ES_X^p = \frac{1}{1-p}\mathbb{E}[X\mathbb{1}_{\{X \geq V_X^p\}}] + \frac{V_X^p}{1-p}\Big(1 - p - P(X \geq V_X^p)\Big)$$

Proof (Proposition 4.3).

We have:

$$\begin{aligned}
ES_X^p &= V_X^p + \frac{1}{1-p}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X \geq V_X^p\}}\Big] \\
&= V_X^p + \frac{1}{1-p}\mathbb{E}\Big[X - V_X^p \Big| X \geq V_X^p\Big]P(X \geq V_X^p) \\
&= V_X^p + \frac{P(X \geq V_X^p)}{1-p}\Bigg\{\mathbb{E}\Big[X|X \geq V_X^p\Big] - V_X^p\Bigg\} \\
&= V_X^p + \frac{P(X \geq V_X^p)}{1-p}\Bigg\{\frac{1}{P(X \geq V_X^p)}\mathbb{E}\Big[X\mathbb{1}_{\{X \geq V_X^p\}}\Big] - V_X^p\Bigg\} \\
&= \frac{1}{1-p}\mathbb{E}\Big[X\mathbb{1}_{\{X \geq V_X^p\}}\Big] + V_X^p\Big(1 - \frac{P(X \geq V_X^p)}{1-p}\Big) \\
&= \frac{1}{1-p}\mathbb{E}\Big[X\mathbb{1}_{\{X \geq V_X^p\}}\Big] + \frac{V_X^p}{1-p}\Big(1 - p - P(X \geq V_X^p)\Big)
\end{aligned}$$

$\square$.

Corollary 4.1: ES_X^p when $P(X = V_X^p) = 0$

As a direct consequence of Proposition 4.3, we have:

$$P(X = V_X^p) = 0 \iff ES_X^p = CTE_X^p$$

By proposition 4.2, we can even have:

$$P(X = V_X^p) = 0 \iff ES_X^p = CTE_X^p = TV_X^p$$

Proof (Corollary 4.1).

When $P(X = V_X^p) = 0$, we have $p = F_X(V_X^p) = P(X \leq V_X^p) = P(X < V_X^p)$. Hence, by the definition of expected shortfall, we have:

$$\begin{aligned}
ES_X^p &= V_X^p + \frac{1}{1-p}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X \geq V_X^p\}}\Big] \\
&= V_X^p + \frac{1}{1 - P(X \leq V_X^p)}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X \geq V_X^p\}}\Big] \\
&= V_X^p + \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[(X - V_X^p)\mathbb{1}_{\{X \geq V_X^p\}}\Big] \\
&= V_X^p + \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[X - V_X^p \Big| X \geq V_X^p\Big]P(X \geq V_X^p) \\
&= V_X^p + \frac{1}{P(X > V_X^p)}\mathbb{E}\Big[X - V_X^p \Big| X > V_X^p\Big]P(X > V_X^p) \\
&= V_X^p + \mathbb{E}\Big[X - V_X^p \Big| X > V_X^p\Big] \\
&= \mathbb{E}\Big[X \Big| X > V_X^p\Big] = CTE_X^p
\end{aligned}$$

$\square$.

Proposition 4.4: ES_X^p and TV_X^p for any $p \in (0,1)$

This proposition proves a much stronger result than Corollary 4.1. For any $p \in (0,1)$ (not just when $P(X = V_X^p) = 0$), we have:

$$ES_X^p = TV_X^p = \frac{1}{1-p}\int_p^1 V_X^q dq$$

Proof (Proposition 4.4).

We first prove the following claim:

Claim 1 : $V_X^p\Big(1 - p - P(X \geq V_X^p)\Big) = -\mathbb{E}\Big[X\mathbb{1}_{\{(X\geq V_X^p)\cap(U<p)\}}\Big]$

For $U \sim Uniform(0,1)$, note that:

$$P(U \geq p) = 1 - p = \mathbb{E}\Big[\mathbb{1}_{\{U\geq p\}}\Big]$$
$$P(X \geq V_X^p) = \mathbb{E}\Big[\mathbb{1}_{\{X\geq V_X^p\}}\Big]$$

(Note that it is tempting to conclude that $P(X \geq V_X^p) = P(V_X^U \geq V_X^p) = P(U \geq p)$. However, this is not true by lemma 3.1 because we can have $V_X^{p_1} \geq V_X^{p_2}$ for $p_1 < p_2$ if there is discontinuity between p_1 and p_2).

Now, we have:

$$\begin{aligned}
V_X^p\Big(1 - p - P(X \geq V_X^p)\Big) &= V_X^p\left(\mathbb{E}\Big[\mathbb{1}_{\{U\geq p\}}\Big] - \mathbb{E}\Big[\mathbb{1}_{\{X\geq V_X^p\}}\Big]\right)\\
&= -V_X^p\mathbb{E}\Big[\mathbb{1}_{\{X\geq V_X^p\}} - \mathbb{1}_{\{U\geq p\}}\Big]\\
&= -V_X^p\mathbb{E}\Big[\mathbb{1}_{\{(X\geq V_X^p)\setminus(U\geq p)\}}\Big]\\
&= -V_X^p\mathbb{E}\Big[\mathbb{1}_{\{(X\geq V_X^p)\cap(U<p)\}}\Big]
\end{aligned}$$

As illustrated in figure 3, when the event $(V_X^U \geq V_X^p) \cap (U < p)$ occurs, we have $X = V_X^p$. Hence, we can also write:

$$\begin{aligned}
V_X^p\Big(1 - p - P(X \geq V_X^p)\Big) &= -V_X^p\mathbb{E}\Big[\mathbb{1}_{\{(X\geq V_X^p)\cap(U<p)\}}\Big]\\
&= -\mathbb{E}\Big[X\mathbb{1}_{\{(X\geq V_X^p)\cap(U<p)\}}\Big]
\end{aligned}$$

Claim 2 : $ES_X^p = TV_X^p$

From the above, we have:

$$\begin{aligned}
ES_X^p &= \frac{1}{1-p}\mathbb{E}\Big[X\mathbb{1}_{\{X\geq V_X^p\}}\Big] + \frac{V_X^p}{1-p}\Big(1-p-P(X\geq V_X^p)\Big) \quad \textit{(By proposition 4.3)}\\
&= \frac{1}{1-p}\left(\mathbb{E}\Big[X\mathbb{1}_{\{X\geq V_X^p\}}\Big] - \mathbb{E}\Big[X\mathbb{1}_{\{(X\geq V_X^p)\cap(U<p)\}}\Big]\right)\\
&= \frac{1}{1-p}\mathbb{E}\Big[X\mathbb{1}_{\{(X\geq V_X^p)\cap(X\geq V_X^p)^c\cup(U\geq p)\}}\Big]\\
&= \frac{1}{1-p}\mathbb{E}\Big[X\mathbb{1}_{\{U\geq p\}}\Big]\\
&= \frac{1}{1-p}\mathbb{E}\Big[V_X^U\mathbb{1}_{\{U\geq p\}}\Big]\\
&= \frac{1}{1-p}\int_p^1 V_X^q dq = TV_X^p
\end{aligned}$$

$\square$.

Theorem 4.1: Coherence of ES_X^p and TV_X^p

Expected shortfall (ES) and **Tail Value at Risk (TVaR)** are coherent risk measures. The **Conditional Tail Expectation (CTE)** is generally not coherent (incoherent when $P(X = V_X^p) > 0$).

Proof (Theorem 4.1).

Since Expected Shortfall (ES) and Tail Value at Risk (TVaR) are the same for any $p \in (0,1)$, we can use either measure to prove coherence. In this proof, we will use Tail Value at Risk (TVaR).

(i) Monotonicity

Let X, Y be random variables such that $X \leq Y$. We have:

$$TV_X^p = \frac{1}{1-p}\int_p^1 V_X^q dq \leq \frac{1}{1-p}\int_p^1 V_Y^q dq = TV_Y^p \quad \textit{(Since VaR is monotone)}$$

(ii) Positive homogeneity and translation invariance

Let $\lambda > 0$ and $\mu \geq 0$, we have:

$$\begin{aligned}
TV_{\mu+\lambda X}^p &= \frac{1}{1-p}\int_p^1 V_{\mu+\lambda X}^q dq\\
&= \int_p^1 \Big(\mu + \lambda V_X^q\Big) dq\\
&= \mu + \frac{\lambda}{1-p}\int_p^1 V_X^q dq\\
&= \mu + \lambda TV_X^p
\end{aligned}$$

(iii) Sub-additivity

Since the Expected Shortfall (ES) can be written as a Distortion Risk Measure with the following distortion function:

$$f_X(x) = \frac{1}{1-p}\left(\mathbb{1}_{\{X>V_X^p\}} + \mathbb{1}_{\{P(X=V_X^p)>0\}}\mathbb{1}_{\{X=V_X^p\}}\frac{1-p-P(X>V_X^p)}{P(X=V_X^p)}\right)$$

We have:

$$\begin{aligned}(1-p)(ES^p_{X+Y}-ES^p_X-ES^p_Y) &= (1-p)\Big(\mathbb{E}\Big[(X+Y)f_{X+Y}(X+Y)\Big]-\mathbb{E}\Big[Xf_X(X)\Big]-\mathbb{E}\Big[Yf_Y(Y)\Big]\Big)\\ &= (1-p)\Big(\mathbb{E}\Big[X\Big(f_{X+Y}(X+Y)-f_X(X)\Big)\Big]+\mathbb{E}\Big[Y\Big(f_{X+Y}(X+Y)-f_Y(Y)\Big)\Big]\Big)\\ &\le V^p_X\mathbb{E}\Big[f_{X+Y}(X+Y)-f_X(X)\Big]+V^p_Y\mathbb{E}\Big[f_{X+Y}(X+Y)-f_Y(Y)\Big]\\ &= V^p_X(1-1)+V^p_Y(1-1)=0\\ \implies ES^p_{X+Y} &\le ES^p_X+ES^p_Y\end{aligned}$$

$\square$.

4.4 Conclusion (Asset risk measures)

Remark : As a concluding remark, we have the following

$$\text{When } P(X=V^p_X)\begin{cases}=0 & \implies TV^p_X=ES^p_X=CTE^p_X\\ \neq 0 & \implies TV^p_X=ES^p_X\neq CTE^p_X\end{cases}$$

Summary of asset risk measures :

Risk measure	Definition	Gaussian	Others
VaR (V^p_X)	$\inf\Big\{x\in\mathbb{R}:P(X\le x)\ge p\Big\}$	$\mu_X+\sigma_X q^p_Z$	N/A
TVaR (TV^p_X)	$\frac{1}{1-p}\int_p^1 V^q_X dq$	N/A	N/A
ES (ES^p_X)	$V^p_X+\frac{1}{1-p}\mathbb{E}\Big[(X-V^p_X)\mathbb{1}_{\{X\ge V^p_X\}}\Big]$	N/A	$\frac{1}{1-p}\Big(\mathbb{E}[X\mathbb{1}_{\{X\ge V^p_X\}}]+V^p_X(1-p-P(X\ge V^p_X))\Big)$
CTE (CTE^p_X)	$\mathbb{E}\Big[X\Big\vert X>V^p_X\Big]$	$\mu_X+\frac{\sigma_X}{1-p}\varphi(q^p_Z)$	$V^p_X+\mathbb{E}\Big[X-V^p_X\vert X>V^p_X\Big]$

Coherence of asset risk measures :

Risk measure	Monotonicity	Homogeneity	Sub-additivity	Coherence
VaR	Yes	Yes	No	No
TVaR	Yes	Yes	No	No
ES	Yes	Yes	Yes	Yes
CTE	Yes	Yes	Yes	Yes

5 Time-series for financial data

5.1 Autoregressive Moving Average

5.1.1 MA and AR models

Definition 5.1 (White Noise).
*A **White noise sequence** $(Z_n)_{n\in\mathbb{Z}}$ is a sequence of i.i.d centered, unit variance random variables with*

$$\mathbb{E}[Z_n] = 0,\ Cov(Z_n, Z_m) = \mathbb{1}_{\{n=m\}}$$

Definition 5.2 (Moving Average (MA) model).
*In the $MA(q)$ model of order $q \geq 1$, the current state of the system is expressed as the **linear (independent) combination**:*

$$\boxed{X_n = Z_n + \sum_{k=1}^{q} \beta_k Z_{n-k}}$$

of q previous states $Z_{n-1}, Z_{n-2}, \dots, Z_{n-q}$. $\beta_1, \dots, \beta_q$ is a sequence of deterministic coefficients. Define the lag operator L as followed:

$$LZ_n = Z_{n-1},\ L^k Z_n = Z_{n-k}$$

We can then rewrite the $MA(q)$ model as:

$$\boxed{X_n = Z_n + \sum_{k=1}^{q} \beta_k L^k Z_n = Z_n + \Psi(L) Z_n}$$

*Where $\Psi(L) = \sum_{k=1}^{q} \beta_k L^k$ is called the **moving average operator**.*

Definition 5.3 (Autoregressive (AR) model).
*In the $AR(p)$ model of order $p \geq 1$, the current state of the system is expressed as the **Linear (feedback) combination**:*

$$\boxed{X_n = Z_n + \sum_{k=1}^{p} \alpha_k X_{n-k}}$$

Again, we can rewrite the $AR(p)$ model as:

$$\boxed{X_n = Z_n + \sum_{k=1}^{p} \alpha_k L^k X_n = Z_n + \phi(L) X_n}$$

Where the operator $\phi(L) = \sum_{k=1}^{p} \alpha_k L^k$.

Proposition 5.1: Recursive solution of $AR(1)$

The $AR(1)$ process:

$$X_n = Z_n + \alpha_1 X_{n-1}$$

can be solved recursively in the following cases:

- When $|\alpha_1| < 1 \implies$ **Causal** moving average solution:

$$X_n = \sum_{k\geq 0} \alpha_1^k Z_{n-k},\ n \in \mathbb{Z}$$

- When $|\alpha_1| > 1 \implies$ **Non-causal** moving average solution:

$$X_n = -\sum_{k\geq 1} \alpha_1^{-k} Z_{n+k},\ n \in \mathbb{Z}$$

With the variance of X_n in both cases are:

$$Var(X_n) = \frac{1}{|1-\alpha_1^2|}$$

No such converging solutions exist when $|\alpha_1| = 1$.

Proof (Proposition 5.1).

Proving each case, we have

(*i*) $|\alpha_1| < 1$

Solving using backward induction, we have:

$$\begin{aligned} X_n &= Z_n + \alpha_1(Z_{n-1} + \alpha_1 X_{n-2}) \\ &= Z_n + \alpha_1(Z_{n-1} + \alpha_1(Z_{n-2}\alpha_1 X_{n-3})) \\ &= Z_n + \alpha_1 Z_{n-1} + \alpha_1^2 Z_{n-2} + \alpha_1^3 X_{n-3} \\ &\vdots \\ &= \sum_{k\geq 0} \alpha_1^k Z_{n-k} \end{aligned}$$

Which converges when the equation $\phi(z) = \alpha_1 z = 1$ satisfies $|\alpha_1| < 1 \implies |z| > 1$ (Causal). We also have:

$$Var(X_n) = \sum_{k\geq 0} \alpha_1^{2k} Var(Z_{n-k}) = \sum_{k\geq 0} \alpha_1^{2k} = \frac{1}{1-|\alpha_1|^2}$$

(*ii*) $|\alpha_1| > 1$

We can write:

$$X_{n+1} = Z_{n+1} + \alpha_1 X_n \implies X_n = -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1} X_{n+1}$$

Solving using forward induction, we have:

$$\begin{aligned}
X_n &= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1} X_{n+1} \\
&= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+2} + \alpha_1^{-1} X_{n+2}) \\
&= -\alpha_1^{-1} Z_{n+1} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+2} + \alpha_1^{-1}(-\alpha_1^{-1} Z_{n+3} + \alpha_1^{-1} X_{n+3})) \\
&= -\alpha_1^{-1} Z_{n+1} - \alpha_1^{-2} Z_{n+2} - \alpha_1^{-3} Z_{n+3} - \alpha_1^{-3} X_{n+3} \\
&\vdots \\
&= -\sum_{k \geq 1} \alpha_1^{-k} Z_{n+k}
\end{aligned}$$

Which converges when the equation $\phi(z) = \alpha_1 z = 1$ satisfies $|\alpha_1| > 1 \implies |z| < 1$ (Non-Causal). We also have:

$$Var(X_n) = \sum_{k \geq 0} \alpha_1^{-2k} Var(Z_{n-k}) = \sum_{k \geq 0} \alpha_1^{-2k} = \frac{1}{|\alpha_1|^2 - 1}$$

□.

5.1.2 ARMA model

Definition 5.4 (Autoregressive Moving Average (ARMA) model).
In the $ARMA(p, q)$ model with orders $p \geq 1, q \geq 1$, the current state X_n is expressed as the following linear combination:

$$\boxed{X_n = Z_n + \sum_{k=1}^{p} \alpha_k X_{n-k} + \sum_{k=1}^{q} \beta_k Z_{n-k}}$$

Making use of the ***moving average operator*** *$\Psi(L)$ and the $\phi(L)$ operator, we have:*

$$\boxed{X_n = Z_n + \phi(L) X_n + \Psi(L) Z_n}$$

5.1.3 ARIMA model

Definition 5.5 (Difference operator).
Consider the difference operator ∇ defined as:

$$\nabla := I - L$$

Where I is the identity operator, so that:

$$\nabla X_n = X_n - X_{n-1}, \; n \geq 1$$

Proposition 5.2: Iteration of ∇ operator

The difference operator can be iterated as followed:

$$\nabla^d X_n = \sum_{k=0}^{d} \binom{d}{k} (-1)^k X_{n-k}, \quad d \geq k \geq 0$$

Proof (Proposition 5.2).
Using the Binomial expansion formula, we have:

$$\begin{aligned}
\nabla^d &= (I - L)^d \\
&= \sum_{k=0}^{d} I^{d-k}(-L)^k \\
&= \sum_{k=0}^{d} (-L)^k = \sum_{k=0}^{d} (-1)^k L^k \\
\implies \nabla^d X_n &= \sum_{k=0}^{d} (-1)^k L^k X_n = \sum_{k=0}^{d} (-1)^k X_{n-k}
\end{aligned}$$

$\square$.

Proposition 5.3: Recovery of X_n using ∇ operator

For any time-series $(X_n)_{n\geq 1}$, we can recover X_n using the difference operator via:

$$X_n = \sum_{k=0}^{d} \binom{d}{k} \nabla^k X_{n-d+k}$$

Proof (Proposition 5.3).
Applying the Binomial expansion, we have:

$$\begin{aligned}
I &= (L + I - L)^d = (L + \nabla)^d \\
&= \sum_{k=0}^{d} \binom{d}{k} L^{d-k} \nabla^k \\
\implies X_n &= \sum_{k=0}^{d} \binom{d}{k} L^{d-k} \nabla^k X_n = \sum_{k=0}^{d} \binom{d}{k} \nabla^k X_{n-d+k}
\end{aligned}$$

$\square$.

Definition 5.6 (Autoregressive Integrated Moving Average (ARIMA) model).
In the $ARIMA(p, d, q)$ model, the iterated difference process $(\nabla^d X_n)_{n\geq 0}$ is modelled as the $ARMA(p, q)$ time-series:

$$\boxed{\nabla^d X_n = Z_n + \sum_{k=0}^{p} \alpha_k \nabla^d X_{n-k} + \sum_{k=0}^{q} \beta_k Z_{n-k}}$$

Making use of the ***moving average operator*** *$\Psi(L)$ and the operator $\phi(L)$, we have:*

$$\boxed{\nabla^d X_n = Z_n + \phi(L)\nabla^d X_n + \Psi(L) Z_n}$$

5.2 Time-series stationarity

5.2.1 Strict & Weak stationarity

Definition 5.7 (Strict stationarity).

*A time-series $(X_n)_{n\in\mathbb{Z}}$ is **strictly stationary** if the equality:*

$$(X_n, X_{n-1}, \dots, X_{n-p}) \simeq (X_{n+m}, X_{n+m-1}, \dots, X_{n+m-p})$$

*holds **in distribution** for all $n \in \mathbb{Z}$ and $m, p \geq 0$.*

Definition 5.8 (Weak stationarity).

*A time-series $(X_n)_{n\in\mathbb{Z}}$ is **weakly stationary** if the following holds:*

- *(i)* $\mathbb{E}[X_n] = \mathbb{E}[X_0], \ n \geq 0.$
- *(ii) The auto-covariance:*

$$(n, m) \to Cov(X_n, X_m)$$

depends only on the absolute difference $|n-m|, \ n, m \geq 0$. The covariance is defined as $Cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$.

Theorem 5.1: Unit root test

Consider the $AR(p)$ time-series $(X_n)_{n\geq 0}$ solution of:

$$X_n = Z_n + \phi(L)X_n = Z_n + \alpha_1 X_{n-1} + \cdots + \alpha_p X_{n-p}$$

with the characteristic polynomial:

$$\phi(z) = \alpha_1 z + \cdots + \alpha_q z^q, \ z \in \mathbb{C}$$

Then the time-series $(X_n)_{n\geq 0}$ is called:

- **Weakly stationary** : if no solution of $\phi(z) = 1$ **lies on** the complex unit circle.
- **Causality** : if no solution of $\phi(z) = 1$ **lies inside** the complex unit circle.

The complex unit circle is defined on the complex plane as:

$$\{z \in \mathbb{C} : |z| \leq 1\}$$

5.2.2 Stationarity test

Definition 5.9 (Dickey-Fuller test).

*The Dickey-Fuller test allows us to test the null hypothesis that "The time-series is **non-stationary**":*

$$\begin{cases} H_0 : |\alpha_1| = 1 & \textit{(Non-stationarity)} \\ \\ H_1 : |\alpha_1| \neq 1 & \textit{(Stationarity)} \end{cases}$$

6 Credit Scoring

6.1 Discriminant Analysis

Overview : The problem of credit scoring arises when we have a pool of credit applicants (Ω) and among the applicants, there are good (G) and bad (B) applicants such that:

$$\Omega = G \cup B; \; G \cap B = \emptyset$$
$$P(G) + P(B) = 1$$

Definition 6.1 (TPR and FPR).
*Given a random variable X representing the **credit score**. We have:*

- ***True Positive Rate (TPR)*** *is the tail distribution function:*

$$\bar{F}_G(x) = P(X > x|G) = \int_x^\infty f_X(y|G)dy$$

- ***False Positive Rate (FPR)*** *is the tail distribution function:*

$$\bar{F}_B(x) = P(X > x|B) = \int_x^\infty f_X(y|B)dy$$

Intuitively, TPR represents the rate of accepting good applicants whereas FPR represents the rate of accepting bad applicant.

Definition 6.2 (Acceptance & Default curve).
Given a random variable X representing the credit score. We have:

- ***Probability default curve*** *is given by:*

$$P(B|X = x) = \frac{f_X(x|B)P(B)}{f_X(x)} = \frac{f_X(x|B)P(B)}{f_X(x|B)P(B) + f_X(x|G)P(G)}$$

- ***Probability acceptance curve*** *is given by:*

$$P(G|X = x) = \frac{f_X(x|G)P(G)}{f_X(x)} = \frac{f_X(x|G)P(G)}{f_X(x|G)P(G) + f_X(x|B)P(B)}$$

*The **default curve** is the probability that an applicant is bad given his credit score whereas the **acceptance curve** is the probability that an applicant is good given his credit score.*

6.2 Decision Rule

Theorem 6.1: Optimal acceptance set

Given X as the random variable representing credit scores of applicants and

- $L_G(X)$ is the loss for rejecting potential good applicants.
- $L_B(X)$ is the loss of accepting bad applications.

We need a decision rule, represented by an **acceptance set** $\mathcal{A}$, such that the overall loss, given by:

$$\mathbb{E}\Big[L_G(X)\mathbb{1}_{\{(X\in\mathcal{A}^c)\cap G\}} + L_B(X)\mathbb{1}_{\{(X\in\mathcal{A})\cap B\}}\Big]$$

is minimized. The **optimal acceptance set** is then given by:

$$\boxed{\mathcal{A}^* = \left\{x\in\mathbb{R} : \lambda(x) \geq \frac{P(B)}{P(G)}\cdot\frac{L_B(x)}{L_G(x)}\right\}}$$

Where $\lambda(x)$ is the **likelihood ratio** :

$$\lambda(x) = \frac{f_X(x|G)}{f_X(x|B)} = \frac{P(G|X=x)}{P(B|X=x)}\cdot\frac{P(B)}{P(G)}$$

Proof (Theorem 6.1).

Let $h:\mathbb{R}\to\{G,B\}$ be a classifier associated to the acceptance set $\mathcal{A}$ such that:

$$h(X) = \begin{cases} G, & \text{if } X\in\mathcal{A} \\ \\ B, & \text{if } X\in\mathcal{A}^c \end{cases}$$

Hence, we can rewrite the loss function as followed:

$$\begin{aligned}&\mathbb{E}_{x\sim X}\left[L_G(x)\mathbb{1}_{\{h(x)=B,G\}} + L_B(x)\mathbb{1}_{\{h(x)=G,B\}}\right]\\ &\quad= \mathbb{E}_{x\sim X}\left[L_G(x)\mathbb{1}_{\{h(x)=B\}}P(G|h(x)=B) + L_B(x)\mathbb{1}_{\{h(x)=G\}}P(B|h(x)=G)\right]\end{aligned}$$

Now, we have:

$$\mathbb{1}_{\{h(x)=B\}}P(G|h(x)=B) = \mathbb{1}_{\{h(x)=B\}}P(G|X=x)$$

Because the case where $h(x)=G$ is already erased by the indicator function. Similarly, we have:

$$\mathbb{1}_{\{h(x)=G\}}P(B|h(x)=G) = \mathbb{1}_{\{h(x)=G\}}P(B|X=x)$$

Define the following function:

$$\eta(x) = P(G|X=x) \implies 1-\eta(x) = P(B|X=x)$$

We have:

$$\begin{aligned}&\mathbb{E}_{x\sim X}\left[L_G(x)\mathbb{1}_{\{h(x)=B,G\}} + L_B(x)\mathbb{1}_{\{h(x)=G,B\}}\right]\\ &\quad= \mathbb{E}_{x\sim X}\left[\eta(x)L_G(x)\mathbb{1}_{\{h(x)=B\}} + (1-\eta(x))L_B(x)\mathbb{1}_{\{h(x)=G\}}\right]\end{aligned}$$

To minimize the above cost, we have to minimize the integrand inside the expectation. Notice that $\mathbb{1}_{\{h(x)=B\}}$ *and* $\mathbb{1}_{\{h(x)=G\}}$ *are mutually exclusive. Hence, we have the optimal classifier* $h^* : \mathbb{R} \to \{G, B\}$ *such that:*

$$h^*(x) = \begin{cases} G, & \textit{if } (1-\eta(x))L_B(x) \leq \eta(x)L_G(x) \\ \\ B, & \textit{Otherwise} \end{cases}$$

Therefore, we have the following optimal acceptance set:

$$\begin{aligned} \mathcal{A}^* &= \Big\{ x \in \mathbb{R} : (1-\eta(x)L_B(x) \leq \eta(x)L_G(x) \Big\} \\ &= \left\{ x \in \mathbb{R} : \frac{\eta(x)}{1-\eta(x)} \geq \frac{L_B(x)}{L_G(x)} \right\} \\ &= \left\{ x \in \mathbb{R} : \frac{P(G|X=x)}{P(B|X=x)} \geq \frac{L_B(x)}{L_G(x)} \right\} \\ &= \left\{ x \in \mathbb{R} : \lambda(x) \geq \frac{P(B)}{P(G)} \cdot \frac{L_B(x)}{L_G(x)} \right\} \end{aligned}$$

$\square$.

Proposition 6.1: Optimal acceptance set for Gaussian scores

Let $X|G \sim \mathcal{N}(\mu_G, \sigma^2)$ and $X|B \sim \mathcal{N}(\mu_B, \sigma^2)$. Intuitively, the scores of good and bad applicants are normally distributed with different means but same variance. The optimal acceptance set is given by:

$$\mathcal{A}^* = \left[\frac{\mu_G + \mu_B}{2} + \frac{1}{\beta} \log\left(\frac{L_B P(B)}{L_G P(G)} \right), \infty \right)$$

Where $\beta = \frac{\mu_G - \mu_B}{\sigma^2} > 0$. Note that this bound is under the condition that the mean score of good applicants is strictly greater than that of bad applicants:

$$\mathbb{E}\Big[X|G\Big] = \mu_G > \mu_B = \mathbb{E}\Big[X|B\Big]$$

6.3 ROC curve

Definition 6.3 (ROC curve)**.**
The ***Receiver Operating Characteristic (ROC)*** *curve is a function of the threshold* $p \in [0, 1]$*, defined as:*

$$\boxed{ROC(p) = \bar{F}_G\Big(\bar{F}_B^{-1}(p)\Big)}$$

Proposition 6.2: ROC curve integral formula

The ROC curve can be rewritten as an integ ral:

$$\boxed{ROC(p) = \bar{F}_G\left(\bar{F}_B^{-1}(p)\right) = \int_0^p \lambda\left(\bar{F}_B^{-1}(q)\right)dq}$$

Where the likelihood ratio $\lambda(x)$ is given by:

$$\lambda(x) = \frac{f_X(x|G)}{f_X(x|B)}$$

Proof (Proposition 6.2)**.**

We have:

$$\begin{aligned}
\frac{d}{dq}\bar{F}_G\left(\bar{F}_B^{-1}(p)\right) &= \frac{\bar{F}'_G\left(\bar{F}_B^{-1}(p)\right)}{\bar{F}'_B\left(\bar{F}_B^{-1}(p)\right)} \\
&= \frac{f_X(F_B^{-1}(p)|G)}{f_X(F_B^{-1}(p)|B)} \\
&= \lambda\left(F_B^{-1}(p)\right) \\
\implies ROC(p) &= \int_0^p \lambda\left(F_B^{-1}(q)\right)dq
\end{aligned}$$

$\square$.

7 Insurance Risk

7.1 Poisson Process

Definition 7.1 (Poisson process).

A ***Poisson process*** $(N_t)_{t\geq 0}$ *is a counting process with jump size equals to* $+1$ *only and the path is constant between two jumps. The value of count* N_t *is defined as:*

$$N_t = \sum_{k\geq 1} \mathbb{1}_{\{t\geq T_k\}}$$

Where $(T_k)_{k\geq 1}$ *is an* ***increasing jump-time family*** *such that:*

$$\lim_{k\to\infty} T_k = +\infty$$

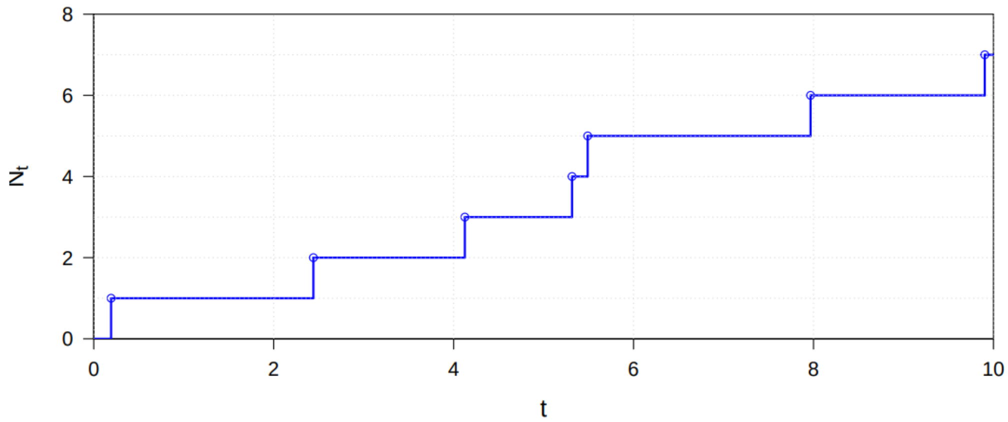

Figure 5: Homogeneous Poisson Process with constant jump size (figure sampled from [3])

Proposition 7.1: Properties of Poisson Process

In order for the counting process $(N_t)_{t\geq 0}$ to be a Poisson process, it has to satisfy the following properties:

- **Independence of increments** : For all $0 \leq t_0 < t_1 < \cdots < t_n$ and $n \geq 1$ increments, we have

$$N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \ldots, N_{t_n} - N_{t_{n-1}}$$

  are mutually independent random variables.

- **Stationarity of increments** : For all $0 \leq s \leq t$ and some $h > 0$, the increment $N_t - N_s$ has the same distribution as $N_{t+h} - N_{s+h}$. In other words, for some $k \geq 0$, we have:

$$P(N_{t+h} - N_{s+h} = k) = P(N_t - N_s = k)$$

Theorem 7.1: Poisson increment

Let $(N_t)_{t\geq 0}$ be a Poisson process. For any $0 \leq s \leq t$, the increment $N_t - N_s$ follows a Poisson distribution with parameter $(t-s)\lambda$ for some $\lambda > 0$.

$$N_t - N_s \sim Poisson\Big((t-s)\lambda\Big)$$

The constant λ is called the **intensity of Poisson process** $(N_t)_{t\geq 0}$ and is given by:

$$\lambda = \lim_{h\to 0} \frac{1}{h} P(N_h = 1)$$

Intuitively, the parameter λ reflects how soon it is the get the first count. The higher the intensity, the closer the gap in between jumps.

Proof (Theorem 7.1).

(The proof for theorem 7.1 is too technical and will not be included. Refer to [1] for more information). □.

Corollary 7.1: Distribution of N_t

As a direct consequence of theorem 7.1, we have:

$$N_t \sim Poisson\Big(\lambda t\Big) \implies P(N_t = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

Corollary 7.2: Short-time asymptotics of Poisson process

For a current state N_t of the Poisson process, the behaviour of the process in a short time $h > 0$ ahead is expressed in the following probability:

$$P(N_{t+h} - N_t) \approx \frac{\lambda^k h^k}{k!},\ h \to 0$$

Proof (Corollary 7.2).

By theorem 7.1, we have:

$$N_{t+h} - N_t \sim Poisson(h\lambda)$$

Hence, we have:

$$\begin{aligned} P(N_{t+h} - N_t) &= \frac{(\lambda h)^k e^{-\lambda h}}{k!} \\ &\approx \frac{\lambda^k h^k}{k!} \quad \textit{(When } h \approx 0\textit{)} \end{aligned}$$

Because $e^{-\lambda h} \to 1$ *as* $h \to 0$. □.

Proposition 7.2: Distribution of jump-time T_n

For all $n \geq 1$, the jump-time T_n has the gamma distribution:

$$T_n \sim Gamma(n, 1/\lambda)$$

With the probability density function:

$$f_{T_n}(t) = \lambda^n e^{-\lambda t} \frac{t^{n-1}}{(n-1)!} = \lambda^n e^{-\lambda t} \frac{t^{n-1}}{\Gamma(n)}$$

For all $t > 0$, the probability that $T_n \geq t$ is given by:

$$P(T_n \geq t) = \lambda^n \int_t^\infty e^{-\lambda s} \frac{s^{n-1}}{(n-1)!} ds$$

Proof (Proposition 7.2).

Proving by induction, for base case, we have:

$$P(T_1 \geq t) = P(T_1 > t) = P(N_t = 0) = e^{-\lambda t}$$

For inductive case, suppose that we have:

$$P(T_{n-1} > t) = \lambda^{n-1} \int_t^\infty e^{-\lambda s} \frac{s^{n-2}}{(n-2)!} ds$$

For T_n, we have:

$$\begin{aligned}
P(T_n > t) &= P(T_n > t \geq T_{n-1}) + P(T_{n-1} > t) \\
&= P(N_t = n-1) + P(T_{n-1} > t) \\
&= \frac{(\lambda t)^{n-1} e^{-\lambda t}}{(n-1)!} + \lambda^{n-1} \int_t^\infty e^{-\lambda s} \frac{s^{n-2}}{(n-2)!} ds \\
&= \lambda \int_t e^{-\lambda s} \frac{(\lambda s)^{n-1}}{(n-1)!} ds \\
&= \lambda^n \int_t^\infty \frac{e^{-\lambda s} s^{n-1}}{(n-1)!} ds
\end{aligned}$$

$\square$.

7.2 Compound Poisson Process

Definition 7.2 (Compound Poisson Process).

Let $(Z_k)_{k\geq 1}$ be a sequence of i.i.d square integrable random variables with a probability distribution $\nu(.)$ independent of the Poisson Process $(N_t)_{t\geq 0}$.

The process $(Y_t)_{t\geq 0}$ is called a ***Compound Poisson Process*** *if it is given by the random sum:*

$$Y_t = \sum_{k=1}^{N_t} Z_k$$

The Compound Poisson Process is indeed a Poisson process because it satisfies both properties:

- ***Independent increments****.*
- ***Stationarity of increments****.*

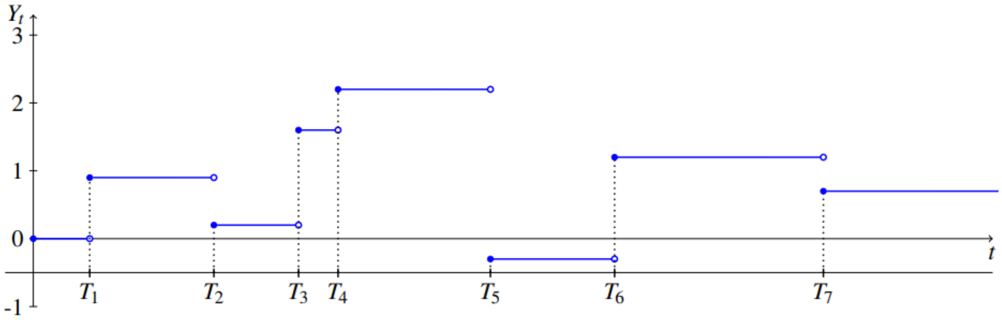

Figure 6: Compound Poisson Process with non-constant jump size (figure sampled from [3])

Proposition 7.3: Mean and Variance of $(Y_t)_{t\geq 0}$

Let $\mathbb{E}\Big[N_t\Big]$ be the mean number of jump times and $\mathbb{E}\Big[Z\Big]$ be the mean jump size. We have:

- $\mathbb{E}[Y_t] = \mathbb{E}[N_t]\mathbb{E}[Z] = \lambda t\mathbb{E}[Z]$.
- $Var(Y_t) = \mathbb{E}[N_t]\mathbb{E}[Z^2] = \lambda t\mathbb{E}[Z^2]$.

7.3 Claim and Reserve Process

Overview : In an insurance risk settings, we have

- **Number of claims** (N_t) : modelled by the homogeneous (constant jump size) Poisson process $(N_t)_{t\geq 0}$ with intensity $\lambda > 0$.
- **Claim amounts** (Z_k) : A sequence of non-negative, i.i.d random variables.

Definition 7.3 (Aggregated claims).
*The **Aggregated claim** amount up to time t is defined by the Compound Poisson Process:*

$$S(t) = \sum_{k=1}^{N_t} Z_k$$

Definition 7.4 (Standard compound risk model).
The reserve process is defined by

$$R_x(t) = x + f(t) - S(t)$$

Where $x \geq 0$ is the initial reserve, $f(t)$ is an income function up to time $t > 0$ and $S(t)$ is the aggregated claims.

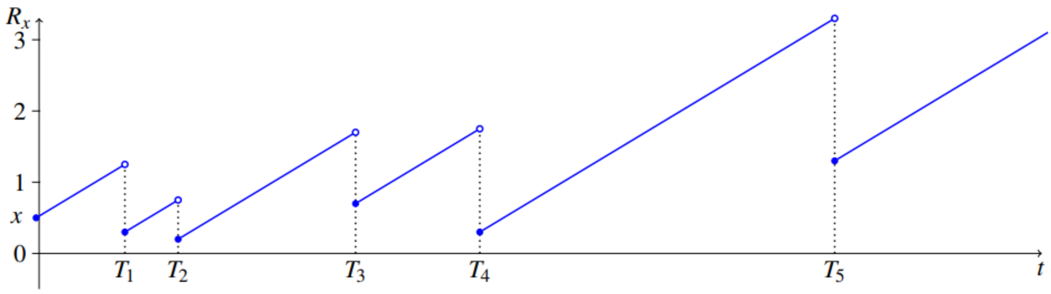

Figure 7: Sample reserve process with non-zero initial reserve and a linear income function (figure sampled from [3])

7.4 Ruin Probability

Definition 7.5 (Ruin probability).
There are two types of ruin probabilities

- ***Infinite-time ruin probability*** *:*

$$\Psi(x) = P(\exists t \in [0, \infty) : R_x(t) < 0)$$

- ***Finite-time ruin probability*** *:*

$$\Psi_T(x) = P(\exists t \in [0, T] : R_x(t) < 0)$$

Theorem 7.2: Cramer Lundberg model

Assume that the income function is linear with a positive slope $f(t) = ct$. We have:

- **Zero initial reserve** :

$$\Psi(x) = \frac{\lambda\mu}{c}$$

Provided that $c > \lambda\mu$ where $\mu = \mathbb{E}[Z]$.

- **Non-zero initial reserve** : Suppose that $Z_k \sim \mathit{Exponential}(1/\mu)$. then,

$$\Psi(x) = \frac{\lambda\mu}{c} \exp\left(\left(\frac{\lambda}{c} - \frac{1}{\mu}\right)x\right)$$

Either case, $\Psi(x) = 1$ if $c \leq \lambda\mu$.

A List of Definitions

B Important Theorems

C Important Corollaries

D Important Propositions

E References

References

[1] Denis Bosq and Hung T. Nguyen. *A Course in Stochastic Processes*. Springer Netherlands, 1996. ISBN: 9789401587693. DOI: 10.1007/978-94-015-8769-3. URL: http://dx.doi.org/10.1007/978-94-015-8769-3.

[2] Paul Embrechts and Marius Hofert. "A note on generalized inverses". In: *Mathematical Methods of Operations Research* 77.3 (Apr. 2013), pp. 423–432. ISSN: 1432-5217. DOI: 10.1007/s00186-013-0436-7. URL: http://dx.doi.org/10.1007/s00186-013-0436-7.

[3] Nicolas Privault. *MH6331 Financial risk & analytics (2023-2024) - Nanyang Technological University*. URL: https://personal.ntu.edu.sg/nprivault/MH8331/financial_risk_analytics.pdf.

[4] Keiichi Tanaka, Takeshi Yamada, and Toshiaki Watanabe. "Applications of Gram–Charlier expansion and bond moments for pricing of interest rates and credit risk". en. In: *Quant. Finance* 10.6 (June 2010), pp. 645–662.