1 Probability theory overview

1.1 Cumulative distribution function (cdf)

Every real-valued random variable X can be described by its cumulative distribution function $(cdf) F_X : \mathbb{R} \to [0,1]$ defined as

$$F_X(x) = \mathbb{P}[X \le x]$$

Notice that, as mentioned above, $\{X \le x\} = X^{-1}((-\infty, x]) \in \mathcal{F}$.

Discrete random variable

We call var. discrete, if its image (pmf) is **finite** or **countably finite**.

Continuos random variables

We call radom var continuos if its cdf is contunous. It implies the image of X (pdf) is "uncountably infite".

1.2 Quantile functions and quantiles

Quantile is **generalized inverse of the cdf F_x of X, defined as:

$$Q_X(p) = \inf\{x \in \mathbb{R} \mid p \le F_X(x)\}$$

For some probability level $\alpha \epsilon (0,1)$, the $\alpha - quantile$ is just the value $Q_x(\alpha)$

1.3 Expected Value

or expectation, mean, average, first moment- characterises the tendency or long-run average value

For absolutely continuous

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

For discrete

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} x_k \mathrm{P}[X = x_k] = \sum_{k=1}^{\infty} x_k p_k$$

Main take-home Example. Let $X \sim \mathcal{U}(a,b)$, (uniformly distributed) then the expected value of X is

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f_X(x) \mathrm{d}x = \int_a^b \frac{x}{b-a} \,\mathrm{d}x = \left.\frac{x^2}{2(b-a)}\right|_{x=a}^{b}$$
$$= \frac{b^2 - a^2}{2(b-a)} = \frac{a+b}{2}$$

$$\mathbb{E}[X] = \frac{a+b}{2}$$

1.4 Variance and SD

- variance: distance to the mean, expecteation of the squared devation from its mean
- standard devation: sqare root of variance

$$\operatorname{Var}(X) = \mathbb{V}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

$$\operatorname{Var}(X) = \frac{(b-a)^2}{12}$$

1.5 Covariance

expected value of the product of their SD.

$$\operatorname{cov}(X,Y) = \sigma_{XY} = \sigma(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$
$$\operatorname{cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

1.6 Pearson's correlation

Measure of linear correlation betwewen two RV.

$$\rho_{X,Y} = \frac{\operatorname{cov}(X,Y)}{\sqrt{\mathbb{V}[X]}\sqrt{\mathbb{V}[Y]}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

! If two vars are independent, they are also uncorrelated, the opposite must not be true.

$$\mathrm{V}[X+Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2\operatorname{cov}(X,Y)$$

Note that variance of a sum is sum of variance only if vars are uncorrelated.

1.7 Skewness

Measure of the assymetry of the distribution around its mean. Third standardized moment.

$$\gamma_1(X) = \mathbb{E}\left[\left(\frac{X - \mathbb{E}[X]}{\sqrt{\mathbb{V}[X]}}\right)^3\right] = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

1.8 Kurtosis

Measure of the width of its tails. Fourth standadized moment. Kurtosis = 3 is average, above (below) is excess kurtosis.

$$\operatorname{Kurt}[X] = \gamma_2(X) = \kappa(X) = \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

1.9 Independence

We say two elements A and B of a σ -algebra are independent and we write $A \perp B$ if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$$

To see why we call this property independence, we recall the definition of the conditional probability and use 'independence',

$$\mathbb{P}[A \mid B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]} = \frac{\mathbb{P}[A]\mathbb{P}[B]}{\mathbb{P}[B]} = \mathbb{P}[A]$$

So the probability of A stays the same independent of the occurrence of B. The same is true the other way around

$$\mathbb{P}[B \mid A] = \frac{\mathbb{P}[B \cap A]}{\mathbb{P}[A]} = \mathbb{P}[B]$$

1.10 Normal distribution

probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

density function , cdf

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

Combined

$$f_X(x) = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$

1.11 Log-normal distribution

We say X is log-normally distributed and write $X \sim \log\mathcal{N}\left(\mu, \sigma^2\right)$ if

$Y = \log(X) \sim \mathcal{N}\left(\mu, \sigma^2\right)$

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right]$$

1.12 t-distribution

Like normal but with heavier tails.

cdf

$$t_\nu(x) = \int_{-\infty}^{x} \frac{\Gamma\left(\frac{\nu+1}{2}\right)\nu^{\frac{\nu}{2}}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)} \left(\nu + t^2\right)^{-\frac{\nu+1}{2}} \,\mathrm{d}t$$

Where ν denotes the degrees of freedom.

2 Bonds

2.1 Present value

$$PV(y) = \frac{C_T}{(1+y)^T}$$

Notes:

- market value of zero-coupon bonds decreases with longer maturities
- keeping T fixed, the market value of the zero-coupon bond decreases as the yield increases

Semi-annual compounding (e.g. used in the U.S. Treasury bond market): interest rate y_s is derived from:

$$P(y_s) = \frac{C_T}{\left(1 + \frac{y_s}{2}\right)^{2T}}$$

Continuous compounding (used ubiquitously in the quantitative finance literature) interest rate y_c is derived from:

$$P(y_c) = \frac{C_T}{\exp(y_c T)} = \mathrm{e}^{-y_c T} C_T$$

$$yield = \frac{\frac{CT}{P(y)}}{T}$$

$$P(y) = \frac{cF}{y}\left(1 - \frac{1}{(1+y)^T}\right) + \frac{F}{(1+y)^T}$$

Price of a perpetual bond/consol: $P(y) = \frac{c}{y}F$

2.2 Taylor series

$$T_{f,x_0}(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$$
$$= f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + \ldots$$

$$\sum_{n=0}^{2} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2$$

a Taylor polynomial of degree two and so on.

2.3 Duation

▷ The Dollar-Duration (DD) is the negative of the first derivative

$$DD(y) = -P'(y)$$

▷ The Modified Duration (D^*) is a normalized version of the Dollar-Duration:

$$D^*(y) = \frac{DD(y)}{P(y)} = -\frac{P'(y)}{P(y)}$$

▷ The Macaulay Duration (D_M), yet another normalized version:

$$D_M(y) = (1+y)D^*(y) = -(1+y)\frac{P'(y)}{P(y)}$$

Duration is average cash-flow during maturity, or average time at which cash is received. Measures sensitivity to change in yields.

2.4 Convexity

Improved approximation of bond price sensitivity given by duration. Taylor series up to the second derivative. Hence: ▷ Again, we take the second derivative of each term in the sum separately. We get for all $t \in \{1, \ldots, T\}$,

$$\frac{\mathrm{d}^2}{\mathrm{d}y^2}\left(\frac{C_t}{(1+y)^t}\right) = \frac{\mathrm{d}}{\mathrm{d}y}\left(-\frac{tC_t}{(1+y)^{t+1}}\right) = \frac{t(t+1)C_t}{(1+y)^{t+2}}$$

⊵ Hence, the second derivative of the bond price is given by

$$P''(y) = \frac{\mathrm{d}^2 P}{\mathrm{d}y^2}(y) = \sum_{t=1}^{T} \frac{t(t+1)C_t}{(1+y)^{t+2}} = \frac{1}{(1+y)^2}\sum_{t=1}^{T}\frac{t(t+1)C_t}{(1+y)^t}$$
$$= \frac{1}{(1+y)^2}\left(\frac{2C_1}{1+y} + \frac{6C_2}{(1+y)^2} + \ldots + \frac{T(T+1)C_n}{(1+y)^T}\right).$$

⊵ Dollar convexity (DC) is defined as the second derivative of the bond price with respect to the yield, that means we set

$$DC(y) = \frac{\mathrm{d}^2 P}{\mathrm{d}y^2}(y) = P''(y)$$

$$ Bond convexity (κ) is defined to be the second derivative of the bond price with respect to the yield, divided by the bond price, that is

$$\kappa(y) = \frac{DC(y)}{P(y)} = \frac{P''(y)}{P(y)}$$

Note that some practitioners refer to convexity as the second derivative divided by the present value and additionally multiplied by the factor $\frac{1}{2}$

Conclusions:

- for small movements, duration (linear approx.) is reasonably good.
- dollar duration measures the negative slope of the tanget to the price-yield curve at the point y_0
- duration is the **average time to wait for each payment weighted by their present values**

3 Econometrics (volatility)

- vol alludes to the fact that asset price, thus the return, is **random**
- vol is a measure of uncertainty of the stock returns
- typical vol of stocks is between 20-50%
- matters for rebalancing positions

3.1 Stylized facts

- returns of stocks exhibit **heteroscedasticity and positive autocorrelation**
- volatility clusters - large (small) vol is followed by also large(small) vol, of either sign
- autocorrelations are not large and decrease more or less slowly with the lag order

3.2 Intro

Unbiased estimator of the variance

$$\widehat{v} = \frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2$$

The annualized volatility $\widehat{\sigma}$, under the assumption of Gaussianity, can be estimated as

$$\widehat{\sigma} = \frac{\sqrt{v}}{\sqrt{\tau}}$$

$SE = \frac{\widehat{\sigma}}{\sqrt{2n}}$.

3.3 ARCH(m)

- captures dependence in the form of positive autocorrelations in the square residuals
- estimate of the variance is based on LT average variance and m observations. The older it is, less weight it's given.

Creation

$$\sigma_n^2 = \sum_{i=1}^{m} \alpha u_{n-i}^2$$

▷ The coefficient $\alpha_i > 0$ is the weight given to the observation i days ago. **Alpha serves to give higher weights to more recent observations**
Extension is the assumption there is a long-run variance rate and that it should be given some weight.

$$\sigma_n^2 = \gamma V_L + \sum_{i=1}^{m} \alpha_i u_{n-i}^2$$

where V_L is the long-run variance rate and γ is the weight assigned to V_L. Often, $\omega = \gamma V_L$. Because the weights must sum to unity, we have

$$\gamma + \sum_{i=1}^{m} \alpha_i = 1.$$

EWMA Model

- designed to track changes in the volatility

Weights α_i **decrease exponentially** as we move back through time.

$$\alpha_{i+1} = \lambda \alpha_i, \text{ where } \lambda \in [0, 1]$$

It turns out that this weighting scheme leads to a particularly simple formula for updating volatility estimates:

$$\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2$$

$$\sigma_n^2 = (1 - \lambda) \sum_{i=1}^{m} \lambda^{i-1} u_{n-i}^2 + \lambda^m \sigma_{n-m}^2$$

Lambda is empirically at 0.94.

GARCH (1,1)

- vol. clustering often generates more extreme values and less central values compared to case when returns are independent, i.e. $\alpha = \beta = 0$
- incorporates mean-reversion, whereas EWMA does not

The conditional variance σ_t^2 is calculated from a long-run average variance rate V_L, the past variance σ_{t-1}^2, and the past return u_{t-1} :

$$\sigma_t^2 = \underbrace{\gamma \cdot V_L}_{\omega} + \beta \cdot \sigma_{t-1}^2 + \alpha \cdot u_{t-1}^2,$$

Where:

- conditional variance σ_{t-1}^2 is calculated from the LT average V_L, past return u_{t-1}
- $1 = \gamma + \beta + \alpha$

4 Copulas

Correlation measures only **linear dependency**- copulas solve this problem.

Definition Copulas are functions that separate the **marginal distributions** from the **dependency structure** of a multivar distribution
Copulas express dependence on a **quantile scale-** useful for dependence of **extreme outcomes**. They facilite **bottom-up approach-** that is better because we often understand marginal behavior of individual risk factors than the aggregate, whole dependence structure.

Definition 2 Coupula is a function with 2 vars and has 3 properties:

- groundedness
- margins
- increasing

4.1 Grounded functions

Definition. We say a function $H : X \times Y \to \mathbb{R}$ is grounded if

$$H(x, m_Y) = H(m_X, y) = 0 \text{ for all } (x, y) \in X \times Y$$

Example. Let $H : [-1, 1] \times [0, \infty] \to \mathbb{R}$ be defined by

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}$$

Then H is grounded, since $H(x, 0) = H(-1, y) = 0$ for all $x \in [-1, 1]$ and $y \in [0, \infty]$.

4.2 Margins

Suppose now that $M_X = \max X$ and $M_Y = \max Y$ exist.

Definition. The margins of a function $H : X \times Y \to \mathbb{R}$ are the functions

$$F : X \to \mathbb{R} \text{ defined by } F(x) = H(x, M_Y) \text{ and}$$
$$G : Y \to \mathbb{R} \text{ defined by } G(y) = H(M_X, y)$$

4.3 2-increasing functions

Definition. Let $A = [x_1, x_2] \times [y_1, y_2] \subset X \times Y$. The H -volume of the rectangle A is defined as

$$V_H(A) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

We abbreviate this as

$$V_H(A) = \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y)$$

for the first order differences of H on A. $\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y)$ and $\Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1)$. We call a function $H : X \times Y \to \mathbb{R}$ 2-increasing 1 if $V_H(A) \geq 0$ for all rectangles $A \subset X \times Y$

4.4 Summary of copulas with properties

Definition. A function $C : [0, 1]^2 \to [0, 1]$ is called a copula (or more precisely, a two-dimensional copula, or a 2-copula) if 1. C is grounded, that means $C(u, 0) = C(0, v) = 0$, 2. Its margins are the identity, that means for every $u, v \in [0, 1]$

$$C(u, 1) = u \text{ and } C(1, v) = v$$

3. C is 2 -increasing, that means for every $u_1, u_2, v_1, v_2 \in [0, 1]$ with $u_1 \leq u_2, v_1 \leq v_2$ we have

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

4.5 Types of copulas

Example. The fundamental examples are: ▷ Product or independence copula: $\Pi(u, v) = uv$, ▷ Lower Fréchet-Hoeffding bound: $W(u, v) = \max(u + v - 1, 0)$, ▷ Upper Fréchet-Hoeffding bound: $M(u, v) = \min(u, v)$.

4.6 Sklar's theorem

Theorem. (Sklar's theorem) 1. Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all $x, y \in \overline{\mathbb{R}}$

$$H(x, y) = C(F(x), G(y))$$

If F and G are continuous, then C is unique. Otherwise, C is uniquely determined on $\operatorname{Ran} F \times \operatorname{Ran} G$, where $\operatorname{Ran} F$ denotes the range of F. 2. Conversely, if C is a copula and F and G are distribution functions, then the function H, given in Equation (1), defines a joint distribution function with margins F and G.
THEN THE ASSOCIATED COPULA IS:

$$C(u, v) = \frac{uv}{u + v - uv}$$

Examples of copulas

- fundamental couplas: independent copula, upper and lower bounds
- implicit copulas: extracted from known cdf's F via Sklar's theorem. Cannot write C=..... directly.
- explicit copulas: can write C=... directly, but their use is very limited and they are quite unnecessary.

Parametric copulas

We can divide parametric couplas into *implicit* and *explicit* ones. Sklar's theorem states that we can always find a copula in a parametric distribution function. Let H be the distribution function and let F_1, F_2 its continuous margins. Then the implied copula is

$$C(u_1, u_2) = H\left(F_1^{(-1)}(u_1), F_2^{(-1)}(u_2)\right)$$

Such a copula may not have a simple closed form.

Financial interpretation ▷ Consider 2 daily returns $X = (X_1, X_2)$ with pairwise correlations all equal to $\rho = 0.5$.
However, we are unsure about the best joint model.
D If the copula of X is $C_{\rho=0.5}^{\text{Gauss}}$, the probability that on a given day both returns lie below their $\alpha = 0.01$ quantiles is (this is a coupla)

$$\mathbb{P}\left[X_1 \leq F_1^{-1}(\alpha), X_2 \leq F_2^{-1}(\alpha)\right] = \mathbb{P}[U_1 \leq \alpha, U_2 \leq \alpha] \approx 1.29 \times 10^{-3}$$

In the long run such an event will happen approximately once every 775 trading days ($\approx$ once every 3 years
▷ If the copula of X is $C_{\nu=4,\rho=0.5}^{t}$ (student distribution) however, such an event will happen approximately 2.22 times more often, i.e. $\approx$ once every 1.3 years.!!!!!