Case Study 1

1. Moments of distribution

2. Graphical comparison between empirical return distribution and Gaussian distribution

3. **Maximum Likelihood** calibration of power-law distribution on empirical return distribution (with bootstrap)

4. Maximum likelihood calibration of **“body-tail”** distribution

Moments and their estimators

Mean $$m = \mathbb{E}[r] \quad \longrightarrow \quad \hat{m} = \frac{1}{N}\sum_{i=1}^{N} r_i$$

Variance $$\sigma^2 = \mathbb{E}[(r-m)^2] \quad \longrightarrow \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(r_i - \hat{m})^2$$

Skewness $$\xi = \frac{1}{\sigma^3}\mathbb{E}[(r-m)^3] \quad \longrightarrow \quad \hat{\xi} = \frac{1}{N\sigma^3}\sum_{i=1}^{N}(r_i - \hat{m})^3$$

Kurtosis $$\kappa = \frac{1}{\sigma^4}\mathbb{E}[(r-m)^4] \quad \longrightarrow \quad \hat{\kappa} = \frac{1}{N\sigma^4}\sum_{i=1}^{N}(r_i - \hat{m})^4$$

Gaussian Distribution

Probability density function (PDF)

$$p(r) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(r-m)^2}{2\sigma^2}}$$

Cumulative distribution function (CDF)

$$F(r) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{r-m}{\sqrt{2}\sigma}\right)\right]$$

Complementary cumulative distribution function (CCDF)

$$C(r) = 1 - F(r) = \frac{1}{2}\left[1 - \operatorname{erf}\left(\frac{r-m}{\sqrt{2}\sigma}\right)\right]$$

Maximum likelihood

The general Maximum Likelihood (ML) problem consists of 2 steps

1. **Assuming** a probability distribution function (PDF) to model the data

2. Finding the PDF's parameters that maximise the probability of observing the data **if they were generated by the chosen PDF**

Maximum likelihood (ML)

Let us consider a set of data points $\{r\} = (r_1, r_2, \ldots, r_T)$,
and let us consider a PDF $p(r; \{\beta\})$ depending on a set of parameters $\{\beta\} = (\beta_1, \beta_2, \ldots)$

We define the likelihood of observing the data under the chosen PDF as the product of the PDF's values on the data points

$$\mathcal{L}(\{r\}|\{\beta\}) = \prod_{i=1}^{T} p(r_i|\{\beta\})$$

and define the maximum likelihood problem as

$$\{\beta^*\} = \arg\max \mathcal{L}(\{r\}|\{\beta\})$$

In practice, it is always preferable to work with the **log-likelihood**, i.e., the logarithm of the likelihood function

ML example with the Gaussian distribution

Let us assume that our data could be well described by a Gaussian distribution: let us use ML to determine its mean and standard deviation.
The likelihood and log-likelihood functions read:

$$\begin{aligned} \mathcal{L}(\{r\}|\{\mu,\sigma\}) &= \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(r_i-\mu)^2}{\sigma^2}\right) \\ &= \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{T}\frac{(r_i-\mu)^2}{\sigma^2}\right) \end{aligned}$$

$$\log(\mathcal{L}(\{r\},\{\mu,\sigma\})) = -\tfrac{T}{2}\log(2\pi\sigma^2) - \tfrac{1}{2}\textstyle\sum_{i=1}^{T}\tfrac{(r_i-\mu)^2}{\sigma^2}$$

In order to maximise the log-likelihood we must set its derivatives with respect to the parameters to zero, and solve the corresponding system of equations

ML example with the Gaussian distribution

$$0 = \frac{\partial \log(\mathcal{L})}{\partial \mu} = -\sum_{i=1}^{T} \frac{r_i - \mu}{\sigma^2} \implies \mu^* = \frac{1}{T} \sum_{i=1}^{T} r_i$$

$$0 = \frac{\partial \log(\mathcal{L})}{\partial \sigma} = -\frac{T}{\sigma} + \sum_{i=1}^{T} \frac{(r_i - \mu)^2}{\sigma^3} \implies \sigma^* = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (r_i - \mu)^2}$$

The ML solution gives us the familiar expressions for the sample mean and standard deviation!

Maximum likelihood for heavy-tailed data

Only very rarely an entire dataset follows a power-law distribution. In most practical situations empirical data follow a power law only in their tail region (i.e., large absolute values). This can be tackled via maximum likelihood with two different strategies

1. Assuming power-law behaviour only for values higher than a chosen threshold

2. Assuming the data are described by a combination of two distributions: a thin-tailed distribution for the central part of the data and a heavy-tailed one for the tails (**body-tail fitting**)

Power-law fitting with maximum likelihood

Let us start with the first option: let us assume that our data are well described by a power-law in the tail region and let us forget about the rest of the data.
In practice, we will assume that the data follow a power-law for values larger than a threshold $r_{\min}$

With these assumptions, we can use the following power-law PDF:

$$p(r) = \frac{\alpha}{r_{\min}} \left(\frac{r}{r_{\min}} \right)^{-(\alpha+1)} \qquad \text{for} \quad r \in [r_{\min}, \infty)$$

Solving the maximum likelihood for the only parameter (the **tail exponent**) gives:

$$\alpha^* = T / \sum_{i=1}^{T} \log(r_i / r_{\min})$$

Bootstrap

Maximum likelihood only provides us with one estimate for the tail exponent. We can complement this with a confidence interval via bootstrap.

In a nutshell, bootstrap is a **random resampling of portions of the same data.** For instance, we could randomly select 70% or 80% of the data, and compute a slightly different estimate of the tail exponent from those.

Repeating the process many times (i.e., selecting many different random portions of the data) ultimately provides us with a confidence interval

Body-tail fitting

1. Calibrating 2 different distributions to capture different regimes in the body and tail of an empirical distribution

2. A possible way to do it is to maximize the likelihood in order to find an optimal point r^* such that

$$p(r) = p_{\mathrm{body}}(r) \qquad \text{for} \qquad r < r^*$$

$$p(r) = p_{\mathrm{tail}}(r) \qquad \text{for} \qquad r \geq r^*$$

$$p_{\mathrm{body}}(r^*) = p_{\mathrm{tail}}(r^*)$$

Example (one tail fit)

Let us assume an exponential PDF for the body and a power law PDF for the tail of the distribution

Body PDF

$$p_{\text{body}}(r) = C^{-1}\lambda e^{-\lambda r}$$

Tail PDF

$$p_{\text{tail}}(r) = C^{-1}\frac{\alpha}{r_{\min}}\left(\frac{r}{r_{\min}}\right)^{-(\alpha+1)}$$

Normalization:

$$C = \int_0^{r_{\min}} p_{\text{body}}(r)dr + \int_{r_{\min}}^{\infty} p_{\text{tail}}(r)dr = 2 - e^{-\lambda r_{\min}}$$

The likelihood maximisation problem

We want to maximise the likelihood function

$$\mathcal{L}(\{r_i\}|\alpha,\lambda,r_{\min}) = \prod_{r_i<r_{\min}} p_{\text{body}}(r_i) \prod_{r_i\geq r_{\min}} p_{\text{tail}}(r_i)$$

To improve numerical accuracy, it is usually better to maximise the log-likelihood, which in our case reads

$$\log\mathcal{L}(\{r_i\}|\alpha,\lambda,r_{\min}) = \sum_{r_i<r_{\min}} \log p_{\text{body}}(r_i) + \sum_{r_i\geq r_{\min}} \log p_{\text{tail}}(r_i)$$

$$= \sum_{r_i<r_{\min}}(\log(\lambda) - \lambda r_i - \log(C))$$

$$+ \sum_{r_i>r_{\min}}(\log(\alpha/r_{\min}) - (\alpha+1)\log(r_i/r_{\min}) - \log(C))$$

Ensuring continuity

In principle, nothing guarantees that the PDF resulting from the tail-body maximum likelihood fitting will be a continuous function in $r_{\min}$

We have to manually ensure continuity by adding a **penalty term** to the log-likelihood function, which penalises discontinuities in $r_{\min}$.
A possible choice is:

$$\log \mathcal{L}(\{r_i\}|\alpha, \lambda, r_{\min}) \to \log \mathcal{L}(\{r_i\}|\alpha, \lambda, r_{\min}) - \Phi \left(p_{\mathrm{body}}(r_{\min}) - p_{\mathrm{tail}}(r_{\min})\right)^2$$

with $\Phi > 0$