Projet Statistique

June 4, 2019

Préambule

Etude statistique du return du stock d'Apple entre le 1er janvier 2012 et le 30 Mai 2019.
Nous utiliserons la base de données "YahooFinance" de l'action Apple.
Return(t)=$\frac{S_t - S_{t-1}}{S_t}$

- Hypothèse 1:

The two distributions most commonly used in the analysis of financial asset returns and prices are the normal distribution and its cousin the lognormal distributions. "In practice, the lognormal distribution has been found to be a usefully accurate description of the distribution of prices for many financial assets…the normal distribution is often a good approximation for returns."[1]

Figure: Extrait d'une recherche doctorale sur la distribution des index returns

- Hypothèse 2 : Les returns sont i.i.d.

Test sur la loi de notre observation

Hyptohèse : notre modèle est $\mathscr{P} = \{\mathcal{N}(\theta, \sigma^2), \theta \in R, \sigma > 0\}$, on le suppose exact et notons $\mathscr{F}$ la famille de fonction de répartition associée.
Test de Kolmogorov-Smirnov : Règle de décision : on rejette $\mathcal{H}_0$ si $\forall F \in \mathscr{F}, ||\hat{F}_n - F||_\infty > \frac{z_\alpha}{\sqrt{n}}$.

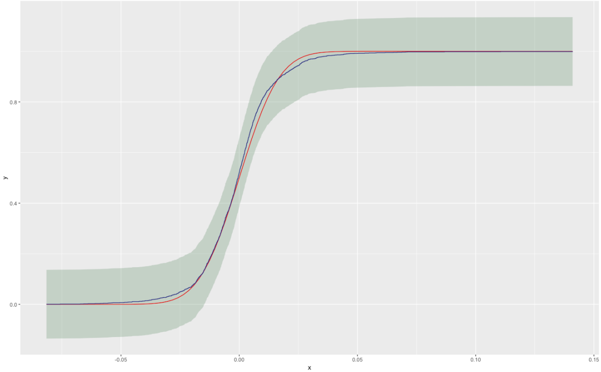


Figure: Fonction de répartition empirique et bande de confiance

Test sur la loi de notre observation

On trouve qu'il existe une fonction de répartition $F_{a,b} \in \mathscr{F}$ tel que $||\hat{F}_n - F_{a,b}||_\infty > \frac{z_\alpha}{\sqrt{n}}$ $(\alpha = 5\%)$ pour a=75 et b=0 (correspondant à une $\mathscr{N}(0, 0.0001777)$). On accepte donc l'hypothèse nulle.

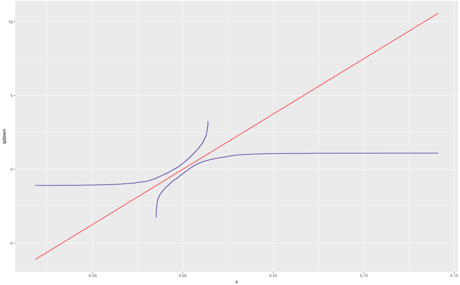
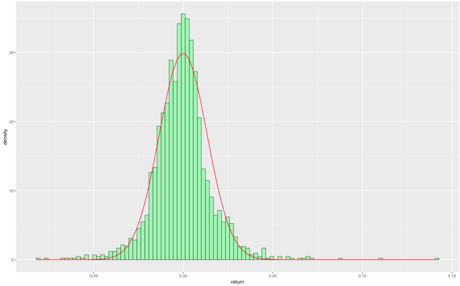


Figure: Obtention d'une loi normale appartenant à la bande de confiance en passant par la fonction quantile

Test sur la loi de notre observation



Fact (heavy tails)

The probability distribution of return often exhibits **heavier tails** *than those of a normal distribution.*

Fact (asymmetry)

The distribution of return is **asymmetric** *and often* **negatively skewed**, *reflecting the fact that the downturns of financial markets are often much steeper than the recoveries. Investors tend to react more strongly to negative news than to positive news*

Figure: Distribution de nos observations

Comparaison médiane et moyenne empirique

Médiane empirique : -0.000512078
Moyenne empirique : -0.0006809

Modèle linéaire gaussien

On peut modéliser notre étude en se ramenant au modèle linéaire gaussien suivant:

- $\begin{pmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \dots \\ Y_n \end{pmatrix} = \theta \begin{pmatrix} 1 \\ 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} + \sigma\varepsilon$
- $\varepsilon \sim \mathcal{N}(0,1)$
- $\sigma \in \mathbb{R}_+^*$
- $\hat{\theta}_n = \bar{Y}_n$
- $\hat{\sigma}_n^2 = \frac{\sum (Y_i - \hat{\theta}_n)^2}{n-1}$
- $\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\hat{\sigma}_n} \backsim \Gamma(n-1)$
- $\frac{\hat{\sigma}_n}{\sigma^2}(n-1) \backsim \chi^2(n-1)$

Test et Intervalle de confiance sur θ

- On peut effectuer le test : $\mathcal{H}_0 : \theta < 0$ contre $\mathcal{H}_1 : \theta \geq 0$
- Règle de décision : $\hat{\theta_n} > t$ (t à déterminer) On trouve t=$\frac{z_\alpha \hat{\sigma}_n}{\sqrt{n}} = 6.29 \times 10^{-3}$ où z_α est le $(1-\alpha)$ quantile d 'une Student(n-1)
- Or $\hat{\theta_n} = \bar{X}_n = -0.0006809$, donc on ne peut pas rejeter l'hypothèse nulle.
- Nous construisons un intervalle de confiance bilatère à 95% pour θ afin de pouvoir au jour le jour avoir une approximation
- $\theta \in \left[\hat{X}_n \pm \frac{z_{\alpha/2} \times \hat{\sigma}_n}{\sqrt{n}}\right]$ avec $z_{\alpha/2}$le $1-\frac{\alpha}{2}$ quantile d'une student (n-1=1862)
- Avec notre jeu de données on obtient l'intervalle suivant $\left[-0.000143; 7 \times 10^{-5}\right]$

Test statistique

- On souhaite effectuer le test suivant:
  $\mathcal{H}_0 : \sigma^2 > 0.0002$ contre $\mathcal{H}_1 : \sigma^2 \leq 0.0002$
- Règle de décision : $\hat{\sigma}_n^2 < t$ (t à déterminer pour avoir un test de niveau 5%)
- On trouve $t = \frac{z_\alpha \times 0.0002}{n-1} = 0.000189$ où z_α est le α-quantile d'une chi-deux de paramètre n-1=1862
- On obtient $\hat{\sigma}_n^2 = 0.00026684$ donc on décide d'accepter l'hypothèse nulle

Test statistique version p-valeur

Calculons la p-value du test précédent:
Calcul p-value :

$$\begin{aligned}
\alpha(T((x_1,...x_n)) &= \sup_{\sigma^2 > 0.0002} \mathbb{P}(T(\mathbb{X}) < T(x_1,...x_n)) \\
&= \sup_{\sigma^2 > 0.0002} \mathbb{P}(\hat{\sigma}_n^2 \leq T(x_1,...x_n)) \\
&= \mathbb{P}(\chi^2(n-1) \leqq 2484.28) \\
&= 1
\end{aligned}$$

On retrouve bien le même résultat, on accepte toujours l'hypothèse nulle.
En testant plutôt $\mathcal{H}_0 : \sigma^2 > 0.00025$ contre $\mathcal{H}_1 : \sigma^2 \leq 0.00025$, on obtient une p-value de 0.97
Pour : $\mathcal{H}_0 : \sigma^2 > 0.0003$ contre $\mathcal{H}_1 : \sigma^2 \leq 0.0003$, on obtient une p-value de 1.5×10^{-6}, donc on rejetterai pour presque tous les seuils.

Etude bayésienne

Soient $X_1, ...X_n$ i.i.d de loi $\mathscr{N}(\theta, \sigma^2)$
Loi à priori $\theta \sim \mathscr{N}(0, K^2)$
$\sigma, K \in \mathbb{R}_+^*$
Obtention de la loi à posteriori : $\theta \mid \mathbb{X} \sim \mathscr{N}(\frac{S_n K^2}{\sigma^2+nK^2}, \frac{K^2\sigma^2}{\sigma^2+nK^2})$
D'après le théorème du cours, on obtient l'estimateur bayésien suivant pour θ : $\tilde{\theta}_n = \frac{S_n K^2}{\sigma^2+nK^2}$
Risque quadratique de l'estimateur bayésien : $R(\tilde{\theta}_n, \theta) = \frac{\theta^2\sigma^4+\sigma^2 nK^4}{(\sigma^2+nK^2)^2}$
Risque quadratique de la moyenne empirique : $R(\bar{X}_n, \theta) = \frac{\hat{\sigma}^2}{n}$

Etude bayésienne

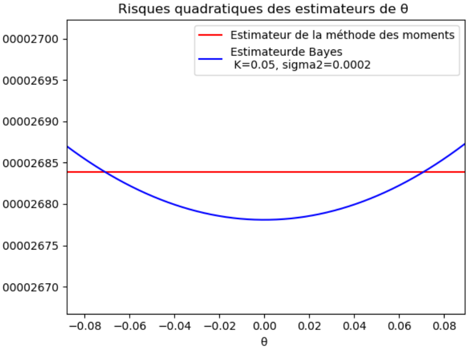


Figure: Risques quadratiques des deux estimateurs

Dans notre jeu de données, la valeur obtenue pour $\hat{\theta}_n$ est -0.0001458, la valeur pour $\bar{X}_n$ est-0.0006809 et $\mid \bar{X}_n - \hat{\theta}_n \mid$=0.0005350
ici notre jeu de données valide l'hypothèse du modèle.