

# Finanças Quantitativas: Lista 2

Lucas Moschen

10 de abril de 2020

## Execício 1

Calcule a média da distribuição GPD e use essa informação para mostrar que a mean excess function  $e(l)$  é linear em  $l$ .

### Resposta:

Seja  $X \sim GPD(m, \lambda, \xi)$ , onde  $m \in \mathbb{R}, \lambda > 0, \xi \in \mathbb{R}$ . Assim, a distribuição acumulada de  $X$  é:

$$F_{m,\lambda,\xi}(x) = \begin{cases} 1 - (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(-\frac{x-m}{\lambda}), & \xi = 0 \end{cases}$$

E a pdf da distribuição é

$$f_{m,\lambda,\xi}(x) = \begin{cases} \frac{1}{\lambda} (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}-1}, & \xi \neq 0 \\ \frac{1}{\lambda} \exp(-\frac{x-m}{\lambda}), & \xi = 0 \end{cases}$$

Considere o caso em que  $1 > \xi > 0$ . Desta forma a média será finita.

$$\mathbb{E}[X] = \int_m^\infty x f_{m,\lambda,\xi}(x) dx \quad (1)$$

$$= \int_m^\infty x \left( \frac{1}{\lambda} (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}-1} \right) dx \quad (2)$$

$$= -x \cdot (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}} \Big|_m^\infty + \int_m^\infty (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}} dx \quad (3)$$

$$= \lim_{x \rightarrow \infty} -\frac{x}{(1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}}} + m + \int_m^\infty (1 + \xi \frac{x-m}{\lambda})^{-\frac{1}{\xi}} dx \quad (4)$$

Note que esse limite é 0, bastando aplicar L'Hôpital, e notando que  $\frac{1}{\xi} > 1$ . Fazendo a substituição  $u = 1 + \xi \frac{x-m}{\lambda} \implies \frac{\lambda}{\xi} du = dx$ . Portanto:

$$\mathbb{E}[X] = m + \frac{\lambda}{\xi} \int_1^\infty u^{-\frac{1}{\xi}} du \quad (5)$$

$$= m + \frac{\lambda}{\xi} \frac{\xi}{-1 + \xi} u^{-\frac{1}{\xi}+1} \Big|_1^\infty \quad (6)$$

$$= m + \frac{\lambda}{-1 + \xi} \left[ \lim_{u \rightarrow \infty} u^{\frac{\xi-1}{\xi}} - 1^{\frac{\xi-1}{\xi}} \right] \quad (7)$$

$$= m + \frac{\lambda}{1 - \xi} \quad (8)$$

No caso em que  $\xi = 0$ , o cálculo fica relativamente mais fácil. Utilizando o mesmo processo, chegamos que  $E[X] = m - \lambda[\exp(-\frac{x-m}{\lambda})]_m^\infty = m + \lambda$ , como esperávamos.

Sabemos que se a distribuição de uma variável aleatória  $X$  é  $GDP(m, \lambda, \xi)$ , então a distribuição excesso  $F_l(x)$ , dado um nível  $l$ , será uma distribuição  $GDP(0, \lambda + \xi(l - m), \xi)$ . Desta maneira, a função de excesso médio é uma função linear em  $l$ , utilizando o valor esperado dessa distribuição, como vimos acima:

$$e(l) = \mathbb{E}[F_l(x)] = 0 + \frac{\lambda + \xi(l - m)}{1 - \xi} = \frac{\lambda - m\xi}{1 - \xi} + \frac{\xi}{1 - \xi} \cdot l = C + D \cdot l \quad (9)$$

Além disso, ela será constante caso  $\xi = 0$ .

## Problema 2.2

1. *For this first question we assume that  $X$  is a random variable with standard Pareto distribution with shape parameter  $\xi$  (location parameter  $m = 0$ , scale parameter  $\lambda = 1$ ).*

- 1.1. *Give a formula for the c.d.f. of  $X$ . Explain*

### Resposta:

Utilizando a distribuição da GDP encontrada no exercício 1, temos que a cdf da distribuição de pareto padrão pode ser escrita como:

$$F_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(-x), & \xi = 0 \end{cases}$$

O suporte para  $\xi \geq 0$  é formado pelos reais não negativos. Caso contrário, será  $0 \leq x \leq -\frac{1}{\xi}$ . Observe que ela será uma CDF, visto que é uma função contínua (combinação de funções contínuas),  $\lim_{x \rightarrow -\infty} F_\xi(x) = 0$  e  $\lim_{x \rightarrow \infty} F_\xi(x) = 1 - \lim_{x \rightarrow \infty} \frac{1}{(1 + \xi x)^{\frac{1}{\xi}}} = 1$ , como desejamos de uma CDF. Em particular já sabíamos que isso aconteceria por ser um caso especial da distribuição GPD.

- 1.2. *Derive a formula for the quantile function of  $X$ .*

### Resposta:

Seja  $Q(p)$  a função quantil. Quando  $\xi = 0$ , basta que  $Q(p) = F_\xi^{-1}(x) = -\log(1 - p) = \log \frac{1}{1-p}$ .

Se  $\xi \neq 0$ , seja  $F_\xi(x) = p$ . Assim  $p = 1 - (1 + \xi x)^{-\frac{1}{\xi}} \implies (1 + \xi x) = (1 - p)^{-\xi} \implies Q(p) = \frac{1}{\xi}((1 - p)^{-\xi} - 1)$ .

Agora podemos agrupar e temos que em  $0 \leq p < 1$ , a função quantil é:

$$Q(p) = \begin{cases} \frac{1}{\xi}((1 - p)^{-\xi} - 1), & \xi \neq 0 \\ \log \frac{1}{1-p}, & \xi = 0 \end{cases}$$

- 1.3. *How would you generate Monte Carlo samples from the distribution of  $X$  if you only had a random generator for the uniform distribution on  $[0, 1]$  at your disposal?*

### Resposta:

Primeiro gere uma amostra de tamanho  $n$  da distribuição uniforme. Ao usar a universalidade da uniforme, podemos aplicar a função quantil encontrada a cada ponto da amostra e assim teremos uma amostra Monte Carlo da distribuição de  $X$ .

2. Give a formula for the density  $f_Y(y)$  of a random variable  $Y$  which is equal to an exponential random variable with mean 2 with probability  $1/3$  and to the negative of a classical Pareto random variable with shape parameter  $\xi = 1/2$  (location  $m = 0$  and scale  $\lambda = 1$ ) with probability  $2/3$ . Explain.

**Resposta:**

Seja  $X_1 \sim \text{Exp}(\frac{1}{2})$  e  $X_2 \sim \text{Pareto}(\frac{1}{2})$ . Assim, pela lei da probabilidade total:

$$\mathbb{P}(Y \leq y) = \mathbb{P}(Y \leq y | Y = X_1) \mathbb{P}(Y = X_1) + \mathbb{P}(Y \leq y | Y = -X_2) \mathbb{P}(Y = -X_2) \quad (10)$$

$$= (1 - \exp(-\frac{1}{2}y)) \cdot \frac{1}{3} + \mathbb{P}(X_2 \geq -y) \cdot \frac{2}{3} \quad (11)$$

$$= \frac{1}{3}(1 - \exp(-\frac{1}{2}y)) + \frac{2}{3}(1 - \frac{y}{2})^{-2} \quad (12)$$

Porém, temos que nos ater aos suportes dessas distribuições. Se  $y < 0$ , teremos que  $\mathbb{P}(X_1 \leq y) = 0$ , logo  $\mathbb{P}(Y \leq y) = \frac{2}{3}(1 - \frac{y}{2})^{-2}$ . Agora, se  $y \geq 0$ , temos que  $\mathbb{P}(X_2 \geq -y) \geq \mathbb{P}(X_2 \geq 0) = 1$ , logo:

$$F_Y(y) = \begin{cases} \frac{2}{3}(1 - \frac{y}{2})^{-2}, & y < 0 \\ \frac{1}{3}(1 - \exp(-\frac{1}{2}y)) + \frac{2}{3}, & y \geq 0 \end{cases}$$

Observe que os limites ao infinito, de ambos os lados ocorrem conforme esperávamos e que essa função é contínua em todos os pontos. Logo é uma CDF bem determinada. Além disso, ela é diferenciável em todos os pontos, com exceção do 0, pois as derivadas laterais são diferentes. Portanto

$$f_Y(y) = \begin{cases} \frac{2}{3}(1 - \frac{y}{2})^{-3}, & y < 0 \\ \frac{1}{6}(\exp(-\frac{1}{2}y)), & y > 0 \end{cases}$$

3. How would you generate Monte Carlo samples from the distribution of  $Y$  ?

**Resposta:**

Gero uma amostra de tamanho  $n$  da distribuição uniforme no intervalo  $[0, 1]$ , inicialmente. Como a densidade é sempre positiva, a distribuição acumulada é invertível. Após a expressão da função quantil, para cada ponto da amostra, calculo a função. Pela universalidade da uniforme, terei uma amostra Monte Carlo dessa distribuição. Confira a função quantil:

$$Q(p) = \begin{cases} 2 - (\frac{8}{3p})^{\frac{1}{2}}, & 0 < p \leq \frac{2}{3} \\ \log(\frac{1}{(3-3p)^2}), & \frac{2}{3} < p < 1 \end{cases}$$

## Problema 2.3

*In this problem, we study the loss distribution of a portfolio over a fixed period whose length does not play any role in the analysis. Loss is understood as the negative part of the return defined as  $L = \max(0, -R)$ . We assume that a fixed level  $\alpha \in (0, 1)$  is given, and we denote by  $\text{VaR}_\alpha$  the Value at Risk (VaR) at the level  $\alpha$  of the portfolio over the period in question. In the present context, this VaR is the  $100(1 - \alpha)$ -percentile of the loss distribution. This is consistent with the definition used in the text. The purpose of the problem is to derive a formula for the expected loss given that the loss is assumed to be larger than the value at risk.*

1. For this question, we assume that the loss distribution is exponential with rate  $r$ .

- 1.1. Give a formula for the c.d.f. of  $L$ . Explain.

**Resposta:**

Se  $L \sim \text{Exp}(r)$ , então a sua cdf, descrita pela função  $F_L$ :

$$F_L(x) = \begin{cases} 1 - \exp(-rx), & x \geq 0 \\ 0, & x < 0 \end{cases}$$

1.2. *Derive a formula for  $VaR_\alpha$ .*

**Resposta:**

Nesse exercício,  $VaR_\alpha$  é  $100(1 - \alpha)$  percentil da distribuição de perda. Desta forma:

$$VaR_\alpha = \inf\{x; F(x) \geq (1 - \alpha)\} = F^{-1}(1 - \alpha) = -\frac{1}{r} \log(\alpha)$$

1.3. *Give a formula for the expected loss given that the loss is larger than  $VaR_\alpha$ .*

**Resposta:**

$$\mathbb{E}[L | L > -\frac{1}{r} \log(\alpha)] = \int_{-\infty}^{\infty} x f_{L|L > VaR_\alpha}(x) dx \quad (13)$$

$$= \int_{-\infty}^{\infty} x f_L(x - VaR_\alpha) dx \quad (14)$$

$$= \int_{-\infty}^{\infty} (u + VaR_\alpha) f_L(u) du \quad (15)$$

$$= \mathbb{E}[L] + VaR_\alpha = \frac{1}{r} [1 - \log(\alpha)] \quad (16)$$

A segunda igualdade é válida, pois a distribuição exponencial apenas foi deslocada em  $VaR_\alpha$ , inalterando o formato da curva em si. Essa propriedade é conhecida por *memoryless*.

2. *For this question, we assume that the loss distribution is the standard Pareto distribution with shape parameter  $\xi$ , location parameter  $m = 0$  and scale parameter  $\lambda = 1$ .*

2.1. *Give a formula for the c.d.f. of  $L$ . Explain.*

**Resposta:**

Como já destacado anteriormente:

$$F_L(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp(-x), & \xi = 0 \end{cases}$$

2.2. *Derive a formula for  $VaR_\alpha$ .*

**Resposta:**

Utilizando a função quantile da Distribuição de Pareto encontrada no exercício anterior:

$$VaR_\alpha = \inf\{x; F_L(x) \geq (1 - \alpha)\} = F^{-1}(1 - \alpha) = \frac{1}{\xi} (\alpha^{-\xi} - 1)$$

2.3. *Give a formula for the expected loss given that the loss is larger than  $VaR_\alpha$ .*

**Resposta:**

$$\mathbb{E}[L|L > VaR_\alpha] = \int_0^\infty x f_{L|L > VaR_\alpha}(x) dx \quad (17)$$

$$= \int_{VaR_\alpha}^\infty x f_L(x) \cdot \frac{1}{\mathbb{P}[L > VaR_\alpha]} dx \quad (18)$$

$$= \frac{1}{\alpha} \cdot [VaR_\alpha \cdot \alpha - \int_{VaR_\alpha}^\infty (1 - F_L(x)) dx] \quad (19)$$

$$= \frac{1}{\alpha} \cdot [VaR_\alpha \cdot \alpha - (1 + \xi VaR_\alpha)^{1-\frac{1}{\xi}} \cdot \frac{1}{\xi - 1}] \quad (20)$$

$$= VaR_\alpha - (1 + \xi VaR_\alpha) \cdot \frac{1}{\xi - 1} \quad (21)$$

$$= \frac{VaR_\alpha + 1}{1 - \xi} \quad (22)$$

$$= \frac{1 - \xi - \alpha^{-\xi}}{\xi(\xi - 1)} \quad (23)$$

3. The expected short fall (also known as the conditional VaR) at the level  $\alpha$  is the expected loss conditioned by the fact that the loss is greater than or equal to  $VaR_\alpha$ . The goal of this question is to quantify the differences obtained when using it as a measure of risk in the two loss models considered in questions 1 and 2.

3.1. For each  $\alpha \in (0, 1)$ , derive an equation that the rate parameter  $r$  and the shape parameter  $\xi$  must satisfy in order for the values of  $VaR_\alpha$  computed in questions 1.2 and 2.2 to be the same.

**Resposta:**

Queremos que  $\frac{\alpha^\xi - 1}{\xi} = -\frac{\log(\alpha)}{r} \implies r = \xi \frac{\log(\alpha)}{1 - \alpha^{-\xi}}$ . De fato, para cada  $\alpha \in (0, 1)$ , essa é uma equação que, para cada  $\xi$ ,  $r$  deve satisfazer, e vice-versa.

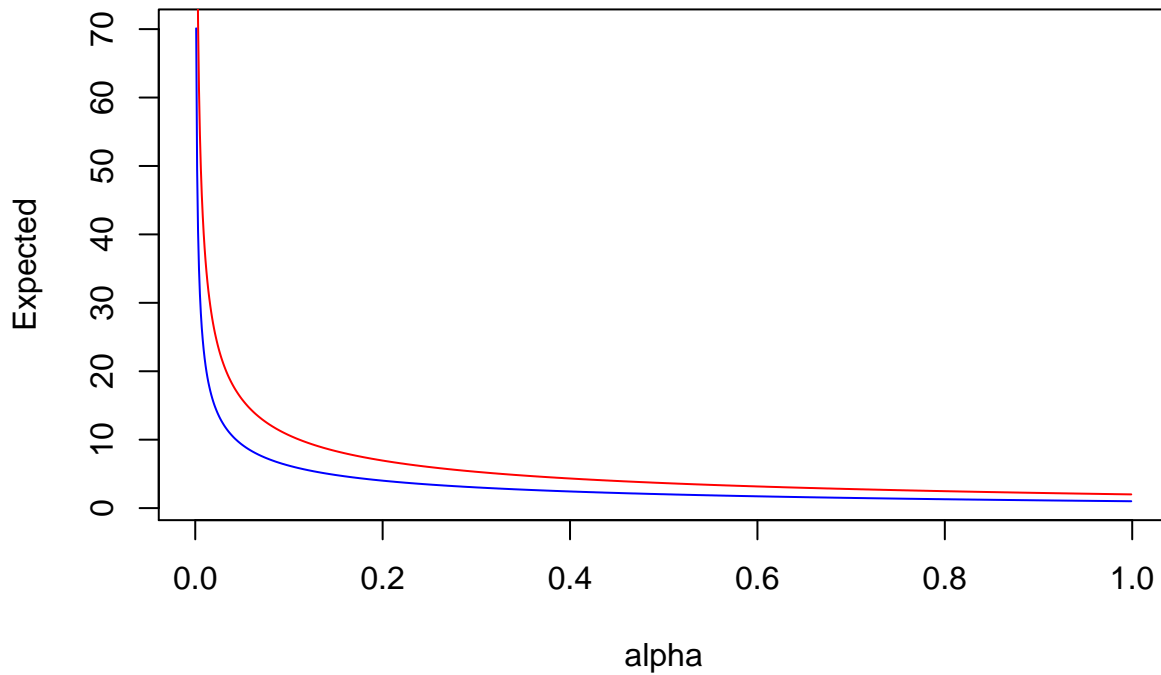
3.2. Assuming that the parameters  $r$  and  $\xi$  satisfy the relationship derived in question 3.1 above, compare the corresponding values of the expected short fall in the models of questions 1 and 2 and comment on the differences.

**Resposta:**

Suponha que  $r$  satisfaça a equação acima. Então, no caso da questão 1, se definirmos  $E_1$  como o “expected short fall”,  $E_1 = \frac{1}{r}[1 - \log(\alpha)] = \frac{1 - \alpha^{-\xi}}{\xi \log(\alpha)}[1 - \log(\alpha)]$ . Da mesma forma  $E_2 = \frac{1 - \xi - \alpha^{-\xi}}{\xi(\xi - 1)}$ . No primeiro caso, teremos uma componente que cresce menos com  $\alpha$ , isto é, se  $\alpha$  crescer, o  $E_1$  vai decrescer mais do que  $E_2$ . Observo também que as expressões são bem diferentes, sendo a segunda linear em  $\alpha^{-\xi}$ . Além disso, considere o gráfico:

```
E1 <- function(a, xi) {return((1 - log(a))*(1 - a**(-xi))/(xi*log(a)))}
E2 <- function(a, xi) {return((1 - xi - a**(-xi))/(xi*(xi - 1)))}
a <- seq(from = 0.001, to = 0.999, by = 0.001)
plot(a, E1(a, 0.5), type = 'l', main = "Comparing E1 and E2", ylab = "Expected", col = "blue", xlab = "alpha")
points(a, E2(a, 0.5), type = 'l', col = "red")
```

## Comparing E1 and E2



### Problema 2.4

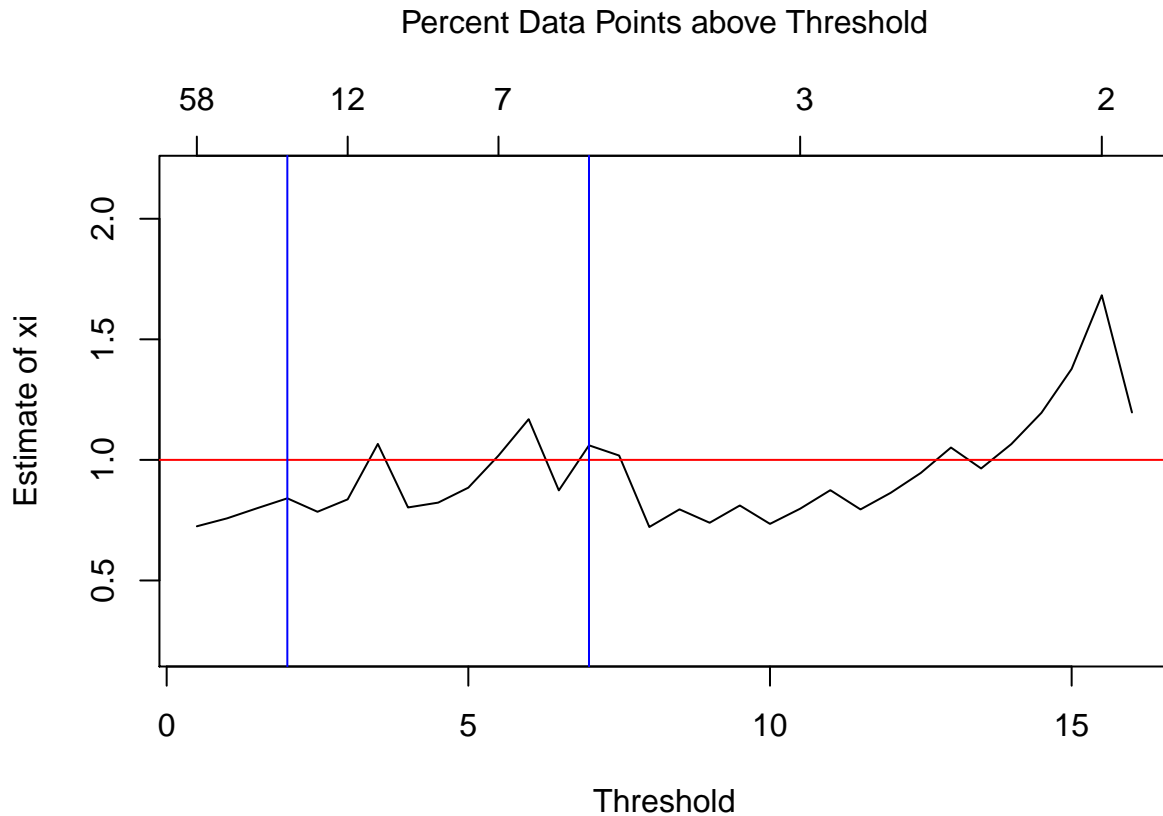
*This problem attempts to quantify the goodness of fit resulting from our GPD analysis of samples with heavy tails.*

- 1. Use the method described in the text to fit a GPD to the PCS index, and generate a Monte Carlo random sample from the fitted distribution five times the size of the original data sample.*

#### Resposta:

Considero inicialmente o gráfico gerado por *sample.plot* para analisar a variação de  $\xi$  dado o limiar. As linhas azuis indicam limites que ou tomam muitos dados da amostra, ou tomam poucos, considerando valores razoáveis. O limiar vermelho é a estimada considerada razoável para o parâmetro  $\xi$ .

```
# Getting the data and ordering it
library(Rsafd)
data(PCS, "PCS", package = "Rsafd")
# Seeing the sample.plot graphic.
shape.plot(PCS$Col12)
abline(a = 1, b = 0, col = "red")
abline(v = 2, col = "blue")
abline(v = 7, col = "blue")
```



Agora façamos as estimativas utilizando método POT descrito no capítulo:

```
# Using the method gpd.tail to estimate in the tail the parameters.
PCS.est <- gpd.tail(PCS$Col2, one.tail = TRUE, upper = 4, plot = FALSE, upper.method = "ml")
print(PCS.est$upper.par.ests)
```

```
##      lambda      xi
## 4.5014927 0.8027783
```

Por fim, façamos uma amostra Monte Carlo usando o método *gqpd*, que é o mesmo princípio da inversa da distribuição acumulada, como já destacada nos exercícios anteriores. Logo:

```
PCSsample <- gpd.1q(runif(5*PCS.est$n), PCS.est)
```

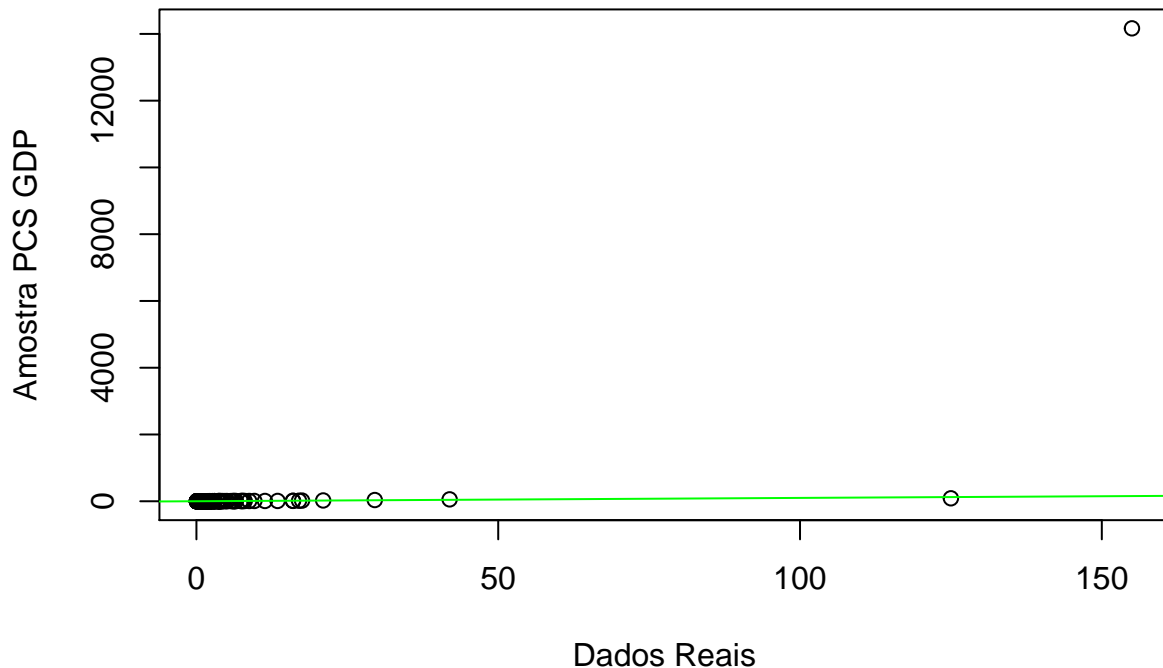
2. Produce a Q-Q plot to compare graphically the two samples and comment.

**Resposta:**

Considere o Q-Q plot. A linha em verde representa a reta  $y = x$ . Note que existe dados extremamente longe da média em nossa amostra. Isso faz sentido, porque esperávamos isso de uma GDP, então não precisamos nos importar tanto com esses valores extremos. Os pontos aparentam ter se encaixado com a reta disposta, o que significa que as duas distribuições são muito semelhantes. De fato, os maiores erros se localizam em dados nos extremos.

```
qqplot(PCS$Col2, PCSsample, ylab = "Amostra PCS GDP", xlab = "Dados Reais", main = "Q-Q Plot das amostras")
abline(a = 0, b = 1, col = "green")
```

## Q-Q Plot das amostras



3. Use a two-sample Kolmogorov-Smirnov goodness-of-fit test to quantify the qualitative results of the previous question. NB: Such a test is performed in R with the command `ks.test`. Check the help files for details on its use and the returned values.

### Resposta:

Usando método indicado, obtemos que, sendo  $D$  a estatística de teste e o  $p$  – valor quase 0:

```
ks.test(PCS$Col2, PCSsample, alternative = "t")
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: PCS$Col2 and PCSsample
## D = 0.018898, p-value = 0.9999
## alternative hypothesis: two-sided
```

A definição de  $p$ -valor dada no livro do DeGroot é

*"p-value . In general, the p-value is the smallest level  $\alpha_0$  such that we would reject the null-hypothesis at level  $\alpha_0$  with the observed data."*

Desta maneira, para qualquer teste com tamanho maior que  $p$  – value, nós rejeitamos a hipótese nula com os dados observados. Nesse caso, o  $p$ -valor é alto, o que significa que não rejeitamos a hipótese nula. Além de olhar para o  $p$ -valor, vemos que a estatística de teste é menor do que o referencial adotado pelo teste utilizado. Como a hipótese nula não foi rejeitada, quantificamos a ideia das amostras terem sido geradas a partir da mesma distribuição.

4. Same questions as above for the weekly log-returns on the S&P data.

### Resposta:

Façamos o mesmo processo para S&P data.



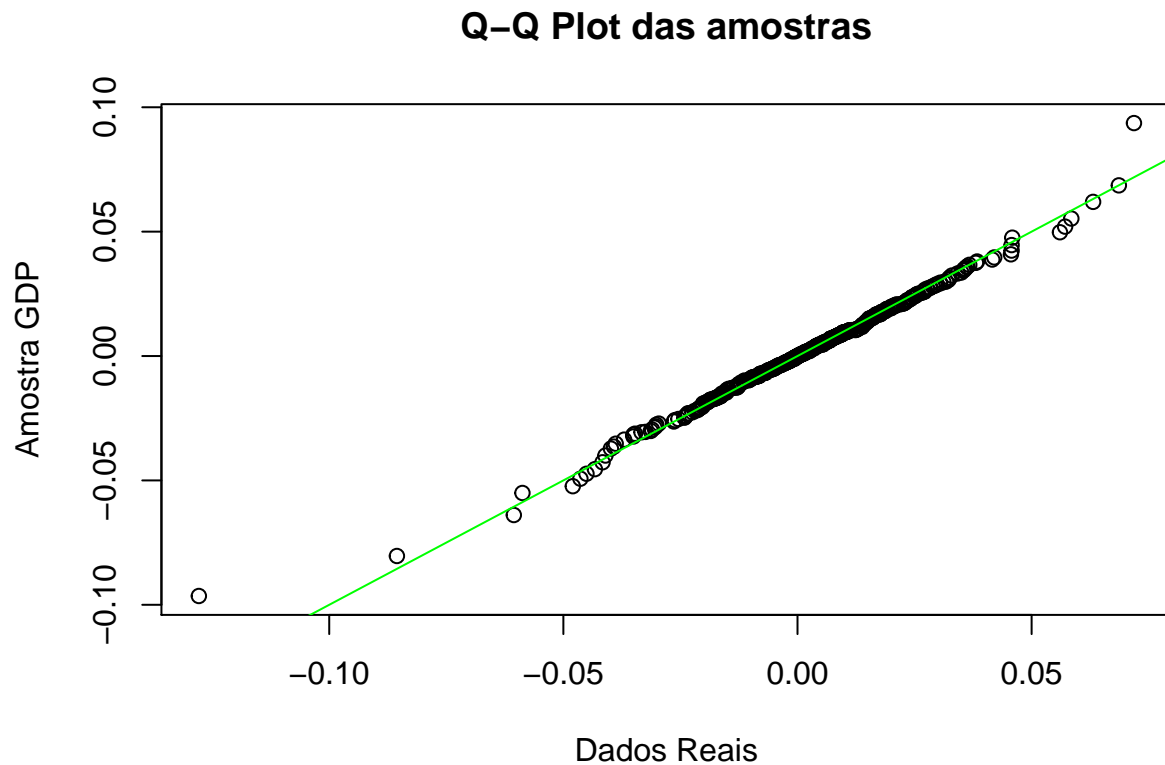
```
data("WSPLRet", WSPLRet, package = "Rsafd")

WSPLRet.est <- gpd.tail(WSPLRet, one.tail = FALSE, lower = -0.02, upper = 0.02, upper.method = "ml", pl
print(WSPLRet.est$upper.par.est)

##      lambda      xi
## 0.01182778 -0.01058769

WSPLRetsample <- gpd.2q(runif(5*WSPLRet.est$n), WSPLRet.est)

qqplot(WSPLRet.est$data, WSPLRetsample, xlab = "Dados Reais", ylab = "Amostra GDP", main = "Q-Q Plot da
abline(a = 0, b = 1, col = "green")
```



```
print(ks.test(WSPLRet, WSPLRetsample, alternative = "t"))
```

```
##
## Two-sample Kolmogorov-Smirnov test
##
## data: WSPLRet and WSPLRetsample
## D = 0.02381, p-value = 0.9812
## alternative hypothesis: two-sided
```

Podemos inferir que os dados podem ser modelados pelo método POT, dado que o p-valor é bem alto. Também observamos no Q-Q Plot que as amostras tem distribuições similares, diferindo apenas por alguns pontos nos extremos.

## Problema 2.6

*This problem requires the data set DSP. The entries of this numeric vector represent the daily closing values of the S&P 500 index between the beginning of January 1960 and September 18, 2001.*

```
data("DSP", DSP, package = "Rsafd")
data("MSP", MSP, package = "Rsafd")
```

1. Compute the vector of log-returns and call it *DSPLRet*.

**Resposta:**

```
DSPLRet = diff(log(DSP), lag = 1)
```

2. We now use the data set *MSP*. The entries of this numeric vector represent minute by minute quotes of the S&P 500 on September 10, 1998. Compute the corresponding log-return vector and call it *MSPLRet*.

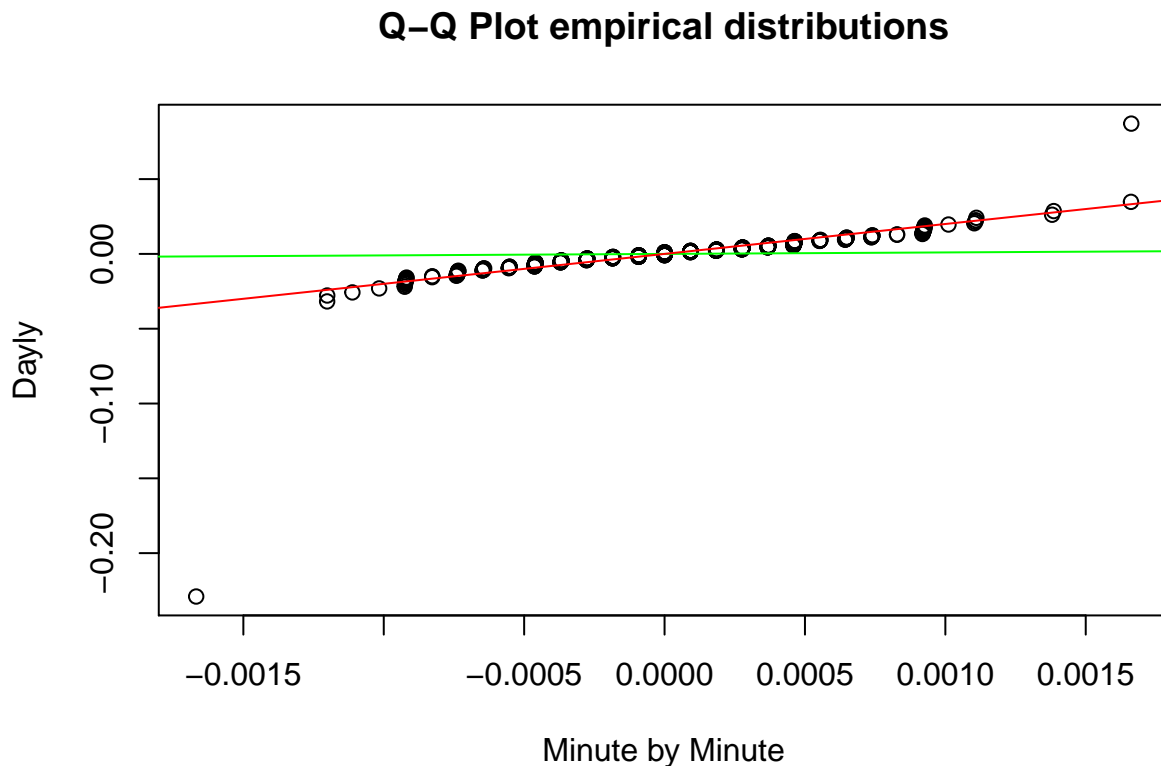
**Resposta:**

```
MSPLRet = diff(log(MSP), lag = 1)
```

3. Produce a Q-Q plot of the empirical distributions of the two log-return vectors, and comment. In particular, say if what you see is consistent with the claim that the properties of the daily series are shared by the minute by minute series. Such an invariance property is called self-similarity. It is often encountered when dealing with fractal objects.

**Resposta:**

```
qqplot(MSPLRet, DSPLRet, ylab = "Daily", xlab = "Minute by Minute", main = "Q-Q Plot empirical distributions")
abline(a = 0, b = 1, col = "green")
abline(a = 0, b = 20, col = "red")
```



De fato essa propriedade é observada. A escala de longe não é a mesma, o que faz muito sentido, visto que proporcionalmente você pode ganhar muito mais em um dia do que no outro, sem contar que os outliers

podem ser muito mais distantes. A reta em verde é desenhada apenas para ter uma noção que não é a mesma distribuição que rege ambos os dados, enquanto a linha em vermelho evidencia que é a mesma distribuição, com diferença apenas em um fator constante.

4. *Compute the empirical means and variances of the DSPLRet and MSPLRet data. Assuming that these data sets are Gaussian, would you say that the two distributions are the same in view of these two statistics?*

**Resposta:**

```
DSPLRet.mean = mean(DSPLRet)
DSPLRet.var = var(DSPLRet)
```

```
MSPLRet.mean = mean(MSPLRet)
MSPLRet.var = var(MSPLRet)
```

```
cat("Mean DSPLRet: ", DSPLRet.mean, "\n")
```

```
## Mean DSPLRet: 0.0002717871
```

```
cat("Variance DSPLRet: ", DSPLRet.var, "\n")
```

```
## Variance DSPLRet: 8.418932e-05
```

```
cat("Mean MSPLRet: ", MSPLRet.mean, "\n")
```

```
## Mean MSPLRet: 2.446315e-05
```

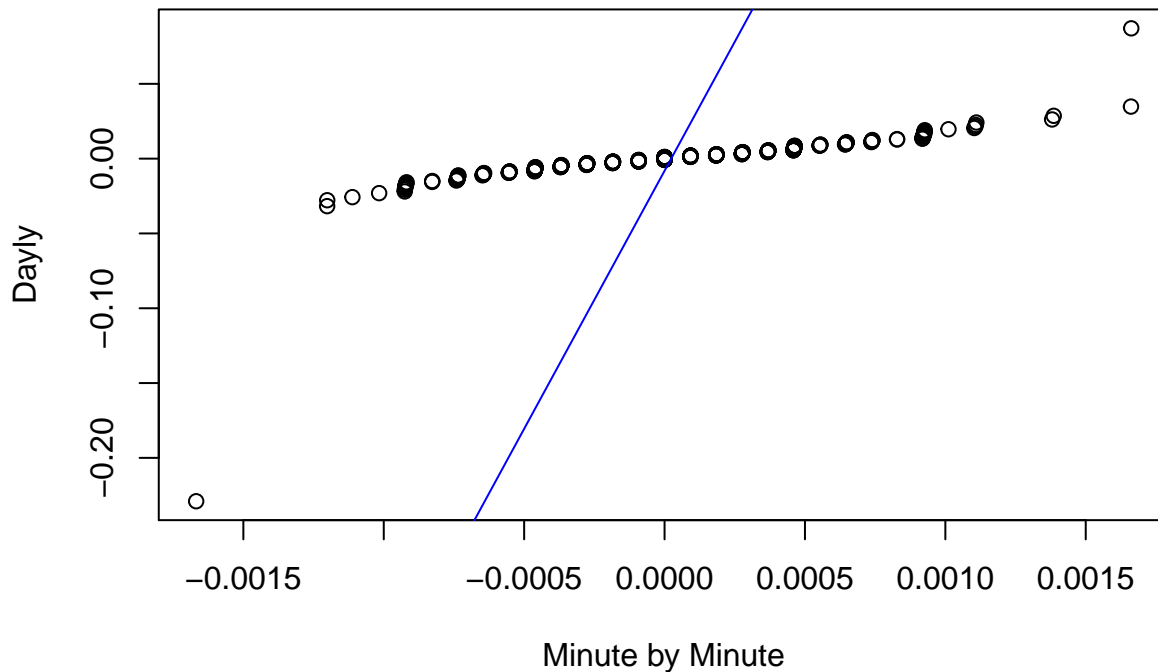
```
cat("Variance MSPLRet: ", MSPLRet.var)
```

```
## Variance MSPLRet: 2.4412e-07
```

Se esse fosse o caso, saberíamos que  $\frac{y-\mu_1}{\sigma_1} = \frac{x-\mu_2}{\sigma_2}$ , já que ambos seriam uma normal padrão. Assim,  $y = \frac{\sigma_1}{\sigma_2}(x - \mu_2) + \mu_1$  seria a reta que esperaríamos encontrar. Mas como observamos no gráfico, essa reta de longe está dos dados.

```
qqplot(MSPLRet, DSPLRet, ylab = "Dayly", xlab = "Minute by Minute", main = "Q-Q Plot empirical distribu
abline(a = -MSPLRet.mean*(DSPLRet.var/MSPLRet.var) + DSPLRet.mean,
      b = DSPLRet.var/MSPLRet.var, col = "blue")
```

## Q-Q Plot empirical distributions



5. Fit GPDs to the DSPLRet and MSPLRet data, and compare the distributions one more time by comparing the shape parameters.

Resposta:

```
DSPLRet.est <- gpd.tail(DSPLRet, plot = FALSE)
MSPLRet.est <- gpd.tail(MSPLRet, lower = -0.0005, upper = 0.0005, plot = FALSE)

cat("Shape Parameter Upper DSPLRet: ", DSPLRet.est$upper.par.ests[2], "\n")

## Shape Parameter Upper DSPLRet: 0.1679115
cat("Shape Parameter Lower DSPLRet: ", DSPLRet.est$lower.par.ests[2], "\n")

## Shape Parameter Lower DSPLRet: 0.4492701
cat("Shape Parameter Upper MSPLRet: ", MSPLRet.est$upper.par.ests[2], "\n")

## Shape Parameter Upper MSPLRet: -0.5419938
cat("Shape Parameter Lower MSPLRet: ", MSPLRet.est$lower.par.ests[2], "\n")

## Shape Parameter Lower MSPLRet: -0.5775632
```

## Problema 2.8

The goal of this problem is to highlight some of the properties of the estimates obtained with the command `fit.gpd` when fitting a GPD to a data sample  $x_1, \dots, x_n$ . We assume that the distribution of the data has two tails (one extending to  $-\infty$  and the other one to  $\infty$ , and we are interested in understanding the effect of the choice of the thresholds *lower* and *upper*.

1. What should you expect from the estimate  $\hat{\xi}_+$  given by the function `fit.gpd` if you use a threshold *upper*:

## Respostas:

### 1.1. *Exactly equal to 2.*

Eu esperaria que ele estimasse o parâmetro muito próximo do verdadeiro, visto que ele usaria os dados da distribuição da cauda de Pareto.

### 1.2. *Greater than 5.*

A função teria menos pontos para estimar a curva, e essa quantidade seria bem menor, por que a maior parte da cauda está concentrada entre 2 e 5. A variância da distribuição acabaria sendo aumentada.

### 1.3. *Between 0 and 1*

Nesse caso, muitos pontos seriam adicionados à estimativa. O parâmetro tentaria encaixar uma cauda em uma distribuição com uma descontinuidade severa, assim ela ficaria, provavelmente, abaixo da curva original.

## 2. *What should you expect from the estimate $\hat{\xi}_-$ given by the function `fit.gpd` if you use a threshold lower:*

### 2.1. *Exactly equal to $-2$ .*

Eu esperaria que ele estimasse o parâmetro muito próximo do verdadeiro, visto que ele usaria os dados da distribuição da cauda de Pareto, agora para o lado esquerdo. Nesse caso, a estimativa do percentil 1 seria muito próxima da verdadeira.

### 2.2. *Smaller than $-8$ .*

A quantidade de dados seria tão pequena, que não seria uma estimativa consistente. A parte principal da cauda à esquerda seria perdida, que fica entre  $-2$  e  $-5$ . Como em 1.2, a variância seria aumentada. Nesse caso, a estimativa do primeiro percentil seria menor do que o desejado, devido ao aumento da variância da distribuição estimada.

### 2.3. *Between 0 and $-1$ .*

Da mesma forma que comentei no item 1.3, essa curva seria rebaixada para que se encaixasse na descontinuidade central da distribuição. O parâmetro ficaria menor do que o desejado, já que tentaria encaixar uma cauda a partir de valores já baixos. Nesse caso, a estimativa do percentil 1 seria maior, pois a curva seria um pouco mais achatada, e o valor seria mais próximo de 0.