Assignmetn 3

Anton Linnér and Arvid Gramer

2023-10-12

Theoretical questions

1. The characterization of the BEVD with unit Fréchet margins $()G_*(x,y) = \exp(-(\frac{1}{x}+\frac{1}{y})A(\frac{x}{x+y}))$ where A is dependence function. For this function it holds that $(G_*(tx,ty))^t = G_*(x,y)$ which makes it max-stable.

2. The dependence function $A(\omega)$ fulfills the following properties:
   - $A(0) = 1$
   - $A(1) = 1$
   - $A'(1) \geq 1$
   - $A'(0) \geq -1$
   - A is convex

3. The formula for extreme value copulas is $C(u,v) = (uv)^{A\left(\frac{\log u}{\log uv}\right)}$. Some examples of these are (with $\tilde{u} - \log(u), \tilde{v} = -\log(v)$
   - Gumbel copula: $C(u,v) = \exp\left(-(\tilde{u}^\delta + \tilde{v}^\delta)^{1/\delta}\right)$
   - Galambos copula: $C(u,v) = uv\exp(-(\tilde{u}^{-\delta} + \tilde{v}^{-\delta})^{-1/\delta})$
   - Hüsler-Reiss copula: $C(u,v) = \exp\left(-\tilde{u}\Phi\left(\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{\tilde{u}}{\tilde{v}})\right) - \tilde{v}\Phi\left(\frac{1}{\theta} + \frac{\theta}{2}\log(\frac{\tilde{u}}{\tilde{v}})\right)\right)$, where Φ is the standard normal cumulative distribution function.

   Any copula that satisfies $C(u,v)^t = C(u^t, v^t)$ is an extreme value copula, and we can thus check that a given copula fulfills this condition to see if it is an extreme value copula.

4. If we want to create a bivariate extreme value distribution with our choice of arbitrary univariate extreme value margins we can use the extreme value copula and insert $(u,v) = (F_1^{\leftarrow}(x), F_2^{\leftarrow}(y))$, where $F^{\leftarrow}$ is the generalised inverse $(F^{\leftarrow}(t) = \inf\{x : t < F(x)\})$ of the distribution function we want to have as our margins. The distribution by $C(F_1^{\leftarrow}(x), F_2^{\leftarrow}(y))$ will have the wanted properties.

5. To derive the formula for Pickand's estimator of the dependence function $A(\omega)$ we use that if $(X,Y) \sim (\mathrm{e}^{-1/x}, \mathrm{e}^{-1/y})$ then

$$P\left[\min\left(\frac{1}{(1-\omega)X}, \frac{1}{\omega Y}\right) \geq z\right] = P\left[X \leq \frac{1}{(1-\omega)z}, Y \leq \frac{1}{\omega z}\right] \tag{1}$$

   Using that x, y have unit Fréchet distribution we get that the probability is:

$$= \exp\left\{-(z(1-\omega) + \omega z)\mathrm{A}\left(\frac{1/z(1+\omega)}{1/(z(1-\omega)) + 1/\omega z}\right)\right\} = \mathrm{e}^{-z\mathrm{A}(\omega)} \tag{2}$$

   We can thus estimate A from our data by taking the bivariate samples (x_i, y_i), $i = 1, ..., n$ and computing $z_i(\omega) = \min\left(\frac{1}{(1-\omega)x_i}, \frac{1}{\omega y_i}\right)$ and since we then have $P(z_i \leq z) = \mathrm{e}^{-z\mathrm{A}(\omega)}$ an estimate of A is

$$\hat{A}(\omega) = \frac{n}{\sum_{i=1}^n (z_i(\omega))} \tag{3}$$

Random number generation from bivariate EVD

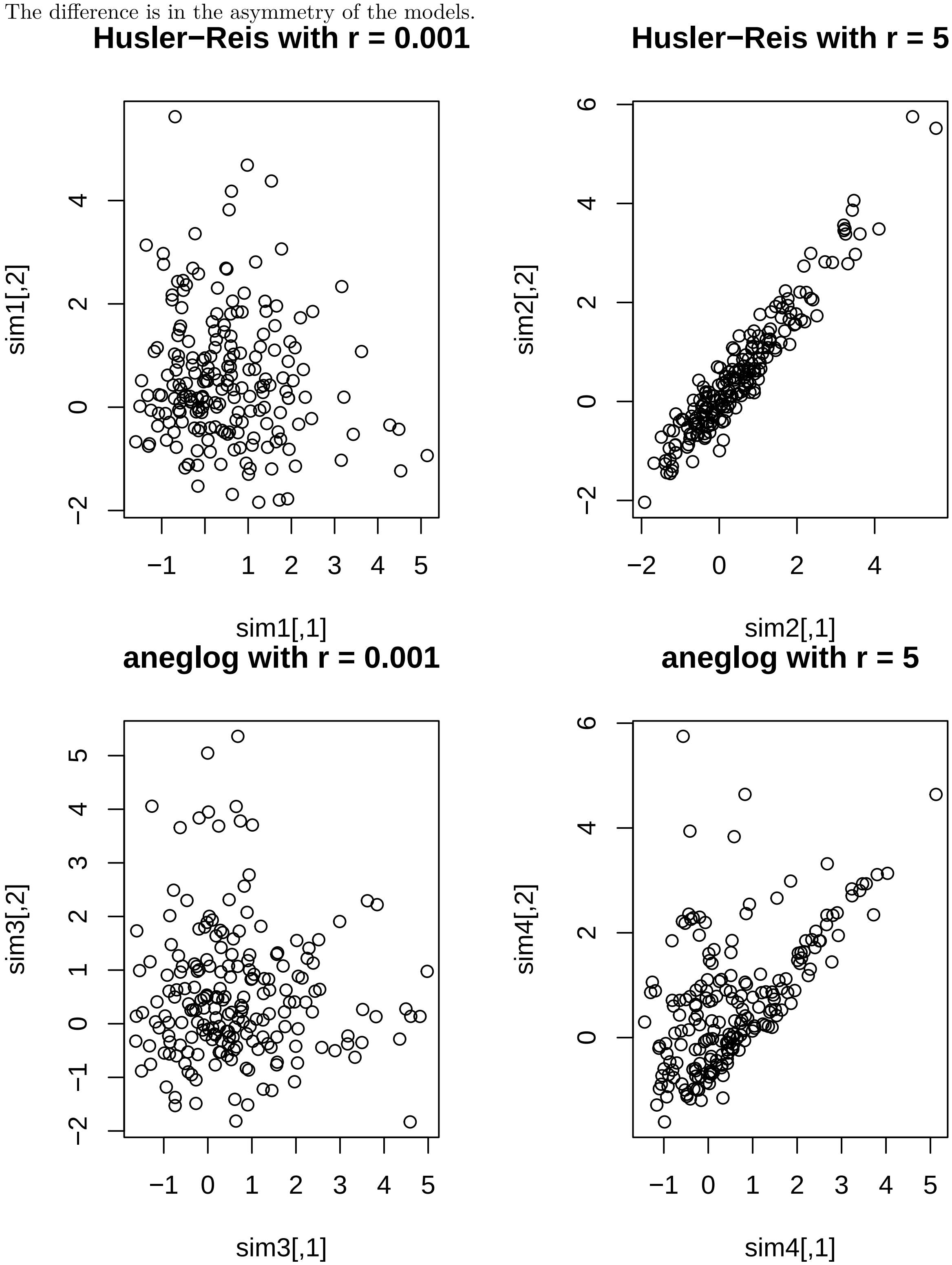
Parametric bivariate EV models

In the package `evd` the nine models and their respective dependence and asymmetry parameters are:

- Logistic. Dependence parameter `r` between (0,1]. Smaller `r` implies higher dependence.
- Asymmetrix logistic. Dependence parameter r, as in (symmetric) Logistic. Asymmetry parameters are $\texttt{t}_1$ and $\texttt{t}_2$. Indepence if any of $\texttt{t}_1, \texttt{t}_2$ are 0 or `r = 1`. For complete dependence $\texttt{t}_1 = \texttt{t}_2$=1 and $\texttt{r} \to 0$.
- Husler-Reiss. Dependence parameter $\texttt{r} \in (0, \infty)$. Full dependence as $\texttt{r} \to \infty$, and independence as $\texttt{r} \to 0$
- Negative logistic. Dependence parameter $\texttt{r} > 0$. Higher `r` implies higher dependence.
- Asymmetric negative logistic. Dependence parameter $\texttt{r} > 0$ and asymmetry parameters $\texttt{t}_1, \texttt{t}_2 \in (0, 1]$. Indepence if any of $\texttt{t}_1, \texttt{t}_2, \texttt{r}$ approaches 0. Complete dependence if $\texttt{t}_1, \texttt{t}_2 = 1, 1$ and $\texttt{r} \to \infty$.
- Bilogistic. Parameters α, β. When $\alpha = \beta$ the model is equivalent to logistic with dependence parameter $\texttt{r} = \alpha$. As in logistic, when $\alpha = \beta = \texttt{r} \to 0$ the model tends to complete dependence. Independence as either both tends to 1, or one is fix and other tends to 1.
- Negative bilogistic Parameters α, β. When $\alpha = \beta$ the model is equivalent to negative bilogistic with dependence parameter $\texttt{r} = 1\alpha$. When $\alpha = \beta \to 0$ the model tends to complete dependence. Independence as either both tends to ∞, or one is fix and other tends to ∞.
- Coles-Tawn. Parameters $\alpha, \beta > (0, 0)$. As $\alpha = \beta \to \infty$ the model shows complete dependence. Independence as either both tends to 0, or one is fix and other tends to 0.
- Asymmetric mixed distribution. Parameters α, β fulfill the following conditions: α and $\alpha + 3\beta > 0$, and $\alpha + 2\beta$, $\alpha + \beta \leq 1$ As β is fix, the strength of dependence increases with α. Complete dependence is not achievable. Independence as $\alpha = \beta = 0$.

Exercise 2.2

As we can see from the plot below, the dependence parameter does what one might assume it does. Both the assymmetrical and symmetrical model shows higher dependence for a higher dependence parameter. The difference is in the asymmetry of the models.



Dependence measures for different r and models

We can also see that the dependence measures change when the dependence parameter changes.

```
## [1] "For HR model, we have:"

## [1] "DEPENDANCE PARAM:     r=0.001                r=5"

## [1] "Correlation         " "-0.113"               "0.956"

## [1] "Kendall             " "-0.048"               "0.782"

## [1] "Spearmans Rho       " "-0.08"                "0.934"

## [1] " "

## [1] "For aneglog model, we have:"

## [1] "DEPENDANCE PARAM:     r=0.001                r=5"

## [1] "Correlation         " "-0.044"               "0.567"

## [1] "Kendall             " "0.005"                "0.407"

## [1] "Spearmans Rho       " "0.008"                "0.529"
```

Exercise 2.3

For the bivariate cases we have:

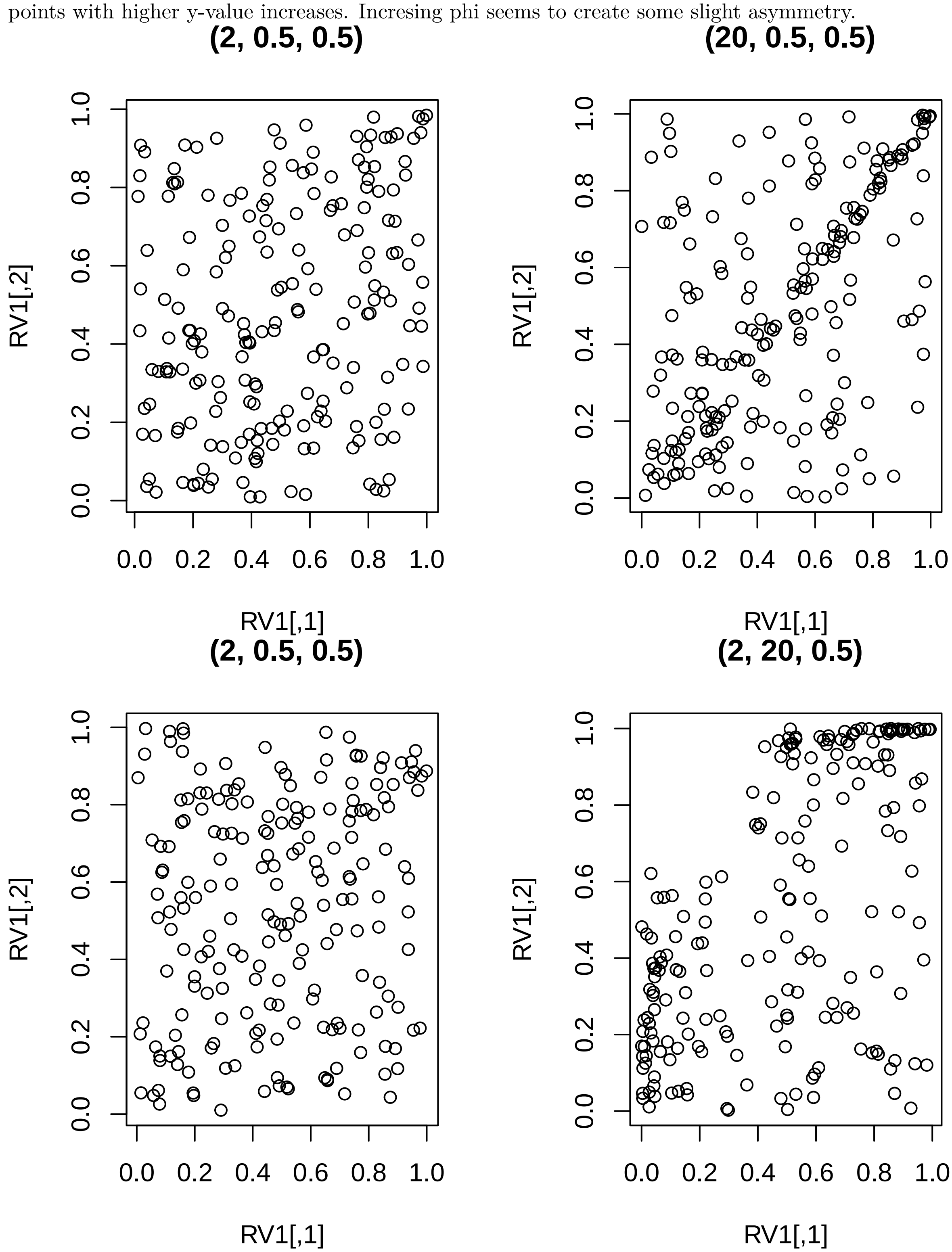
- Bivariate Extreme Value Logistic Copula: dependence parameter $r \geq 1$.
- Bivariate Extreme Value Asymmetric Logistic Copula dependence parameter $r \leq 1$, asymmetry parameters $\theta, \phi \in [0, 1]$. If both are 1 then this is a symmetric logistic.
- Bivariate Extreme Value Mixed Model Copula: parameter $0 \leq \theta \leq 1$
- Bivariate Extreme Value Asymmetric Mixed Model Copula: parameters θ, ϕ. These fulfill the following conditions, $\theta \geq 0$, $\theta + 3\phi \geq 0$, $\theta + \phi \leq 1$, $\theta + 2\phi \leq 1$.
- Bivariate Extreme Value Spline Copula. Parameter: a spline function for the dependence function.

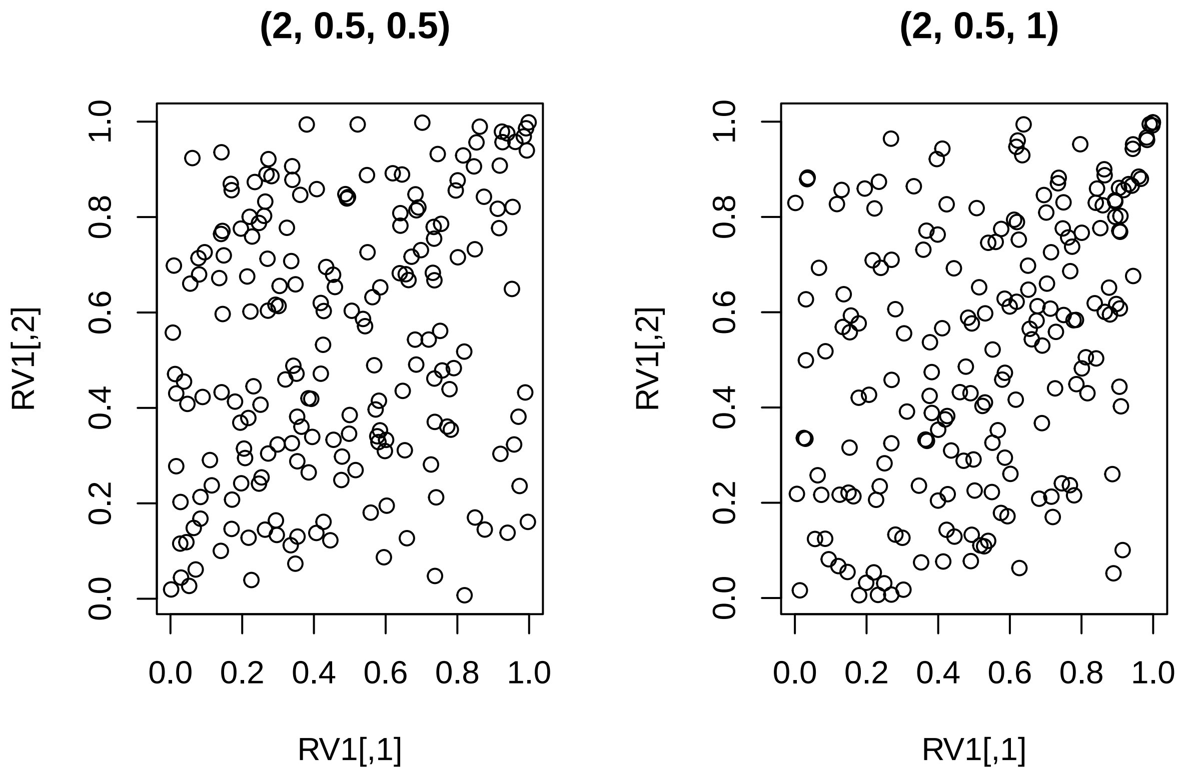
As for the multivariate copulas we have

- Multivariate Extreme Value Asymmetric Logistic Copula: θ, matrix with asymmetry parameters. $\mathbf{r}$, vector with dependency parameters. Terms and conditions apply.
- Multivariate Extreme Value Gumbel Copula. Dependence parameter $r \geq 1$.
- Multivariate Clayton Copula. Dependence parameter $\theta > 0$
- Multivariate Frank Copula. Parameter $\alpha > 0$

Exercise 2.4

The parameters are (r, theta, phi) and we compare to the case (2, 0.5, 0.5) in the plots below. We can clearly see that increasing the dependence parameter increase the dependence. Increasing theta seems to create almost a step-function. It seems like after that after a certain value on the x-axis the number of points with higher y-value increases. Incresing phi seems to create some slight asymmetry.




(2, 0.5, 0.5)
(2, 0.5, 1)
RV1[,2]
RV1[,1]
0.0
0.2
0.4
0.6
0.8
1.0

Exercise 3

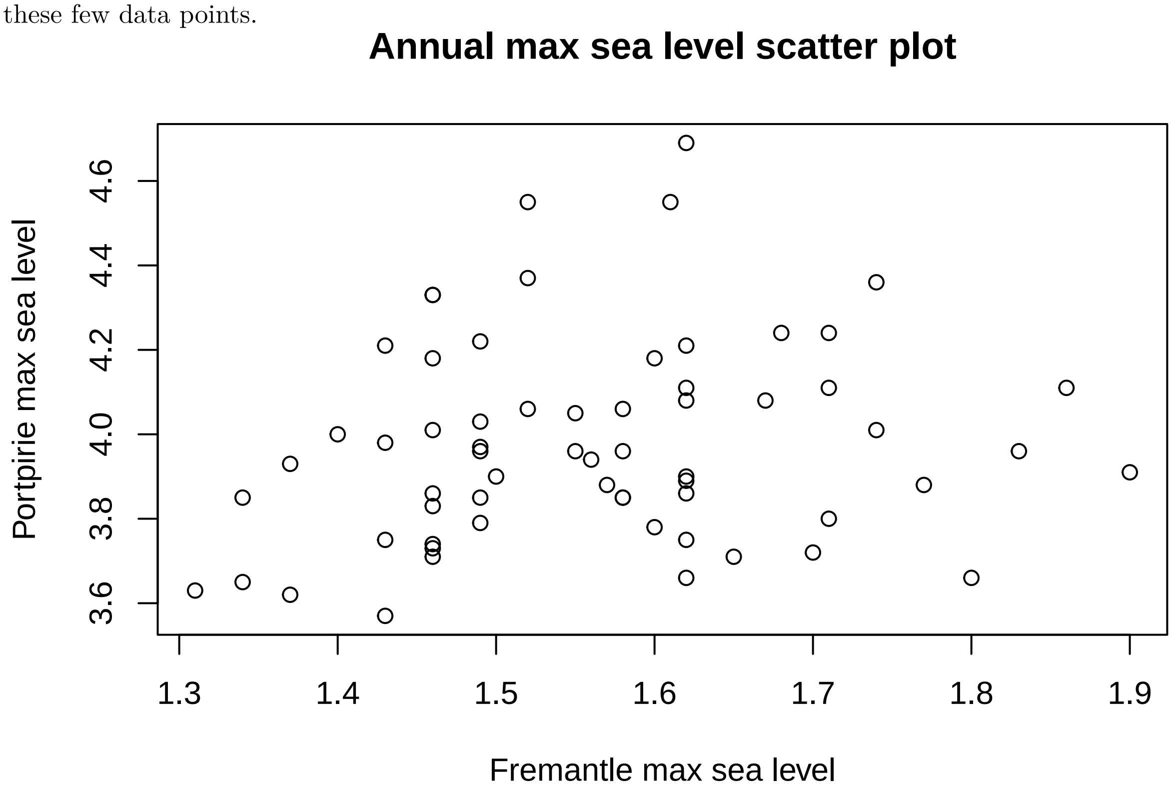
Finding the appropriate years

```
fre.years = fre$value$Year
port.years = port$value$Year
```

```
tot.years = merge(x = fre$value,y = port$value, by = "Year")
tot.years = tot.years[,-1]
tot.years = tot.years[,-2]
```

3.3

The scatter plot for the relevant years is shown below. A slight dependence can be suspected even from these few data points.



3.4

The parameters obtained from the FML method are (for model "aneglog" and "hr"):

```
#Aneglog:
(ml.fml1$estimate)
```

```
##          loc1        scale1        shape1          loc2        scale2        shape2
##   1.508781100   0.116653786  -0.129925772   3.873135899   0.201917082  -0.003309747
##          asy1          asy2           dep
##   0.196874200   0.999450028   0.607262945
```

```r
#Husler-Reiss
(ml.fml2$estimate)
```

```
##        loc1      scale1      shape1        loc2      scale2      shape2
##  1.50772409  0.11683488 -0.11597394  3.87005155  0.19716004 -0.01667925
##         dep
##  0.66560318
```

And for the IFM method we have:

```r
#Parameters for Marginal distribution of Fremantle data:
(marg.fit1)
```

```
##        loc      scale      shape
##  1.5089080  0.1174061 -0.1489734
```

```r
#Parameters for Marginal distribution of Portpirie data:
(marg.fit2)
```

```
##         loc       scale       shape
##  3.87112446  0.19746288 -0.04287117
```

```r
#Parameters for the "aneglog" model:
(ml.ifm1$estimate)
```

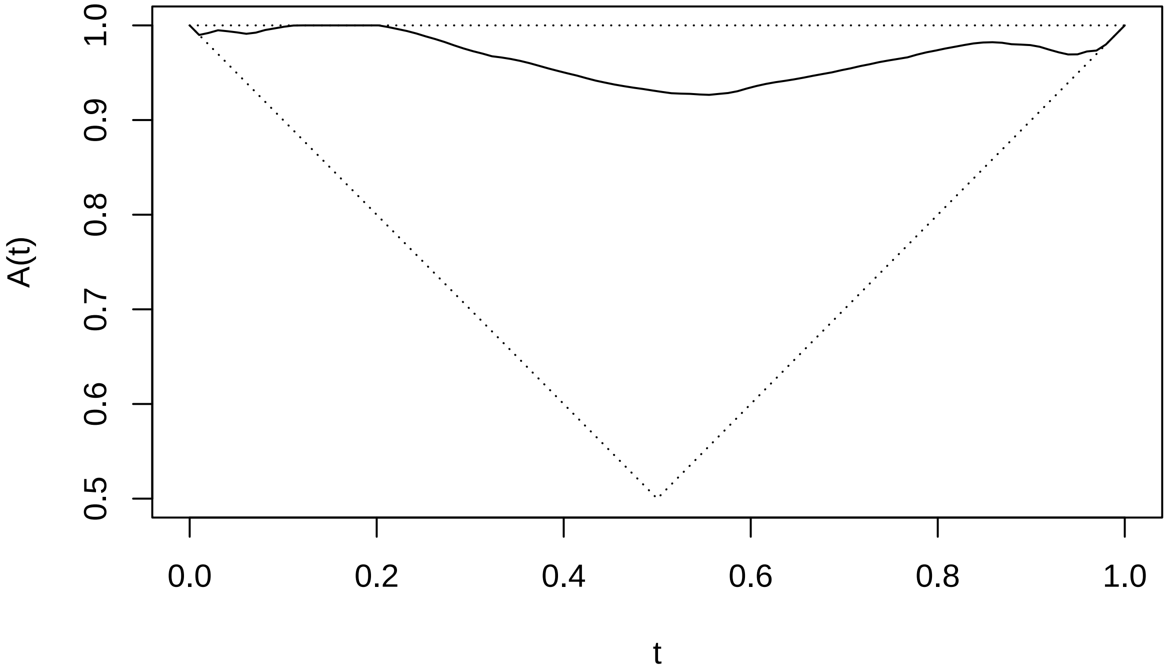
```
##      asy1      asy2       dep
## 0.1821782 0.9995316 0.5572431
```

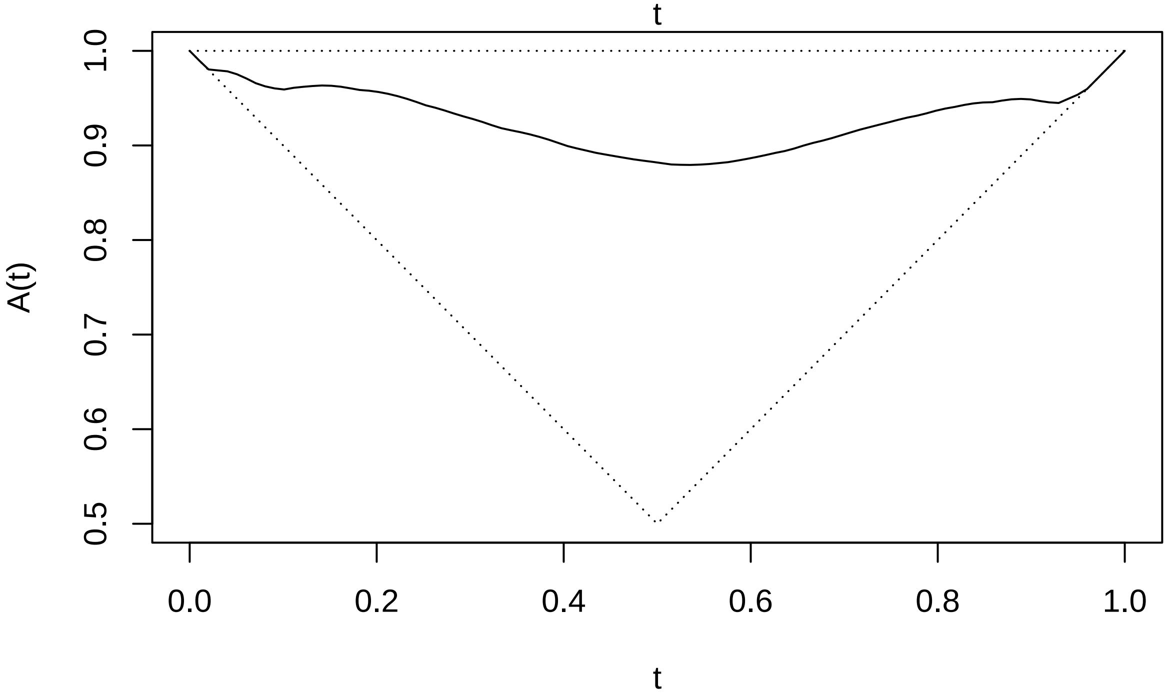
```r
#Parameters for the "hr" model:
(ml.ifm2$estimate)
```

```
##       dep
## 0.6219562
```

Non parametric plots

We can see that the parametric estimation of the marginals (when epmar is FALSE), gives us higher independance than when we use empircal transformation of the marginals.





Exercise 3.6

Using the "aneglog" model, the three probabilities are:

```
#a)
(prob1)
```

```
## [1] 0.04006492
```

```
#b)
(prob2)
```

```
## [1] 0.008952153
```

```
#c)
(prob3)
```

```
## [1] 0.8969128
```

And when we use the "hr" model we get

```
#a)
(prob1.sym)
```

```
## [1] 0.041262
```

```
#b)
(prob2.sym)
```

```
## [1] 0.00999359
```

```
#c)
(prob3.sym)
```

```
## [1] 0.8928402
```

These are close to the empirical estimates, which are:

```
#a)
(prob1.emp)
```

```
## [1] 0.03174603
```

```
#b)
(prob3.emp)
```

```
## [1] 0.8571429
```

```r
#c)
(prob3.emp)
```

```
## [1] 0.8571429
```

The second probability is 0 in the empirical case due to the fact that it is has a probability estimate of less than one in one hundred, which means that it is plausible that there is no such data point in our 63 data points.

Non-parametric

The non parametric estimates are also reasonable when compared with the empirical data:

```
#a)
(prob1.nonpar)
```

```
## [1] 0.04273503
```

```
#b)
(prob2.nonpar)
```

```
## [1] 0.01075644
```

```
#c)
(prob3.nonpar)
```

```
## [1] 0.8934205
```

Exercise 3.7

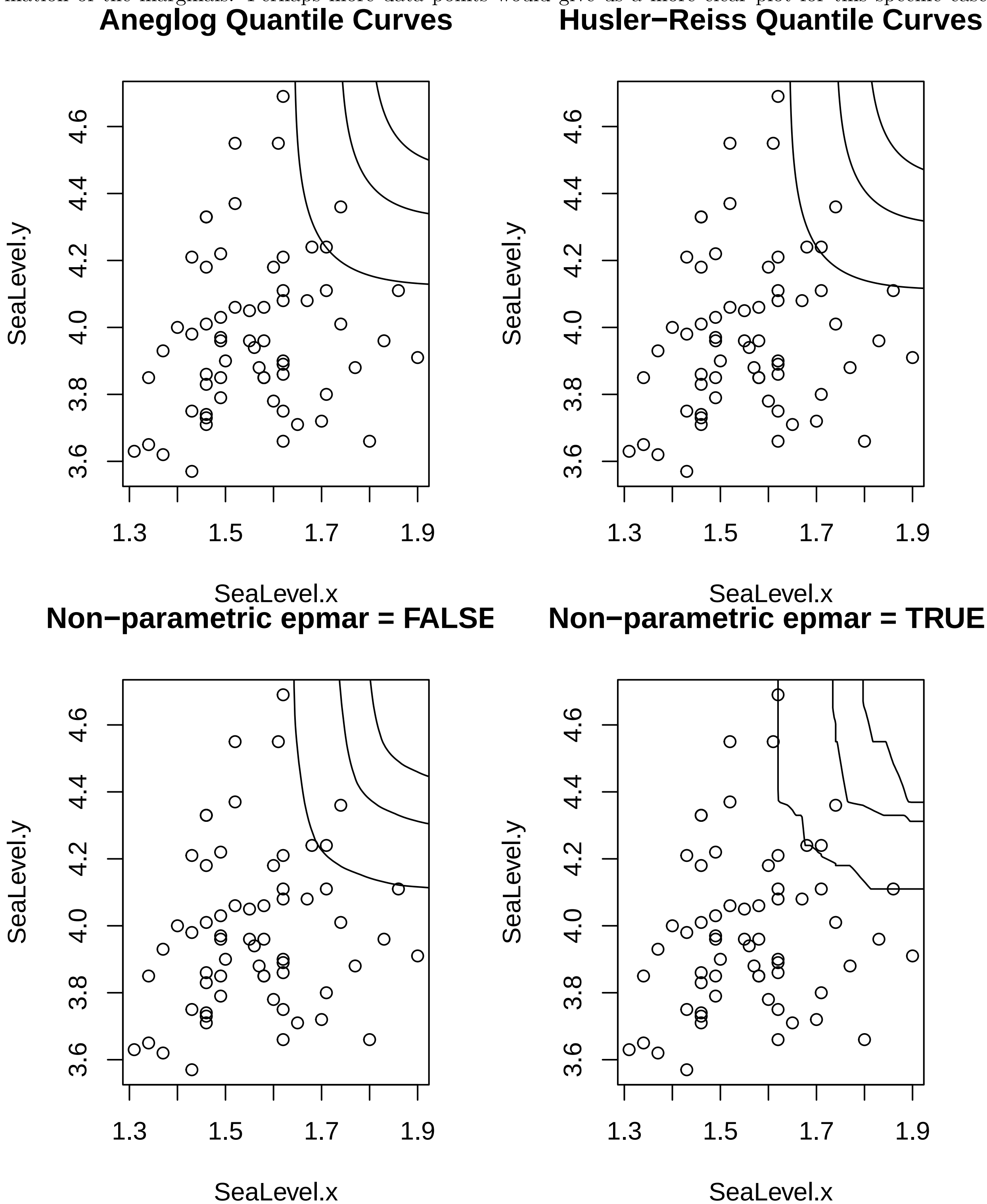
Here we need to use the relationship:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}. \tag{4}$$

In this case A is when we are below 1.95 and 4.8 and B is not being below 1.478 and 3.850 (Freemantle and Portpirie respectively). Using this, the probabilites are 0.9904152 for the non parametric and 0.9995249 for the parametric "hr" model

Exercise 3.8

We can see from the plot below that we have similar quantiles, except for when we use empirical transformation of the marginals. Perhaps more data points would give us a more clear plot for this specific case.



Exercise 3.9

When doing the statistical tests to see if SOI or YEAR plays a role, either seperately or together, we can see that only SOI shows a significant difference compared to the model without it:

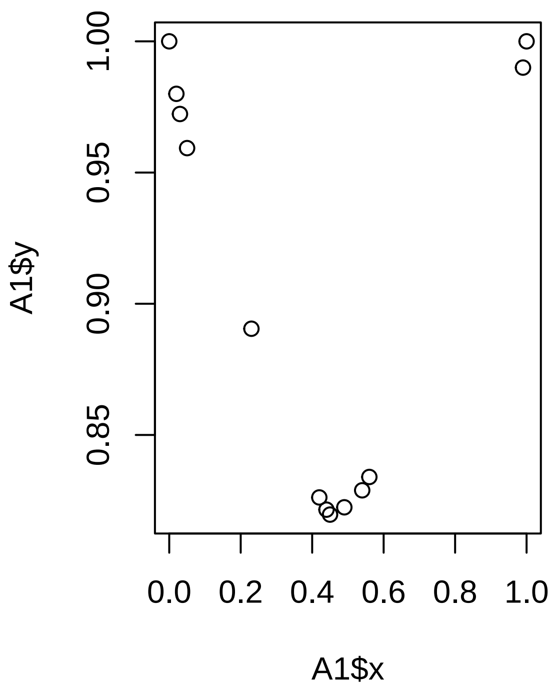
```
## [1] "The p value for including SOI in aneglog model is:"
## [2] "0.0071584418587456"
```

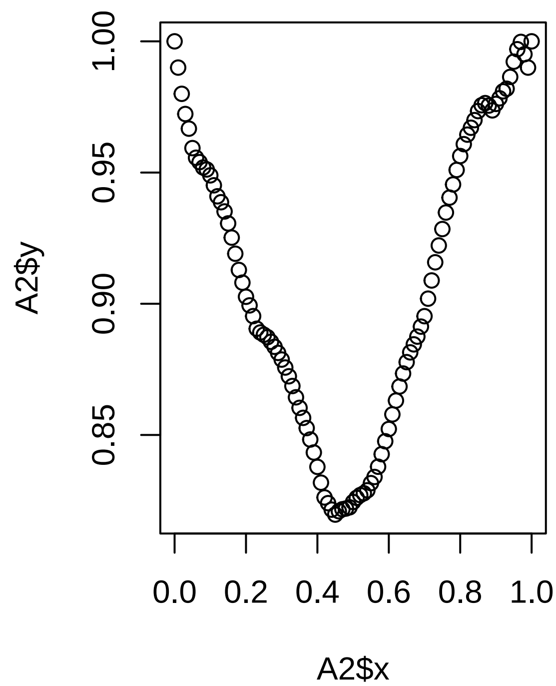
```
## [1] "The p value for including SOI in hr model is:"
## [2] "0.00594756545264922"
```

Exercise 5

Exercise 5.1

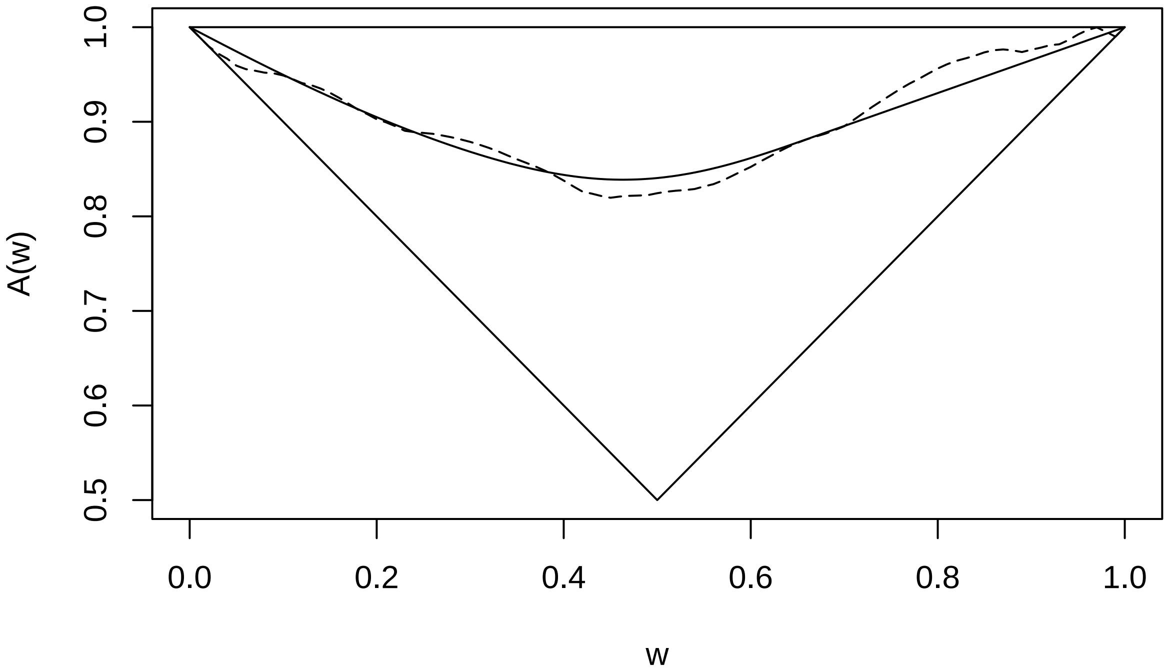
We see that we don't force convexity the result plot is not convex. How ever, enforcing convexity only leaves 13 points withing the hull.





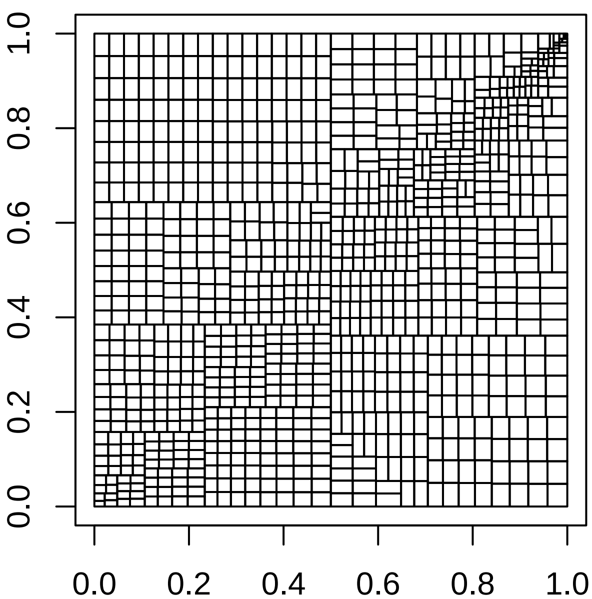
Exercise 5.2

Using the spline gives us a much nicer and smoother (and more convex) approximation of the dependence function. This can be seen in the plot below where the solid line is when spline fit is used.



Exercise 5.3

The plot of the copula is shown below. We can see some clear dependence in the diagonal.



Exercise 5.4

We first simulate the values, which gives us pairs of probabilites from the copula. We use these to obtain the values we want using the GEV quantile function. Then we can simply use these samples to estimate the probabilites. The results are shown below

```
## [1] "The probabilities are:"
```

```
## [1] "a)"     "0.0621"
```

```
## [1] "b)"     "0.0169"
```

```
## [1] "c)"     "0.8735"
```

```
## [1] "d)"      "0.98832"
```