Основы статистического анализа в «R» Оценка рыночных рисков с помощью обобщённого гиперболического распределения

ЦМФ. Количественная аналитика

Затабулированные распределения

Название	Обозначение в R	Параметры
Нормальное	norm	mean, sd
t-распределение	t	df
Равномерное	unif	min, max
Хи-квадрат	chisq	df
F-распределение	f	df1, df2
Гамма	gamma	shape, scale
…	…	…

```
# пример со стандартным нормальным распределением
N <- 100; x <- seq(-5,5,by=0.1); alpha <- 0.95
rnorm(n=N,mean=0,sd=1)            # генератор случайных чисел
qnorm(alpha,mean=0,sd=1)          # квантиль
pnorm(x,mean=0,sd=1)              # функция распределения
dnorm(x,mean=0,sd=1)              # функция плотности
```

Исходные данные

```
library(datasets)
dax <- EuStockMarkets[,"DAX"]

T <- length(dax) - 1
dax <- dax[2:(T+1)]/dax[1:T] - 1
```

Сравнение с затабулированным распределением

эмпирическая плотность распределения

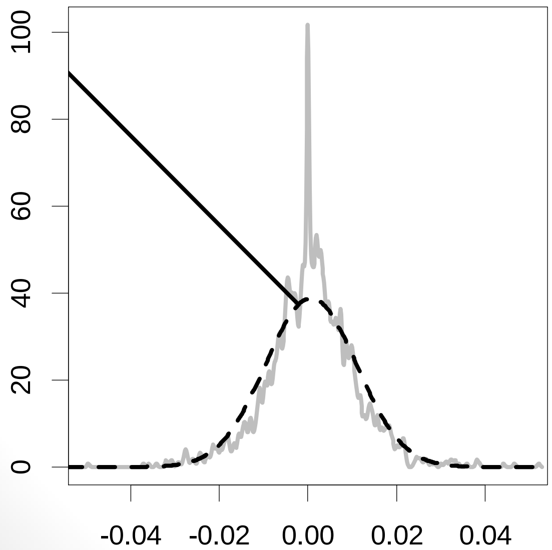
```
dens <- density(dax, bw = "ucv", n = 1024)
plot(dens$x, dens$y, type = "l", lwd = 5)
```

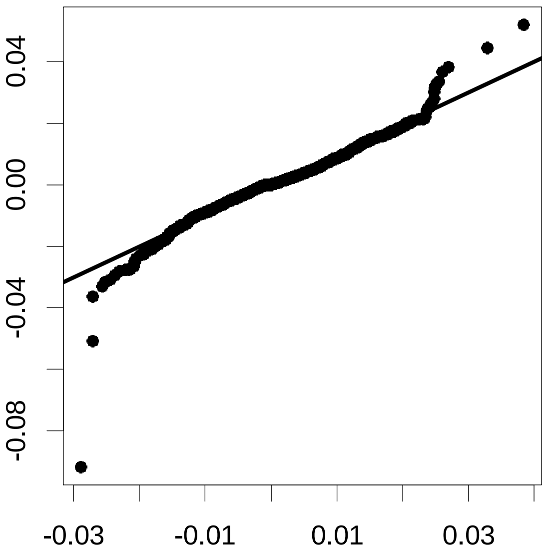
плотность нормального распределения

```
lines(dens$x, dnorm(dens$x, mean = mean(dax), sd = sd(dax)),
lty = "dashed", lwd = 5)
```

график квантиль–квантиль

```
qqplot(rnorm(n=10^3,mean=mean(dax),sd=sd(dax)),dax,pch=16)
abline(0,1,lwd=5)
```





Тесты на нормальность

\# Шапиро–Уилка

\# гипотеза: $H_0: x \sim N(\mu, \sigma)$

\# статистика: $W = \frac{\left(\sum a_i x_{(i)}\right)^2}{\sum (x_i - \bar{x})^2}$, $(a_1, \ldots, a_n) = \frac{m' V^{-1}}{(m' V^{-1} V^{-1} m)^{0.5}}$,

\# $m_i = E\left(x_{(i)} \middle| x \sim N(0,1)\right)$, $V = cov(m)$

```
shapiro.test(dax)
```

\# Колмогорова–Смирнова

\# гипотеза: $H_0: x \sim F(x)$

\# статистика: $D = \sup_x |y.cdf(x) - F(x)|$

```
ks.test(dax, "pnorm", mean = mean(dax), sd = sd(dax))
```

Домашнее задание

- скачать данные о дневных доходностях акции или биржевого индекса с сайтов finam.ru, finance.yahoo.com или др. за последние 2 года
- провести тесты на нормальность их распределения
- рассмотреть график «квантиль–квантиль» для эмпирического распределения доходностей и нормального распределения, сделать выводы о лёгкости или тяжести эмпирических хвостов
- написать комментарии

Обобщённое гиперболическое распределение (GHD)

$$f_{GHD}(x;\mu,\sigma,\gamma,\lambda,\chi,\psi) =$$

$$\frac{(\psi\chi^2)\psi^{\lambda}\left(\psi+\frac{\gamma^2}{\sigma^2}\right)^{\frac{1}{2}-\lambda}K_{\lambda-\frac{1}{2}}\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)e^{\frac{\gamma(x-\mu)}{\sigma^2}}}{(2\pi)^{\frac{1}{2}}\sigma K_{\lambda}(\psi\chi)^{\frac{1}{2}}\left(\sqrt{\left(\chi+\frac{(x-\mu)^2}{\sigma^2}\right)\left(\psi+\frac{\gamma^2}{\sigma^2}\right)}\right)^{\frac{1}{2}-\lambda}}$$

$K_u(v)$ — модифицированная функция Бесселя второго рода

Оценка параметров распределения

```
install.packages("ghyp")
library(ghyp)
fit.[...]uv(dax, symmetric = FALSE, silent = TRUE)
```

\# если symmetric == FALSE, то оценивается скошенное

\# распределение, иначе — симметричное;

\# вместо [...] следует подставить название распределения:

\#   ghyp — обобщённое гиперболическое

\#   hyp — гиперболическое

\#   NIG — нормально-обратное гауссовское

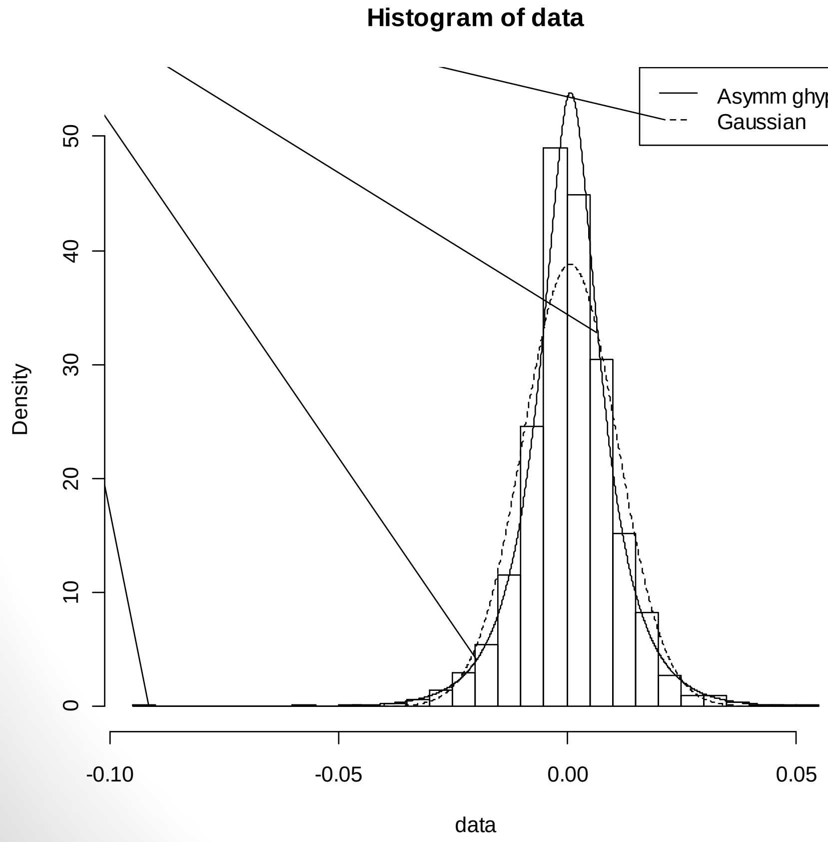
\#   VG — Variance-Gamma

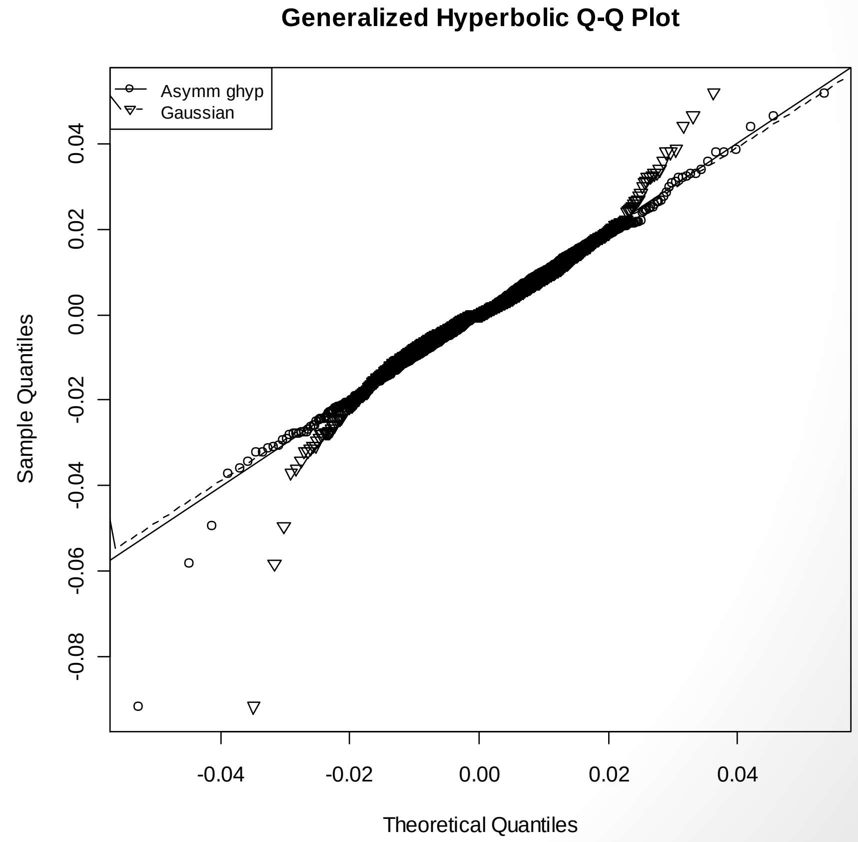
\#   t — t-распределение Стьюдента

\#   gauss — нормальное

Графический анализ модели

```
dax.ghyp <- fit.ghypuv(dax,symmetric=FALSE,silent=TRUE)
hist(dax.ghyp)   # гистограмма
qqghyp(dax.ghyp)   # график квантиль-квантиль
```





Выбор наилучшей модели

Отношение правдоподобия

H_0: более общая модель обладает той же объясняющей силой, что и её частный случай

$$LR = -2ln\frac{L_{H_0}}{L_{H_{alt}}} \sim \chi^2(\nu), \qquad \nu = df_{H_0} - df_{H_{alt}}$$

```
dax.t <- fit.tuv(dax,symmetric=FALSE,silent=TRUE)
lik.ratio.test(dax.ghyp,dax.t,conf.level=0.95)
```

Информационный критерий Акаике

$AIC = 2k - 2\ln(L) \rightarrow min,\ k$ — количество параметров модели

```
aic.uv <- stepAIC.ghyp(dax,dist=c("gauss","t","ghyp"),
symmetric=NULL,silent=TRUE)
summary(aic.uv$best.model)   # статистики по модели
```

Оценка финансового риска

Меры риска:

- граница потерь (Value-at-Risk)

$$P(x < VaR_{1-\alpha}) = 1 - \alpha$$

- ожидаемые потери (Expected Shortfall)

$$ES_{1-\alpha} = E(x|\, x < VaR_{1-\alpha})$$

Метод Монте-Карло

```
alpha <- 0.1; N <- 10^6
dax.sim <- rghyp(n=N,object=aic.uv$best.model)
dax.sim <- sort(dax.sim)
VaR <- dax.sim[alpha*N]
```

другой вариант: VaR <- qghyp(alpha,object=aic.uv$best.model)

```
ES <- mean(dax.sim[1:(alpha*N-1)])
```

VaR	-0.011
ES	-0.018

Кривая VaR

Используется для тестирования качества оценок риска

Кривая VaR — набор последовательных во времени значений VaR

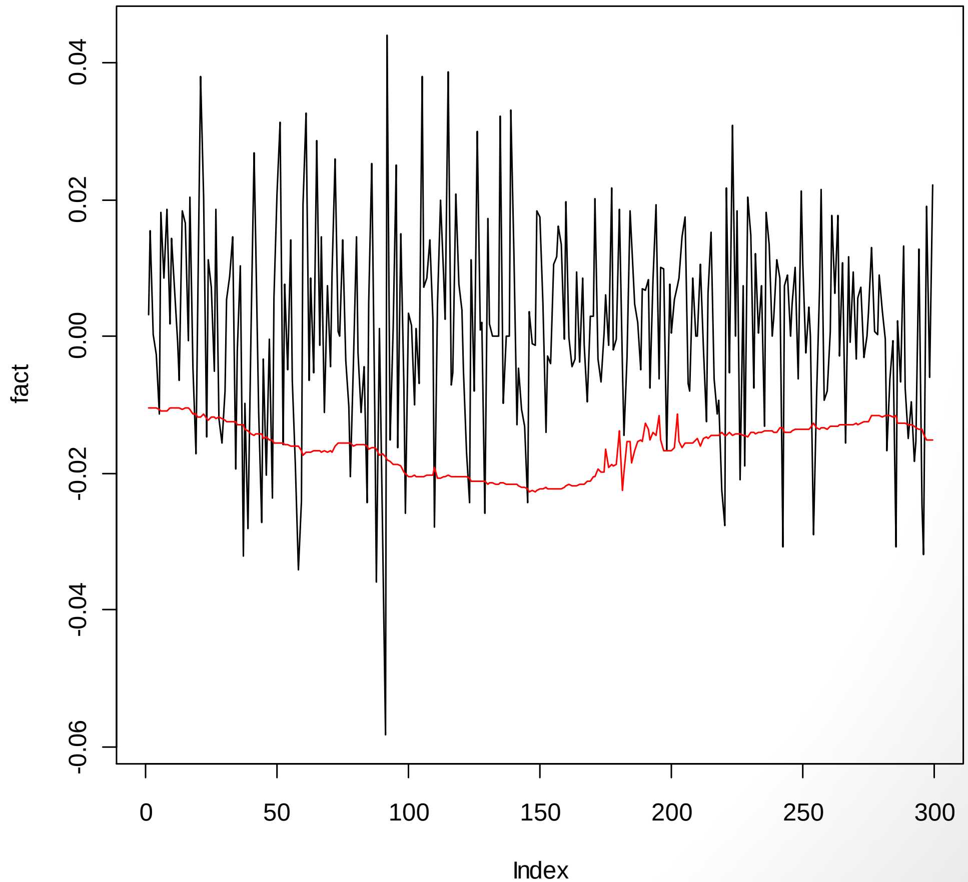
```
# разделим выборку на обучающую и экзаменующую
T1 <- 6*260; T2 <- T - T1

# на пространстве экзаменующей выборки построим набор
# последовательных значений VaR
VaR <- numeric()
h <- 0.5 * 260  # длина обучающей выборки
for (i in (T1+1):(T1+T2)) {
  h.dax <- dax[(i-h):(i-1)]
  dax.fit <- stepAIC.ghyp(h.dax,dist=c("gauss","t","ghyp"),
  symmetric=NULL,silent=TRUE)
  VaR[i-T1] <- qghyp(alpha,object=dax.fit$best.model)
}
```

Кривая VaR

\# сравнение оценок риска с фактом

```
fact <- dax[(T1+1):(T1+T2)]
plot(fact,type="l")
lines(VaR,col="red")
```



Кривая VaR

Тест Купика

Идея состоит в сравнении модельной и эмпирической частот превышений фактическими убытками границы VaR

$$K = \sum I(x_t < VaR_t), \;\; \alpha_0 = \frac{K}{T_2}$$

$$H_0: \;\; \alpha_0 = \alpha$$

Статистика:

$$S = -2\ln\left((1-\alpha)^{T_2-K}\alpha^K\right) + 2\ln\left((1-\alpha_0)^{T_2-K}\alpha_0^K\right) \sim \chi^2(1)$$

тест Купика в R:

```
K <- sum(fact<VaR); alpha0 <- K/T2
S <- -2*log((1-alpha)^(T2-K)*alpha^K)+
2*log((1-alpha0)^(T2-K)*alpha0^K)
p.value <- 1-pchisq(S,df=1)
```

alpha	0.100
alpha0	0.130
p.value	0.092

Кривая VaR

Функции потерь

Величина функции потерь измеряет глубину пробоев кривой VaR и интерпретируется как размер понесённых потерь

Функция потерь Лопеса:

$$L_{Lo} = \frac{1}{K}\sum_{t=T_1+1}^{T_2}\left((x_t - VaR_t)^2 \cdot I(x_t < VaR_t)\right)$$

Функция потерь Бланко-Ила:

$$L_{BI} = \frac{1}{K}\sum_{t=T_1+1}^{T_2}\left(\frac{x_t - VaR_t}{VaR_t} \cdot I(x_t < VaR_t)\right)$$

\# функции потерь в R:

```
L.Lo <- sum((fact-VaR)^2*(fact<VaR))/K
L.BI <- sum((fact-VaR)/VaR*(fact<VaR))/K
```

L.Lo*10^4	1.399
L.BI	0.611

Домашнее задание

- рассчитать оценки риска для финансового актива по всей совокупности наблюдений на основе наилучшей модели
- построить кривую VaR на основе ОГР и проверить качество оценок

Исходные данные — дневные котировки акций и биржевых индексов за период с 2014 г. по н.в. с сайтов finam.ru, finance.yahoo.com