Новиков Лев Ильич
Домашнее задание

1.

В качестве данных я взял с сайта finam.ru 200 наблюдений по ценам обыкновенных акций Сбербанка в период с 09 июня 2018 года по 22 марта 2019 года.

Нужно было оценить, какая функция распределения подходит для этих данных, определить наиболее вероятный вид распределения, и оценить его параметры.

Построим ядерную оценку плотности и посмотрим на ее график:

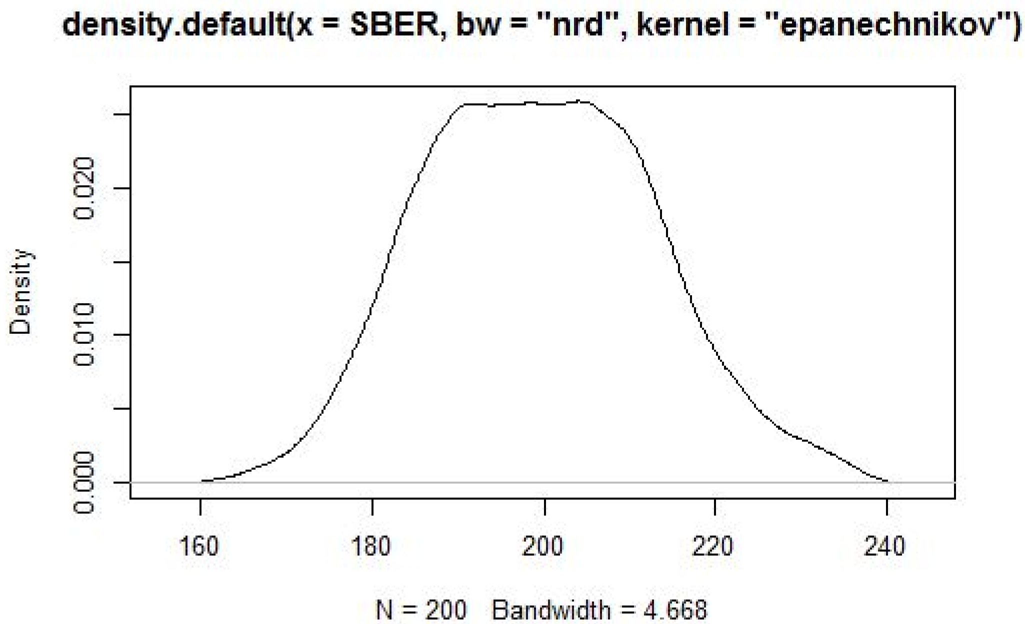

Рис. 1: Ядерная оценка плотности распределения

Видно, что правый хвост нашего распределения чуть более тяжелый, поэтому подозреваем логнормальное распределение. С помощью пакета fitdistrplus подбираем оптимальные параметры для этого логнормального распределения:

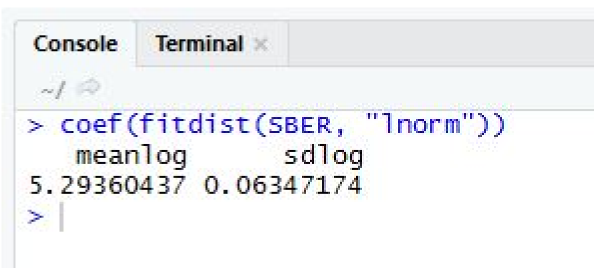

Рис. 2: Подбор оптимальных параметров логнормального распределения

Итак, оптимальные параметры $\mu^* = 5.29360437$ и $\sigma^* = 0.06347174$.

Проверим, насколько сильно отличается функция распределения логнормального распределения с такими параметрами от эмпирической функции распределения (проведём тест Колмогорова-Смирнова):

```
> ks.test(SBER, 'plnorm', meanlog = 5.29360437, sdlog = 0.06347174)

        One-sample Kolmogorov-Smirnov test

data:  SBER
D = 0.059451, p-value = 0.4795
alternative hypothesis: two-sided
```

Рис. 3: Тест Колмогорова-Смирнова

Как мы можем видеть, значение статистики Колмогорова-Смирнова невелико ($D \approx 0.059$), p-значение больше 0.1, поэтому на любом разумном уровне значимости мы не отвергаем нулевую гипотезу о том, что наши данные подчиняются логнормальному закону.

Графики эмпирических и теоретических плотностей и функций только подтверждают нашу гипотезу: всё практически в точности совпадает.

```
> plot(fitdist(SBER, "lnorm"))
```

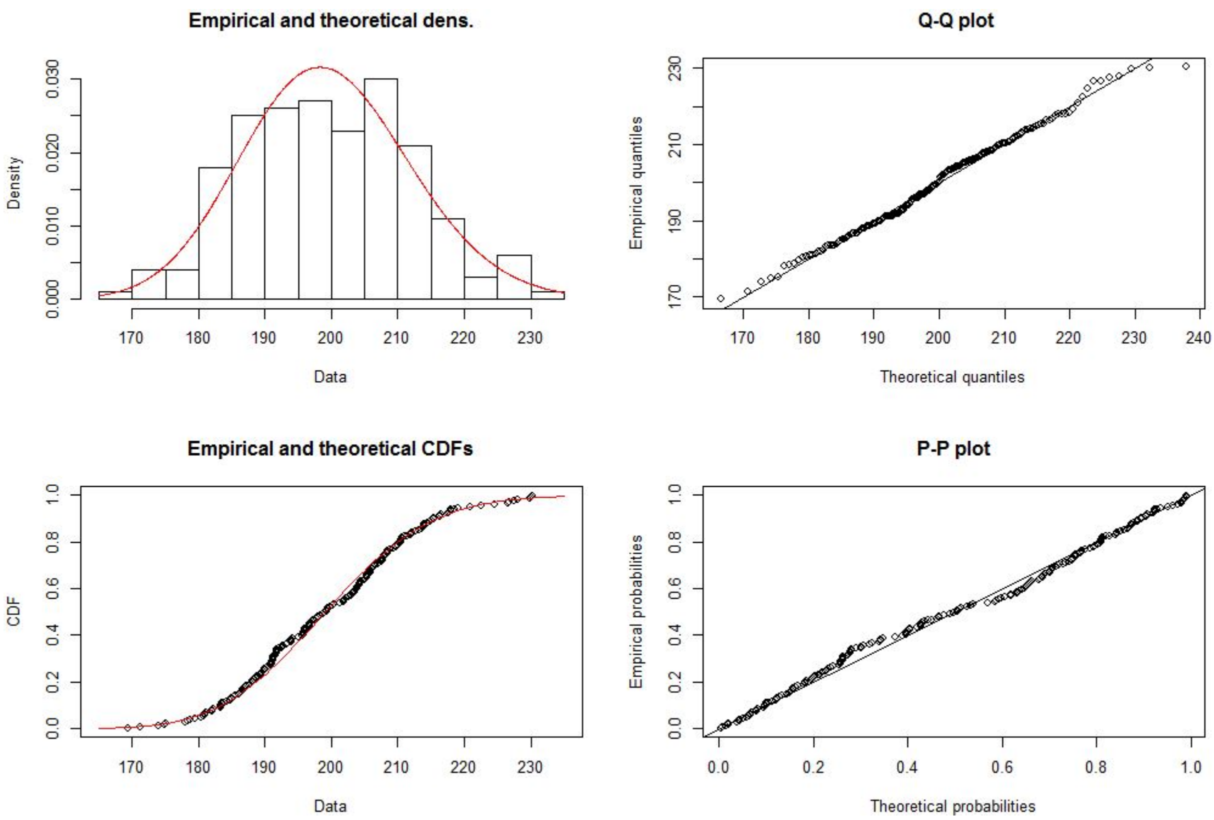

Рис. 4: Эмпирические и теоретические плотности и функции распр-ия; QQ график

2.

Нужно выяснить, в области притяжения какого закона находится наше логнормальное распределение. Начнем доказательство того, что логнормальное распределение лежит в области притяжения закона Гумбеля $\Lambda(x)$ с определений:

Определение 1 (Область притяжения максимума, MDA)**.**
Пусть $M_n = \max(X_1, X_2, \dots X_n)$ для некоторой случайной величины X.
Говорят, что случайная величина X и её функция распределения $F(x)$ принадлежат области притяжения (MDA) для экстремального распределения H, если $\exists c_n > 0,\ d_n \in \mathbb{R}$ такие, что $\dfrac{1}{c_n}(M_n - d_n) \xrightarrow{d} H$.
В дальнейшем будем это обозначать как $X \in MDA(H)$ или $F \in MDA(H)$, в зависимости от того, о чем идет речь — случайной величине или законе распределения.

Следствие 1 (Описание $MDA(H)$)**.**
Функция распределения F принадлежит $MDA(H)$ с нормирующими константами $c_n > 0,\ d_n \in \mathbb{R}$ тогда и только тогда, когда $\lim\limits_{n\to\infty} n\,\overline{F}(c_n x + d_n) = -\ln H(x),\ x \in \mathbb{R}$.
(для $H(x) = 0$ предел следует понимать как $+\infty$).
Данный факт следует напрямую из эквивалентности определений:

$$\frac{1}{c_n}(M_n - d_n) \xrightarrow{d} H \ \Leftrightarrow\ \lim_{n\to\infty} P(M_n \leqslant c_n x + d_n) = \lim_{n\to\infty} F^n(c_n x + d_n) = H(x)$$

Определение 2 (Функция фон Мизеса)**.**
Пусть F — функция распределения, являющаяся непрерывной на своём правом конце x_F. Предположим, что существует число $z < x_F$ такое, что:

$$\overline{F}(x) = c \cdot \exp\left(-\int_z^x \frac{1}{a(t)}dt\right), \quad z < x < x_F$$

(здесь $\overline{F}(x) = 1 - F(x)$ — функция дожития, $c > 0$ — константа, $a(t)$ — положительная, абсолютно непрерывная функция с плотностью $a'(t)$, для которой выполнено $\lim\limits_{x\to x_F} a'(x) = 0$.) Тогда функция F называется функцией фон Мизеса, а ф-ия a — вспомогательной функцией для F.

Докажем две следующих леммы:

Лемма 1 (Дифференцируемость на правом конце)**.** *Пусть F — функция распределения с правым концом x_F и допустим что $\exists z < x_F$ такое, что F дважды дифференцируема на всем интервале (z, x_F) с положительной плотностью $f = F'$ и $F''(x) < 0, \quad \forall x \in (z, x_F)$. Тогда F является функцией фон Мизеса с вспомогательной функцией $a = \dfrac{\overline{F}}{f}$ тогда и только тогда, когда $\lim\limits_{x\to x_F} \dfrac{\overline{F}(x)\cdot F''(x)}{f^2(x)} = -1$.*

Доказательство:

□
Для $x \in (z, x_F)$ положим $B(x) = -\ln \overline{F}(x)$ и $a(x) = \frac{1}{B'(x)}$.
Тогда $a(x) = \dfrac{\overline{F}(x)}{f(x)} > 0$ и $a'(x) = -\dfrac{\overline{F}(x)\cdot F''(x)}{f^2(x)} - 1$.
То есть утверждение леммы 1 эквивалентно определению $a'(x) \xrightarrow{x\to x_F} 0$. ■

Лемма 2 (Функции фон Мизеса и $MDA(\Lambda)$)**.** *Предположим, что функция распределения F является функцией фон Мизеса. Тогда распределение $F \in MDA(\Lambda)$, а нормирующие константы* $d_n = \inf\limits_x \left(F(x) \geqslant 1 - \dfrac{1}{n} \right)$ *и* $c_n = a(d_n)$.

Доказательство:

□
Из определения функции фон Мизеса следует, что

$$\frac{\overline{F}(x+ta(x))}{\overline{F}(x)} = \exp\left(-\int_x^{x+ta(x)} \frac{1}{a(u)}\,du \right) \quad \text{для любых } x, \text{ достаточно близких к } x_F,$$

$$\text{тогда имеем} \quad \frac{\overline{F}(x+ta(x))}{\overline{F}(x)} = \exp\left(-\int_0^t \frac{a(x)}{a(x+va(x))}\,dv \right) \quad \left(\text{здесь } v = \frac{u-x}{a(x)} \right).$$

Поскольку $a'(x) \overset{x\to x_F}{\longrightarrow} 0$, для $\varepsilon > 0$ и $x \geqslant x_0(\varepsilon)$, имеем оценку сверху на интеграл:

$$|a\left(x+va(x)\right) - a(x)| = \left| \int_x^{x+va(x)} a'(s)\,ds \right| \leqslant \varepsilon \cdot |v| \cdot a(x) \leqslant \varepsilon \cdot |t| \cdot a(x).$$

Отсюда при том же предположении $x \geqslant x_0(\varepsilon)$ следует, что $\left| \dfrac{a\left(x+va(x)\right)}{a(x)} - 1 \right| \leqslant \varepsilon |t|$.

Получаем $\lim\limits_{x\to x_F} \left| \dfrac{a\left(x+va(x)\right)}{a(x)} - 1 \right| = \lim\limits_{x\to x_F} \varepsilon |t| \iff \lim\limits_{x\to x_F} \left| \dfrac{a\left(x+va(x)\right)}{a(x)} - 1 \right| = 0$
(так как правая часть может быть сделана сколь угодно маленькой)
$\iff \lim\limits_{x\to x_F} \dfrac{a\left(x+va(x)\right)}{a(x)} = 1 = \lim\limits_{x\to x_F} \dfrac{a(x)}{a\left(x+va(x)\right)}$ (равномерная сходимость).

Вспоминаем, что $\dfrac{\overline{F}(x+ta(x))}{\overline{F}(x)} = \exp\left(-\displaystyle\int_0^t \frac{a(x)}{a(x+va(x))}dv \right)$ и наконец получаем

$\lim\limits_{x\to x_F} \dfrac{\overline{F}(x+ta(x))}{\overline{F}(x)} = e^{-t}$. Получаем, что $\lim\limits_{n\to\infty} n\,\overline{F}(d_n + tc_n) = e^{-t} = -\ln \Lambda(t), \ \forall t \in \mathbb{R}$.
Из следствия 1 наконец получаем, что $F \in MDA(\Lambda)$. ■

Лемма 3. *Обозначим для удобства за Φ и φ функцию распределения и плотность стандартного нормального распределения.*
Утверждается, что Φ является функцией фон Мизеса и что $\Phi(x) \in MDA(\Lambda)$.

Доказательство:

□
Для того, чтобы доказать, что Φ является функцией фон Мизеса, воспользуемся леммой 1: для этого отметим, что $\lim\limits_{x\to\infty} \dfrac{\overline{\Phi}(x)}{\frac{1}{x}\varphi(x)}$ равен $\lim\limits_{x\to\infty} \dfrac{-\varphi(x)}{-\frac{1}{x^2}\varphi(x) + \frac{1}{x}\varphi'(x)}$ (применяем правило Лопиталя). $\lim\limits_{x\to\infty} \dfrac{-\varphi(x)}{-\frac{1}{x^2}\varphi(x) + \frac{1}{x}\varphi'(x)} = \lim\limits_{x\to\infty} \dfrac{\varphi(x)}{\varphi(x)\cdot\left(\frac{1}{x^2} + \frac{x}{x}\right)}$.

Получаем, что $\overline{\Phi}(x) \sim \frac{1}{x}\varphi(x)$ (это отношение Милла, то есть частное этих двух функций сходится к 1 при $x \to \infty$).

Поскольку $\varphi'(x) = -x\varphi(x)$, $\varphi'(x) < 0$ и мы получаем $\lim\limits_{x\to\infty} \dfrac{\overline{\Phi}(x)\cdot\varphi'(x)}{\varphi^2(x)} = -1$, что, согласно леммам 1 и 2, означает, что $\Phi \in MDA(\Lambda)$. ■

В нашей задаче мы получили не нормальное распределение, а логнормальное; покажем, что и оно лежит в $MDA(\Lambda)$:

Пусть X — стандартное нормальное распределение, а $f(x) = e^{\mu+\sigma x}$ ($\mu \in \mathbb{R}$, $\sigma > 0$). Тогда $Y = f(X) = e^{\mu+\sigma X}$ определяет логнормальное распределение с выбранными средним и дисперсией.

Поскольку $X \in MDA(\Lambda)$, имеем $\lim\limits_{n\to\infty} P\left(M_{Yn} \leqslant e^{\mu+\sigma(c_n x+d_n)}\right) = \Lambda(x)$, $x \in \mathbb{R}$, где $M_{Yn} = \max(Y_1, Y_2, \ldots, Y_n)$, c_n и d_n — нормирующие коэффициенты для стандартного нормального распределения.

Отсюда $\lim\limits_{n\to\infty} P\left(e^{-\mu-\sigma d_n} \cdot M_{Yn} \leqslant 1 + \sigma c_n x + o(c_n)\right) = \Lambda(x)$, $x \in \mathbb{R}$. Поскольку $c_n \to 0$ (для $N(0,1)$, $c_n = a(d_n) \sim \frac{1}{\sqrt{2\ln n}}$), получаем $\dfrac{e^{-\mu-\sigma d_n}}{\sigma c_n} \cdot \left(M_{Yn} - e^{\mu+\sigma d_n}\right) \xrightarrow{d} \Lambda$, то есть логнормальное распределение $Y \in MDA(\Lambda)$.

Итого в моей задаче при рассмотрении цены обыкновенных акций Сбербанка получилось, что данные распределены по логнормальному закону, с законом притяжения Гумбеля.