

Сравнение нормального и логнормального распределений

Shapoval Egor feat Vladimir Popov (+neg)

03.10.2021

Нормальное распределение: плотность, математическое ожидание и дисперсия

$\xi \sim N(a, \sigma^2)$, Плотность $p_\xi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}$.

Математическое ожидание $E\xi = a$.

Дисперсия $D\xi = \sigma^2$.

Медиана, мода

Нормальное распределение симметрично, поэтому медиана $\text{med}\xi = E\xi = a$.

Экстремум функции плотности достигается в точке, где $e^{-\frac{(x-a)^2}{2\sigma^2}} = 1 \iff x = a$, соответственно нормальное распределение унимодальное.

Ассиметрия, эксцесс

Коэффициент асимметрии $A\xi = \frac{E(\xi-a)^3}{\sigma^3} = 0$, т.е. распределение симметричное.

Эксцесс (коэффициент крутизны, характеризующий скорость убывания плотности на хвостах, kurtosis)

$K = \frac{E(\xi-a)^4}{\sigma^4} - 3 = 0$.

Аддитивность

$\xi_1 \sim N(a_1, \sigma_1^2)$, $\xi_2 \sim N(a_2, \sigma_2^2)$, тогда $\xi_1 + \xi_2 \sim N(a_1 + a_2, \sigma_1^2 + \sigma_2^2)$.

Центральная предельная теорема

$\{\xi_n\}_{n=1}^\infty$ — последовательность независимых, одинаково распределённых случайных величин с $E\xi_i = a$ и $D\xi_i = \sigma^2$, тогда $\sqrt{n} \cdot \frac{\sum_{i=1}^n \xi_i/n - a}{\sigma} \Rightarrow N(0, 1)$.

Примеры из жизни

Рост человека, отклонение при стрельбе по мишени, ошибка в измерениях имеют нормальное распределение (рост взрослого человека хорошо аппроксимируется нормальным распределением, но он не может иметь в точности нормальное распределение, так как у него область значений включает отрицательные числа).

Логнормальное распределение: плотность, математическое ожидание и дисперсия

Пусть $\xi \sim \text{LogN}(\mu, \sigma^2)$, тогда плотность ξ имеет вид:

$$p_{\xi}(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp(-(\ln x - \mu)^2/2\sigma^2).$$

$$E\xi = e^{\mu+\sigma^2/2}, D\xi = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2}, \text{med}\xi = e^{\mu}.$$

Как видим, математическое ожидание зависит от σ , а значит, что если два нормальных распределения отличались только дисперсиями, то соответствующие логнормальные будут отличаться и по мат. ожиданию. Медиана, тем не менее, не зависит от σ .

Утверждение

Пусть $\zeta \sim N(a, \sigma^2)$. Тогда $\xi = e^{\zeta} \sim \text{LogN}(a, \sigma^2)$.

И наоборот, если $\xi \sim \text{LogN}(a, \sigma^2)$, то $\zeta = \ln \xi \sim N(a, \sigma^2)$.

Асимметрия, эксцесс

$$A\xi = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}, K\xi = e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 6.$$

Распределение, очевидно, асимметрично (носитель $(0, +\infty)$) и имеет хвост справа (асимметрия положительна). Крутизна тоже положительная, т.е. плотность “круче”, чем у нормального распределения.

Мультипликативность

Если $\xi_1, \dots, \xi_n$ — независимые логнормальные случайные величины, такие что $\xi_i \sim \text{LogN}(\mu, \sigma_i^2)$, то их произведение также логнормально: $\eta = \prod_{i=1}^n \xi_i \sim \text{LogN}\left(n\mu, \sum_{i=1}^n \sigma_i^2\right)$.

Примеры из жизни

Логнормальное распределение описывает многие естественные явления, например: продолжительность шахматной партии, число госпитализаций в случае эпидемии, численность населения городов, уровень дохода населения и многое другое.

Нормальное vs Логнормальное

Нормальное	Логнормальное
Аддитивно (сложение — естественная операция)	Мультипликативно (умножение — естественная операция)
Симметрично	Асимметрично (“хвост” значений вправо)
Мат. ожидание не зависит от σ	Мат. ожидание зависит от σ
Его экспонента имеет логнормальное распределение	Логарифм имеет нормальное распределение
Эксцесс равен нулю	Эксцесс не равен нулю (положителен)

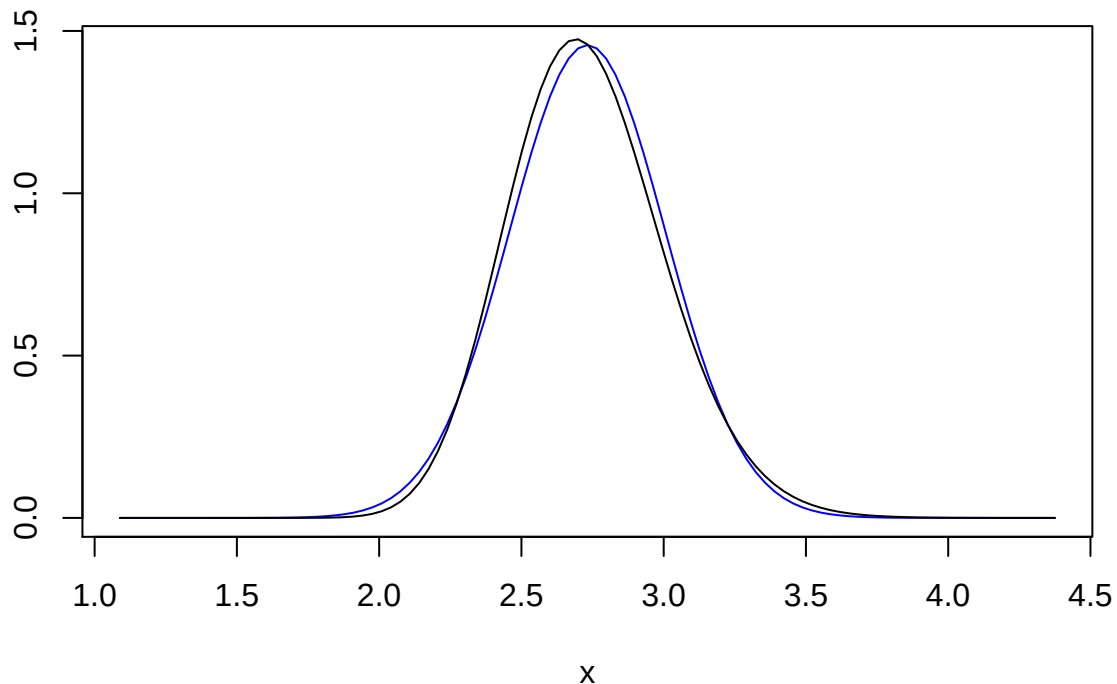
Графики плотностей нормального и логнормального распределения

Выбрать нормальное и логнормальное распределение с одинаковыми матем. ожиданием и дисперсией несложно, так как мы знаем связь этих характеристик у распределений.

Можно ли на глаз отличить плотность нормального распределения от плотности логнормального, зависит, в первую очередь, от коэффициентов асимметрии и эксцесса логнормального распределения. Чем ближе они к нулю, тем более похожи плотности (при одинаковых мат.ож. и дисперсии). Посмотрев на формулы для логнормального распределения, можно увидеть, что асимметрия и эксцесс равны нулю, только если параметр σ^2 близок к нулю. На этом и основаны след. примеры.

Одинаковые математические ожидания, одинаковые дисперсии, похожие плотности.

```
a <- 1
sigma <- 0.1
a.log <- exp(a+sigma^2/2) #the same mean as for LNorm
sigma.log <- sqrt((exp(sigma^2)-1))*a.log #the same standard as for LNorm
f2 <- function(x)dlnorm(x,a,sigma)
f1 <- function(x)dnorm(x,a.log,sigma.log)
k <- 6
curve(f1,xlim=c(a.log-k*sigma.log,a.log + k*sigma.log),col='blue',ylab='')
curve(f2,add=TRUE)
```



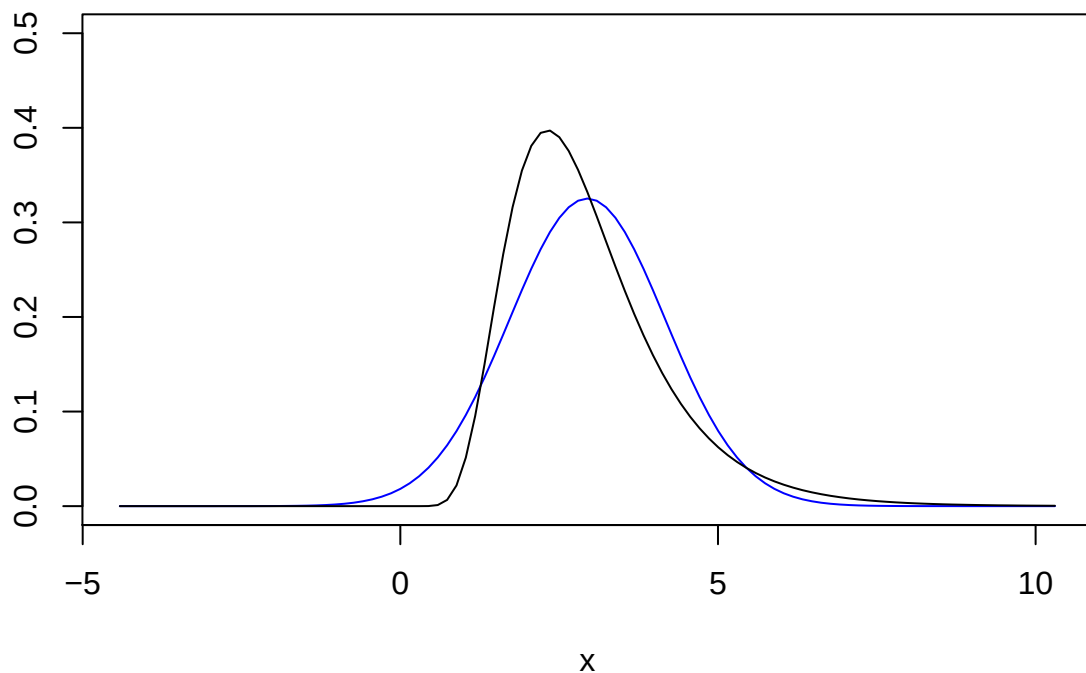
Одинаковые математические ожидания, одинаковые дисперсии, сильно различающиеся плотности.

```
a<-1
sigma<-0.4
a.log <- exp(a+sigma^2/2)
sigma.log <- sqrt((exp(sigma^2)-1))*a.log
f2<-function(x)dlnorm(x,a,sigma)
f1<-function(x)dnorm(x,a.log,sigma.log)
```

```

k <- 6
curve(f1,xlim=c(a.log-k*sigma.log,a.log + k*sigma.log),col='blue',
      ylab='', ylim = c(0,0.5))
curve(f2,add=TRUE)

```



Моделирование нормального распределения

Промоделируем выборку объёмом $n = 1000$ из нормального распределения $N(a, \sigma^2)$ со средним $a = 0$ и $\sigma = 1$

```

n <- 10000
a <- 0
sigma <- 1
Sample <- rnorm(n, mean = a, sd = sigma)

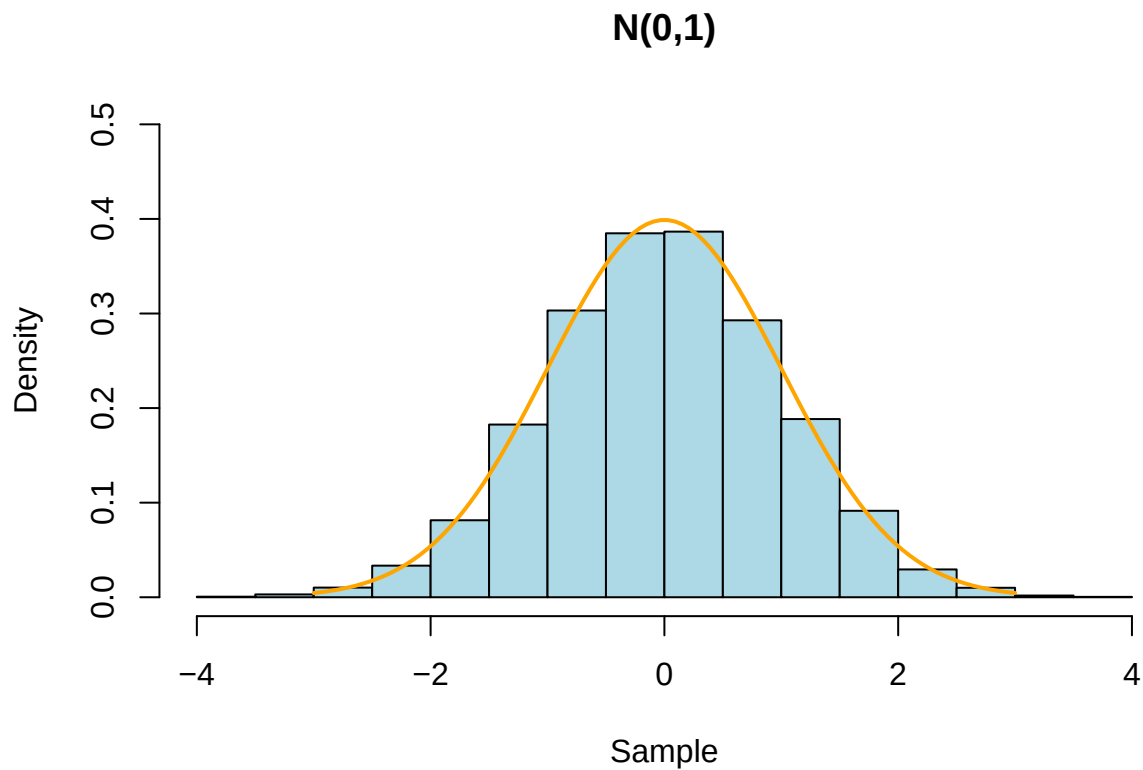
```

Построим гистограмму на фоне плотности

```

hist(Sample, main = "N(0,1)", ylab = "Density", col = "light blue", freq = FALSE, ylim = c(0, 0.5))
x <- seq(-3, 3, length.out = 100) #строим последовательность из 100 чисел,
                                   #заполняющих промежутки от -3 до 3 сигм
y <- dnorm(x, a, sigma) #подставляем x в функцию плотности и получаем вектор
lines(x, y, lwd = 2, col = "orange") #строим кривую поверх гистограммы

```



Найдём выборочное среднее, медиану и дисперсию и сравним их с генеральными характеристиками, которые должны быть равны 0, 0 и 1, соответственно.

```
mean(Sample)
```

```
## [1] 0.001384408
```

```
median(Sample)
```

```
## [1] 0.0007553686
```

```
var(Sample)
```

```
## [1] 0.9954965
```

Моделирование логнормального распределения

Промоделируем выборку объёмом $n = 1000$ из логнормального распределения $\log N(a, \sigma^2)$ с параметрами $a = 0$ и $\sigma = 1$. В этом случае должны получиться

```
n <- 10000
a <- 0
sigma <- 0.2
a.log <- exp(a+sigma^2/2)
med.log <- exp(a)
sigma.log <- sqrt((exp(sigma^2)-1))*a.log
print(c(a.log, med.log, sigma.log))
```

```
## [1] 1.0202013 1.0000000 0.2060978
```

```
Sample <- rlnorm(n, a, sigma)
mean(Sample)
```

```
## [1] 1.016291
```

```
median(Sample)
```

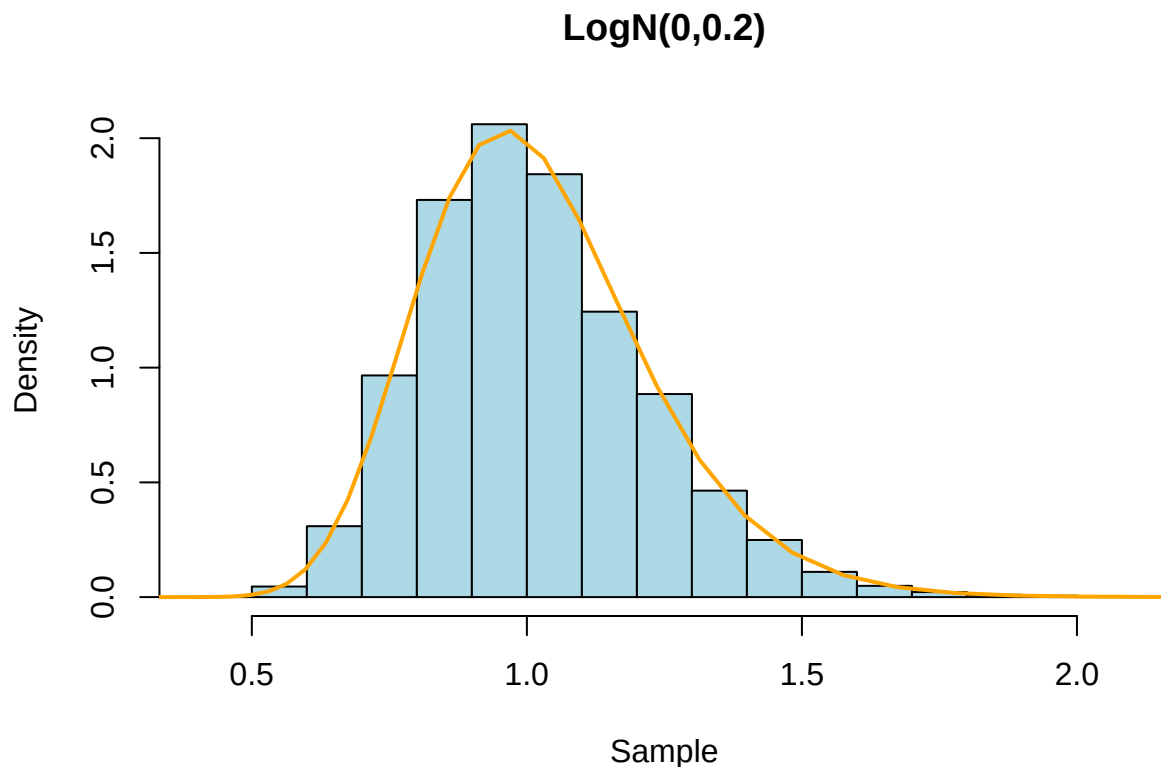
```
## [1] 0.994
```

```
var(Sample)
```

```
## [1] 0.04175021
```

Построим гистограмму на фоне плотности

```
hist(Sample, main = "LogN(0,0.2)", ylab = "Density", col = "light blue", freq = FALSE)
x <- exp(seq(-3, 3, length.out = 100))
y <- dlnorm(x, a, sigma) #подставляем x в функцию плотности и получаем вектор
lines(x, y, lwd = 2, col = "orange") #строим кривую поверх гистограммы
```



Проверим свойство, что логарифм логнормально распределенной случайной величины имеет нормальное распределение. Прологарифмируем выборку, стандартизуем ее и сравним с $N(0,1)$

```
logSample <- scale(log(Sample))
hist(logSample, main = "Standardized log of LogN(0,0.2)",
     ylab = "Density", col = "light blue", freq = FALSE)
x <- seq(-3, 3, length.out = 100)
y <- dnorm(x, 0, 1) #подставляем x в функцию плотности и получаем вектор
lines(x, y, lwd = 2, col = "orange") #строим кривую поверх гистограммы
```

