
 This repository Search Pull requests Issues Gist

NikitaIT / org.stepik.math.statistics Unwatch 1 Star 0 Fork 0

[Code](#) [Issues 0](#) [Pull requests 0](#) [Projects 0](#) [Wiki](#) [Pulse](#) [Graphs](#) [Settings](#)

Branch: master org.stepik.math.statistics / IDZ_1 / IDZ_1.md Find file Copy path

 NikitaIT Update IDZ_1.md 9ee09ab 13 minutes ago 1 contributor

207 lines (151 sloc) 9.03 KB Raw Blame History

Студент: Федоров Н.С.
Группа: 5362
Преподаватель: Медведев А.Н.

Репозиторий Статистика

Из этой первой лекции по теории вероятностей я запомнил только полужнакомый термин «математическое ожидание». Незнакомец употреблял этот термин неоднократно, и каждый раз я представлял себе большое помещение, вроде зала ожидания, с кафельным полом, где сидят люди с портфелями и бьюрами и, подбрасывая время от времени к потолку монетки и бутерброды, сосредоточенно чего-то ожидают. До сих пор я часто вижу это во сне. Но тут незнакомец оглушил меня звонким термином «предельная теорема Муавра — Лапласа» и сказал, что всё это к делу не относится.

Аркадий и Борис Стругацкие, Стажёры

Отчеты по ИДЗ

- > 1. Нормальное, Гамма и N-Биомиальное распределения {отчет}
- > 2. {отчет}
- > 3. {отчет}
- > 4. {отчет}
- > 5. {отчет}

ИДЗ №1

- -> Код {file}
- -> Данные {file}
- -> Задание {file}

Данные

Даны следующие распределения и параметры:

Тип Распределения	Параметры
Гамма	$p=10, b=5.3$
Нормальное	$a=2.3, \sigma=0.3$
Отрицательное Биномиальное	$m=32, p=1/5$

Распределение в файле: [file](#)

Выполнение работы

Необходимо для каждого распределения:

- а. сгенерировать выборку длины 1000 из данного распределения (см. стр. 19 методички)

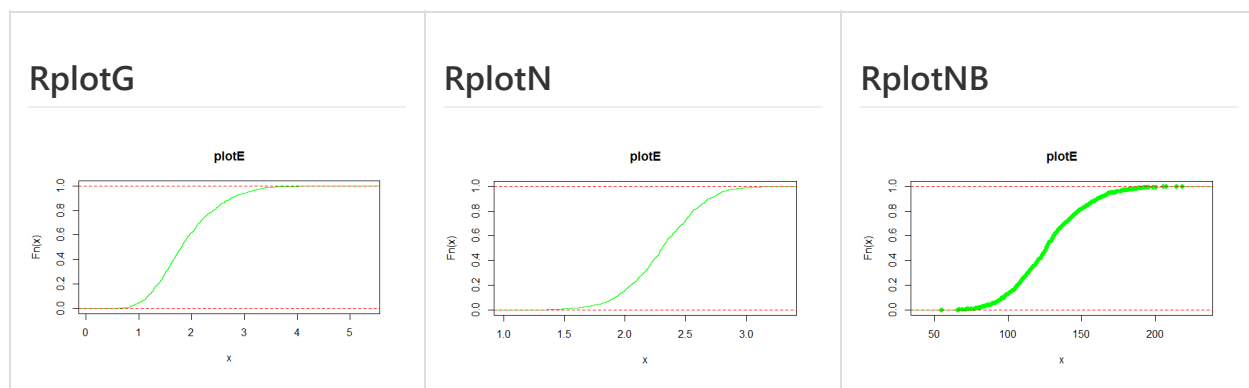
```
#а. сгенерировать выборку длины 1000 из данного распределения (стр. 19)
n <- 1000;
rG <-- rgamma(n = n, shape = 10, rate = 5.3);
...
```

- б. построить по данной выборке эмпирическую функцию распределения;

```
#б. построить по данной выборке эмпирическую функцию распределения;
buildEmpiricalPlots <- function(){ empiricalPlot(rG); ... }
```

```
# значения функции распределения в точке x
funP <- list(G = {function(x){ pgamma(q = x, shape = 10, rate = 5.3);}}, ... )
);
```

Эмпирические



Выводы:

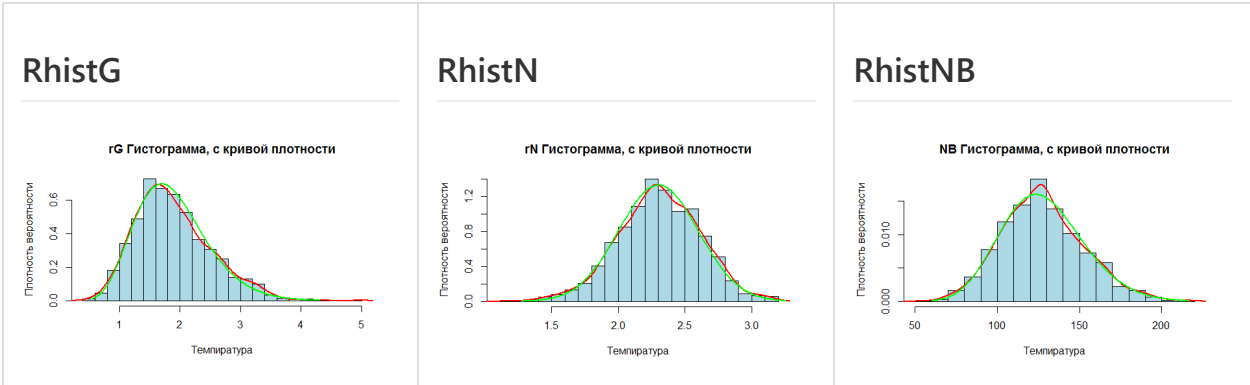
- с. построить гистограмму частот;
- д. сравнить гистограмму частот и реальную плотность данного распределения (вычисление значения плотности в точке в пакете R описано на той же 19 стр.)

```
#с&д. сравнить гистограмму частот и реальную плотность данного распределения
# точки для наложения
ranges <- list(G = (((range(rG)[1]*100):(range(rG)[2]*100))/100), ... )
# плотности
densitys <- list(G = dgamma(x = ranges$G, shape = 10, rate = 5.3), ... )
# построение Зеленый - ген.совок. Красный - выборка
hist3 <- function(){
{
  hist(rG, breaks = 20, freq = F, ... );
  lines(density(rG), col = "red", lwd = 2);
  lines(x = ranges$G, y = densitys$G, col = "green", lwd = 2);
}
```

Реальное распределение - ЗЕЛЕНЫМ

Выборочное распределение - КРАСНЫМ

Гистограммы



Выводы:

По гистограммам видно, что распр. выборки из 1000 наблюдений близко к ген.сов.

Так же очевидно, что работает ЗБЧ и ЦПТ

И гамма и NB иногда могут быть приближены нормальным распределением

- е. вычислить следующие выборочные характеристики: выборочное среднее, выборочную дисперсию, выборочную асимметрию, выборочный эксцесс; (см. стр. 20-22 методички)

```
# е. все характеристики
allProp <- function(x){ data.frame(mean = mean(x),var = var(x),asm = asm(x),exc = exc(x))}
```

- ф. сравнить результаты пункта е с реальными характеристиками распределения

	meansReal	means	varsReal	vars	asmsReal	asms	exc
G	0.64033225	1.88373867	0.1428149	0.35542869	-0.5960285	0.4780694	-1.2
N,	0.48069643	2.31551091	0.16544511	0.09183933	0.0701878	-0.010092957	-1.7
NB	2.75245521612138e-20	127.153	2.2098810347116e-39,	599.19078	2.085205315	0.29146007	2.53

Выводы:

Значения выборки совпали с исходными(в задании),

но не совпали с вычисленными, видимо я вручную сгенерировал для них новое распределение

Распределение из файла:

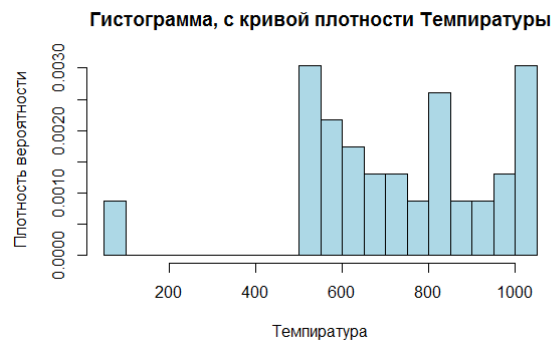
```
# читаем температуры
AnnualDiameter<-as.data.frame(read.csv("IDZ_1/annual-diameter-of-skirt-at-hem-.csv",col.names = c("AnnualDiameter")))
```

Повторяем пункты а-d

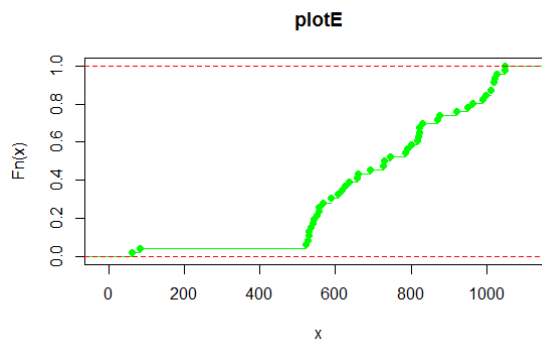
	mean	var	asm	exc
--	------	-----	-----	-----

	731.086956	51786.0811	-0.7350737	0.7727566
--	------------	------------	------------	-----------

Гистограмма



Эмпирические



Выводы:

Данных недостаточно, чтобы делать выводы

```
# сохраняем результаты  
write.csv(AnnualDiameterProp, file = "AnnualDiameterProp.csv")
```

