**Construirea unui pachet R pentru lucrul cu variabile aleatoare continue**

*Proiect realizat de:*

Caruntu Razvan

Ciocan Andrei-Madalin

Matea Marian

Voinea Dragos-Petru

***Cuprins***

*1. Introducere* . . . . . . . . . . . . . . . . . 3

*2. Exercitii* . . . . . . . . . . . . . . . . . . . 3

2.1 Exercitiul 1 . . . . . . . . . . . . . . . . . 3

2.2 Exercitiul 2 . . . . . . . . . . . . . . . . . 6

2.3 Exercitiul 3 . . . . . . . . . . . . . . . . . 7

2.4 Exercitiul 5 . . . . . . . . . . . . . . . . . 10

2.5 Exercitiul 8 . . . . . . . . . . . . . . . . . 12  
 2.6 Exercitiul 4 . . . . . . . . . . . . . . . . . 16

2.7 Exericitul 6 . . . . . . . . . . . . . . . . . 19

2.8 Exercitiul 9 . . . . . . . . . . . . . . . . . 24

3. *Concluzie* . . . . . . . . . . . . . . . . . . 30

1. *Introducere*

Acest proiect se concentrează pe dezvoltarea unui pachet R dedicat lucrului cu variabile aleatoare. Pachetul va oferi utilizatorilor instrumente puternice și ușor de utilizat pentru generarea, manipularea și analiza acestor variabile în cadrul analizei statistice. Scopul final este de a facilita lucrul eficient cu variabile aleatoare în mediul R și de a sprijini cercetarea și analiza datelor în diverse domenii.

1. *Exercitii*

*2.1 Exercitiul 1*

*Verificarea daca o functie introdusa de utilizator este densitate de probabilitate.*

Pentru realizarea acestui exercitiu am cautat documentatia din cadrul cursului cat si a laboratorului astfel am gasit si am folosit 2 proprietati ce caracterizeaza o functie drepta densitate de probabilitate. Cele 2 proprietati folosite pentru verificare sunt:

1. f(x) ≥ 0 (Functia este pozitiva)

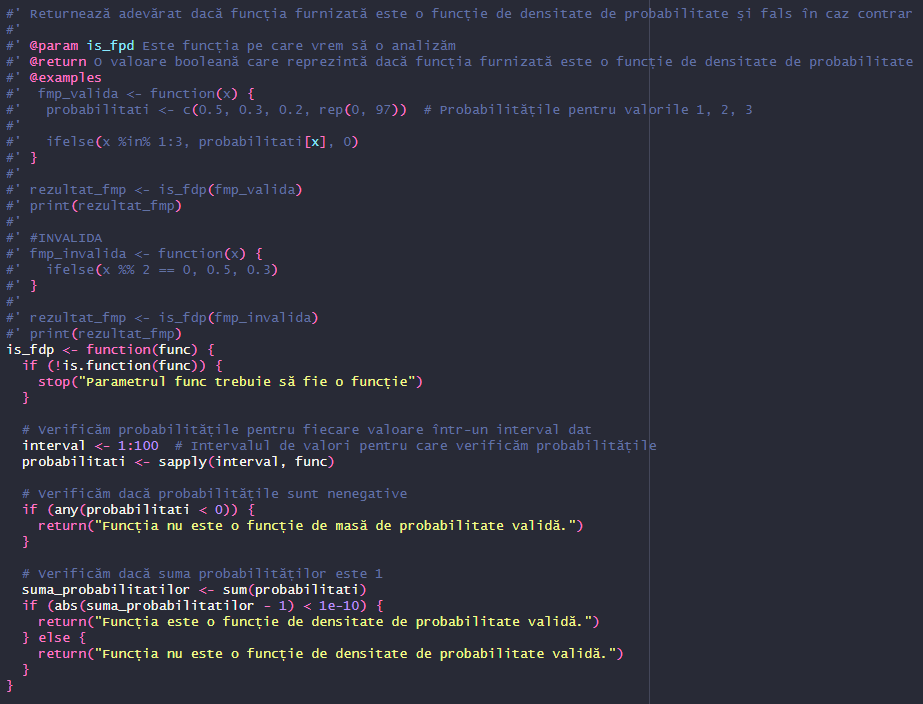
2. = 1

In cadrul verificarii ca functia este pozitiva metoda folosita nu este in totalitate precisa deoarece folosind functia “sequence” ia valorile din cadrul unui interval si calculez functia astfel verificand pozitivitatea, dar din pricina limitarilor functiei aceasta permite un interval finit, si pe langa acesta factor functia permite ca punctele de minim si de maxim ale intervalului sa fie maxim egale cu pas\*10E±10.

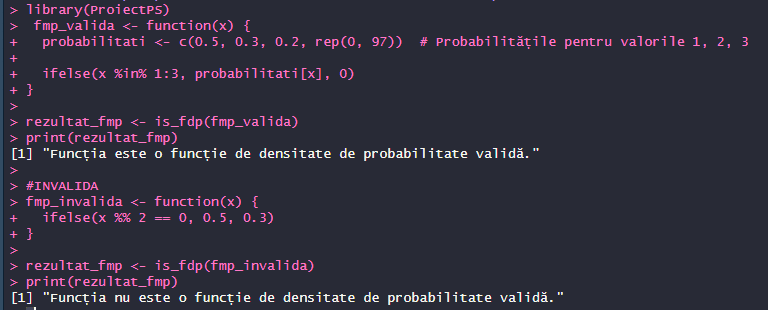
Apoi am calculat integrala si am verificat ca rezultatul sa fie egal cu 1, am decis sa aproximez rezultatul deoarece am intampinat problema ca rezultatul sa fie foarte aproape de 1 dar sa nu fie acceptat pentru verificare, acesta nefiind egal exact cu 1 (ex. 0.999999999999292).

Functia de verificare am creat-o sa primeasca ca parametrii functia de testare, intervalul pentru verificarea pozitivitatii si pasul, toate fiind inputuri de utilizator.

*Cod*



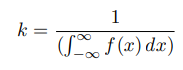
*Exemple*



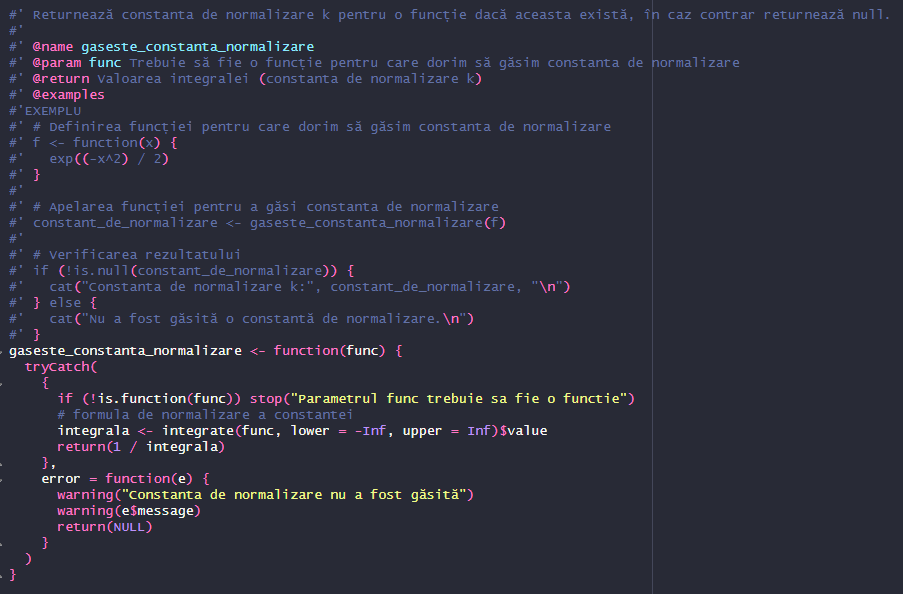
*2.2 Exercitiul 2*

*Fiind dată o funcție f , introdusă de utilizator, determinarea unei constante de normalizare k(cu precizarea dacă această constantă duce la o funcție de masă sau la o densitate de probabilitate). Ȋn cazul ȋn care o asemenea constantă nu există, afișarea unui mesaj corespunzător către utilizator.*

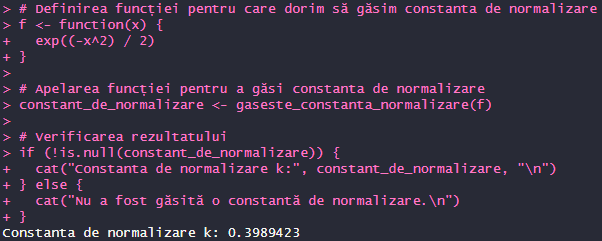
Pentru calcularea constanteri de normalizare k, in raport cu o functie introdusa de catre utilizator, se aplica formula:



*Cod*



*Exemple*



*2.3 Exercitiul 3*

*Reprezentarea grafică a densității de probabilitate/funcției de masă și a funcției de repartiție pentru diferite valori ale parametrilor repartiției. Ȋn cazul ȋn care funcția de repartiție nu este dată ȋntr-o formă explicită(ex. repartiția normală) se acceptă reprezentarea grafică a unei aproximări a acesteia.*

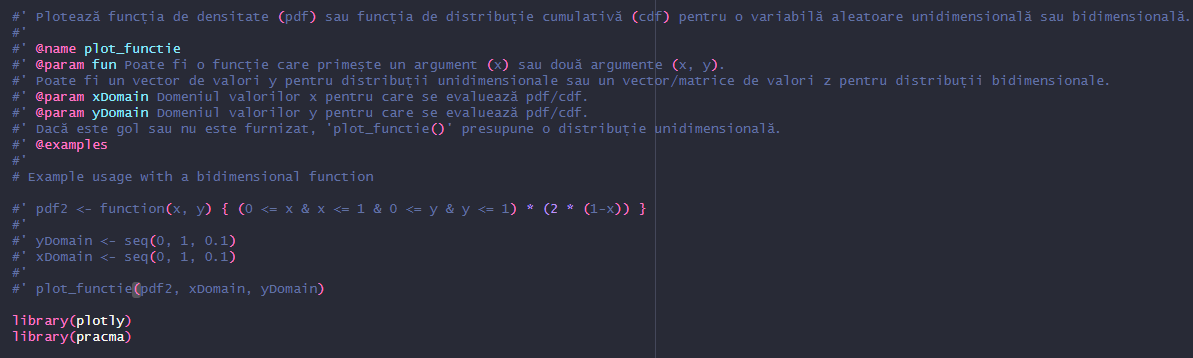
Am creat o singura functie care poate afisa grafice atat pentru un pdf cat si pentru un cdf.

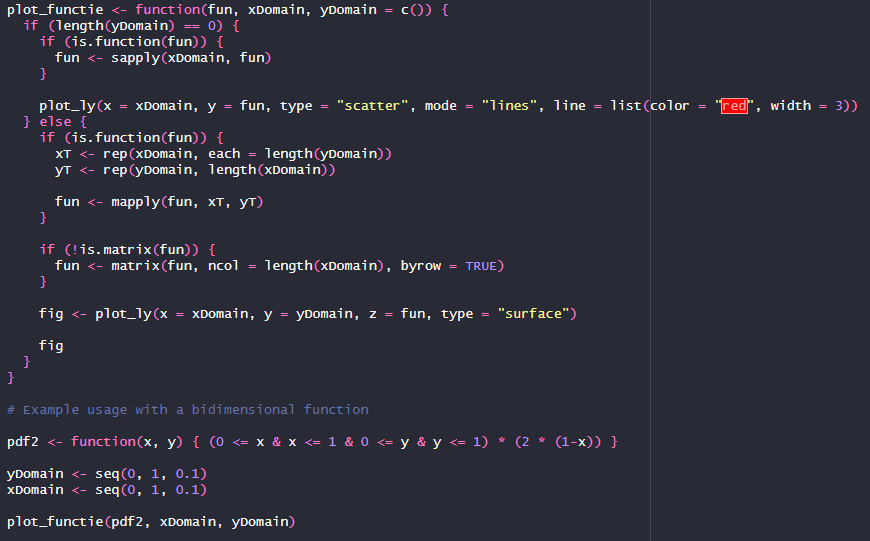
Functia poate genera grafice pentru v.a. unidimensionale / bidimensionale.

Functia poate lua un obiect de tip function sau un set de valori rezultat din utilizarea distributiilor standard din R.

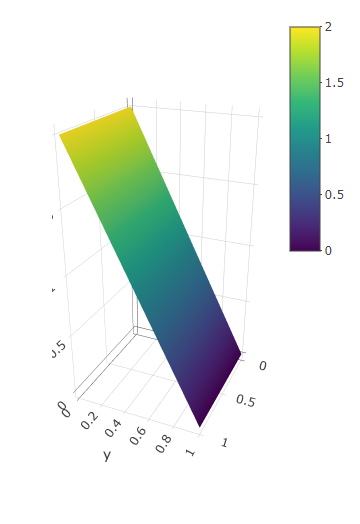
Daca este specificat domeniul pentru y, v.a. data este considerata bidimensionala.

*Cod*





*Exemple*

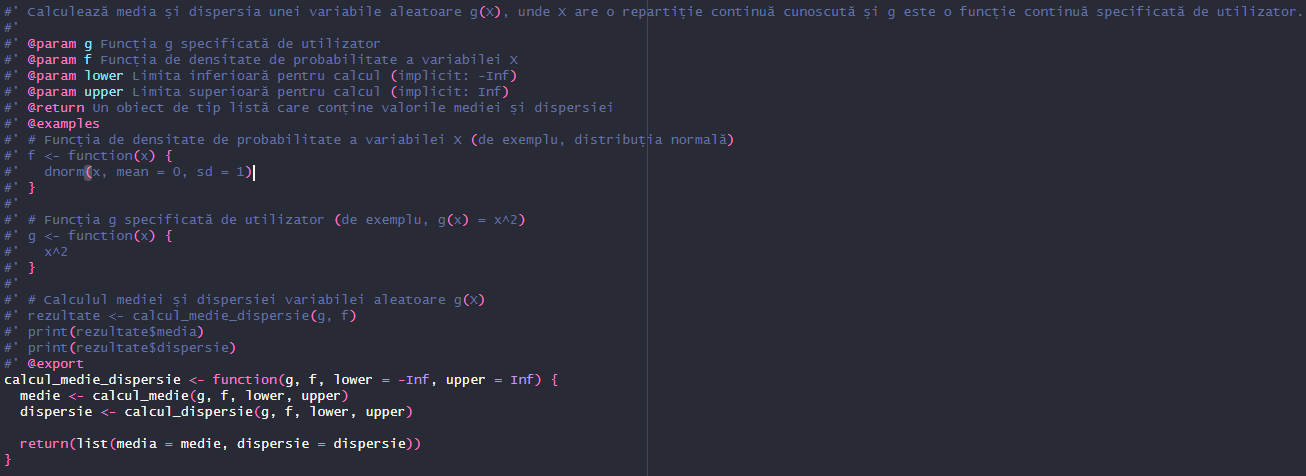


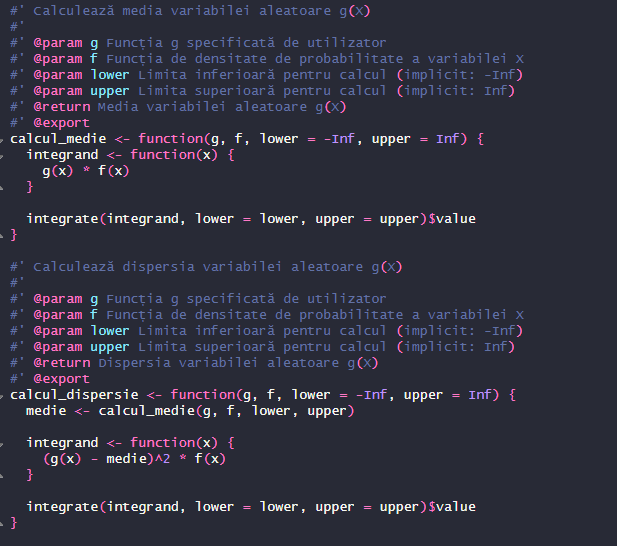
*2.4 Exercitiul 5*

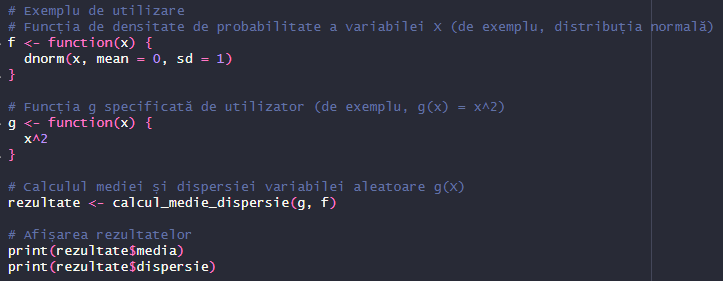
*Calculul mediei și dispersiei unei variabile aleatoare g(X), unde X are o repartiție continuă cunoscută iar g este o funcție continuă precizată de utilizator.*

Pentru calcule, am folosit urmatoarele formule:

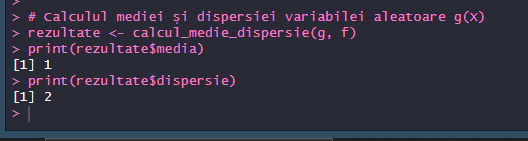


*Cod*





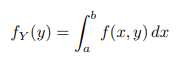
*Exemple*



*2.5 Exercitiul 8*

*Pornind de la funcția de masă/densitatea comună a două variabile aleatoare, construirea funcțiilor de masă/densităților marginale și a funcțiilor de masă/densităților condiționate.*

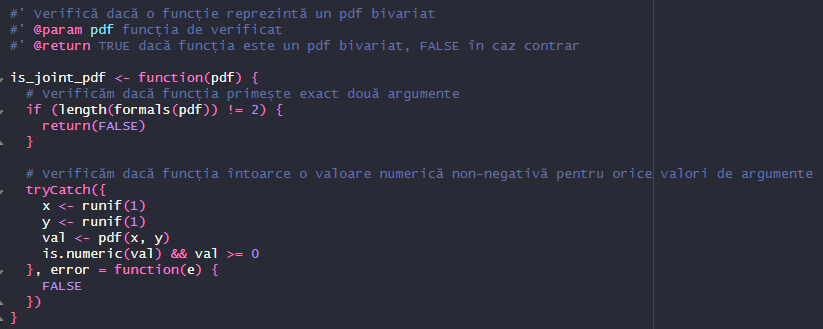
Pentru densitatile marginale, am folosit:

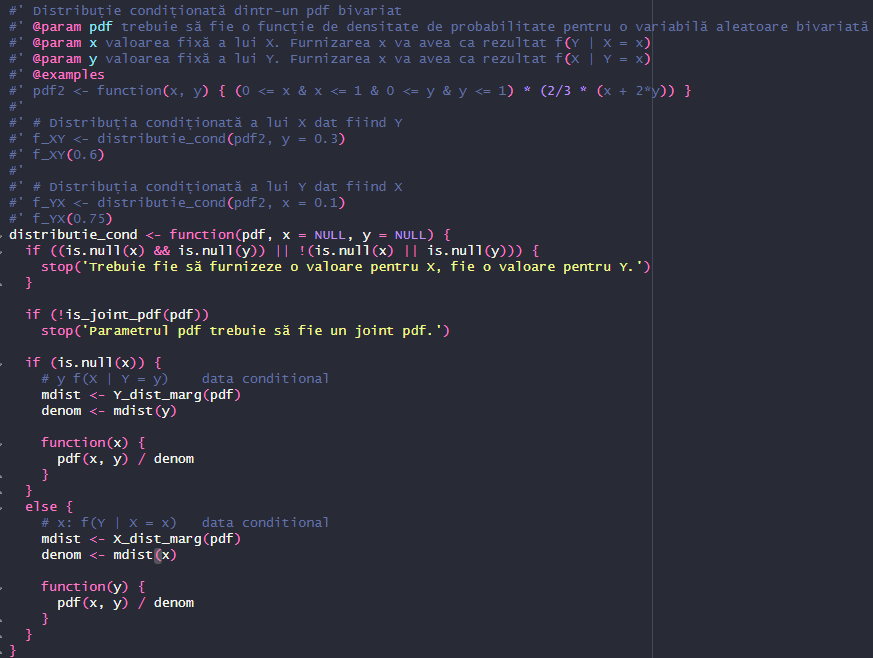
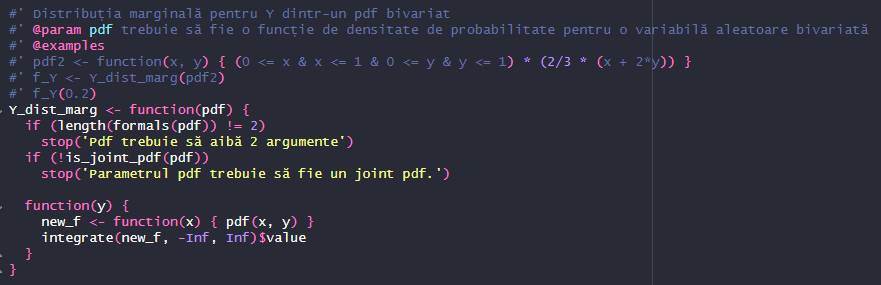
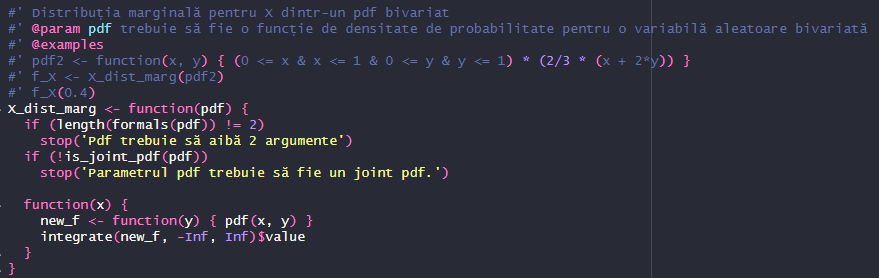


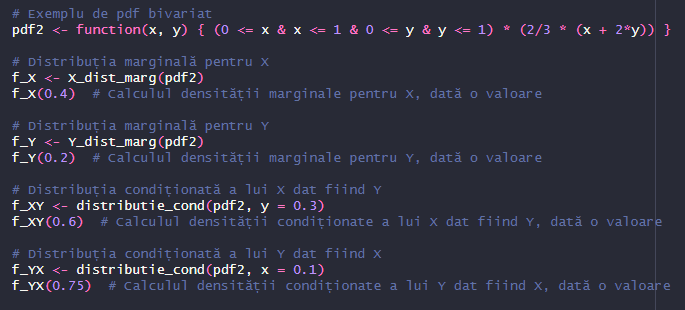
Folosind formulele de mai sus, am calculat si cele doua densitati conditionate:



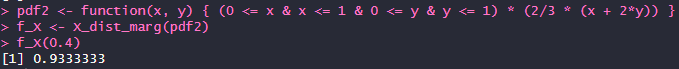
*Cod*

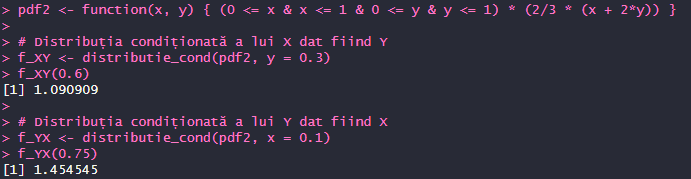
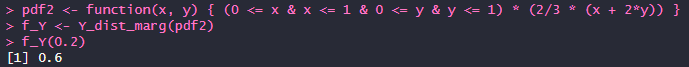




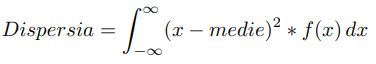
Exemple





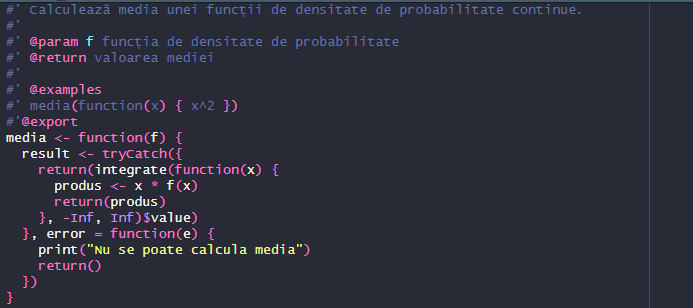
*2.6 Exercitiul 4*

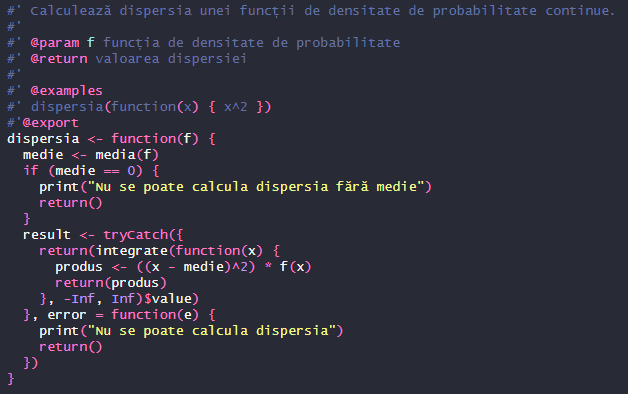
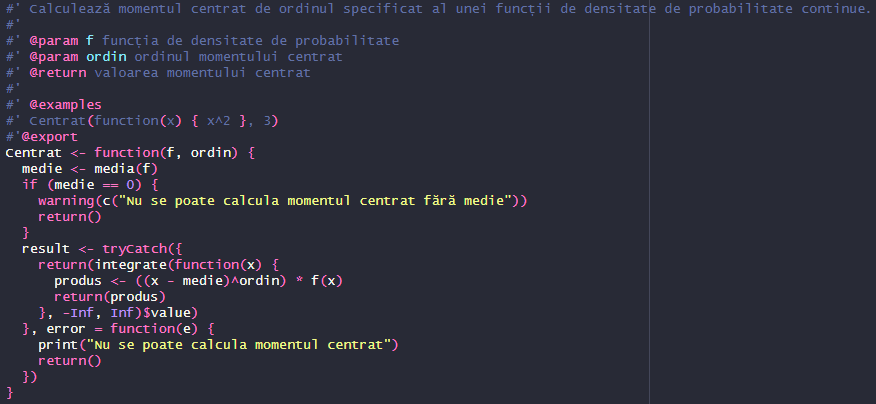
*Calculul mediei, dispersiei și a momentelor inițiale și centrate pȃnă la ordinul 4(dacă există). Atunci cȃnd unul dintre momente nu există, se va afișa un mesaj corespunzător către utilizator.*

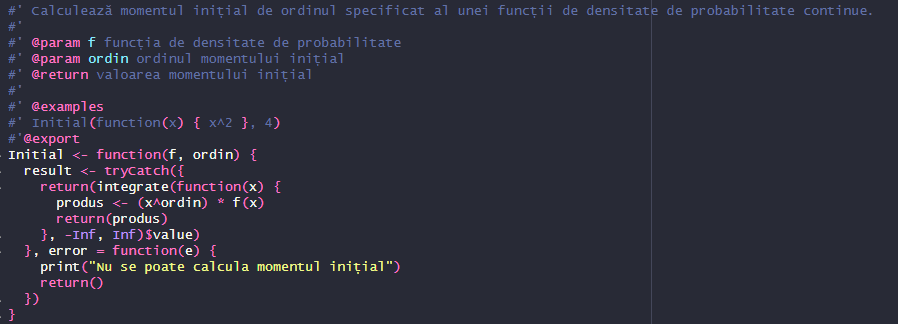
Am calculat cele 4 elemente folosind urmatoarele formule:



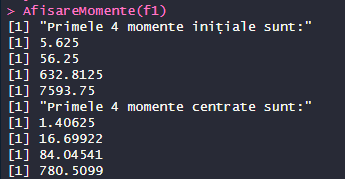
*Cod*







*Exemple*



*2.7 Exercitiul 6*

*Crearea unei funcții P care permite calculul diferitelor tipuri de probabilități asociate unei variabile aleatoare(ex. probabilități condiționate). Exemplificați cu cel puțin 4 timpuri de astfel de situații.*

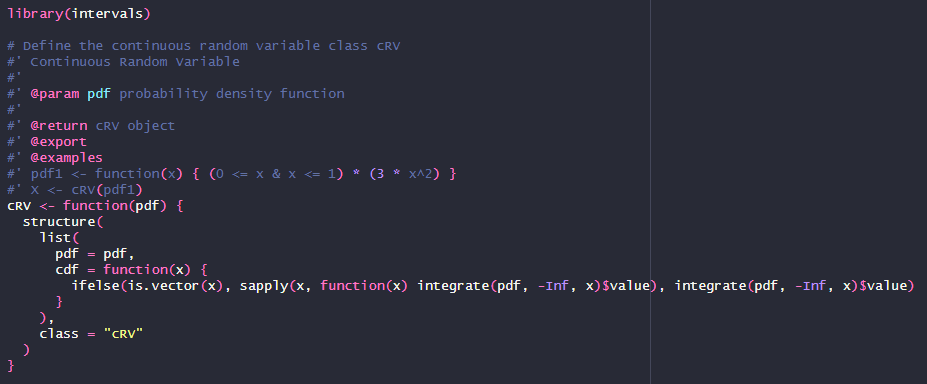
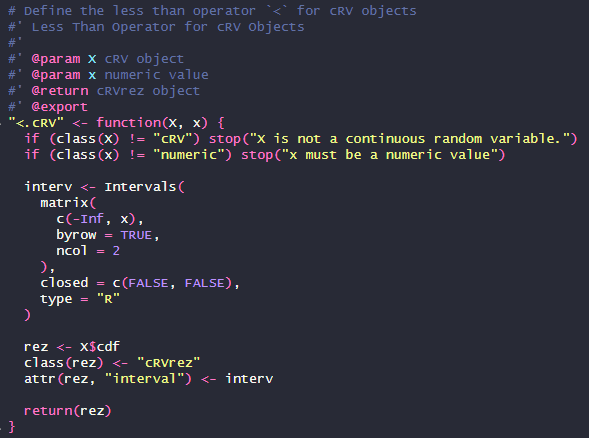
S-au creat operatorii de comparație pentru tipul de date **cRV**.

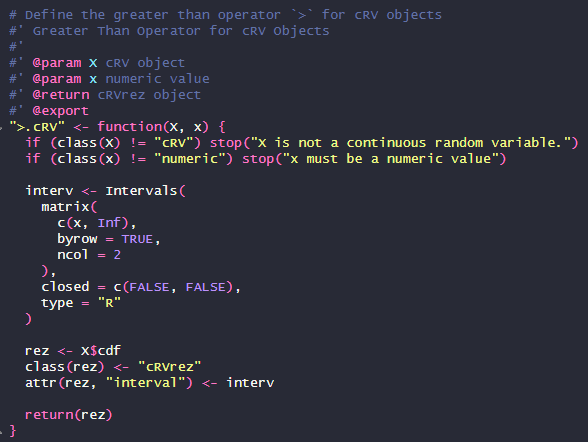
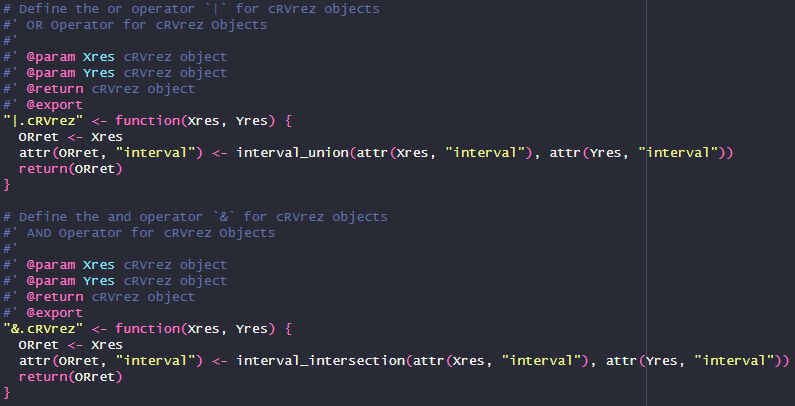
Acești operatori pot avea un obiect de tip **cRV** în stânga și un obiect numeric în dreapta.

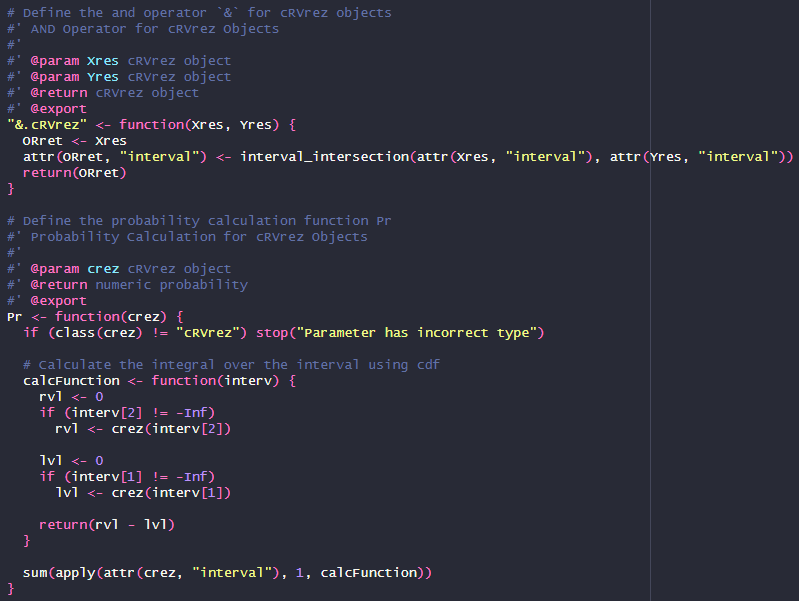
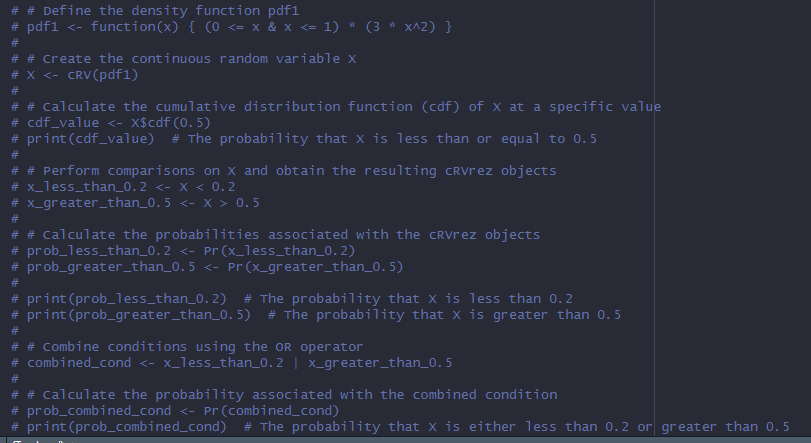
Operatorii returnează un tip nou de date, numit "cRVresult", care stochează atât funcția de distribuție cumulativă a variabilei aleatoare, cât și noul domeniu peste care să fie calculată probabilitatea.

De asemenea, au fost create operatorii logici "AND" și "OR" pentru a lucra cu condiții compuse în calculul probabilităților. Acești operatori iau două obiecte de tip "cRVresult" ca parametri și returnează un obiect de tip "cRVresult".

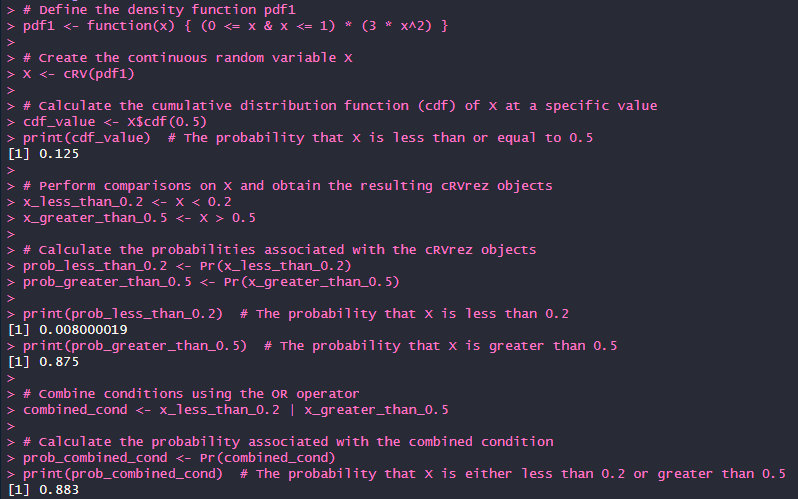
Funcția **Pr** primește un obiect de tip "cRVresult" ca parametru și returnează un număr real cuprins între 0 și 1.

*Cod*





*Exemple*



*2.8 Exercitiul 9*

*Se dau seturile următoare de valori: a) 7 4 2 11 2 1 2 1 6 6 0 1 3 9 7 0 1 14 0 5 1 5 2 4 3 1 0 0 26 1*

*b) -1.91 -0.97 4.59 2.19 -0.86 -0.74 -0.60 -1.29 0.93 1.42 2.14 -2.01 2.60 1.45 2.60 -3.32 -3.62 3.09 2.91 3.60 -0.83 -0.27 1.82 -1.38 -1.76 1.43 -0.59 -1.34 2.07 1.02*

*c) 0.90 8.91 0.06 1.85 1.61 6.50 0.26 0.04 0.62 1.01 3.42 1.45 3.44 0.46 0.55 0.09 2.22 0.65 0.61 6.45 0.27 4.81 2.27 0.34 4.51 0.42 3.71 2.59 0.42 11.18*

*d) 4.83 4.37 5.57 4.22 5.96 5.11 5.52 4.81 5.19 4.19 4.73 5.92 5.63 4.53 4.67 4.84 5.25 5.06 5.98 5.25 4.60 4.11 4.32 5.09 5.25 5.10 4.36 5.40 5.33 4.65*

*e) 11 11 10 10 10 6 5 9 11 10 14 8 11 6 13 9 14 16 14 10 7 7 11 12 9 5 12 15 9 12*

*Pentru fiecare dintre seturile de valori de mai sus efectuați:*

*9.1. Faceți histograma valorilor. Calculați mediana, media și deviația standard și ilustrați pe desen aceste valori.*

*9.2. Găsiți o manieră de a identifica din ce repartiție au fost generate valorile de mai sus, cu identificarea parametrilor și justificarea alegerii făcute.*

*9.3 Presupuneți că valorile au fost extrase dintr-o repartiție normală de parametri necunoscuți. Folosind metoda verosimilității maxime și respectiv metoda momentelor estimați acești parametri ȋn baza celor 5 eșantioane. Comparați estimările obținute prin cele două metode și comentați rezultatele.*

*9.4 Identificați ȋn care din cazurile de mai sus este verosimil ca valorile să fie extrase dintr-o repartiție nomală și justificați răspunsul. Argumentați și ȋn caz contrar.*

9.2

În R, putem utiliza funcția fitdistr din pachetul MASS pentru a estima parametrii distribuțiilor potrivite pentru seturile de valori.

library(MASS)

fit\_a <- fitdistr(a, "poisson")

Aceasta va estima parametrul lambda pentru distribuția Poisson care se potrivește cel mai bine cu datele. Putem utiliza aceeași abordare pentru celelalte seturi de valori și distribuțiile potrivite (cum ar fi distribuția normală, exponențială, etc.).

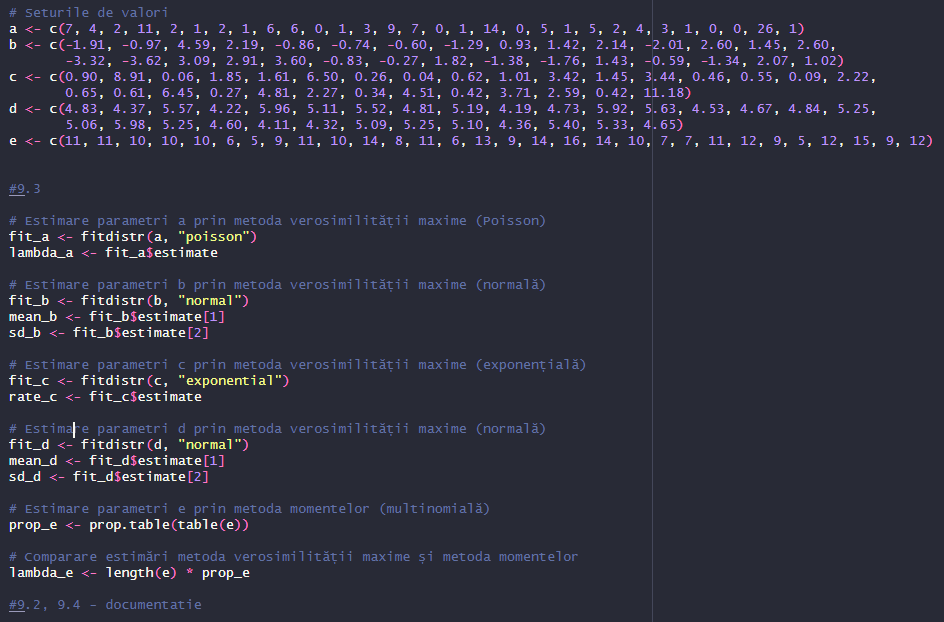
9.4

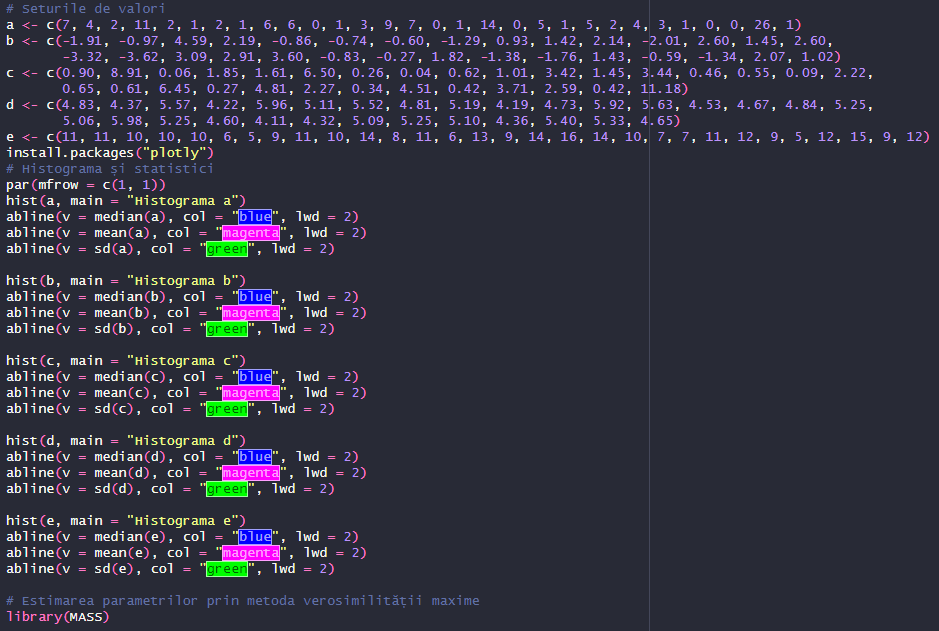
Pentru a determina în care dintre cazuri este verosimil ca valorile să fie extrase dintr-o repartiție normală, putem compara rezultatele estimărilor obținute prin metoda verosimilității maxime și metoda momentelor.

Dacă estimările parametrilor obținute prin cele două metode sunt apropiate și se apropie de valorile medii și deviația standard ale setului de valori, atunci este posibil ca valorile să fie extrase dintr-o repartiție normală.

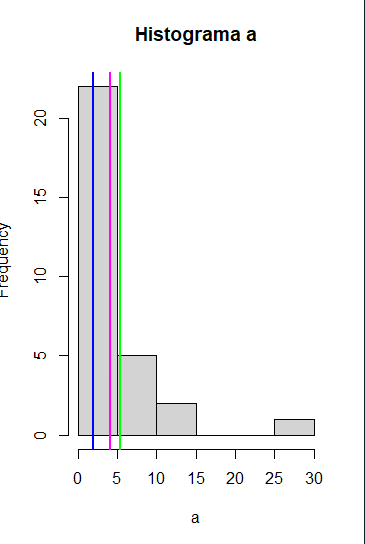
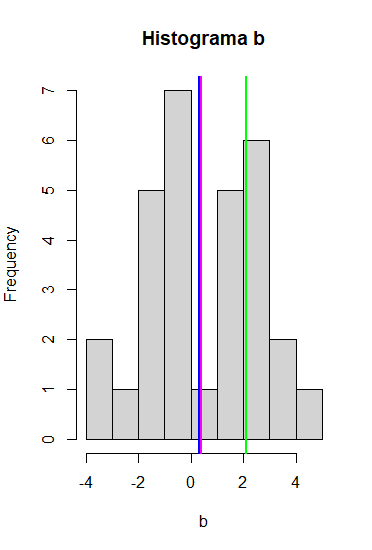
Dacă estimările obținute prin cele două metode diferă semnificativ sau nu se apropie de valorile medii și deviația standard ale setului de valori, atunci este puțin probabil ca valorile să fie extrase dintr-o repartiție normală.

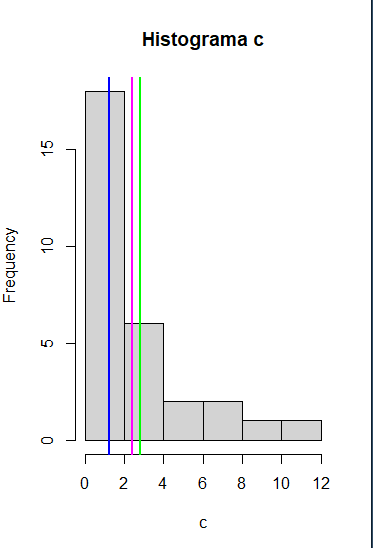
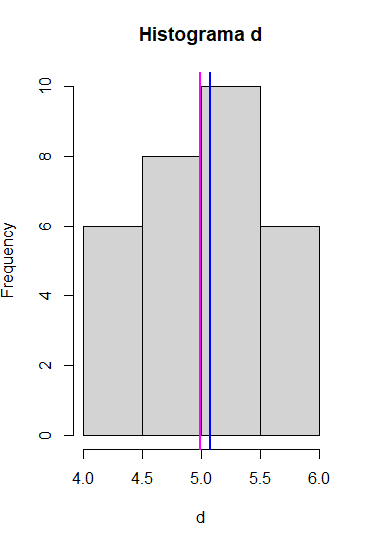
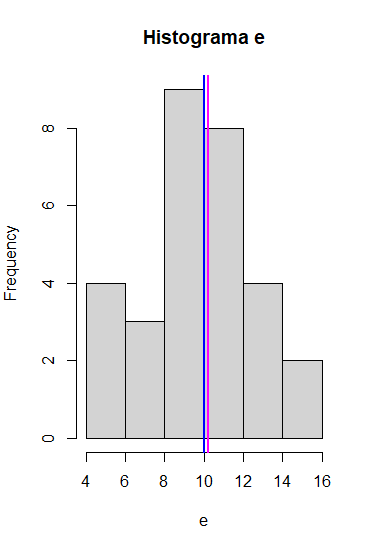
*Cod*



**

*Exemple*



3) *Concluzie*

În concluzie, prin intermediul acestui proiect, am realizat un pachet care oferă soluții pentru a facilita lucrul cu variabile aleatoare. Utilizând acest pachet, am transformat noțiuni teoretice ce pot părea dificile pentru utilizatori în instrumente informatice, ce pot fi folosite în aplicații concrete.