

Documentație Proiect

Probabilitate si Statistica

Membri:

Gheorghe Bogdan-Alexandru

Andrei Cristian-David

Sincari Sebastian-George

Cosor Mihail

Ianuarie 2025

Contents

Cerinta I	2
Descrierea Problemei	2
Exercitiul 1	2
Exercitiul 2	3
Exercitiul 3	3
Exercitiul 4	4
Exercitiul 5	4
Exercitiul 6	4
Cerinta II	7
Aspecte Teoretice	7
Reprezentări Grafice	8
Codul Aplicației	9
app.R	10
ui.R	10
server.R	12
formularea1.R	12
Concluzii	13
Cerinta III	14
Descrierea Problemei	14
Aspecte Teoretice	14
Reprezentări Grafice	15
Pachete Software Folosite	15
Codul Aplicației	16
app.R	16
ui.R	16
server.R	17
utils.R	18
normals.R	18
Dificultăți Întâmpinate	20
Probleme Deschise	20
Concluzii	20

Cerinta I

Descrierea Problemei

Se consideră o activitate care presupune parcurgerea secvențială a n etape. Timpul necesar finalizării etapei i de către o persoană A este o variabilă aleatoare $T_i \sim \text{Exp}(\lambda_i)$.

După finalizarea etapei i , A va trece în etapa $i + 1$ cu probabilitatea α_i sau va opri lucrul cu probabilitatea $1 - \alpha_i$.

Fie T timpul total petrecut de persoana A în realizarea activității respective.

```
1 set.seed(123)
2 n <- 100 # Nr etape
3 lambda <- runif(n, 0.5, 2) # Rata exp pt fiecare etapa
4 alpha <- runif(n-1, 0.8, 1) # Prob trecerii la urmatoarea etapa
5 alpha <- c(1, alpha) # Prob ca o persoana sa participe la prima etapa este 100%
6 nrSimulari <- 1000000 # Nr simulari
7
8
9 simulator <- function() {
10   Ti <- rexp(n, rate = lambda) # Timpul pentru fiecare etapa
11
12   stop_point <- which(runif(n - 1) > alpha[-1])[1] # Determinam momentul opririi
13
14   # Timpul total va fi cel pana la oprire/finalul vectorului daca nu se opreste
15   if (!is.na(stop_point)) {
16     return(sum(Ti[1:stop_point]))
17   } else {
18     return(sum(Ti))
19   }
20 }
21
22 # Simulam 10^6
23 valoriT <- replicate(nrSimulari, simulator())
```

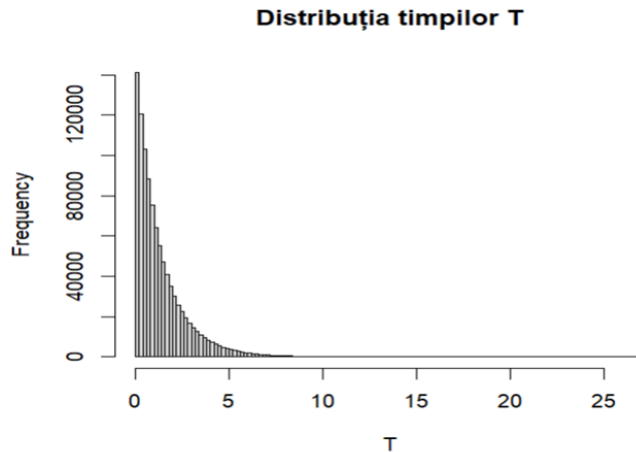
1. Construiți un algoritm în R care simulează 10^6 valori pentru v.a. T și în baza acestora aproximați $E(T)$. Reprezentați grafic într-o manieră adecvată valorile obținute pentru T . Ce puteți spune despre repartiția lui T ?

```
1 approx_E_T <- mean(valoriT)
2
3 hist(valoriT, breaks = 100, main = "Distributia timpilor T", xlab = "T")
```

Explicații:

Pentru calculul aproximativ al mediei am făcut pur și simplu media valorilor rezultate în urma simulării.

Afișând în histograma valorile obținute în `valoriT` rezultă distribuția lui T . Se poate observa că reprezentarea grafică a timpilor T urmează o distribuție exponențială.



Rezultat:

```
1 approx_E_T = 9.51113170988666
```

2. Calculați valoarea exactă a lui $E[T]$ și comparați cu valoarea obținută prin simulare.

```
1 exact_E_T <- sum(1 / \lambda * cumprod(alpha))
```

Explicații:

Valoarea exactă a lui $E[T]$ se calculează după următoarea formulă:

$$E[T] = E[T_1] + \alpha_1 E[T_2] + \alpha_1 \alpha_2 E[T_3] + \dots + \alpha_1 \alpha_2 \dots \alpha_{n-1} E[T_n]$$

Am folosit funcția `cumprod`, care returnează un vector cu produsul cumulativ. α are 99 de valori deoarece există $n - 1$ pași pentru trecerea de la o etapă la alta. Am atașat acestui vector la început valoarea 1, deoarece probabilitatea ca persoana să participe la prima etapă este 100. Apoi am înmulțit produsul cumulativ cu $1/\lambda$, deoarece $X \sim \text{Exp}(\lambda) \Rightarrow E[X] = 1/\lambda$. În final, am făcut suma tuturor mediilor pe fiecare etapă, rezultând $E[T]$.

În urma analizării rezultatului exact observăm că este foarte apropiat de rezultatul aproximativ obținut prin simulare.

Rezultat:

```
1 exact_E_T = 9.51299920302413
```

3. În baza simulărilor de la 1. aproximați probabilitatea ca persoana A să finalizeze activitatea.

```
1 prob_finalizare <- mean(valoriT >= sum(1 / lambda))
```

Explicații:

Am calculat probabilitatea ca o persoana sa finalizeze fiecare etapa luand ca exemplu in simularea noastra un numar de 100 de etape si o probabilitate de a trece de la o etapa la alta $\lambda=80$

In scrierea codului am calculat valoarea teoretica a timpului total pe care o persoana trebuie sa il petreaca in cele 100 de etape pentru a spune ca a finalizat. Astfel am calculat media unui vector de valori booleene care valoarea true corespunde rezultatului unei simulari $\lambda=$ valoarea timpului adica persoanele care au finalizat respectiv false pentru persoanele care nu au reusit sa finalizeze. Rezultatul este mic datorita numarului de etape, acesta ar creste o data cu micsorarea numarului de etape/ ar scadea o data cu cresterea numarului de etape. Totodata, rezultatul va scadea odata cu micsorarea marginii inferioare a valorilor probabilitatilor din alpha si invers.

Rezultat:

```
1      prob_finalizare = 0.00004
```

4. În baza simulărilor de la 1. aproximați probabilitatea ca persoana A să finalizeze activitatea într-un timp mai mic sau egal cu σ .

```
1      sigma <- 94
2      prob_sigma <- mean(valoriT <= sigma & valoriT >= sum(1 / lambda))
```

Explicatii:

Ne-am folosit de modalitatea de rezolvare a exercitiului anterior, astfel am folosit un vector de valori booleene care verifica 2 conditii $j = \text{sigma}$ si $i = \text{valoarea timpului total}$, apoi am facut media si ne-a rezultat valoarea probabilitatii ca o persoana sa finalizeze intr-un timp mai mic decat sigma. Este normal ca valoarea acestei probabilitati sa fie mai mica sau egala cu valoarea probabilitatii de a finaliza.

Rezultat:

```
1      prob_sigma = 0.000007
```

5. În baza simulărilor de la 1. determinați timpul minim și respectiv timpul maxim în care persoana A finalizează activitatea și reprezentați grafic timpii de finalizare a activității din fiecare simulare. Ce puteți spune despre repartiția acestor timpi de finalizare a activității?

```
1      timpMin <- min(valoriT[valoriT >= sum(1 / lambda)] )
2      timpMax <- max(valoriT[valoriT >= sum(1 / lambda)] )
3      cat("Timpul minim:", timpMin, "Timpul maxim:", timpMax, "\n")
4
5      hist(valoriT[valoriT >= sum(1 / lambda)], breaks=50,
6           main="Distributia timpilor de finalizare",
7           xlab="Timpul de finalizare (T)", ylab="Frecventa",
8           col="lightblue", border="black")
```

Explicatii:

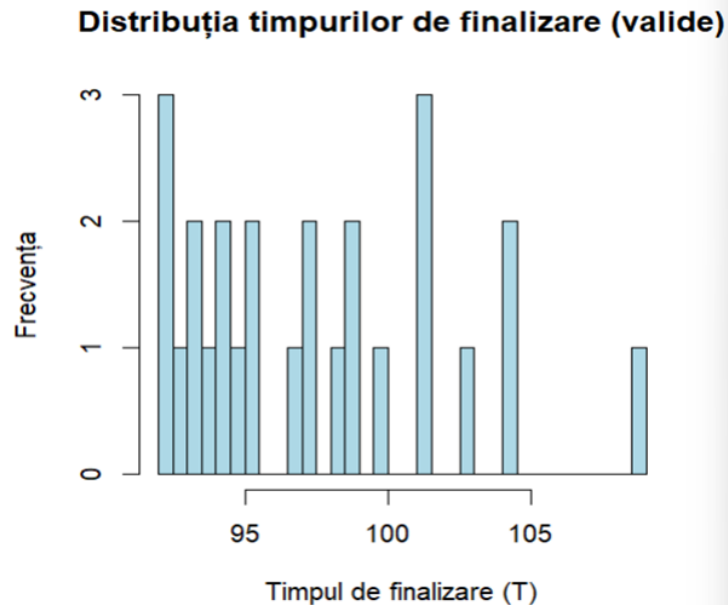
Am considerat variabilele `timpMin` și `timpMax` pentru valorile minime și maxime ale timpilor persoanelor care au finalizat evenimentul. Am folosit funcția `min` pentru minim și, ca parametru, am dat lista de timpi simulați, filtrată de valorile mai mari decât media timpului total. Am folosit din nou, pentru reprezentarea grafică, o histogramă, deoarece aceasta ajută la identificarea repartiției. Două observații sunt faptul că numărul de persoane care au reușit să finalizeze evenimentul este foarte mic în comparație cu numărul de simulări și că se observă că repartiția este uniformă, deoarece pentru alegerea valorilor random am folosit `runif`.

Rezultat:

```
1      timpMax = 108.54443085275
2      timpMin = 92.0171680316328
```

6. În baza simulărilor de la 1. aproximați probabilitatea ca persoana A să se oprească din lucru înainte de etapa k , unde $1 < k \leq n$. Reprezentați grafic probabilitățile obținute într-o manieră corespunzătoare. Ce puteți spune despre repartiția probabilităților obținute?

```
1      k <- 5
2      PStopK <- mean(sapply(valoriT, function(x) {
3      if(x <= sum(1/lambda[1:k-1]))
4      return (TRUE)
```



```

5     else
6         return(FALSE)
7     )))

```

Explicații:

Am creat o variabilă numită `PstopK` care primește probabilitatea ca persoana A să se oprească din lucru înainte de etapa k . Pentru această probabilitate am aplicat ca parametru unei funcții fiecare valoare simulată. Funcția returnează TRUE dacă timpul este mai mic decât media timpului total până la pasul $k-1$ (adică poate face maxim $k-1$ pași), și FALSE altfel. Variabila primește media valorilor din lista booleană.

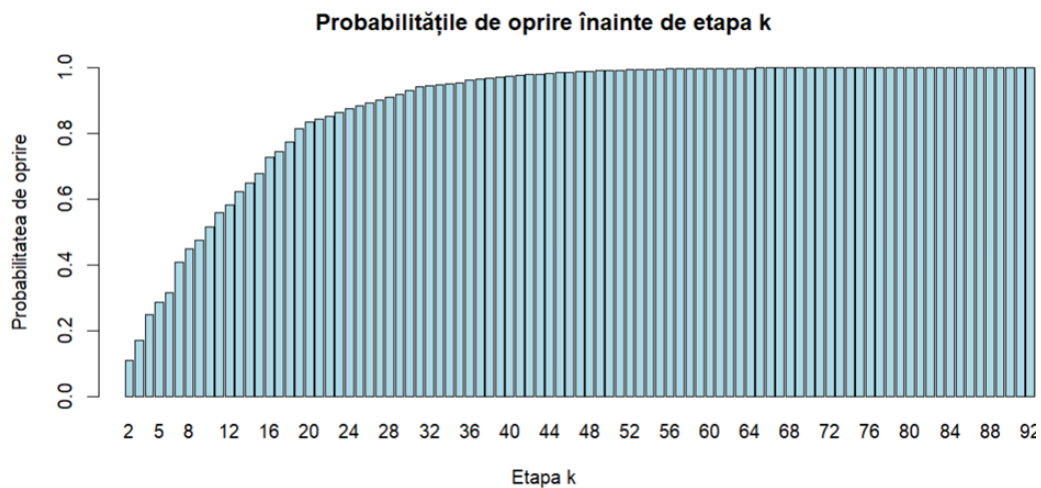
Rezultat:

```

1     PstopK = 0.284293  pentru k=5
2     PstopK = 0.51482  pentru k=10
3     PstopK = 0.991332 pentru k=50

1     calculate_PStopBeforeK <- function(valoriT, lambda, alpha, n) {
2     # Calculam timpii cumulativi pentru fiecare etapa
3     cumulative_times <- cumsum(1 / lambda)
4
5     # Calculam probabilitatile pentru fiecare k
6     PStopBeforeK <- sapply(2:n, function(k) {
7         mean(valoriT <= cumulative_times[k - 1])
8         # Probabilitatea de oprire inainte de etapa k
9     })
10
11     return(PStopBeforeK)
12 }
13 PStopBeforeK <- calculate_PStopBeforeK(valoriT, lambda, alpha, n)
14
15 barplot(PStopBeforeK, names.arg = 2:n, col = "lightblue",
16         main = "Probabilitatile de oprire inainte de etapa k",
17         xlab = "Etapa k", ylab = "Probabilitatea de oprire",
18         xlim = c(1, n), ylim = c(0, 1))

```



Observații:

Se poate observa că o persoană, cu cât ajunge mai departe în etape, cu atât are șanse mai mari să finalizeze evenimentul, iar în etapele inițiale șansele sunt foarte mici. Aceste observații sunt date datorită pantei abrupte în faza inițială, respectiv panta lină de la final.

Cerinta II

Descrierea Problemei

Scopul acestui proiect este de a construi o aplicație web interactivă folosind framework-ul **Shiny** din **R**, pentru a reprezenta cele 5 formulări ale repartiției Negativ Binomiale, graficele animate și situații concrete în care se pot aplica aceste formulări. Aplicația permite vizualizarea repartiției Negativ Binomiale pentru diferite valori ale parametrilor r și p .

Aspecte Teoretice

Pentru dezvoltarea aplicației, am utilizat următoarele aspecte teoretice importante:

Distribuția Negativ Binomială: Este definită ca o variabilă aleatoare discretă $X \sim \text{NB}(r, p)$, unde r reprezintă numărul de succese dorite, iar p este probabilitatea de succes pentru o singură încercare.

Repartiția Negativ Binomială poate fi definită în mai multe moduri, în funcție de context. Iată cele 5 formulări principale:

1. **k eșecuri, dat fiind r succese**

Probabilitatea de a avea k eșecuri înainte de a obține r succese:

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k+r-1}{k} p^r (1-p)^k$$

Alternativ:

$$f(k; r, p) = \binom{k+r-1}{r-1} p^r (1-p)^k$$

Suport: $k = 0, 1, 2, \dots$

2. **n încercări, dat fiind r succese**

Probabilitatea de a avea n încercări pentru a obține r succese:

$$f(n; r, p) \equiv \Pr(X = n) = \binom{n-1}{r-1} p^r (1-p)^{n-r}$$

Alternativ:

$$f(n; r, p) = \binom{n-1}{n-r} p^r (1-p)^{n-r}$$

Suport: $n = r, r+1, r+2, \dots$

3. **n încercări, dat fiind r eșecuri**

Probabilitatea de a avea n încercări pentru a obține r eșecuri:

$$f(n; r, p) \equiv \Pr(X = n) = \binom{n-1}{r-1} p^{n-r} (1-p)^r$$

Alternativ:

$$f(n; r, p) = \binom{n-1}{n-r} p^{n-r} (1-p)^r$$

Suport: $n = r, r+1, r+2, \dots$

4. **k succese, dat fiind r eșecuri**

Probabilitatea de a avea k succese înainte de a obține r eșecuri:

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k+r-1}{k} p^k (1-p)^r$$

Alternativ:

$$f(k; r, p) = \binom{k+r-1}{r-1} p^k (1-p)^r$$

Suport: $k = 0, 1, 2, \dots$

5. k succese, dat fiind n încercări

Aceasta este de fapt repartiția binomială, nu cea negativ binomială:

$$f(k; n, p) \equiv \Pr(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

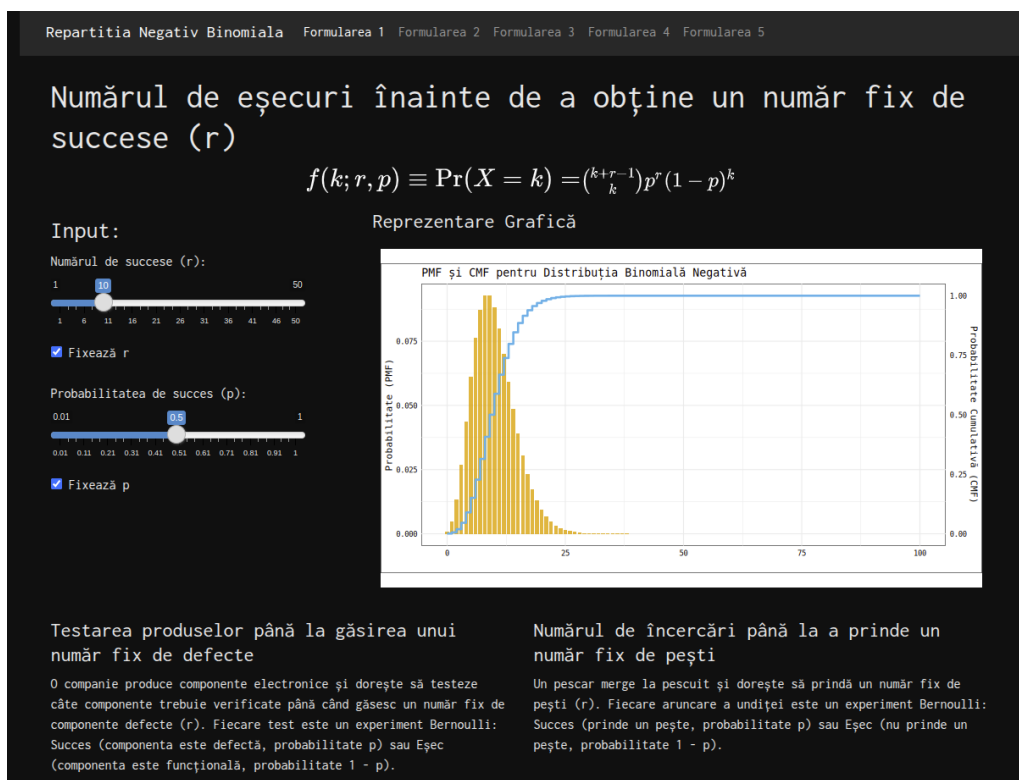
Alternativ:

$$f(k; n, p) = \binom{n}{n-k} p^k (1 - p)^{n-k}$$

Suport: $k = 0, 1, 2, \dots, n$

Reprezentări Grafice

Aplicația oferă reprezentări grafice ale funcției de masă și a funcțiilor de repartiție. În cele ce urmează, vom ilustra cateva capturi de ecran din aplicație.



Numărul total de încercări necesare pentru a obține un număr fix de eșecuri (r)

$$f(n; r, p) \equiv \Pr(X = n) = \binom{n-1}{r-1} p^{n-r} (1-p)^r$$

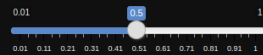
Input:

Numărul de succese (r):



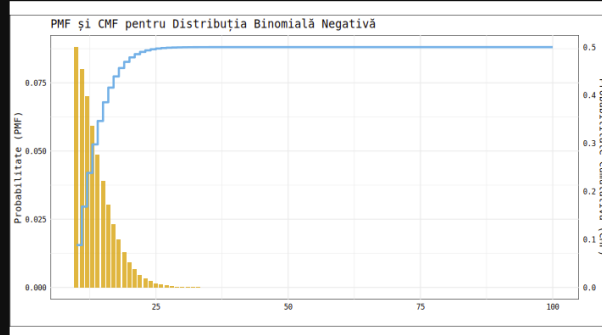
☒ Fixează r

Probabilitatea de succes (p):



☒ Fixează p

Reprezentare Grafică



Testarea produselor până la găsirea unui număr fix de produse neconforme

Un controlor de calitate testează produse și dorește să găsească un număr fix de produse neconforme. Fiecare test este un experiment Bernoulli: Succes (produsul este conform, probabilitate p) sau Eșec (produsul este neconform, probabilitate $1 - p$).

Numărul de întrebări până la obținerea unui număr fix de răspunsuri greșite

Un student răspunde la întrebări și dorește să obțină un număr fix de răspunsuri greșite. Fiecare întrebare este un experiment Bernoulli: Succes (răspunsul este corect, probabilitate p) sau Eșec (răspunsul este greșit, probabilitate $1 - p$).

Numărul de succese într-un număr fix de încercări (n)

$$f(k; n, p) \equiv \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{n-k} p^k (1-p)^{n-k} = \binom{n}{k} p^k (1-p)^r$$

Input:

Numărul de succese (r):



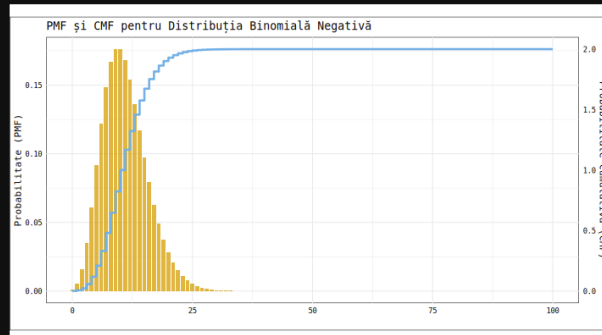
☒ Fixează r

Probabilitatea de succes (p):



☒ Fixează p

Reprezentare Grafică



Numărul de răspunsuri corecte la un test cu un număr fix de întrebări

Un student răspunde la un test cu un număr fix de întrebări. Fiecare întrebare este un experiment Bernoulli: Succes (răspunsul este corect, probabilitate p) sau Eșec (răspunsul este greșit, probabilitate $1 - p$).

Numărul de produse conform într-un lot de dimensiune fixă

Un lot de produse este testat, iar fiecare produs este un experiment Bernoulli: Succes (produsul este conform, probabilitate p) sau Eșec (produsul este neconform, probabilitate $1 - p$).

Codul Aplicației

Aplicația a fost organizată în mai multe fișiere pentru a separa clar logica și pentru a facilita reutilizarea funcțiilor. Această structurare are ca scop eliminarea redundanței și creșterea clarității codului, permițând adăugarea sau modificarea distribuțiilor fără a perturba structura generală a aplicației.

app.R

Fișierul principal **app.R** conține doar partea de inițializare a aplicației Shiny, inclusiv încărcarea fișierelor **ui.R** și **server.R**. Aceasta facilitează gestionarea centralizată a componentelor aplicației. Fiecare formulare a Repartiției Negativ Binomiale are propriile fișiere în directorul **server**, în care sunt definite logicile de generare a graficelor corespunzătoare.

```
1 library(shiny)
2 library(bslib)
3 library(ggplot2)
4
5 source("ui.R")
6 source("server.R")
7
8 shinyApp(ui = ui, server = server)
```

ui.R

Interfața utilizatorului este creată folosind funcția **fluidPage**, care organizează interfața pe tab-uri, câte unul pentru fiecare formulare a Repartiției Negativ Binomiale. Fiecare tab este generat folosind funcția personalizată **create_tab**, care asociază un input specific formularii respective, o zonă de afișare a graficului și situații ipotetice în care se poate aplica.

```
1 create_tab <- function(tab_title, title, img1_src, img2_src, tab_number,
2                               titlu_ex_1, ex_1, titlu_ex_2, ex_2) {
3   tabPanel(
4     tab_title,
5     div(
6       class = "container",
7       h1(title),
8       div(
9         class = "d-flex justify-content-center",
10        img(src = img1_src, style = "filter: invert(1);height:4vh;"),
11        img(src = img2_src, style = "filter: invert(1);height:4vh;")
12      ),
13      div(
14        class = "row",
15        div(
16          class = "col-4",
17          tags$h3("Input:"),
18          sliderInput(paste0("r", tab_number), "Numarul de succese (r):", min = 1,
19                                max = 50, value = 10, step = 1),
20          checkboxInput(paste0("fix_r", tab_number), "Fixeaza r", value = TRUE),
21          sliderInput(paste0("p", tab_number), "Probabilitatea de succes (p):",
22                                min = 0.01, max = 1, value = 0.5, step = 0.01),
23          checkboxInput(paste0("fix_p", tab_number), "Fixeaza p", value = FALSE)
24        ),
25        div(
26          class = "col-8",
27          h4("Reprezentare Grafica"),
28          div(
29            style = "padding: 10px; margin-bottom: 10px;",
30            plotOutput(outputId = paste0("mass_function_plot_", tab_number)),
31          )
32        )
33      ),
34      div(
35        class = "row",
36        div(
37          class = "col-6",
38          h4(titlu_ex_1),
39          p(ex_1)
40        ),
```

```

41     div(
42       class = "col-6",
43       h4(titlu_ex_2),
44       p(ex_2)
45     )
46   )
47 )
48 )
49 }
50 ui <- fluidPage(
51   theme = bs_theme(version = 4, bg = "#101010", fg = "#FFF", primary = "#E69F00",
52     base_font = font_google("Inconsolata")),
53   navbarPage(
54     "Repartitia Negativ Binomiala",
55     create_tab(
56       "Formulara 1",
57       "Numarul de esecuri inainte de a obtine un numar fix de succese (r)",
58       "https://wikimedia.org/api/rest_v1/media/math/render/svg/1a26b86be5f2...",
59       "https://wikimedia.org/api/rest_v1/media/math/render/svg/29944ccb6e33...",
60       1,
61       "Testarea produselor pana la gasirea unui numar fix de defecte",
62       "0 companie produce componente electronice si doreste sa testeze cate ...",
63       "Numarul de incercari pana la a prinde un numar fix de pesti",
64       "Un pescar merge la pescuit si doreste sa prinda un numar fix de pesti..."),
65     create_tab(
66       "Formulara 2",
67       "Numarul total de incercari necesare pentru a obtine un numar fix ...",
68       "https://wikimedia.org/api/rest_v1/media/math/render/svg/ccc5e37984e75...",
69       "https://wikimedia.org/api/rest_v1/media/math/render/svg/122b97516ba76...",
70       2,
71       "Testarea medicamentelor pana la obtinerea unui numar fix de ...",
72       "Un cercetator testeaza un nou medicament si doreste sa obtina un numar ...",
73       "Numarul de aruncari ale unui zar pana la obtinerea unui numar fix de sase",
74       "Un jucator arunca un zar si doreste sa obtina un numar fix de sase..."),
75     create_tab(
76       "Formulara 3",
77       "Numarul total de incercari necesare pentru a obtine un numar ...",
78       "https://wikimedia.org/api/rest_v1/media/math/render/svg/ccc5e37984e...",
79       "https://wikimedia.org/api/rest_v1/media/math/render/svg/7fd538f62a6...",
80       3,
81       "Testarea produselor pana la gasirea unui numar fix de produse neconforme",
82       "Un controlor de calitate testeaza produse si doreste sa gaseasca un ...",
83       "Numarul de intrebari pana la obtinerea unui numar fix de raspunsuri gresite",
84       "Un student raspunde la intrebari si doreste sa obtina un numar fix ..."),
85     create_tab(
86       "Formulara 4",
87       "Numarul de succese inainte de a obtine un numar fix de esecuri (r)",
88       "https://wikimedia.org/api/rest_v1/media/math/render/svg/1a26b86be5f...",
89       "https://wikimedia.org/api/rest_v1/media/math/render/svg/5d39d1344ae...",
90       4,
91       "Numarul de clienti satisfacuti inainte de a obtine un numar fix de ...",
92       "Un magazin doreste sa obtina un numar fix de reclamatii de la clienti. ...",
93       "Numarul de lovituri la tinta inainte de a obtine un numar fix de rateuri",
94       "Un sportiv trage la tinta si doreste sa obtina un numar fix de rateuri..."),
95     create_tab(
96       "Formulara 5",
97       "Numarul de succese intr-un numar fix de incercari (n)",
98       "https://wikimedia.org/api/rest_v1/media/math/render/svg/d23e3eb017f...",
99       "https://wikimedia.org/api/rest_v1/media/math/render/svg/09ed5ae47e20d...",
100      5,
101      "Numarul de raspunsuri corecte la un test cu un numar fix de intrebari",
102      "Un student raspunde la un test cu un numar fix de intrebari. Fiecare ...",
103      "Numarul de produse conform intr-un lot de dimensiune fixa",
104      "Un lot de produse este testat, iar fiecare produs este un ...")
105   )
106 )

```

server.R

Acesta este punctul central al logicii serverului.

```
1 source("server/form_1.R")
2 source("server/form_2.R")
3 source("server/form_3.R")
4 source("server/form_4.R")
5 source("server/form_5.R")
6 source("server/global.R")
7
8
9 server <- function(input, output, session) {
10   # Creare reactive pentru fiecare tab
11   reactive_values1 <- create_reactive_values(input, session, "1")
12   reactive_values2 <- create_reactive_values(input, session, "2")
13   reactive_values3 <- create_reactive_values(input, session, "3")
14   reactive_values4 <- create_reactive_values(input, session, "4")
15   reactive_values5 <- create_reactive_values(input, session, "5")
16
17   output$mass_function_plot_1 <- render_plot_1(reactive_values1$r_value,
18                                                reactive_values1$p_value)
19   output$mass_function_plot_2 <- render_plot_2(reactive_values2$r_value,
20                                                reactive_values2$p_value)
21   output$mass_function_plot_3 <- render_plot_3(reactive_values3$r_value,
22                                                reactive_values3$p_value)
23   output$mass_function_plot_4 <- render_plot_4(reactive_values4$r_value,
24                                                reactive_values4$p_value)
25   output$mass_function_plot_5 <- render_plot_5(reactive_values5$r_value,
26                                                reactive_values5$p_value)
27 }
```

formulare1.R

Modul în care am calculat funcția de masă este manual, cu formula specifică pentru fiecare formulare, iar funcția de repartiție este calculată ca sumă parțială, folosind funcția `cumsum`.

```
1 negative_binomial_pmf_1 <- function(k, r, p) {
2   binomial_coefficient <- choose(k + r - 1, k)
3   probability <- binomial_coefficient * p^r * (1 - p)^k
4   return(probability)
5 }
6
7 render_plot_1 <- function(r_value, p_value) {
8   renderPlot({
9     r <- r_value()
10    p <- p_value()
11    x <- 0:100
12
13    pmf_values <- sapply(x, function(k) negative_binomial_pmf_1(k, r, p))
14    cmf_values <- cumsum(pmf_values)
15
16    df <- data.frame(x = x, pmf = pmf_values, cmf = cmf_values)
17    scaling_factor <- max(df$pmf) / max(df$cmf)
18
19    render_plot(df, scaling_factor)
20  })
21 }
```

Concluzii

Aplicația web dezvoltată folosind Shiny oferă o modalitate interactivă și intuitivă de a explora și înțelege repartiția binomială negativă. Utilizatorii pot vizualiza grafice animate ale funcției de masă (PMF) și funcției de repartiție (CDF), ajustând parametri precum numărul de succese (r) și probabilitatea de succes (p). Aceste vizualizări dinamice ajută la înțelegerea impactului parametrilor asupra distribuției. Aplicația include și exemple practice din viața reală, cum ar fi testarea produselor până la găsirea unui număr fix de defecte sau numărul de încercări necesare pentru a prinde un anumit număr de pești. Prin această abordare, utilizatorii pot vedea cum repartiția binomială negativă se aplică în diverse contexte, consolidând înțelegerea teoretică și practică a conceptului. Aplicația este un instrument educativ valoros, care combină teoria statistică cu aplicații practice, facilitând învățarea interactivă.

Cerinta III

Descrierea Problemei

Scopul acestui proiect este de a construi o aplicație web interactivă folosind framework-ul **Shiny** din **R**, pentru a reprezenta grafic funcțiile de repartiție (CDF) ale unor variabile aleatoare definite conform cerinței din enunț. Aplicația permite vizualizarea CDF-urilor pentru distribuții normale, exponențiale, binomiale, Poisson.

Aspecte Teoretice

Pentru dezvoltarea aplicației, am utilizat următoarele aspecte teoretice importante:

- **Distribuția Normală:** Este definită ca o variabilă aleatoare continuă $X \sim N(\mu, \sigma^2)$, unde μ reprezintă media, iar σ^2 varianța.

Formula funcției de repartiție cumulativă:

$$F(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$

- **Distribuția Exponențială:** O variabilă aleatoare $X \sim \text{Exp}(\lambda)$ este folosită pentru modelarea timpilor de așteptare între evenimente.

Formula funcției de repartiție cumulativă:

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & \text{pentru } x \geq 0 \\ 0, & \text{pentru } x < 0 \end{cases}$$

- **Distribuția Poisson:** Reprezintă numărul de evenimente într-un interval fix de timp și este notată $X \sim \text{Poisson}(\lambda)$.

Formula funcției de repartiție cumulativă:

$$F(x) = P(X \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \frac{\lambda^k e^{-\lambda}}{k!}$$

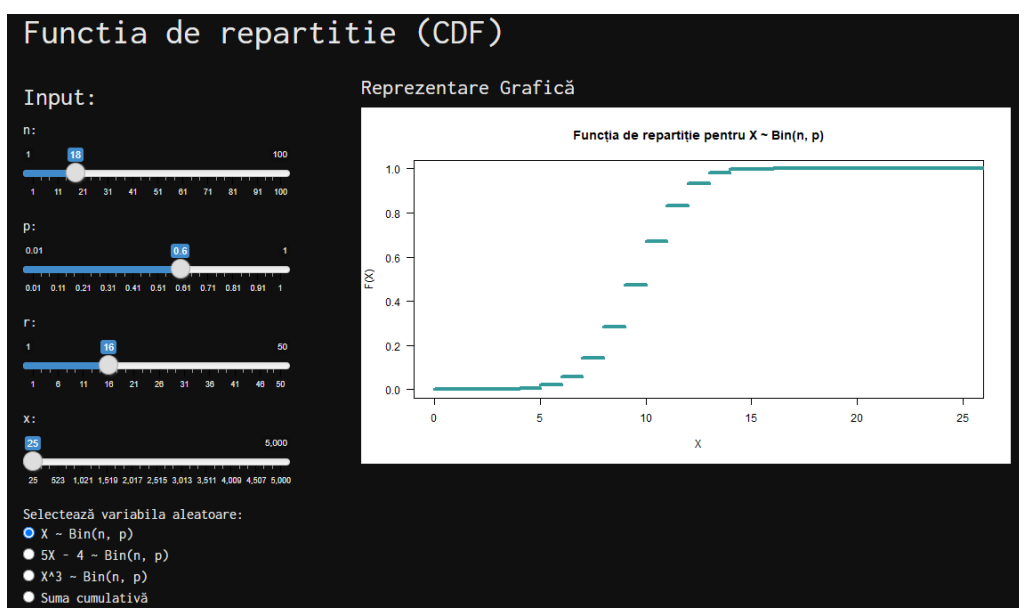
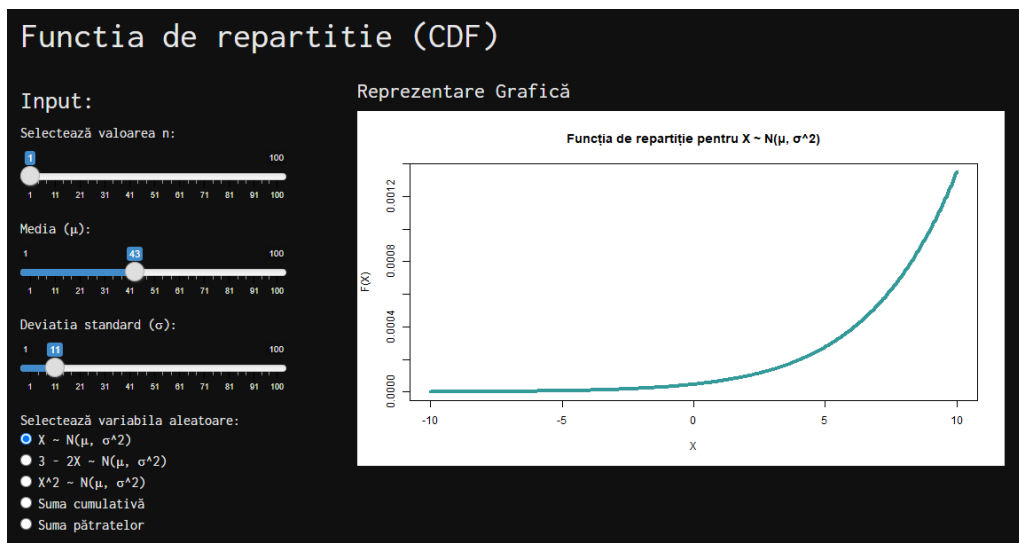
- **Distribuția Binomială:** Este definită de parametrii n și p și descrie numărul de succese în n experimente independente.

Formula funcției de repartiție cumulativă:

$$F(x) = P(X \leq x) = \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k (1-p)^{n-k}$$

Reprezentări Grafice

Aplicația oferă reprezentări grafice ale funcțiilor de repartiție pentru fiecare dintre variabilele aleatoare definite în cerințe. În cele ce urmează, vom ilustra câteva capturi de ecran din aplicație.



Pachete Software Folosite

Am folosit următoarele pachete software pentru implementarea proiectului:

- **shiny**: Crearea aplicației interactive.
- **bslib**: Personalizarea interfeței grafice a aplicației folosind teme Bootstrap.
- **graphics**: Generarea reprezentărilor grafice ale funcțiilor de repartiție.

Codul Aplicației

Codul aplicației a fost structurat în mai multe fișiere pentru a asigura o separare clară a logicii și o reutilizare ușoară a funcțiilor. Această abordare are ca scop evitarea redundanței și îmbunătățirea clarității codului, permițând adăugarea sau modificarea distribuțiilor fără a afecta structura generală a aplicației.

app.R

Fișierul principal **app.R** conține doar partea de inițializare a aplicației Shiny, inclusiv încărcarea fișierelor **ui.R**, **utils.R** și **server/server.R**. Aceasta facilitează gestionarea centralizată a componentelor aplicației. Fiecare distribuție are propriile fișiere în directorul **server**, în care sunt definite logicile de generare a graficelor corespunzătoare.

```
1 library(shiny)
2 library(bslib)
3
4 source("utils.R")
5 source("ui.R")
6 source("server/server.R")
7
8 shinyApp(ui = ui, server = server)
```

ui.R

Interfața utilizatorului este creată folosind funcția **fluidPage**, care organizează interfața pe tab-uri, câte unul pentru fiecare distribuție (normală, exponențială, binomială, Poisson). Fiecare tab este generat folosind funcția personalizată **create_tab**, care asociază un input specific distribuției respective și o zonă de afișare a graficului.

```
1 source("server/server.R")
2
3 create_tab <- function(tab_title, title, distribution) {
4   tabPanel(
5     tab_title,
6     div(
7       class = "container",
8       h1(title),
9       div(
10        class = "row",
11        div(
12          class = "col-4",
13          tags$h3("Input:"),
14          switch(
15            distribution,
16            BINOMIALA = create_binom_slider(),
17            NORMALA_STANDARD = create_std_normal_slider(),
18            NORMALA = create_normal_slider(),
19            EXPONENTIALA = create_exponential_slider(),
20            POISSON = create_pois_slider()
21          )
22        ),
23        div(
24          class = "col-8",
25          h4("Reprezentare Grafica"),
26          get_output_distribution(distribution)
27        )
28      )
29    )
30  }
31
32 }
```



```

33 ui <- fluidPage(
34   theme = bs_theme(
35     version = 4,
36     bg = "#101010",
37     fg = "#FFF",
38     primary = "#E69F00",
39     base_font = font_google("Inconsolata")
40   ),
41   navbarPage(
42     "CDF",
43     create_tab(
44       "Normala Standard",
45       "Functia de repartitie (CDF)",
46       NORMALA_STANDARD
47     ),
48     create_tab(
49       "Normala",
50       "Functia de repartitie (CDF)",
51       NORMALA
52     ),
53     create_tab(
54       "Exponentiala",
55       "Functia de repartitie (CDF)",
56       EXPONENTIALA
57     ),
58     create_tab(
59       "Binomiala",
60       "Functia de repartitie (CDF)",
61       BINOMIALA
62     ),
63     create_tab(
64       "Poisson",
65       "Functia de repartitie (CDF)",
66       POISSON
67     )
68   )
69 )

```

server.R

Acesta este punctul central al logicii serverului. Codul serverului include apeluri către fișierele individuale ale fiecărei distribuții, cum ar fi `std_normal.R`, `normal.R`, `binom.R` etc. Acest lucru permite fiecărei distribuții să fie tratată separat și modular, facilitând adăugarea sau modificarea distribuțiilor fără a afecta alte componente.

```

1  source("server/std_normal.R")
2  source("server/normal.R")
3  source("server/exponential.R")
4  source("server/binom.R")
5  source("server/poisson.R")
6
7  server <- function(input, output, session) {
8    std_normal_server(input, output, session)
9    normal_server(input, output, session)
10   exponential_server(input, output, session)
11   binom_server(input, output, session)
12   pois_server(input, output, session)
13 }

```

utils.R

Acest fișier conține definiții comune, cum ar fi identificatorii fiecărei distribuții și funcția `get_output_distribution` care returnează componenta grafică potrivită pentru distribuția selectată. Aceasta contribuie la eliminarea duplicării codului și la centralizarea logicii de redirectionare a graficelor.

```
1  NORMALA_STANDARD <- "NORMALA_STANDARD"
2  NORMALA <- "NORMALA"
3  EXPONENTIALA <- "EXPONENTIALA"
4  BINOMIALA <- "BINOMIALA"
5  POISSON <- "POISSON"
6
7  get_output_distribution <- function(distribution) {
8    switch(
9      distribution,
10     NORMALA_STANDARD = plotOutput("std_normal_plot"),
11     NORMALA = plotOutput("normal_plot"),
12     EXPONENTIALA = plotOutput("exponential_plot"),
13     BINOMIALA = plotOutput("binom_plot"),
14     POISSON = plotOutput("pois_plot")
15   )
16 }
```

normals.R

Generarea graficelor: Codul pentru reprezentarea grafică este similar pentru toate distribuțiile, însă exemplificarea este prezentată doar pentru distribuția normală în fișierul `normals.R`. Fiecare grafic este generat pe baza parametrilor selectați de utilizator (cum ar fi media μ și deviația standard σ) și utilizează funcția corespunzătoare de calcul al funcției de repartitie cumulativă. Iată principalele tipuri de grafice generate:

- ****Graficul funcției de repartitie cumulativă (CDF)**:** Este reprezentată funcția $F(x) = P(X \leq x)$, calculată cu funcția R corespunzătoare fiecărei distribuții (de exemplu, `pnorm` pentru distribuția normală).
- ****Transformări ale variabilelor**:** De exemplu, pentru $3 - 2X$ sau X^2 , transformările sunt aplicate asupra datelor înainte de a calcula și reprezenta $F(x)$.
- ****Sumă cumulativă**:** Pentru a reprezenta suma cumulativă $\sum X_i$ sau $\sum X_i^2$, se utilizează funcția `cumsum` din R, iar graficul este construit prin afișarea acestor sume pentru fiecare n .

```
1  create_normal_slider <- function() {
2    tagList(
3      sliderInput(
4        inputId = "normal_n",
5        label = "Selectează valoarea n:",
6        min = 1,
7        max = 100,
8        value = 10,
9        step = 1
10     ),
11     sliderInput(
12       inputId = "normal_mu",
13       label = "Media (\u03BC):",
14       min = 1,
15       max = 100,
16       value = 10,
17       step = 1
18     ),
19     sliderInput(
20       inputId = "normal_sigma",
21       label = "Deviația standard (\u03C3):",
22       min = 1,
23       max = 100,
```

```

24     value = 10,
25     step = 1
26   ),
27   radioButtons(
28     inputId = "normal_var",
29     label = "Selecteaza variabila aleatoare:",
30     choices = list(
31       "X ~ N(\u03BC, \u03C3^2)" = "var1",
32       "3 - 2X ~ N(\u03BC, \u03C3^2)" = "var2",
33       "X^2 ~ N(\u03BC, \u03C3^2)" = "var3",
34       "Suma cumulativa" = "var4",
35       "Suma patratelor" = "var5"
36     ),
37     selected = "var1"
38   )
39 )
40 }
41
42 normal_server <- function(input, output, session) {
43   output$normal_plot <- renderPlot({
44     var <- input$normal_var
45
46     if (var == "var1") {
47       x <- seq(-10, 10, length.out = 500)
48       cdf_values <- pnorm(x, mean = input$normal_mu, sd = input$normal_sigma)
49       plot(
50         x, cdf_values,
51         type = "l",
52         lwd = 4,
53         col = "#339999",
54         xlab = "X",
55         ylab = "F(X)",
56         main = "Functia de repartitie pentru X ~ N(\u03BC, \u03C3^2)"
57       )
58     } else if (var == "var2") {
59       x <- seq(-10, 10, length.out = 500)
60       transformed_x <- 3 - 2 * x
61       cdf_values <- pnorm(transformed_x, mean = input$normal_mu, sd = input$normal_sigma)
62       plot(
63         x, cdf_values,
64         type = "l",
65         lwd = 4,
66         col = "#FF6666",
67         xlab = "3 - 2X",
68         ylab = "F(3 - 2X)",
69         main = "Functia de repartitie pentru 3 - 2X ~ N(\u03BC, \u03C3^2)"
70       )
71     } else if (var == "var3") {
72       x <- seq(-10, 10, length.out = 500)
73       transformed_x <- x^2
74       cdf_values <- pnorm(transformed_x, mean = input$normal_mu, sd = input$normal_sigma)
75       plot(
76         x, cdf_values,
77         type = "l",
78         lwd = 4,
79         col = "#3399FF",
80         xlab = "X^2",
81         ylab = "F(X^2)",
82         main = "Functia de repartitie pentru X^2 ~ N(\u03BC, \u03C3^2)"
83       )
84     } else if (var == "var4") {
85       n <- input$normal_n
86       X <- rnorm(n, mean = input$normal_mu, sd = input$normal_sigma)
87       S_n <- cumsum(X)
88       plot(
89         1:n, S_n,

```

```

90     type = "l",
91     lwd = 2,
92     col = "#FF9900",
93     xlab = "n",
94     ylab = "\u2211 X_i",
95     main = "Suma cumulativa a variabilelor aleatoare X_i"
96   )
97 } else if (var == "var5") {
98   n <- input$normal_n
99   X <- rnorm(n, mean = input$normal_mu, sd = input$normal_sigma)
100   S_n2 <- cumsum(X^2)
101   plot(
102     1:n, S_n2,
103     type = "l",
104     lwd = 2,
105     col = "#33CC33",
106     xlab = "n",
107     ylab = "\u2211 X_i^2",
108     main = "Suma cumulativa a patratelor variabilelor X_i"
109   )
110 }
111 })
112 }

```

Dificultăți Întâmpinate

Dezvoltarea aplicației a implicat rezolvarea unor provocări tehnice importante legate de gestionarea parametrilor și a reprezentărilor grafice pentru fiecare distribuție. Principalele dificultăți întâmpinate sunt:

- **Ajustarea automată a parametrilor pentru fiecare distribuție:** Fiecare distribuție are un set unic de parametri care trebuie adaptați corect pentru a asigura o reprezentare grafică precisă a funcției de repartiție cumulativă (CDF). De exemplu, distribuția normală implică media μ și deviația standard σ , în timp ce distribuția binomială depinde de numărul de experimente n și probabilitatea de succes p . Provocarea a fost legată de crearea unui sistem flexibil în care utilizatorul să poată modifica rapid parametrii în interfață, iar graficele să se actualizeze automat în funcție de aceștia. Această problemă a fost rezolvată prin implementarea unor funcții modulare (de exemplu, `create_normal_slider()` și `create_binom_slider()`) care generează dinamic controale de input specifice fiecărei distribuții.

Probleme Deschise

Deși aplicația funcționează corespunzător și oferă o interfață interactivă pentru vizualizarea funcțiilor de repartiție cumulativă, există câteva direcții de dezvoltare viitoare și probleme deschise:

- **Posibilitatea extinderii aplicației pentru a suporta și alte distribuții:** Momentan, aplicația suportă distribuțiile normală, exponențială, binomială și Poisson. Cu toate acestea, există o varietate de alte distribuții utile în statistică, cum ar fi distribuțiile Bernoulli, binomiale sau geometrice, care ar putea fi integrate cu ușurință în structura modulară a aplicației. O extindere în această direcție ar implica adăugarea fișierelor corespunzătoare în directorul `server` și crearea funcțiilor de input specifice fiecărei noi distribuții.

Concluzii

Funcționalitățile existente permit utilizatorilor să vizualizeze funcțiile de repartiție ale variabilelor normale, exponențiale, binomiale și Poisson, dar structura modulară face posibilă integrarea rapidă a altor distribuții. De asemenea, prin intermediul graficelor dinamice, utilizatorii pot explora nu doar distribuțiile de bază, ci și diverse transformări ale variabilelor aleatoare, cum ar fi transformări liniare și sume cumulative.