Lucrarea nr. 2: Determinarea legii de repartiţie

Tiberiu Laurian, Radu–Florin Mirică

2014

1 Scopul lucrării

Determinarea parametrilor repartiţiei normale şi a repartiţiei de tip Weibull corespunzătoare şirului de rezultate date şi aplicarea testului *Kolmogorov–Smirnov* în vederea verificării concordanţei dintre repartiţiile teoretice (normală şi Weibull) si cea empirică dedusă din rezultatele experimentale.

2 Punerea problemei

S-au observat un număr de 30 autovehicule pe durata a 25000 km. Să se determine parametrii funcţiilor de repartiţie Weibull şi să se testeze concordanţa repartiţiilor teoretice cu cea empirică utilizând testul *Kolmogorov–Smirnov*. Datele, exprimate în km parcurşi, sunt:

$x =$

5140	12450	200	8430	17600	11300	6012	430	2040	540
9710	2830	15870	1635	9000	15700	5400	9111	23130	1000
8200	19300	2280	14080	7510	16102	412	17120	21320	16600

3 Rezolvare

Rezolvarea problemei se va face cu ajutorul calculatorului, utilizând mediul de programare *Octave*. În acest scop se vor parcurge următorii paşi:

1. Introducerea datelor de intrare

2. Determinarea parametrilor repartiţiei normale

3. Determinarea parametrilor repartiţiei Weibull
   (a) utilizând metoda celor mai mici pătrate
   (b) utilizând metoda verosimilităţii maxime
4. Stabilirea domeniilor de încredere ale parametrilor legilor de repartiţie
   (a) pentru repartiţia normală
   (b) pentru repartiţia Weibull
5. Verificarea concordanţei dintre repartiţiile teoretice şi rezultatele experimentale
   (a) pentru repartiţia normală
   (b) pentru repartiţia Weibull – în varianta de la pct. 3.3.1.
   (c) pentru repartiţia Weibull – în varianta de la pct. 3.3.2.
6. Verificarea grafică cu ajutorul reţelelor de probabilitate
7. Concluzii

3.1 Introducerea datelor de intrare

Datele de intrare, specificate în şirul x, se vor scrie într-un fişier ASCII pe o coloană, acesta salvându-se cu numele “x.dat”.

Importarea datelor din fişier se va face cu şirul de comenzi: **fopen**, **fscanf** şi **fclose**.

3.2 Determinarea parametrilor repartiţiei normale (Gauss-Laplace)

Se calculează media aritmetică şi abaterea medie pătratică necorectată şi corectată:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad \sigma = \sqrt{\frac{\sum\limits_{i=1}^{n}(x_i - \overline{x})^2}{n}}; \quad s = \sigma\sqrt{\frac{n}{n-1}}.$$

3.3 Determinarea parametrilor repartiţiei Weibull

Se vor determina parametrii repartiţiei Weibull biparametrice ($\gamma = 0$) utilizând două metode: metoda celor mai mici pătrate şi metoda verosimilităţii maxime.

Se definesc următoarele variabile auxiliare: $r = n$ (nr. de defectări) şi şirul de numere naturale $i = 1 \dots n$, necesare definirii estimatorului repartiţiei empirice:

$$F_i = \frac{i - 0.5}{n}.$$

3.3.1 Metoda celor mai mici pătrate

Se vor defini cei doi parametri ai repartiţiei Weibull utilizând următoarele formule:

$$\beta_1(x) = \frac{r\left(\sum_{i=1}^{r} \ln\left(\ln\frac{1}{1-F_i}\right)\ln(x_i-\gamma)\right) - \left(\sum_{i=1}^{r}\ln\left(\ln\frac{1}{1-F_i}\right)\right)\sum_{i=1}^{r}\ln(x_i-\gamma)}{r\sum_{i=1}^{r}\ln(x_i-\gamma)^2 - \left(\sum_{i=1}^{r}\ln(x_i-\gamma)\right)^2}$$

$$\eta_1(x) = \exp\left[-\frac{\left(\sum_{i=1}^{r}\ln\left(\ln\frac{1}{1-F_i}\right)\right)\left(\sum_{i=1}^{r}\ln(x_i-\gamma)^2\right) - \left(\sum_{i=1}^{r}\ln\left(\ln\frac{1}{1-F_i}\right)\ln(x_i-\gamma)\right)\sum_{i=1}^{r}\ln(x_i-\gamma)}{r\sum_{i=1}^{r}\ln(x_i-\gamma)^2 - \left(\sum_{i=1}^{r}\ln(x_i-\gamma)\right)^2}\cdot\frac{1}{\beta_1(x)}\right]$$

3.3.2 Metoda verosimilităţii maxime

Se definesc două funcţii de variabile β şi η şi, punând condiţia egalităţii cu zero, se determină cele două necunoscute β şi η prin metode numerice.

$$\frac{\sum_{i=1}^{n} x_i^{\beta_2}\ln(x_i)}{\sum_{i=1}^{n} x_i^{\beta_2}} + \frac{\sum_{i=1}^{n} -\ln(x_i)}{n} - \frac{1}{\beta_2} = 0$$

$$\frac{\sum_{i=1}^{n} x_i^{\beta_2}}{n} - \eta_2^{\beta_2} = 0$$

3.4 Stabilirea domeniilor de încredere ale parametrilor legilor de repartiţie

Pentru stabilirea domeniilor de încredere, se va defini nivelul de semnificaţie α =10% sau 5%.

3.4.1 Cazul repartiţiei normale

Se definesc limitele minimă şi maximă pentru cei doi parametri ai repartiţiei normale (media aritmetică şi abaterea medie pătratică).

$$\overline{x}_{min} = \overline{x} - \left| \frac{\sigma}{\sqrt{n}} \cdot \mathrm{t}\left(\frac{\alpha}{2}, n-1\right) \right|; \quad \overline{x}_{max} = \overline{x} + \left| \frac{\sigma}{\sqrt{n}} \cdot \mathrm{t}\left(\frac{\alpha}{2}, n-1\right) \right| \tag{1}$$

$$\sigma_{min} = \sigma \sqrt{\frac{n-1}{\chi^2\left(1-\frac{\alpha}{2}, n-1\right)}}; \quad \sigma_{max} = \sigma \sqrt{\frac{n-1}{\chi^2\left(\frac{\alpha}{2}, n-1\right)}} \tag{2}$$

Se verifică apoi dacă valorile parametrilor (calculate la pct. 3.3) se încadrează în limitele de mai sus.

Funcţiile $\mathrm{t}(\alpha, n)$ şi $\chi^2(\alpha, n)$ reprezintă testele de stabilire a limitelor de încredere, t–Student, respectiv hi pătrat, pentru nivelul de semnificaţie α şi numărul de grade de libertate n.

3.4.2 Cazul repartiţiei Weibull

Se definesc limitele minimă şi maximă între care trebuie să se încadreze valoarea lui η.

$$\eta_{min} = \frac{2\sum\limits_{i=1}^{n}(x_i)^\beta}{\chi^2\left(1-\frac{\alpha}{2}, 2n\right)}; \quad \eta_{max} = \frac{2\sum\limits_{i=1}^{n}(x_i)^\beta}{\chi^2\left(\frac{\alpha}{2}, 2n\right)} \tag{3}$$

Limitele de mai sus se vor defini atât pentru parametrii calculaţi la pct. 3.3.1, cât şi pentru cei calculati la pct. 3.3.2.

Se verifică apoi dacă η se încadreaza între limitele domeniului de încredere (pentru ambele cazuri).

3.5 Verificarea concordanţei dintre repartiţiile teoretice şi rezultatele experimentale

Se va aplica testul Kolmogorov–Smirnov pentru verificarea concordanţei dintre repartiţiile teoretice şi cea empirică dată de rezultatele experimentale.

3.5.1 Pentru repartiţia normală

Se vor defini: o funcţie care determină un şir de valori distribuite conform unei repartiţii normale ($F_n(x)$); diferenţa maximă dintre şirul dat de F_n şi vectorul rezultatelor experimentale (dn_n); parametrul λ_n; funcţia Kolmogorov ($K(\lambda)$); nivelul de semnificaţie α_n.

$$F_n(x) = \Phi\left(\frac{y-\overline{x}}{\sigma}\right) = \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x-\overline{x}}{\sigma\sqrt{2}}\right)\right]$$

unde

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}}\int_{-x}^{x} e^{-t^2}\,dt$$

$$d_n = F_n(x) - F$$

$$dn_n = \max\left(|d_{n1}|, |d_{nn}|\right)$$

$$\lambda_n = dn_n\sqrt{n})$$

$$K(\lambda_n) = \sum_{k=-10}^{10}(-1)^k\exp(-2k^2\lambda_n^2)$$

$$\alpha_n = 1 - K(\lambda_n)$$

Între repartiţia teoretică şi cea empirică există concordanţă dacă se verifică inegalitatea $\alpha_n \geq \alpha$.

3.5.2 Pentru repartiţia Weibull – cazul parametrilor de la pct. 3.3.1.

Se vor redefini relaţiile de la pct. 3.5.1 utilizând alte notaţii. Redefinirea funcţiei Kolmogorov nu este necesară.

$$F_{w1}(y) = 1 - \exp\left[-\left(\frac{x}{\eta_1}\right)^{\beta_1}\right]$$

$$d_{w1} = F_{w1}(x) - F$$

$$dn_{w1} = \max\left(|d_{w1_1}|, |d_{w1_n}|\right)$$

$$\lambda_{w1} = dn_{w1}\sqrt{n}$$

$$K(\lambda_{w1}) = \sum_{k=-10}^{10}(-1)^k\exp(-2k^2\lambda_{w1}^2)$$

$$\alpha_{w1} = 1 - K(\lambda_{w1})$$

3.5.3 Pentru repartiţia Weibull – cazul parametrilor de la pct. 3.3.2.

Se va proceda similar punctului anterior.

$$F_{w2}(y) = 1 - \exp\left[-\left(\frac{x}{\eta_2}\right)^{\beta_2}\right]$$

$$d_{w2} = F_{w2}(x) - F$$

$$dn_{w2} = \max\left(|d_{w2_1}|, |d_{w2_n}|\right)$$

$$\lambda_{w2} = dn_{w2}\sqrt{n}$$

$$K(\lambda_{w2}) = \sum_{k=-10}^{10} (-1)^k \exp(-2k^2\lambda_{w2}^2)$$

$$\alpha_{w2} = 1 - K(\lambda_{w2})$$

3.6 Verificarea grafică cu ajutorul reţelelor de probabilitate

Se definesc două şiruri ce vor da variaţiile de probabilitate pentru repartiţia normală şi pentru cea empirică şi se vor reprezenta grafic cele doua funcţii.

$$\Phi^{-1}(F) \;=\; \sqrt{2}\ \mathrm{erf}^{-1}(2F-1), \quad F \in (0,1).$$

$$x_0 = \frac{x - \overline{x}}{\sigma}$$

Se procedează similar pentru cele două repartiţii Weibull.

$$y = \ln\left(\ln\frac{1}{1-F}\right); \quad w_1 = \beta_1 \ln\frac{x}{\eta_1}; \quad w_2 = \beta_2 \ln\frac{x}{\eta_2}.$$

4 Concluzii

Se vor scrie câteva comentarii referitoare la respectarea domeniilor de încredere, verificarea concordanţei repartiţiilor teoretice cu cea empirică (comparaţie α_n, α_{w1}, α_{w2}), precum şi la reprezentările grafice (suprapunerea sau nu a curbelor).

Lucrarea se încheie cu un raport, în care se vor scrie textul problemei, datele de intrare şi rezultatele obţinute.

5 Listing MATLAB/Octave pentru rezolvarea problemei

```
# ----------------------------
#
# FSM - Lucrare laborator 2
#
# ----------------------------
#
#-----------------------------------------------------------------------------------------
#
clear all;
global x n;
# ------------------------------------------
# 1. Introducerea datelor de intrare
# ------------------------------------------
fname_in = "x.dat";
fname_out = "rezultate.txt"
fid1 = fopen(fname_in, "r");
   x = fscanf(fid1,"%f",Inf);
fclose(fid1);
x = sort(x);
# ------------------------------------------------------------------
# 2. Determinarea parametrilor repartitiei normale (Gauss-Laplace)
# ------------------------------------------------------------------
xm = mean(x,"a");
sigma = std(x,1);
s = std(x,0);
# ----------------------------------------------------
# 3. Determinarea parametrilor repartitiei Weibull
# ----------------------------------------------------
n = length(x);
r = n;
i = (1:1:n)';
F = (i-0.5)/n;
gama = 0;
#
# --- Metoda celor mai mici patrate ---
#
B1 = sum(log(log(1./(1-F))).*log(x-gama));
B2 = sum(log(log(1./(1-F))));
B3 = sum(log(x-gama));
B4 = sum(log(x-gama).^2);
beta = (r.*B1-B2.*B3)./(r.*B4-B3.^2);
#
E1 = B2;
E2 = B4;
E3 = B1;
E4 = B3;
E5 = B3;
eta = exp(-(E1.*E2-E3.*E4)/(r.*E2-E5.^2)*1./beta);
#
# --- metoda verosimilitatii maxime ---
#
options=optimset('MaxFunEvals',2400,'MaxIter',1600,'TolFun',1e-15,'TolX',1e-15);
[w,fval,info] = fsolve(@func,[0.5;5000],options);
beta_2 = w(1);
eta_2 = w(2);
#
h1 = figure(1);
```

```
plot(x,normpdf(x,xm,sigma),'-r','LineWidth',4);
grid on;
hold on;
xlabel("Nr. kilometri parcursi");
ylabel("Densitatea de repartitie");
plot(x,wblpdf(x,eta,beta),'-b','LineWidth',4);
plot(x,wblpdf(x,eta_2,beta_2),'-g','LineWidth',4);
legend("Distr. Normala","Distr. Weibull-met. celor mai mici patrate",
"Distr. Weibull-met. verosimilitatii maxime");
H = 4.5; W = 6;
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','dens_prob.png');
close();
# ------------------------------------------------------------------
# 4. Stabilirea domeniilor de incredere ale parametrilor legilor de repartitie
# ------------------------------------------------------------------
alfa = 0.1;       # nivel de semnificatie
m_min = xm - abs(sigma/sqrt(n)*tinv(alfa/2,n-1));
m_max = xm + abs(sigma/sqrt(n)*tinv(alfa/2,n-1));
sigma_min = sigma*sqrt((n-1)/chi2inv(1-alfa/2,n-1));
sigma_max = sigma*sqrt((n-1)/chi2inv(alfa/2,n-1));
eta_min = 2*sum(x.^beta)./chi2inv(1-alfa/2,2*n);
eta_max = 2*sum(x.^beta)./chi2inv(alfa/2,2*n);
# ----------------------------------------
# 5. Testul de concordanta Kolmogorov - Smirnov
# ----------------------------------------
# Determinarea acuratetei aproximarii valorilor experimentale
# cu legile teoretice de repartitie
Fn  = normcdf(x,xm,sigma);     # distributia cumulata de probabilitate
Fw1 = wblcdf(x,eta,beta);
Fw2 = wblcdf(x,eta_2,beta_2);
dn  = max(abs(Fn-F));
dw1 = max(abs(Fw1-F));
dw2 = max(abs(Fw2-F));
lambda_n  = dn*sqrt(n);
lambda_w1 = dw1*sqrt(n);
lambda_w2 = dw2*sqrt(n);
k = -10:10;
Kn  = sum((-1).^k.*exp(-k.^2.*2.*lambda_n.^2));
Kw1 = sum((-1).^k.*exp(-k.^2.*2.*lambda_w1.^2));
Kw2 = sum((-1).^k.*exp(-k.^2.*2.*lambda_w2.^2));
alfa_n  = 1-Kn;
alfa_w1 = 1-Kw1;
alfa_w2 = 1-Kw2;
#
h1 = figure(1);
plot(x,F,'.r','LineWidth',4);
grid on;
hold on;
plot(x,Fn,'-b','LineWidth',4);
plot(x,Fw1,'-g','LineWidth',4);
plot(x,Fw2,'-m','LineWidth',4);
xlabel("Nr. kilometri parcursi");
ylabel("F");
title("Probabilitatea cumulata");
legend("Repartitia empirica","Rep. Gauss","Rep. Weibull v.1",
"Rep. Weibull v.2");
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
```

```
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','Normala_cumulata.png');
close();
# ----------------------------------------------------------
# 6. Realizarea retelei de probabilitate de tip Gauss
# ----------------------------------------------------------
cuantila = norminv(F,0,1);
x0 = (x-xm)./sigma;
prob = log(log(1./(1-F)));
curbaw1 = beta*log(x./eta);
curbaw2 = beta_2*log(x./eta_2);
#
h1 = figure(1);
plot(x,cuantila,'.r','LineWidth',4);
grid on;
hold on;
plot(x,x0,'-b','LineWidth',4);
xlabel("Nr. kilometri parcursi");
ylabel("Cuantila");
legend("Cuantila","Curba de probabilitate normala");
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','Kolmogorov_n.png');
close();
#
h1=figure(1);
plot(log(x),prob,'.r','LineWidth',4);
grid on;
hold on;
plot(log(x),curbaw1,'-b','LineWidth',4);
plot(log(x),curbaw2,'-g','LineWidth',4);
xlabel("ln(x)");
ylabel("Prob");
legend("Prob. rep. empirice","Curba de probabilitate var.1",
"Curba de probabilitate var.2");
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','Kolmogorov_w.png');
close();
#
h1=figure(1);
[q s] = qqplot(x0);
[qn sn] = qqplot(cuantila);
plot(q,s,".r",qn,sn,"-b",'LineWidth',4);
grid on;
xlabel("Cuantilele distr. normale");
ylabel("Cuantilele datelor");
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','qqplot_n.png');
close();
#
h1=figure(1);
[qw1 sw1] = qqplot(curbaw1);
[qw2 sw2] = qqplot(curbaw2);
```

```
[qp sp]    = qqplot(prob);
plot(qp,sp,".r",qw1,sw1,"-b",qw2,sw2,"-g",'LineWidth',4);
grid on;
xlabel("Cuantilele distr. Weibull");
ylabel("Cuantilele datelor");
set(h1,'PaperUnits','inches')
set(h1,'PaperOrientation','portrait');
set(h1,'PaperSize',[H,W])
set(h1,'PaperPosition',[0,0,W,H])
print(h1,'-dpng','-color','qqplot_w.png');
close();
#
# ----------------------------------------
# Salvare rezultate
# ----------------------------------------
fout = fopen(fname_out,"w");
   fprintf(fout,"DETERMINAREA LEGII DE REPARTITIE\n\n\n");
   fprintf(fout,"1. Parametrii repartitiei normale (Gauss-Laplace)\n\n");
   fprintf(fout," ------------------------------------------------\n");
   fprintf(fout," Media aritmetica         = %10.2f\n",xm);
   fprintf(fout," Abaterea medie patratica = %10.2f\n",sigma);
   fprintf(fout," Ab. med. patr. corectata = %10.2f\n",s);
   fprintf(fout," ------------------------------------------------\n");
   fprintf(fout,"\n\n");
   fprintf(fout,"2. Parametrii repartitiei Weibull\n\n");
   fprintf(fout," ------------------------------------------------\n");
   fprintf(fout," - Metoda celor mai mici patrate - \n\n");
   fprintf(fout," Beta                     = %10.2f\n",beta);
   fprintf(fout," Eta                      = %10.2f\n",eta);
   fprintf(fout," - Metoda verosimilitatii maxime - \n\n");
   fprintf(fout," Beta                     = %10.2f\n",beta_2);
   fprintf(fout," Eta                      = %10.2f\n",eta_2);
   fprintf(fout," ------------------------------------------------\n");
   fprintf(fout,"\n\n");
   fprintf(fout,"3. Stabilirea domeniilor de incredere\n");
   fprintf(fout,"   ale parametrilor legilor de repartitie\n\n");
   fprintf(fout," --------------------------------------------------------\n");
   fprintf(fout," Lim med arit(min/max)= %10.2f  %10.2f\n",m_min,m_max);
   fprintf(fout," Lim ab med pat       = %10.2f  %10.2f\n",sigma_min,sigma_max);
   fprintf(fout," Lim param scara      = %10.2f  %10.2f\n",eta_min,eta_max);
   fprintf(fout," --------------------------------------------------------\n");
   fprintf(fout,"\n\n");
   fprintf(fout,"4. Verificarea concordantei dintre repartitiile teoretice\n");
   fprintf(fout,"   si cele empirice, utilizand testul Kolmogorov-Smirnov\n\n");
   fprintf(fout," ------------------------------------------------\n");
   fprintf(fout," Nivelul de semnificatie pentru:\n");
   fprintf(fout," - repartitia normala         = %10.2f\n",alfa_n);
   fprintf(fout," - repartitia Weibull var.1   = %10.2f\n",alfa_w1);
   fprintf(fout," - repartitia Weibull var.2   = %10.2f\n",alfa_w2);
   fprintf(fout," ------------------------------------------------\n");
fclose(fout);
#
```

Subrutina *func*

```
function y = func(w)
   global x n;
   y = zeros(2,1);
   y(1) = sum(x.^w(1).*log(x))./sum(x.^w(1))+sum(-log(x))./n - 1./w(1);
   y(2) = sum(x.^w(1))./n - w(2).^w(1);
endfunction
```