

Odhad směsi s polovičním normálním rozdělením

Jedná se o normální rozdělení, definované jen pro nezáporné hodnoty (polovina normálního rozdělení). Aby integrál zůstal roven jedné, jeho hustota pravděpodobnosti $f(y)$ je násobena dvěma

$$f(y) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp\left\{-\frac{y^2}{2\sigma^2}\right\}, \quad x \geq 0 \quad (1)$$

Normální rozdělení, ze kterého toto rozdělení vychází má nulovou střední hodnotu a rozptyl σ^2 .

Momenty rozdělení jsou

- střední hodnota $E[Y] = \sigma\sqrt{\frac{2}{\pi}}$
- rozptyl $D[Y] = \sigma^2\left(1 - \frac{2}{\pi}\right)$

Statistiky a jejich přepočet

$$S_t \kappa_t = S_{t-1} + y_t^2, \quad \kappa_t = \kappa_{t-1} + 1 \quad (2)$$

Bodový odhad

$$\hat{\sigma} = \sqrt{\frac{1}{n} \sum_{i=1}^n y_i^2} \quad (3)$$

Simulace

$$y = |z|, \quad z \sim N(0, \sigma^2)$$

Program

Popis následuje za programem

```
// Pokusy s HNor (Half-Normal) proximity jako likelihood
// - dvourozměrná data
// - 3 komponenty
// -----
exec ScIntro.sce, mode(0)

nc=3;
nx=2;
nd=500;
rS=[.5 100 40
     50 130 1];
aS=[.3 .4 .3];
cS(1)=1;

for t=2:nd // ===== SIMULATION =====
```

```

    cS(1,t)=sum(rand(1,1,'u')>cumsum(a1S))+1;
    x(:,t)=abs(sqrt(rS(:,cS(t)).*rand(nx,1,'n')));
end
ka=[1 1 1]'*10;
nu=[.5 .7 .6]';
rI=rS.*(rand(nx,nc,'u')+1);
r=rI;
V=rI;

printf(' ') // ===== TIME LOOP =====
for t=2:nd
    if t/fix(nd/10)==fix(t/fix(nd/10)), printf(' '); end

    for i=1:nc
        [xxx,G(i)]=HnormN(x(:,t),sqrt(r(:,i)));
    end
    Gn=G-max(G);
    g=exp(Gn);
    w=g/sum(g);
    wt(:,t)=w(:);

    // Update of statistic
    for i=1:nc
        V(:,i)=V(:,i)+w(i)*x(:,t)^2;
    end
    ka=ka+w;
    nu=nu+w;

    // Bodové odhady
    for i=1:nc
        r(:,i)=V(:,i)./ka(i);
    end

end
printf('\n')

// VÝSLEDKY
[xxx,Ect]=max(wt,'r');
[q,T]=c2c(cS,Ect);
Ecc=q(Ect);
wrong=sum(cS(:)~=Ecc(:))
from=nd

// sim -----
CS=list();
for i=1:nc
    j=find(cS==i);
    CS(i)=x(:,j);
end

```

```

tx=['b.';'r.';'g.'];
scf(1);
for i=1:nc
    plot(CS(i)(1,:),CS(i)(2,:),tx(i),'markersize',3)
end
title 'Simulated clusters'
set(gca(),'data_bounds',[0 45 0 40])

// est -----
C=list();
for i=1:nc
    j=find(Ecc==i);
    C(i)=x(:,j);
end

set(scf(2),'figure_position',[850 200]);
for i=1:nc
    plot(C(i)(1,:),C(i)(2,:),tx(i),'markersize',3)
end
title 'Estimated clusters'
set(gca(),'data_bounds',[0 45 0 40])

```

Popis programu

Program je prakticky stejný jako předešlé. Liší se jen v

- výpočtu proximity **1**,
- přepočtu statistik **2**,
- výpočtu bodových odhadů **3**.

Simulace se provádí jako pro normální data a bere se jejich absolutní hodnota.