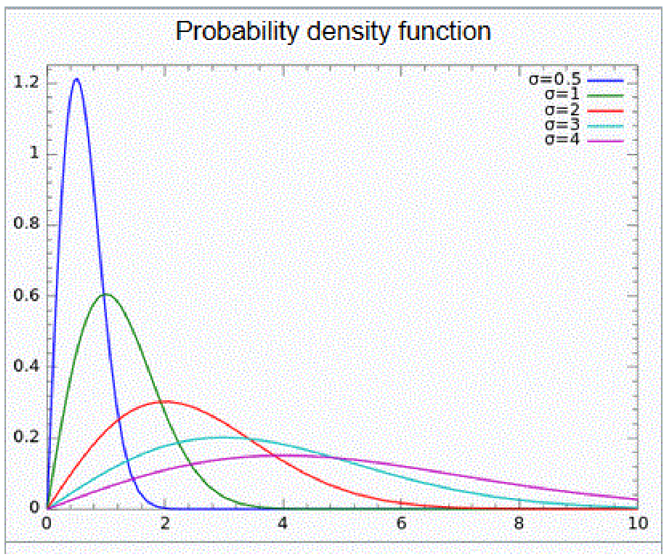
Odhad směsi s Rayleigho komponentami

Rayleigho rozdělení je na pohled podobné rozdělení gama, jeho pokles pro rostoucí hodnoty náhodné veličiny je ale mnohem rychlejší. Klesá s exponenciálou kvadratického argumentu, podobně jako normální rozdělení. Jeho hustota pravděpodobnosti je

$$f(y) = \frac{y}{r} \exp\left\{-\frac{y^2}{2r}\right\} \tag{1}$$

kde r je parametr rozdělení (jako rozptyl). Co je nepříjemné, je, že má opět jen jeden parametr. Tedy střední hodnota a rozptyl jsou navzájem pevně svázány.

Obrázek ukazuje různé průběhy této distribuce (r=σ^2)



Hlavní význam tohoto rozdělení je nezápornost a rychlý pokles pro velké rostoucí hodnoty

Střední hodnota $\quad \sigma\sqrt{\frac{\pi}{2}} \doteq 1.253\sigma, \quad$ kde $\sigma = \sqrt{r}$

Rozptyl $\quad \frac{4-\pi}{2}\sigma^2 \doteq 0.429\sigma^2$

Modus $\quad \sigma$

Věrohodnostní funkce

$$L = \frac{\prod y_i}{r^N} \exp\left\{-\frac{\sum y_i^2}{2r}\right\} = P_N r^{-N} \exp\left\{\frac{S_N}{2r}\right\}$$

Přepočet statistik

$$P_t = P_{t-1} y_t, \qquad S_t = S_{t-1} + y_t^2 \tag{2}$$

Bodový odhad r

$$\hat{r}_N = \frac{1}{2N} \sum y_i^2 = \frac{S_N}{2N} = 0.5\overline{y^2} \tag{3}$$

Generování hodnot

$$y = \sqrt{-2r\ln(u)}$$

kde $u \sim uniform\,(0,1)$.

Odhad směsi v každém čase probíhá podle následujícího schematu

1. Výpočet vah w pro jednotlivé komponenty, kde hlavním prvkem jsou tzv. proximity - tj. hodnoty hustoty pravděpodobnosti příslušné komponenty s dosazenými existujícími bodovými odhady parametrů a aktuálně změřenými daty, tj. $f_c\left(y_t|\hat{\Theta}_{c;t-1}\right)$, kde $c = 1, 2, \cdots, nc$ (pro všechny komponenty).
2. Přepočet statistik, kde data se přidávají s příslušnou váhou.
3. Výpočet bodových odhadů.

Pro tyto kroky potřebujeme

1. Model 1.
2. Vzorec pro přepočet statistik 2.
3. Vzorec pro výpočet bodového odhadu parametrů 3

Všechny tyto kroky jsou specifické pro určité rozdělení (zde gama).

Program (komentáře jsou za programem)

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

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

for t=2:nd     // ============== SIMULATION ============
  cS(1,t)=sum(rand(1,1,'u')>cumsum(alS))+1;
  u=rand(nx,1,'u');
  x(:,t)=sqrt(-2*rS(:,cS(t)).*log(u));
end
//scatt(x)
```

```
//return
V=10*rand(2,3,'u')*10;
ka=[1 1 1]*10;
nu=[.5 .7 .6];
r=[.5 15 55
     1 30 105];

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

  for i=1:nc
    G(1,i)=1;
    for j=1:nx
      G(1,i)=G(1,i)*log(x(j,t))-log(r(j,i))-.5*x(j,t)**2./r(j,i);
    end
  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)=.5*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])
```

Inicializace - dodělat !!!!!