Náhodné procesy – základy

Honza Černocký, FIT VUT Brno

Tato část projektu pokrývá základní práci s náhodnými signály, především **souborové odhady** parametrů na množině (souboru) realizací náhodného procesu. Budeme pracovat výhradně s náhodným procesem s diskrétním časem. O teorii k projektu pojednává přednáška náhodné signály (soubor `nah.pdf`), mějte jej prosím po ruce...

1 Generování náhodného procesu

Náhodný proces bude gaussovský šum se střední hodnotou nula a směrodatnou odchylkou 5, který necháme projít filtrem s přenosovou funkcí:

$$H(z) = \frac{1}{1 - 1.1314z^{-1} + 0.6400z^{-2}}$$

Vygenerujeme $\Omega = 10000$ realizací tohoto náhodného procesu po $N = 200$ vzorcích. Uložíme je do matice `ksi`.

```
Om = 10000; N = 200;   % Om je zkratka Omega
b=[1]; a = [1.0000   -1.1314    0.6400];
nn = 0:N-1;
ksi = zeros(Om,N);
for ii=1:Om,
  x = randn(1,N) * 5 + 0;
  y = filter(b,a,x);
  ksi(ii,:) = y;
%  plot(nn,x,nn,y); pause
end
```

Na začátku odkomentujte linku `plot(nn,x,nn,y); pause` a na několik realizací se podívejte. Pak ji zase zakomentujte a nechte proběhnout celý cyklus.

Q: Jaký pozorujete rozdíl mezi náhodným signálem vyprodukovaným pomocí `randn` a po průchodu filtrem ?

2 Odhad distribuční funkce

Všechny odhady budeme dělat pro určitý čas n, např:

```
n = 50;
```

Na definici a teorii (velmi jednoduchou!) odhadu se podívejte do přednášky. Distribuční funkce v zásadě odpovídá na otázku "jak je pravděpodobné, že hodnota mého signálu pro vzorek n bude menší než nějaká hodnota x". Na začátku musíme určit rozsah hodnot pomocné proměnné x, pro které budeme $F(x, n)$ odhadovat. Můžeme např. použít minimální a maximální hodnotu z dat a mezi nimi umístit 50 hodnot:

```
xmin = min(min(ksi)); xmax = max(max(ksi));
kolik = 50;
x = linspace(xmin,xmax,kolik);
```

a pak už můžeme vesele odhadovat:

```
Fx = zeros(size(x));
for ii = 1:kolik,
  thisx = x(ii);
  % vybereme vsechna pozorovani pro prislusny cas:
  ksin = ksi(:,n);
  % a pocitame odhad Fx jako pomer tech, co jsou pod x a vsech:
  Fxn(ii) = sum(ksin < thisx) / Om;
end
% vysledek
subplot(211); plot (x,Fxn);
```

Q: Jak funguje řádek `Fxn(ii) = sum(ksin < thisx) / Om;` Pokud to není jasné, zkuste si zobrazit výsledek podmínky `ksin < thisx`

Q: splňuje distribuční funkce teoretické předpoklady (0 pro $-\infty$, 1 pro $+\infty$) ?

Q: Jak byste spočítali pravděpodobnost, že náhodná proměnná v čase n bude v intervalu [-10, 5] ?

Q: odhadněte $F(x, n)$ pro jiný čas n (např. 100). Je stejná ? Můžete říci, že signál je stacionární ?

3 Odhad funkce hustoty rozdělení pravděpodobnosti

$p(x, n)$ provedeme pomocí histogramu, opět pro daný čas n:

```
deltax = x(2) - x(1);
pxn = hist(ksin,x) / Om / deltax;
subplot(212); plot (x,pxn);
```

Q: Ověřte, že $\int_{-\infty}^{\infty} p(x, n)dx = 1$.

Q: proč musíme dělit histogram Δx a Ω ?

Q: Jak byste spočítali pravděpodobnost, že náhodná proměnná v čase n bude v intervalu [-10, 5] ?

Q: odhadněte $p(x, n)$ pro jiný čas n (např. 100). Je stejná ? Můžete říci, že signál je stacionární ?

4 Střední hodnota, směrodatná odchylka

pro daný čas n je spočítáme lehce:

```
an = mean(ksin)
stdn = std(ksin)
```

Při výpočtu střední hodnoty a směrodatné odchylky pro **všechny časy** n využijeme s výhodou toho, že funkce `mean` a `std` pracují po sloupcích:

```
aalln = mean(ksi);
stdalln = std(ksi);
subplot(211); plot (nn,aalln);
subplot(212); plot (nn,stdalln);
```

Q: Je podle Vás signál stacionární ?

Q: Proč vidíme pro několik prvních vzorků u směrodatných odchylek divné hodnoty ?

5 Funkce hustoty rozdělení pravděpodobnosti mezi dvěma časy, autokorelační koeficienty

Autokorelační koeficient $R(n_1, n_2)$ udává, jak je "signál sám sobě podobný mezi časy n_1 a n_2". Vypočteme jej jako:

$$R(n_1, n_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x_1 x_2 p(x_1, x_2, n_1, n_2) dx_1 dx_2,$$

takže potřebujeme odhad 2-rozměrné funkce hustoty rozdělení pravděpodobnosti mezi časy n_1 a n_2. Ten získáme pomocí 2-D histogramu. Odhad histogramu, přepočet na 2-D funkci hustoty rozdělení pravděpodobnosti a výpočet autokorelačního koeficientu je implementován ve funkci `hist2opt` (nepoužívejte `hist2` – linka u přednášky – nedočkali byste se výsledku...). Příklad pro $n_1 = 50$ a pro $n_2 = n_1 + 0 \ldots + 20$:

```
n1 = 50;
for n2 = n1:n1+20;
  [h,p,r] = hist2opt(ksi(:,n1),ksi(:,n2),x);
  imagesc (x,x,p); axis xy; colorbar; xlabel('x2'); ylabel('x1');
  [n1 n2 r]
  pause
end
```

Q: Sledujte tvar 2-D funkce hustoty rozdělení pravděpodobnosti a usuďte z toho, jak je signál v časech n_1 a n_2 korelován. Pak se podívejte na hodnotu autokorelačního koeficientu. Odpovídá ?

Q: Proč je hodnota $R(50, 50)$ největší ?

Q: Uložte si autokorelační koeficienty pro všechna $n_2 = n_1 + 0 \ldots + 20$: do vektoru a zobrazte je. Popište, co vidíte.

Q: Zobrazte impulsní odezvu filtru, kterým jsme filtrovali náhodný signál:

```
plot(filter(b,a,[1 zeros(1,255)]))
```

srovnejte s průběhem autokorelačních koeficientů. Komentujte.

Q: Určete autokorelační koeficienty pro jiné n_1, např. pro $n_1 = 100$ a opět pro $n_2 = n_1 + 0 \ldots + 20$. Je signál stacionární ?

Časové odhady, spektra

Tato část projektu pokrývá základní práci s náhodnými signály, především **časové odhady** parametrů na jedné realizaci náhodného procesu. Budeme pracovat výhradně s náhodným procesem s diskrétním časem. O teorii k projektu pojednává přednáška Náhodné signály II. (soubor `nah2.pdf`), mějte jej prosím po ruce...

6 Generování náhodného procesu

Budeme používat stejný náhodný proces jako v minulé části projektu, tedy gaussovský šum prošlý filtrem

$$H(z) = \frac{1}{1 + -1.1314 z^{-1} + 0.6400 z^{-2}},$$

ale vygenerujeme jen jednu realizaci, která bude dostatečně dlouhá: 10000 vzorků:

```
N = 10000;
b=[1]; a = [1.0000   -1.1314    0.6400];
nn = 0:N-1;
aux = randn(1,N) * 5 + 0;
x = filter(b,a,aux);
plot (nn,x);
```

7 Odhad střední hodnoty a směrodatné odchylky

...je opravdu jednoduchý:

```
a = mean(x)
sigma = std(x)
```

Q: Vyšly Vám tyto hodnoty podobně jako v předchozí části projektu ? Pokud ano, můžeme konstatovat, že signál je ergodický ("co se dá odhadnout souborově, dá se odhadnout i z jedné realizace").

8 Odhad distribuční funkce a funkce hustoty rozdělení pravděpodobnosti

provedeme velmi podobně jako minule, ale budeme používat vzorky z jediné realizace, která je k disposici. Drobným problémem je, že signál se jmenuje x, pomocnou proměnnou u všech odhadů tedy nazveme g:

```
gmin = min(min(x)); gmax = max(max(x));
% budeme chtit 50 bodu:
kolik = 50;
g = linspace(gmin,gmax,kolik);
% a jedeme
Fg = zeros(size(g));
for ii = 1:kolik,
  % a pocitame odhad Fg jako pomer tech, co jsou pod g a vsech:
  thisg = g(ii);
  Fg(ii) = sum(x < thisg) / N;
end
% vysledek
subplot(211); plot (g,Fg);
```

Q: Je distribuční funkce podobná té minule ?

Do funkce hustoty pravděpodobnosti zkusíme "dokreslit" Gaussovo rozložení se spočítanou střední hodnotou a směrodatnou odchylkou a podíváme se, zda se mu odhadnuté podobá. Gaussovka je definována jako:

$$p(g) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[x-a]^2}{2\sigma^2}}$$

```
deltag = g(2) - g(1);
pg = hist(x,g) / N / deltag;
% teoreticka Gaussovka
pggaus=1/sqrt(2*pi)/sigma * exp(-(g - a).^2./(2*sigma.^2));
subplot(212); plot (g,pg,g,pggaus);
```

Q: Můžeme o tomto náhodném procesu prohlásit, že je gaussovský ?

Q: Zkuste ověřit (podobně jako minule), že $\int_{-\infty}^{\infty} p(g)dg = 1$.

9 Odhad autokorelačních koeficientů

Podle toho, zda dělíme celkovým počtem vzorků nebo jen počtem vzorků, které se překrývají, rozeznáváme vychýlený a nevychýlený odhad:

$$\hat{R}_v(k) = \frac{1}{N} \sum_{n=0}^{N-1} x[n]x[n+k], \qquad \hat{R}_{nv}(k) = \frac{1}{N-|k|} \sum_{n=0}^{N-1} x[n]x[n+k],$$

V Matlabu je za nás spočítá funkce `xcorr` s přepínačem `'biased'` nebo `'unbiased'`:

```
Rv = xcorr(x,'biased'); k = -N+1:N-1;
subplot(211); plot(k,Rv)
Rnv = xcorr(x,'unbiased'); k = -N+1:N-1;
subplot(212); plot(k,Rnv);
```

Q: Proč jsou okrajové hodnoty nevychýleného odhadu divné ?

Q: Proč jsou okrajové hodnoty vychýleného odhadu malé ?

Q: Zkontrolujte, zda je hodnota nultého autokorelačního koeficientu (pozor, ve vektorech `Rv`, `Rnv` je uprostřed) rovná střednímu výkonu:

$$P_s = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n]$$

10 Spektrální hustota výkonu – power spectral density (PSD)

budeme ji odhadovat pomocí DFT z celé realizace nebo průměrováním přes několik segmentů (vztahy viz. přednáška, slajdy 14 a 15). Průměrování by mělo véest k hladšímu odhadu.

10.1 Odhad z jedné realizace

```
N = 10000;
Gdft = 1/N * abs(fft(x)).^2;
om = (0:N/2-1)/N * 2*pi; Gdft = Gdft(1:N/2);
subplot(211); plot (om,Gdft); grid;
```

Q: Líbí se Vám tento odhad ?

10.2 Odhad z několika realizací průměrováním

zařídí funkce `gprum.m` (je potřeba ještě funkce `trame.m`[1]), která signál rozdělí na segmenty o délce 100 a pak spočítá PSD pro každý a nakonec zprůměruje:

```
[Gprum,om] = gprum(x);
subplot(212); plot (om,Gprum); grid;
```

Q: A líbí se Vám tento ?

11 Průchod náhodného signálu lineárním systémem

Necháme náš náhodný signál projít filtrem s přenosovou funkcí:

$$H(z) = 1 - 1.1314z^{-1} + 0.6400z^{-2}$$

(všimněte si, že to je inverzní filtr k tomu, který jsme použili na začátku). Spektrální hustota výkonu na výstupu by měla být:

$$G_y(e^{j\omega}) = |H(e^{j\omega})|^2 G_x(e^{j\omega}),$$

což se pokusíme ověřit:

[1] Segmentům se někdy říká rámce a “trame” je francouzsky rámec :-)

```
% puvodni PSD - tu mame, ale pro uplnost ...
[Gx,om] = gprum(x);
subplot(131); plot(om,Gx); grid;

% filtr - freq. char na druhou:
a=[1]; b = [1.0000   -1.1314    0.6400];
[h,om] = freqz (b,a,50,2*pi);
h2 = abs(h).^2;
subplot(132); plot(om,h2); grid;

% filtrujeme a zobrazime psd vystupu
% a jeste teoreticky vstupni spek,. hust. vykonu nasobenou
% druhou mocninou H
y = filter(b,a,x);
[Gy,om] = gprum(y); Gyteor = Gx .* h2';
subplot(133); plot(om,Gy, om, Gyteor); grid;
```

Q: Platí vztah $G_y(e^{j\omega}) = |H(e^{j\omega})|^2 G_x(e^{j\omega})$?

Q: Proč je PSD výstupního signálu (skoro) konstantní ?

Q: Jak se takovému náhodnému procesu říká (pomůcka: bílé světlo má konstantní spektrum).

Q: Takový náhodný signál by měl mít pouze nultý korelační koeficient nenulový. Ověřte to.