Математическое моделирование РТУ и С

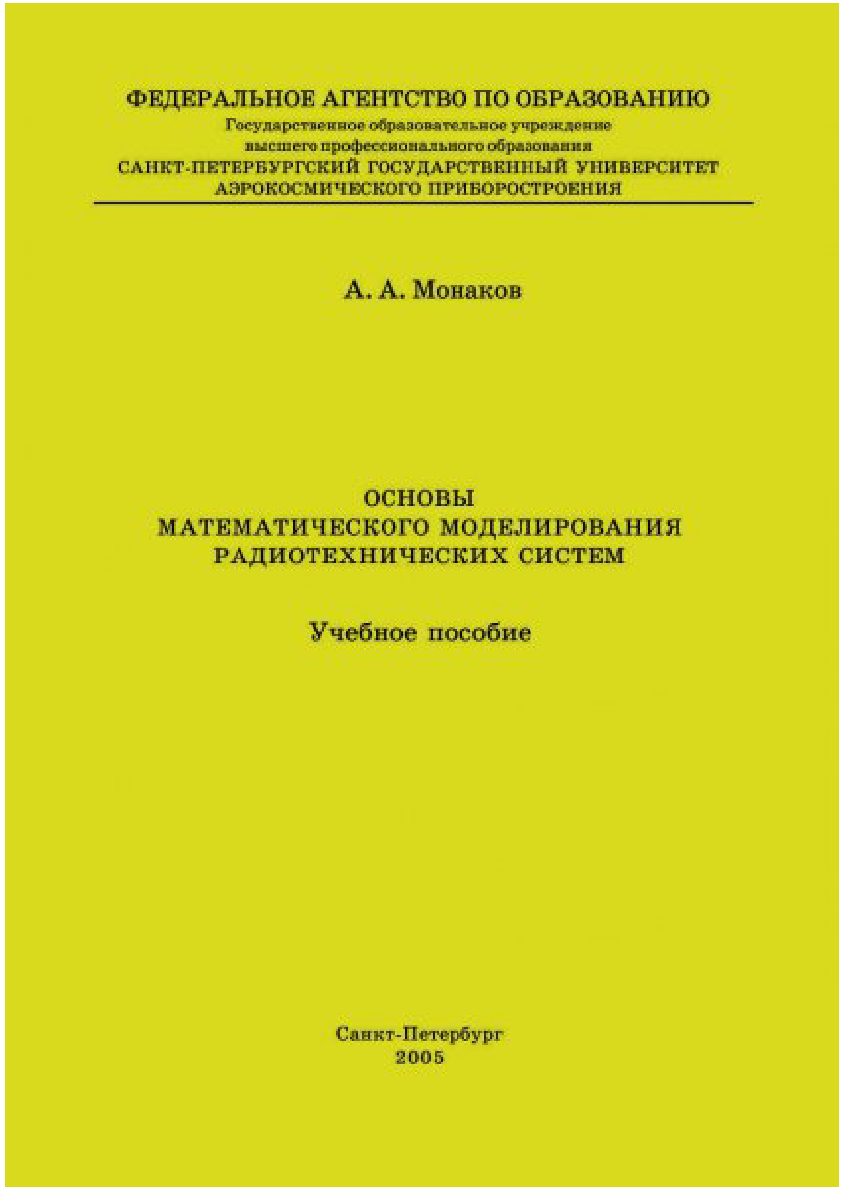
Лекция 14. Обработка результатов статистических экспериментов

Преподаватель:
Корогодин Илья
korogodin@srns.ru

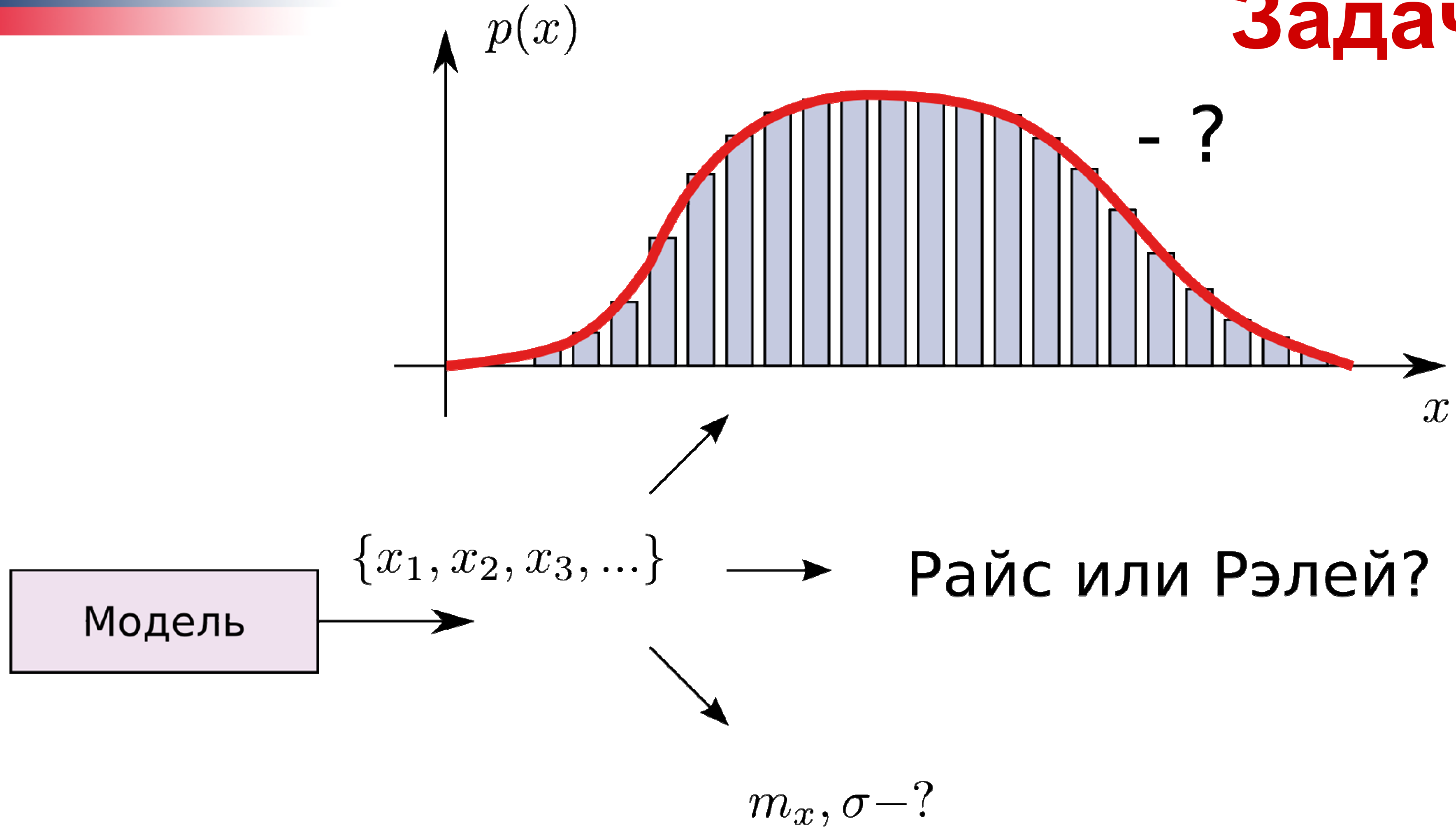
Литература

Монаков А.А. Основы математического моделирования радиотехнических систем. Учебное пособие. – СПб.: ГУАП, 2005. – 100с.

Глава 3. Обработка результатов математического моделирования



Задачи



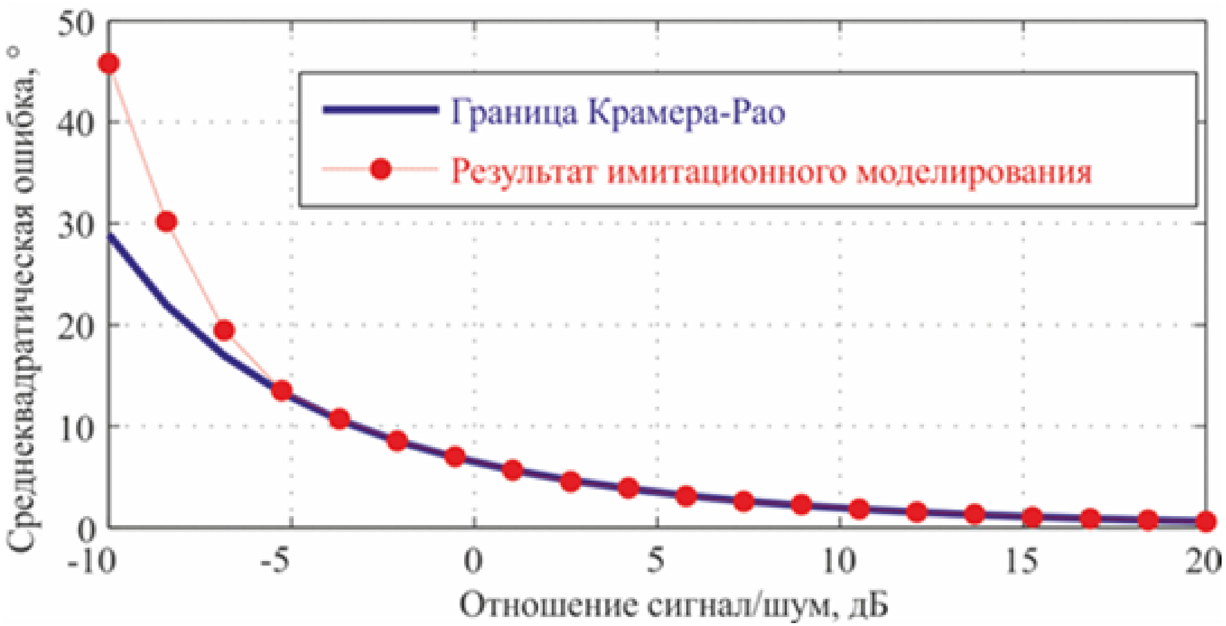
После получения на модели выборок возникают три задачи:
1) Оценить эмпирический закон распределения;
2) Сравнить эмпирический закон с теоретическим (проверить гипотезу);
3) Оценить параметры закона распределения.

Свойства оценок

Состоятельность: $$\lim_{N\to\infty} P\left(\left|\hat{x}-x\right|>\varepsilon\right)=0$$

Смещенность: $$\Delta_x = x - M\left[\hat{x}\right]$$

Эффективность: $$\varepsilon = \frac{D_0}{D\left[\hat{x}\right]} \qquad D_0 = -\frac{1+\Delta'_x}{M\left[\dfrac{\partial^2 \ln\left(p\left(Y\,|\,x\right)\right)}{\partial x^2}\right]}$$



Метод гистограмм

Оцениваем ПВ $p(x)$:

1) Разбиваем область определения на K интервалов $\mathrm{int}_k = \left[x_k;\ x_k + \Delta x_k \right)$

2) Получаем выборку СВ $\left\{ x_1, x_2, \ldots, x_N \right\}$

3) Определяем частоты попадания в интервал $\nu_k = \dfrac{sum\left(x \subset \mathrm{int}_k \right)}{N}$

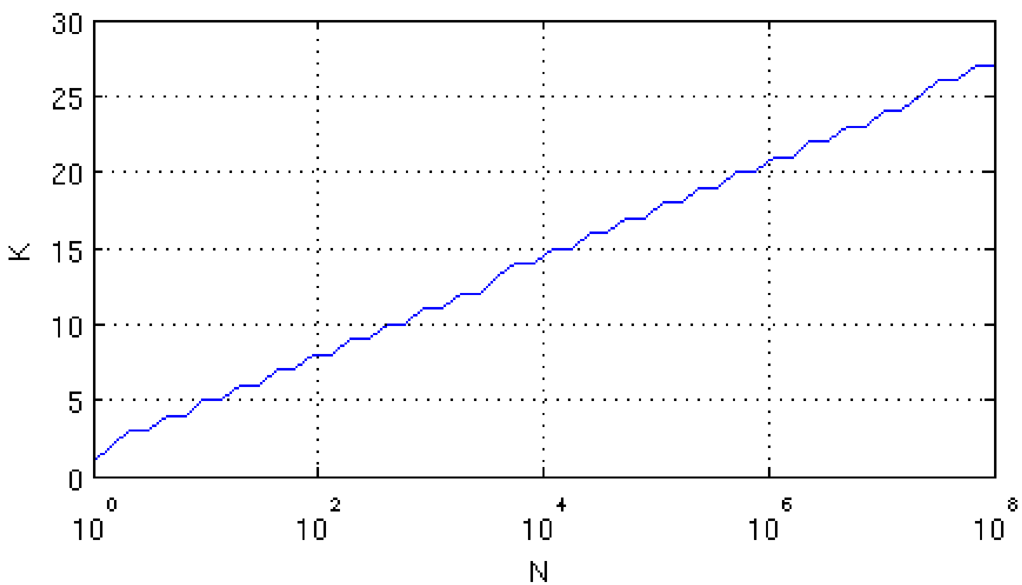
4) Рассчитываем оценку ПВ (Δx_k **- длина k-го интервала)**

$$\hat{p}(x) = \frac{\nu_k}{\Delta x_k}, \quad x_k \le x < x_k + \Delta x_k$$

Такая оценка:
- состоятельная,
- несмещенная,
- эффективная.

Правило Штюргеса:

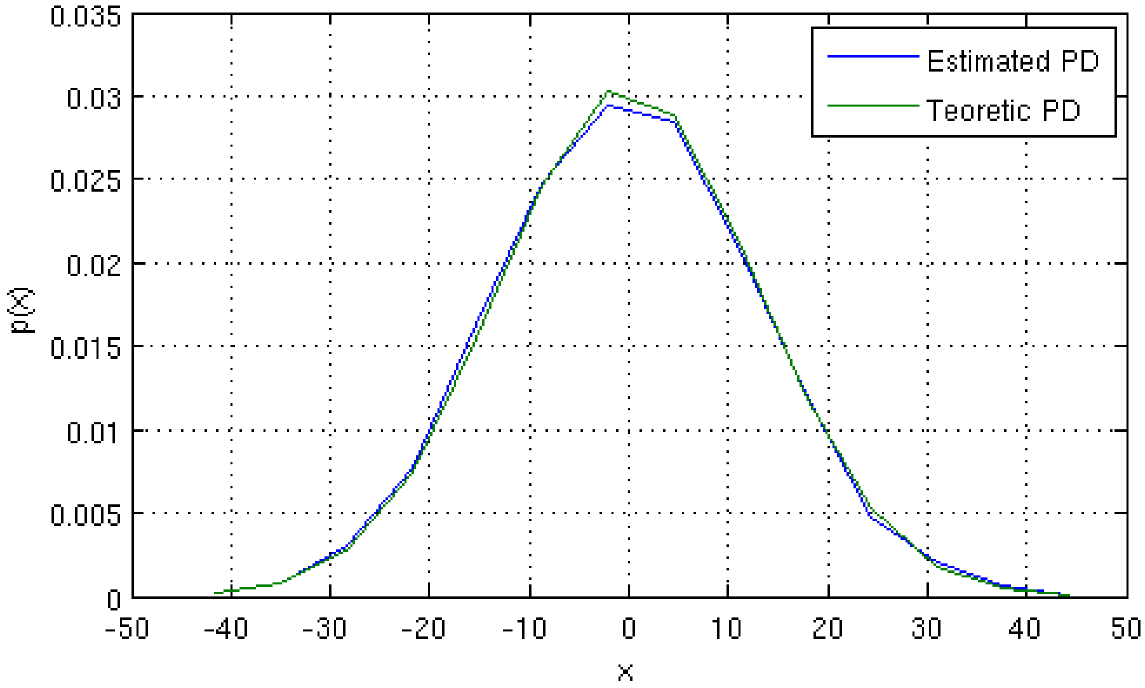
$$K \approx 1 + 3.2 \lg N$$



Метод гистограмм

Пример: Получаем 10 000 реализаций нормальной СВ *N(0, 13*13)*

```
clear all; clc; close all;

N = 10000;
sigma = 13;
x = sigma*randn(1, N);

K = ceil(1 + 3.2 *log10(N));

[sumx, xk] = hist(x, K);
nu = sumx / N;

dx = xk(2) - xk(1);
p_est = nu / dx;

p_teor = 1/sqrt(2*pi)/sigma * exp(-xk.^2 / 2 / sigma^2);

figure(1);
plot(xk, p_est, xk, p_teor);
grid on;
xlabel('x'); ylabel('p(x)');
legend('Estimated PD', 'Teoretic PD');
```




```
>> K

K =

    14
```

Эмпирическая функция распределения

Оцениваем интегральную функцию распределения $F(x)$ **:**

1. Сортируем выборку в порядке возрастания $\left\{x_?^{(1)},\ x_?^{(2)},\ \ldots,\ x_?^{(N)}\right\}$

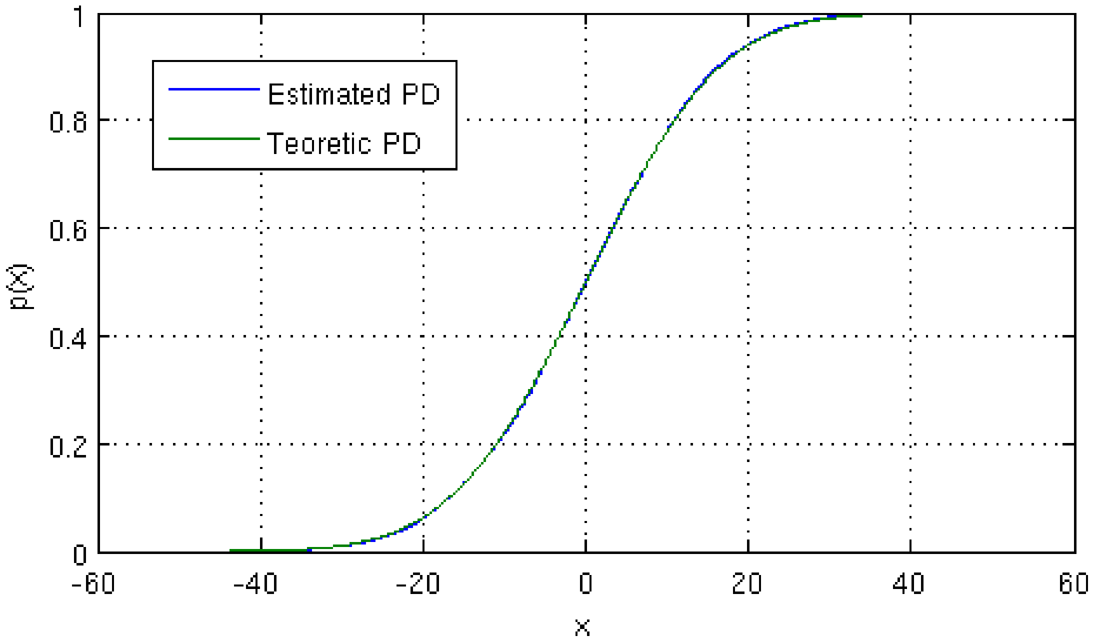
2. Считаем эмпирическую ФР:

$$F(x)=\begin{cases} 0,\ x<x_?^{(1)} \\ \dfrac{n}{N},\ \ x=x_?^{(n)} \\ 1,\ x>x_?^{(1)} \end{cases}$$

Выражение задает точки. Между ними можно провести интерполяцию.

Эмпирическая функция распределения

```
clear all; clc; close all;

N = 10000;
sigma = 13;
x = sigma*randn(1, N);

xs = sort(x);

p_est = (1:N) / N;

p_teor = (erf(xs/sigma/sqrt(2)) + 1)/2;

figure(1);
plot(xs, p_est, xs, p_teor);
grid on;
xlabel('x'); ylabel('p(x)');
legend('Estimated PD', 'Teoretic PD');
```



sort

Sort array elements in ascending or descending order

Syntax

```
B = sort(A)
B = sort(A,dim)
B = sort(...,mode)
[B,IX] = sort(A,...)
```

Description

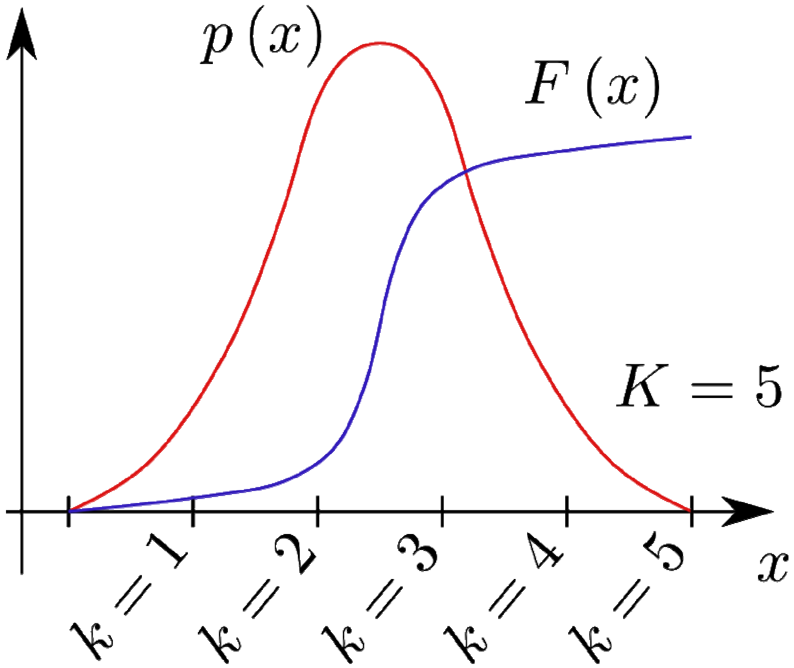
`B = sort(A)` sorts the elements along different dimensions of an array, and arranges those elements in ascending order.

Критерий Пирсона

aka метод Хи-квадрат.

Есть выборка $\left\{x_1, x_2, \ldots, x_N\right\}$

Вопрос: соответствует ли ей $F(x)-?$



Опять разбиваем область значений на K интервалов

Считаем число попаданий в интервалы:

$$n_k = sum\left(x \geq x_{\min}^{(k)},\ x < x_{\max}^{(k)}\right)$$

Считаем вероятности попадания в интервалы:

$$P_k = F\left(x_{\max}^{(k)}\right) - F\left(x_{\min}^{(k)}\right)$$

Идея: посмотреть отклонение теоретической вероятности P_k

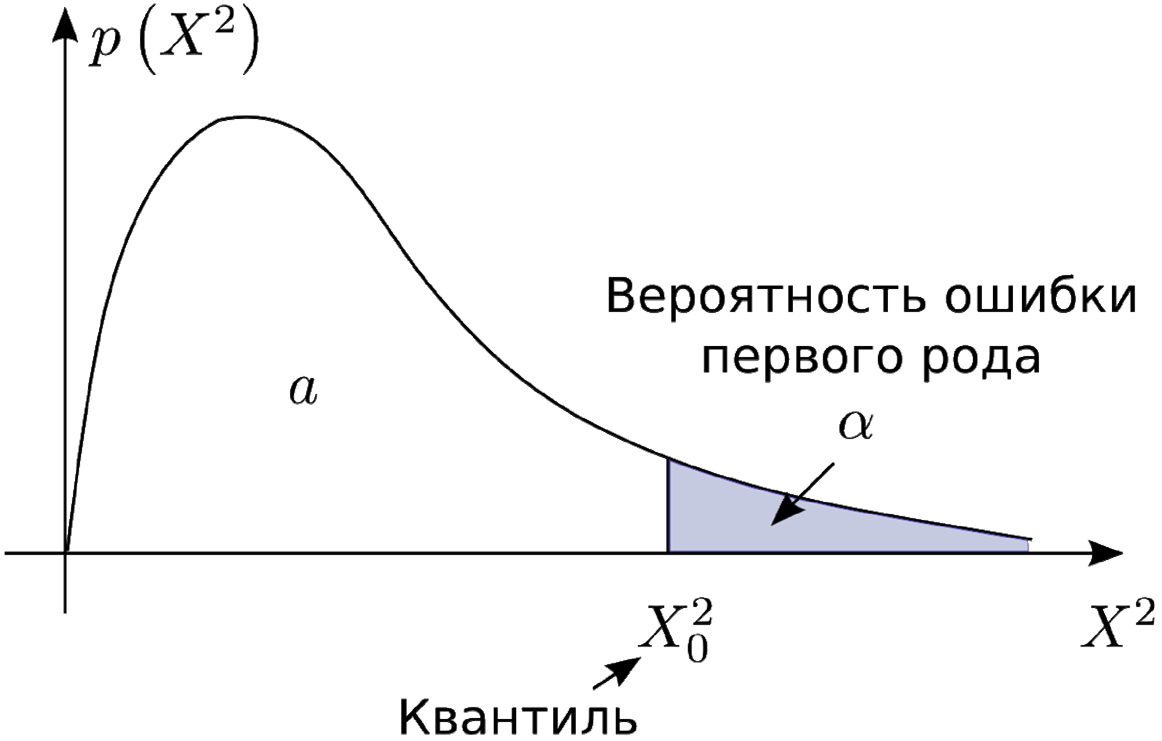
от частоты возникновения событий $\nu_k = n_k / N$

Критерий Пирсона

Статистика критерия согласия Пирсона

$$X_N^2 = N\sum_{k=1}^{K}\frac{\left(\nu_k - P_k\right)^2}{P_k}$$

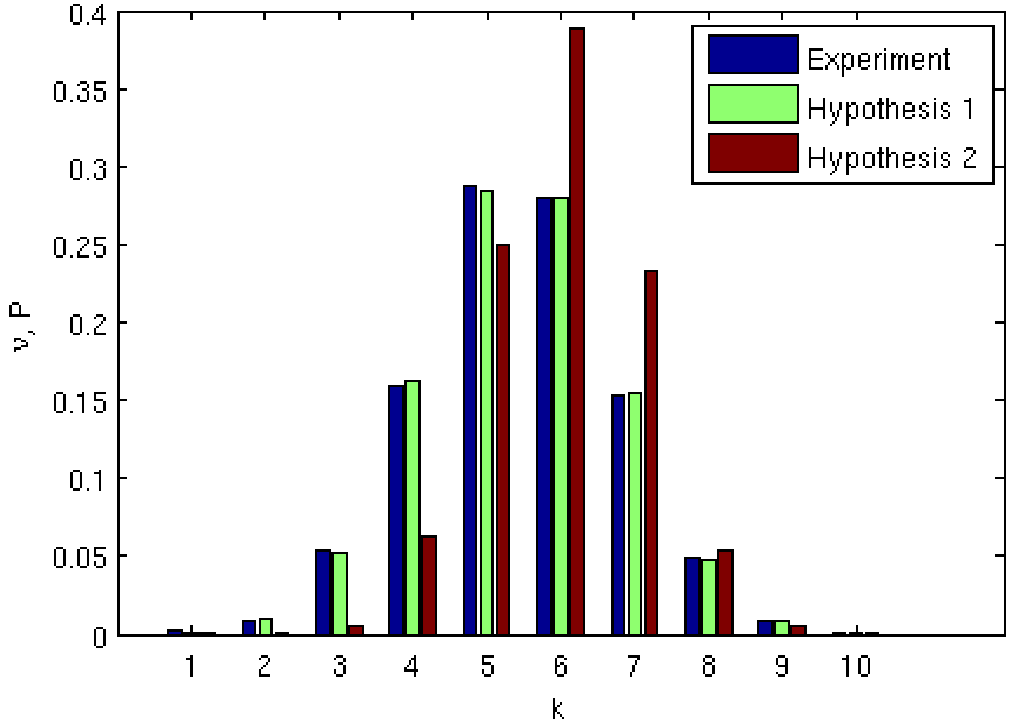
При $N \to \infty$ **СВ** X_N^2 **подчиняется хи-квадрат** χ_r^2 **распределению с** $r = K - 1$ **степенями свободы, <u>если гипотеза о распределении верна</u>**



1. **Считаем** X^2
2. **Сравниваем с квантилем** X_0^2
3. **Если больше квантиля, то гипотеза неверна с вероятностью ошибиться**
$\alpha = 1 - a$

Критерий Пирсона

```
clear all; close all; clc;

N = 10000;
sigma = 13;
mu = 0;
x = randn(1, N) * sigma + mu;

minx = min(x); maxx = max(x);
K = 10; dx = (maxx - minx) / K;
histpoints = (minx + dx/2) + (0:K-1)*dx;
[h, x] = hist(x, histpoints); nu = h/N;

alpha = 0.01; a = 1 - alpha;
X2_0 = chi2inv(a, K-1);

sigma1 = 13; mu1 = 0;
Ph1 = (erf( (x + dx/2 - mu1) /sigma1 / sqrt(2) ) + 1)/2 - ...
        (erf( (x - dx/2 - mu1) /sigma1 / sqrt(2) ) + 1)/2;
X2_1 = N * sum(((nu - Ph1).^2 ./ Ph1));
if X2_1 > X2_0
   fprintf('Гипотеза 1 не верна\n');
else
   fprintf('Гипотеза 1 верна\n‘);
end
```



Гипотеза 1 верна
Гипотеза 2 не верна

```
sigma2 = 10; mu2 = 5;
Ph2 = …

figure(1);
bar([nu; Ph1; Ph2]');
legend('Experiment', …
'Hypothesis 1',  'Hypothesis 2')
ylabel('\nu, P'); xlabel('k');
```

Оценки моментов

Начальные моменты: $$M_p = \int_{-\infty}^{+\infty} x^p p(x)dx, \;\; p = 1, 2, \ldots$$

Центральные моменты: $$m_p = \int_{-\infty}^{+\infty} \left(x - M_1\right)^p p(x)dx, \;\; p = 1, 2, \ldots$$

Мат.ожидание: $$\mu = M_1 = \int_{-\infty}^{+\infty} x \cdot p(x)dx$$

Дисперсия: $$D = m_2 = \int_{-\infty}^{+\infty} (x - \mu)^2 p(x)dx$$

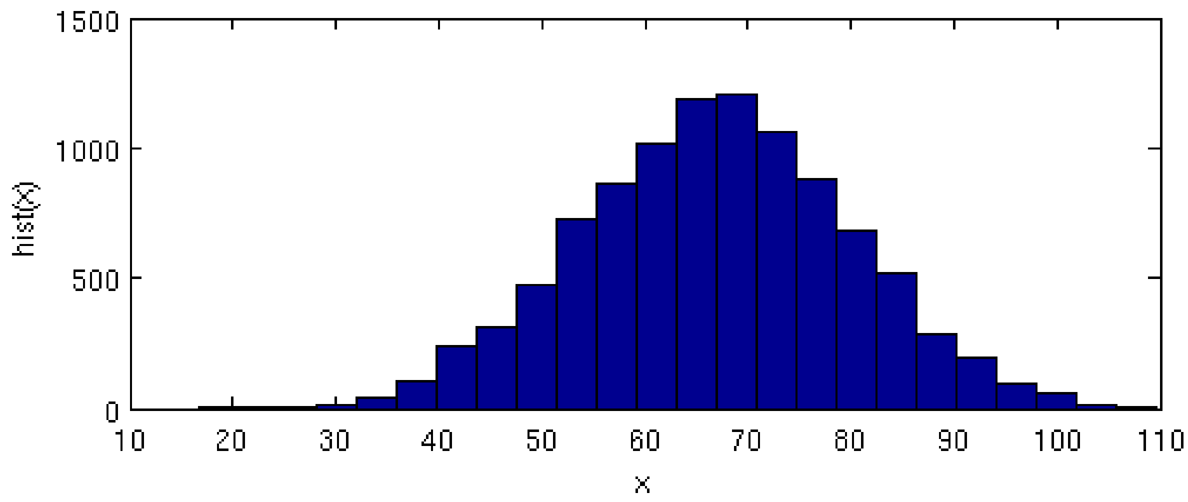
Оценка мат.ожидания: $\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n$ **(в общем случае** $\hat{M}_p = \frac{1}{N}\sum_{n=1}^{N} x_n^p$**)**

Оценка дисперсии: $\hat{D} = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \mu\right)^2$ **(** $m_p = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \mu\right)^p$ **)**

Если неизвестно м.о.: $\hat{D} = \frac{1}{N-1}\sum_{n=1}^{N}\left(x_n - \hat{\mu}\right)^2$ **Оценки несмещенные, состоятельные**

Оценки моментов

```
clear all; clc; close all;

sigma = 13; % 13*13 = 169
mu = 67;

N = 10000;

x = sigma*randn(1, N) + mu;

figure(1);
hist(x, 24);
xlabel('x'); ylabel('hist(x)');

mu_est1 = 1/N * sum(x);
D_est1 = 1/(N-1) * sum( (x - mu_est1).^2 );

fprintf('Formula: mu_est1 = %f, D_est1 = %f\n', mu_est1, D_est1);

mu_est2 = mean(x);
D_est2 = std(x)^2;

fprintf('Functions: mu_est2 = %f, D_est2 = %f\n', mu_est2, D_est2);
```



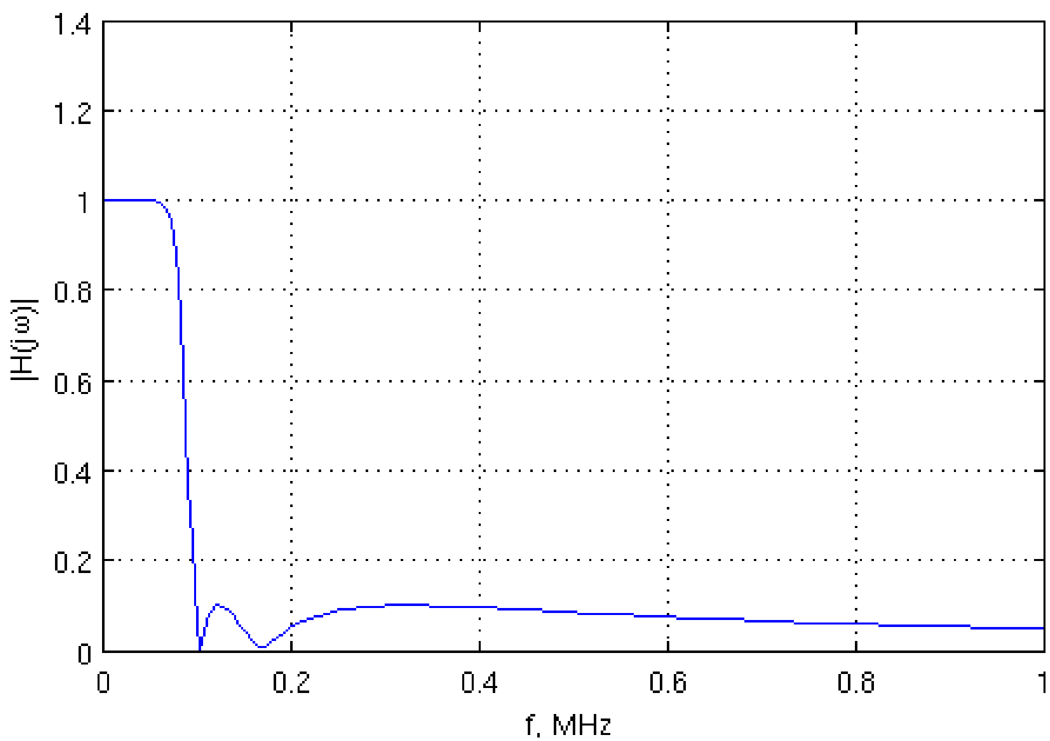
Formula: mu_est1 = 66.877749, D_est1 = 169.408818
Functions: mu_est2 = 66.877749, D_est2 = 169.408818

Оценка корреляционной функции

```
clear all; clc; close all

Rp = 20;
w0 = 2*pi*1e5;

Fs = 15.6e6; Ts = 1/Fs;
N = 1000;
t = ((1:N) - 1) * Ts;
T = max(t);

n = 5;
[z, p, k] = cheb2ap(n, Rp);
[bp, ap] = zp2tf(z, p, k);
[b, a] = lp2lp(bp, ap, w0);
[H, w] = freqs(b, a);

[bz, az] = bilinear(b, a, Fs);

figure(1)
plot(w/2/pi/1e6, abs(H));
xlabel('f, MHz');
ylabel('|H(j\omega)|');
grid on; xlim([0 1]);
```

Несмещенная оценка:

$$\hat{R}\left(\tau = mT_s\right) = \frac{1}{N-|m|}\sum_{n=0}^{N-|m|-1}\left(x_n - \hat{\mu}\right)\left(x_{n+|m|} - \hat{\mu}\right)$$



Оценка корреляционной функции

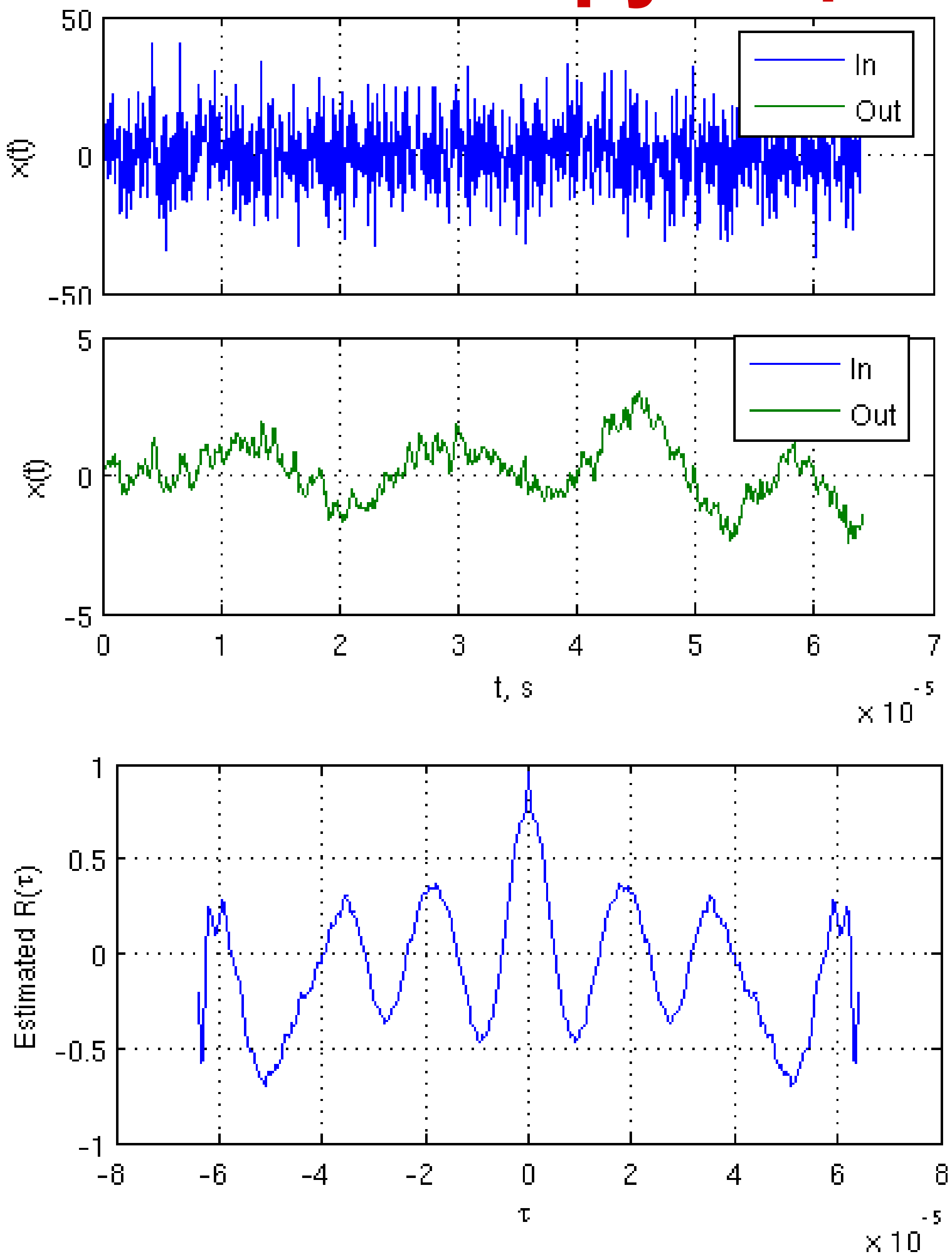
xcorr

Cross-correlation

Syntax

```
c=xcorr(x,y)
c=xcorr(x)
c = xcorr(x,y,'option')
c=xcorr(x,'option')
c=xcorr(x,y,maxlags)
c=xcorr(x,maxlags)
c=xcorr(x,y,maxlags,'option')
```

```
...
sigma = 13;
x_in = randn(1, N) * sigma;

x = filter(bz, az, x_in);

figure(2);
plot(t, x_in, t, x)
xlabel('t, s'); ylabel('x(t)');
legend('In', 'Out'); grid on;

R = xcorr(x - mean(x), x - mean(x), ...
'unbiased');
figure(3)
tau = ( (0:(2*N - 2)) - (N-1)) * Ts;
plot(tau, R); grid on;
xlabel('\tau'); ylabel('Estimated R(\tau)');
```



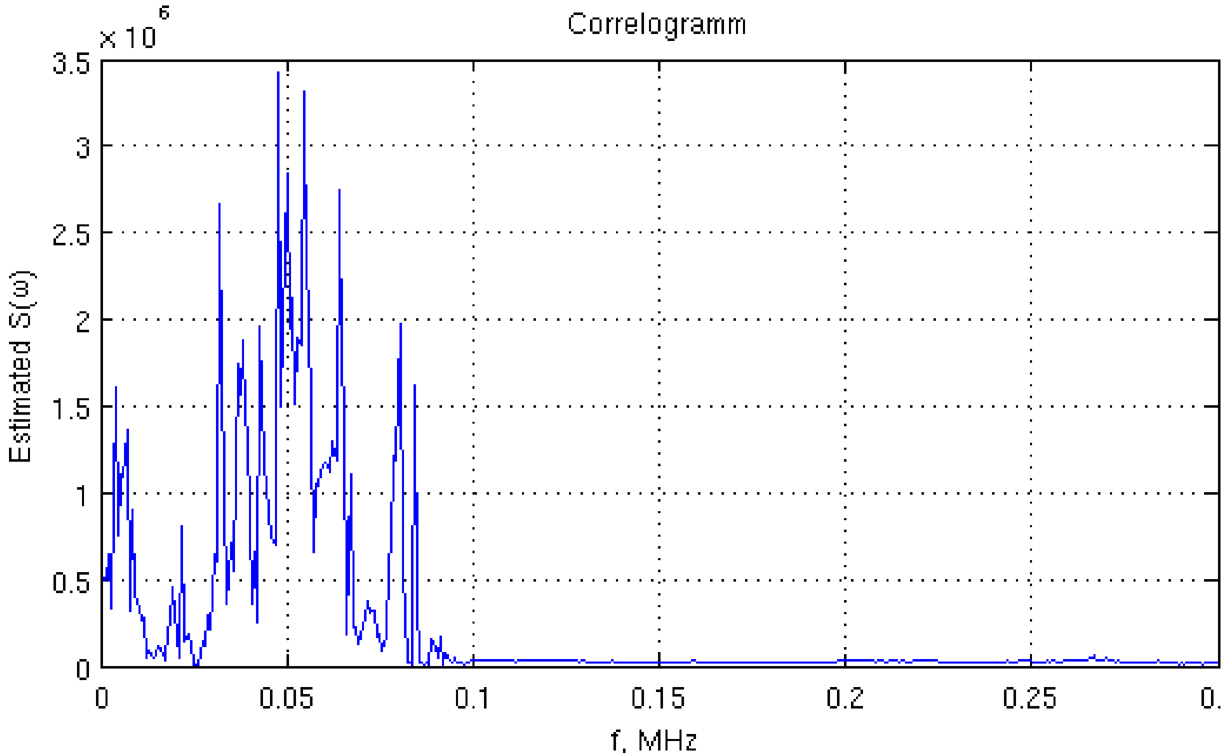
Оценка СПМ

Метод коррелограмм.

Идея - посчитать БПФ от оценки корреляционной функции.

```
figure(4);
f = (0:(length(tau)-1)) * (1/(2*N-1)/Ts);
plot(f/1e6, 2/(2*N-1) / Ts *abs(fft(R)));
xlim([0 0.3]); grid on;
xlabel('f, MHz');
ylabel('Estimated S(\omega)');
title('Correlogramm');
```

$$\hat{S}_k = \frac{2}{N} \cdot \frac{1}{T_d} \cdot \sum_{n=1}^{N} \hat{R}_n e^{-j2\pi \frac{(k-1)(n-1)}{N}}$$



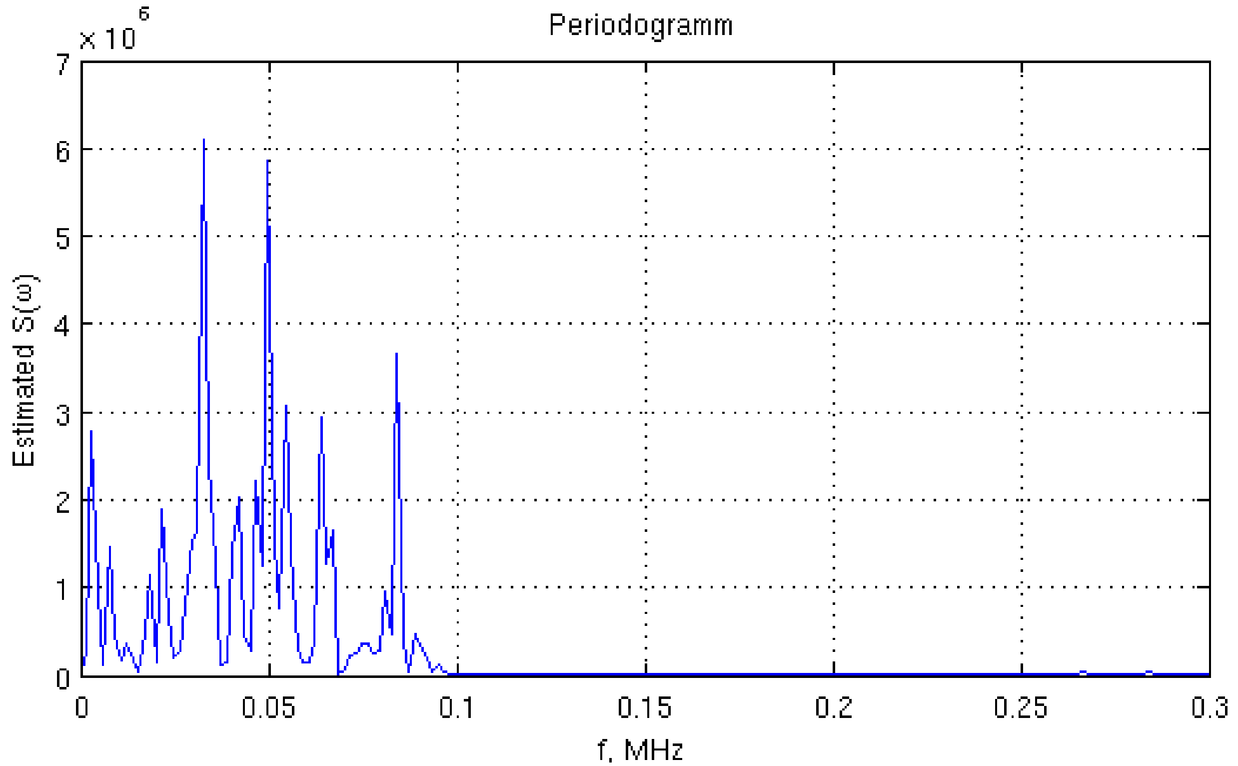
Оценка СПМ

Метод периодограмм.

Идея - посчитать БПФ от процесса, взять квадрат модуля.

```
figure(5);
f = 0:(1/T):(1/Ts);
plot(f/1e6, 4/N^2 / Ts *abs(fft(x)).^2);
xlim([0 0.3]); grid on;
xlabel('f, MHz');
ylabel('Estimated S(\omega)');
title('Periodogramm')
```

$$\hat{S}_k = \left(\frac{2}{N}\right)^2 \cdot \frac{1}{T_d} \cdot \left|\sum_{n=1}^{N} x_n e^{-j2\pi\frac{(k-1)(n-1)}{N}}\right|^2$$



Оценки несостоятельны как для одного, так и для другого метода