

*Рандомизация для задач
анализа информации и
управления в реальном
времени*

Граничин Олег Николаевич

[\(oleg_granichin@mail.ru\)](mailto:oleg_granichin@mail.ru)

Весна 2011, 4 курс,
кафедра Системного программирования

2. Рандомизированные алгоритмы

- «Провалы» детерминированных подходов
- Понятие рандомизированного алгоритма
- Парадигма рандомизированного подхода
- Пример вычисления по Методу Монте-Карло площади «подграфика»
- Пример состоятельности рандомизированного алгоритма в задаче обнаружения сигнала на фоне почти произвольных помех.

Information analysis

As an influence which leads to a transformation

- *information* is something potentially perceived as representation, though not created or presented for that purpose
- information has been perceived by a conscious mind and also interpreted by it, the specific context associated with this interpretation may cause the transformation of the information into *knowledge*
- *data* is signed and saved

Inverse Problems

Information, Data, Knowledge



$$y = \Phi x \quad (= \Phi(x))$$

$$y - \dim m$$

$$x - \dim N$$

How many measurements do we need?

$$y = \Phi x + v$$



$$m > N$$

Information Analytics in Real Time

- Usually in real time environment we have
 - restrictions of resources and
 - insufficient amount of data with necessary varieties

Deterministic approaches often failed

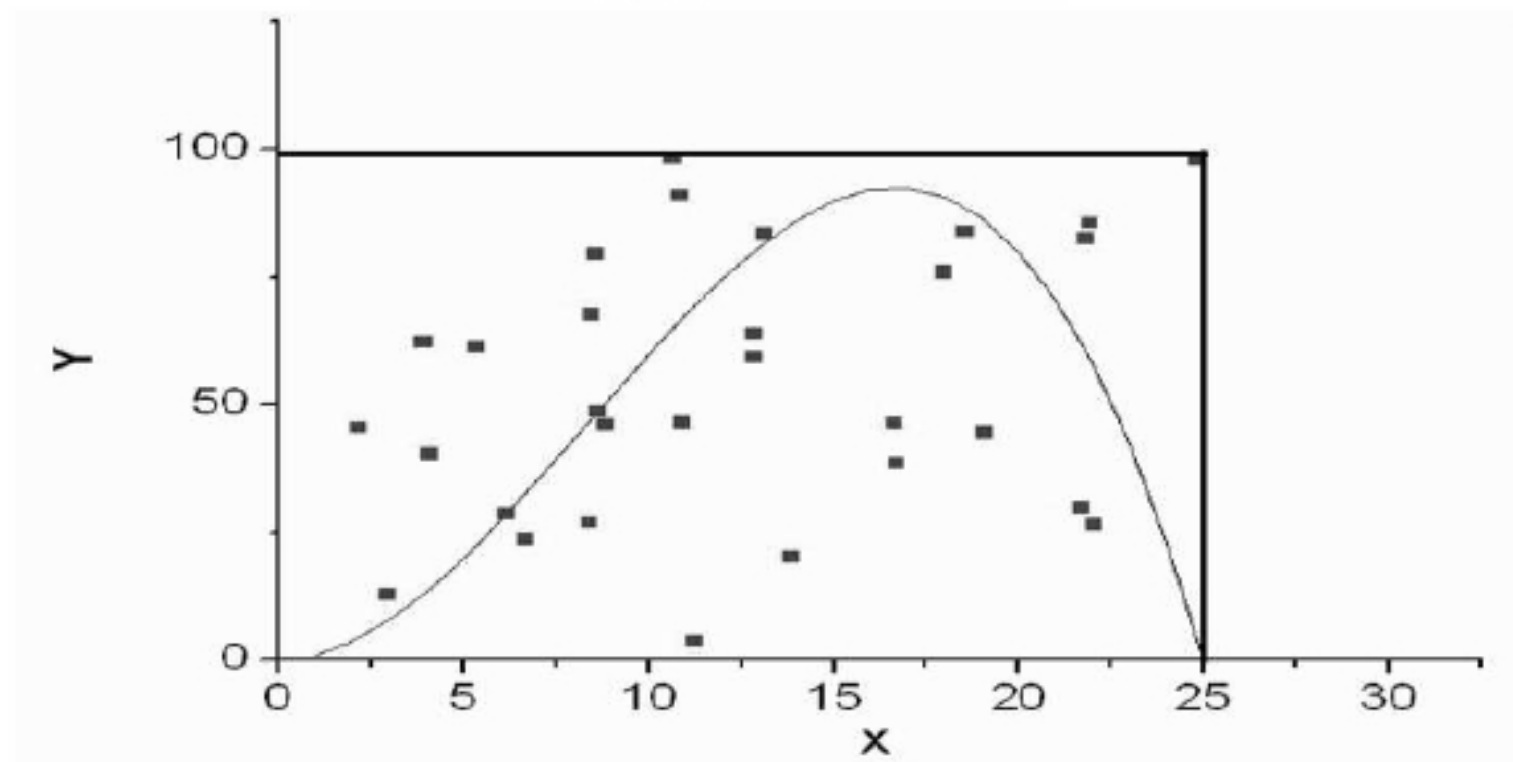
- In theory and practice many difficulties arise when we try to make analytical investigation of “complex” systems. In many practical applications traditionally efficient deterministic methods fail to yield a result when the system is complex. In particular, this led to the notion of *NP*-hard problems.

Randomized Algorithms

- A randomized algorithm is an algorithm where one or more steps are based on a random rule, that is, among many deterministic rules, one rule is selected according to a random scheme. Randomization has turned out to be a powerful tool for solving a number of problems deemed unsolvable with deterministic methods.

Метод Монте-Карло-1

$$f(\theta) = \int_{x \in X} F(\theta, x) dx$$



Интегрирование по Монте-Карло

Метод Монте-Карло-2

Зафиксируем натуральное число N , выберем в параллелепипеде $\Pi \times [0, F_{\max}] \subset \mathbb{R}^{d+1}$ случайно N точек z_1, z_2, \dots, z_N и подсчитаем среди них число K тех, у которых последняя координата не превосходит значения функции F в соответствующей точке, определяемой первыми d координатами. Хорошей оценкой для значения интеграла будет

$$f(\theta) \approx V(\mathbb{X})F_{\max} \frac{K}{N}.$$

Причем необходимую точность приближения можно гарантировать *априори* при выборе N в силу закона больших чисел и неравенства Чебышева

$$\forall \epsilon > 0 \quad \mathbb{P} \left\{ \left| \frac{f(\theta)}{V(\mathbb{X})F_{\max}} - \frac{K}{N} \right| \geq \epsilon \right\} \leq \frac{f(\theta)(V(\mathbb{X})F_{\max} - f(\theta))}{N\epsilon^2(V(\mathbb{X})F_{\max})^2}.$$

При небольшой размерности d множества \mathbb{X} во многих случаях можно при вычислении интегралов пользоваться и традиционными детерминированными приближенными методами, но при $d \gg 1$ на практике фактически ничем нельзя воспользоваться кроме метода Монте-Карло, сложность которого не зависит d .

Decreasing the negative influence of restrictions in real time

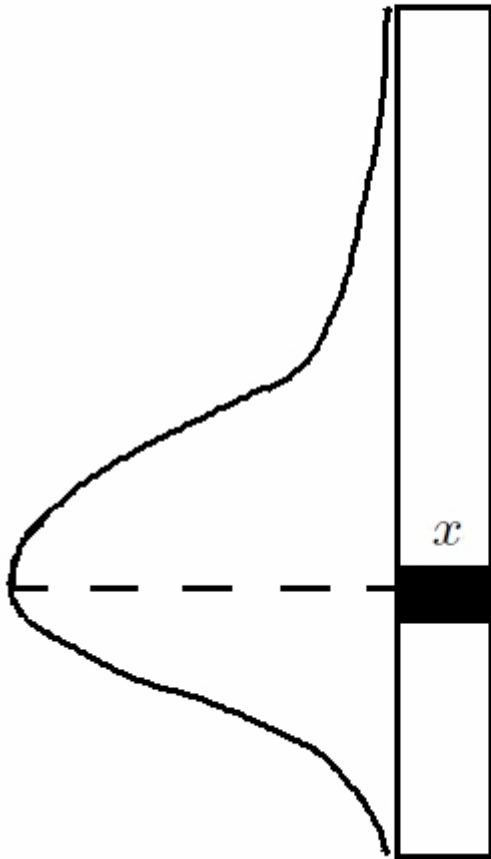
- Randomization can be used for enrichment of the observation's data.
- Randomized scenario approach allows to solve a problem “efficiently with high probability” for almost all conditions, which is often considered sufficient enough.

Enrichment of the Observation's Data

$n = 1, 2, \dots$

$$y_n = \Phi(x) + v_n$$

B — Bayesian distribution



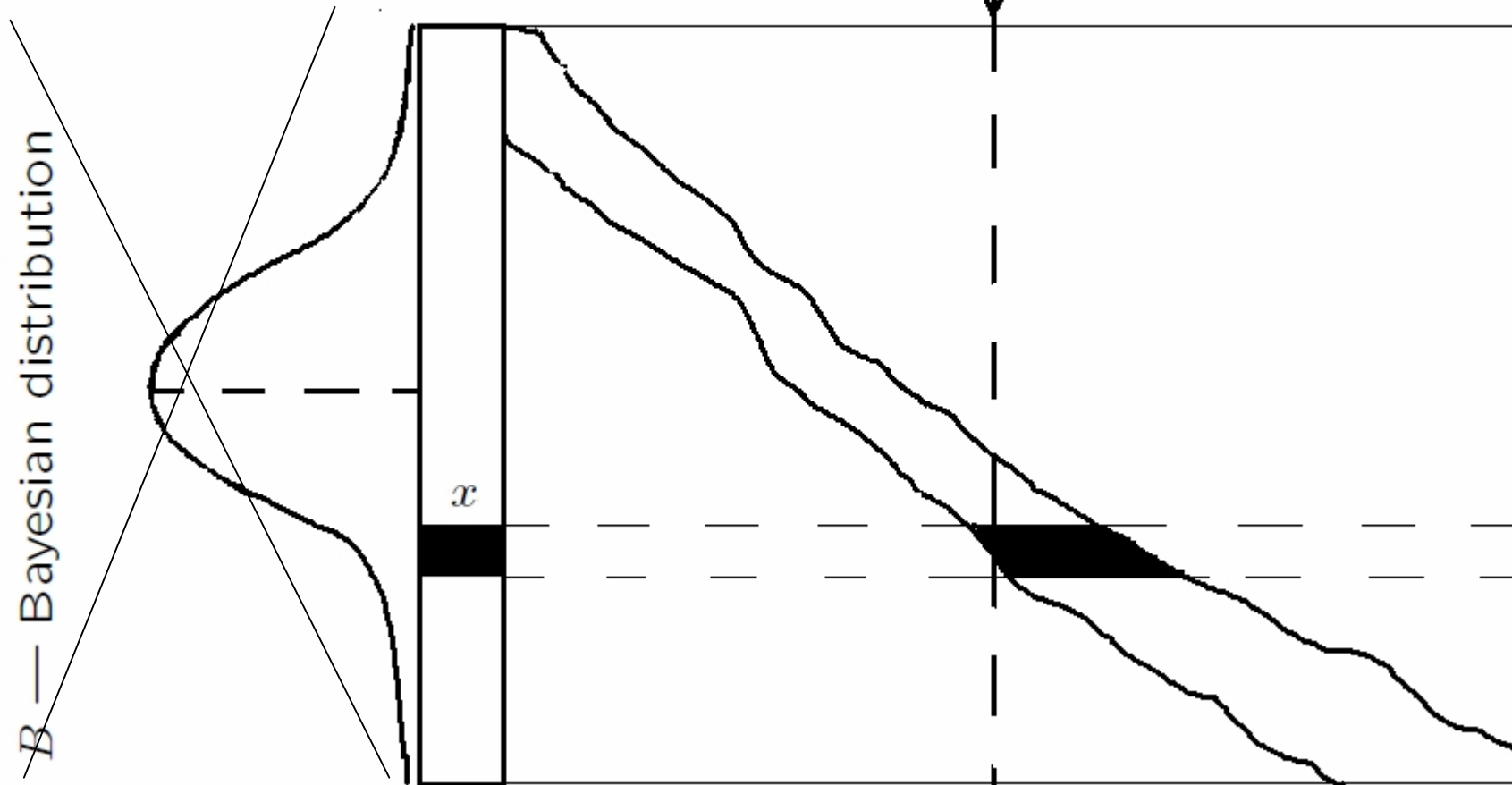
- For time series we can iteratively estimate B

Enrichment of the Observation's Data

$n = 1, 2, \dots$

$$y_n = \Phi_n x_n + v_n$$

P — random Perturbation



Randomization

1930...

- Fisher (remove bias)

1950...

- Metropolis, Ulam (method Monte-Carlo)

1980-90

- Luing, Guffi, Polyak, Thibakov, Spall (fast algorithms)
- Granichin (arbitrary noise)
- Vadiyasagar (randomized Learning Theory)

2000....

- Campi *etc.*

Best Features

- Significantly decreasing the number of operations
- Annihilating the systematic errors (the bias effect or an arbitrary noise)
- Accuracy usually not depend on the dimension of data

Arbitrary Noise

$$y_n = \varphi_n \theta^* + v_n, \quad n = 1, 2, \dots$$

$\{\varphi_n\}$ is known random with mean M_φ and $\sigma_\varphi^2 > 0$

$\{\theta^* = 1\}$ signal is present

$\{\theta^* = 0\}$ no signal

$\{v_n\}$ is unknown but bounded noise

$$(MSE) \quad \hat{\theta}_n = \frac{\sum_{k=1}^n \varphi_k y_k}{\sum_{k=1}^n \varphi_k^2} \rightarrow \theta^* + \frac{M_\varphi M_v}{\sigma_\varphi^2} \quad \text{a.s.}$$

$$\delta = \frac{M_\varphi M_v}{\sigma_\varphi^2} + \frac{1}{2} \quad \text{decision level}$$

If $\hat{\theta}_n < \delta$ then + otherwise —

Randomized Algorithm

$$\Delta_n = \varphi_n - M_\varphi$$

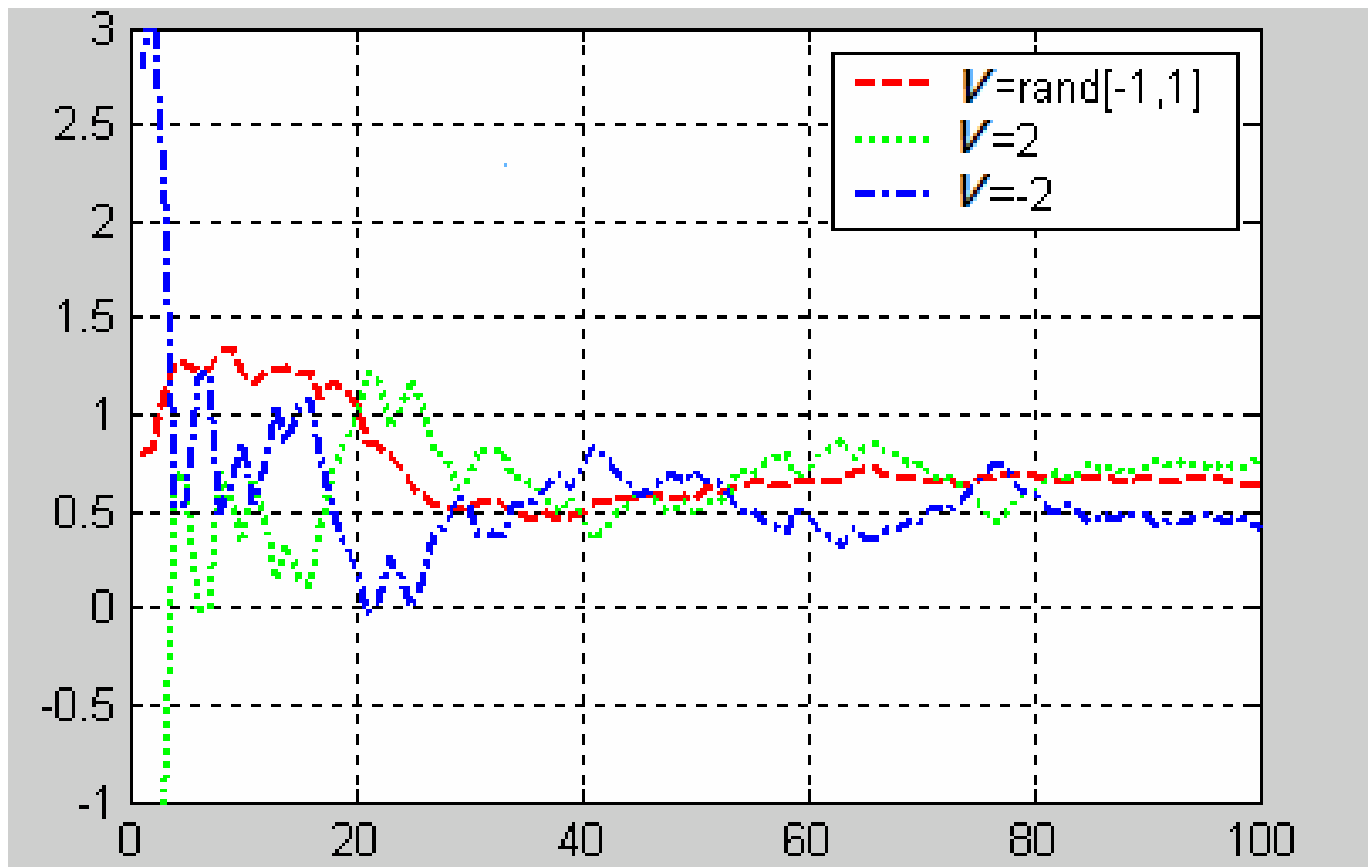
$$\Delta_n y_n = \Delta_n^2 \theta^* + \Delta_n M_\varphi \theta^* + \Delta_n v_n$$

$$\frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} = \theta^* + \frac{\sum_{k=1}^n \Delta_k (M_\varphi \theta^* + v_k)}{\sum_{k=1}^n \Delta_k^2}$$

$$(RMSE) \quad \hat{\theta}_n = \frac{\sum_{k=1}^n \Delta_k y_k}{\sum_{k=1}^n \Delta_k^2} \rightarrow \theta^* \quad \text{a.s.}$$

Simulation

$\{\varphi_n\}$ i.i.d.: $\varphi_n \in [0.5, 1.5]$, $|v_n| \leq 2$



Задание для самостоятельной работы

- Сгенерировать случайный симметрично распределенный вокруг 1 сигнал из интервале $[0,5;1,5]$ на промежутке времени 1:100
- Сгенерировать на промежутке времени 1:100 четыре последовательности помех
 - константа 2
 - константа -2
 - случайно равномерно из интервала $[-2,2]$
 - на промежутках времени 1:25 и 50:75 случайно из интервала $[1.9;2.1]$ и на промежутках 26:49 и 76:100 случайно из интервала $[-2.1;-1.9]$
- Сгенерировать восемь последовательностей наблюдений
- Для всех восьми случаев посчитать оценку МНК и сделать вывод о работоспособности алгоритма
- Для всех восьми случаев посчитать оценку РМНК и сделать вывод о работоспособности рандомизированного алгоритма