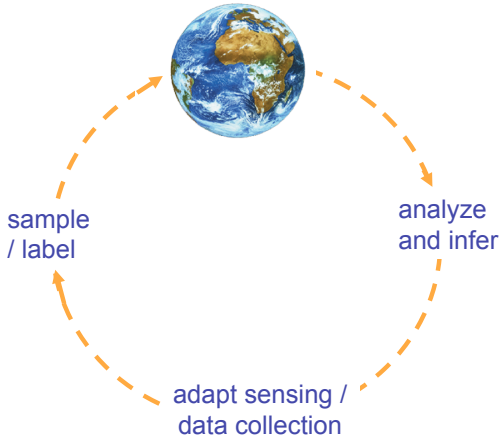
Adaptive Sensing and Information



IPAM Machine Reasoning
Workshop, September 20, 2010

THE UNIVERSITY of WISCONSIN MADISON
Rob Nowak
www.ece.wisc.edu/~nowak

Adaptive Information

Goal: Estimate an unknown object $x \in \mathcal{X}$ from scalar samples

Information: samples of the form $y_1(x), \ldots, y_n(x)$,
the values of certain functionals of x

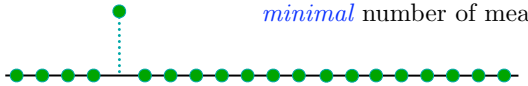
Non-Adaptive Information: $y_1, y_2, \cdots \in \mathcal{Y}$ non-adaptively
chosen (deterministically or randomly) independent of x

Adaptive Information: $y_1, y_2, \cdots \in \mathcal{Y}$ are selected sequentially and y_i can
depend on previously gathered information, i.e., $y_1(x), \ldots, y_{i-1}(x)$

Does adaptivity help?

Does Adaptivity Help ?

identify a sparse signal $x \in \mathbb{R}^n$ from a *minimal* number of measurements



Point measurments: $y = \langle x\,,\ \delta_k \rangle = x_k$

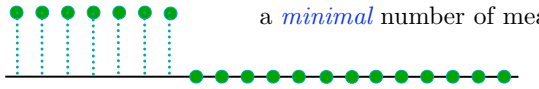
$O(n)$ measurements (random or adaptive) are needed to recover x

Compressed Sensing: $y\ =\ \langle x\,,\, \phi \rangle$ where $\phi \in \{-1, 1\}^n$

$O(\log n)$ measurements (random or adaptive) are needed to recover x

Adaptivity doesn't help

Does Adaptivity Help ?



identify a threshold signal $x \in \mathbb{R}^n$ from
a *minimal* number of measurements

Point measurments: $y = \langle x\,,\ \delta_k \rangle = x_k$

$O(n)$ random measurements are needed to recover x

$O(\log n)$ adaptive measurements are needed to recover x (binary search)

Compressed Sensing: $y\ =\ \langle x\,,\, \phi \rangle$ where $\phi \in \{-1, 1\}^n$

$O(\log n)$ random measurements are needed to recover x

Adaptivity may help, depending on nature of signal and measurements

Does Adaptivity Help ?

identify a sparse signal $x \in \mathbb{R}^n$
from *noisy* measurements

Noisy Compressed Sensing:

$$y \;=\; \langle x, \phi \rangle + \epsilon \text{ , where } \phi \in \{-1, 1\}^n, \epsilon \sim \mathcal{N}(0, 1)$$

$O(\log n)$ measurements suffice if

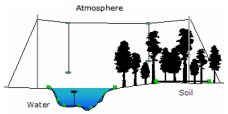
$\max_k x_k > C_0 \sqrt{\log n}$, for random measurements

$\max_k x_k > C_1$, for adaptive measurements

where $C_0, C_1 > 0$ are constants.

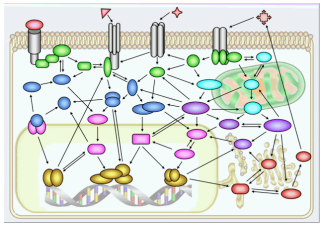
Adaptivity can help in noisy situations

Grand Challenge:
Understanding Large Networked Systems

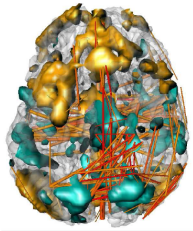


Technological Networks
(Internet Mapping Project, US power grid, UCLA CENS)

Social Networks



Biological Networks
(JMDBase)



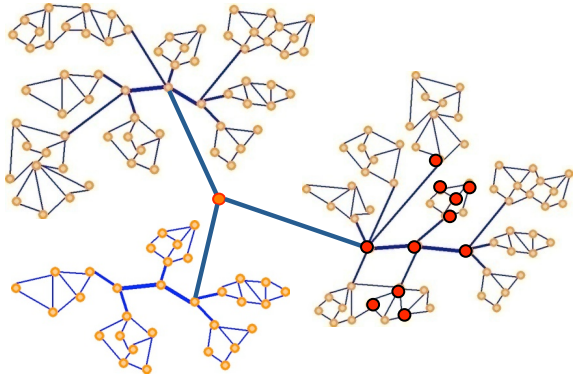
Brain Networks
(Worsley et al, 2005)

Challenges:

- Inferring structure & function of the network
- Pattern analysis & anomaly detection

Laplace defined science as a predicting tool.

Large-scale Network Monitoring



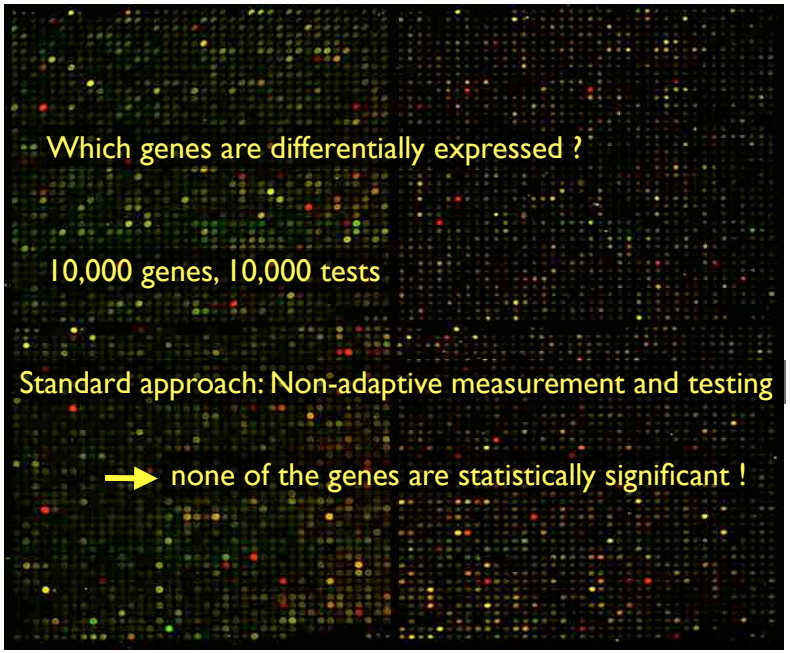
Which nodes are are behaving anomalously/maliciously ?

10,000 nodes, 10,000 tests

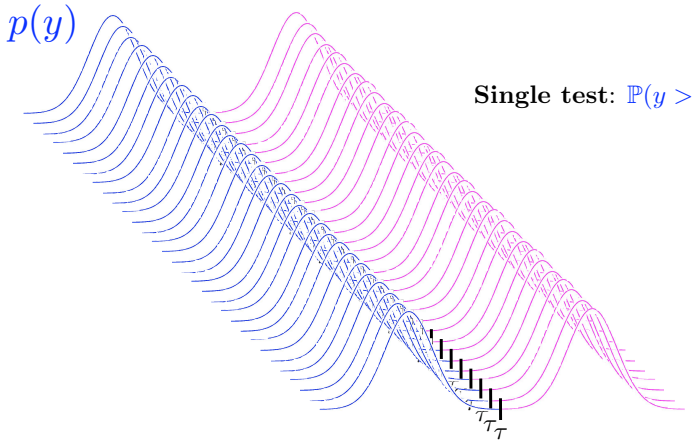
Standard approach: non-adaptive measurement and testing

→ none of the nodes are statistically significant !

Biological Systems



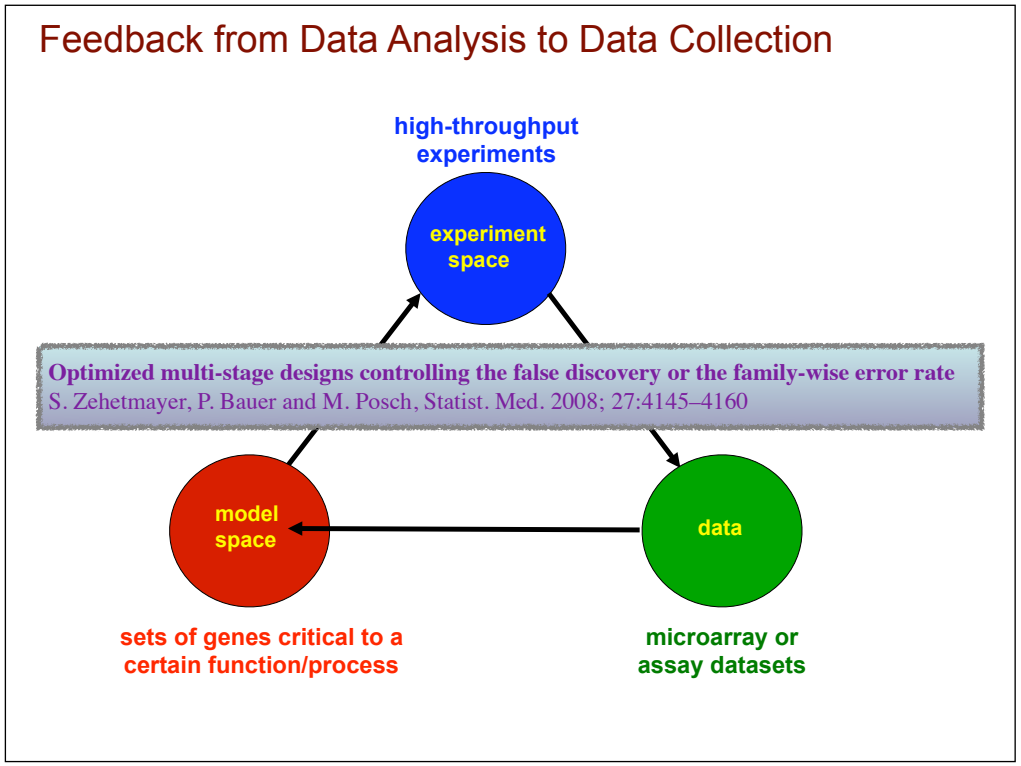
The Multiple Testing Problem



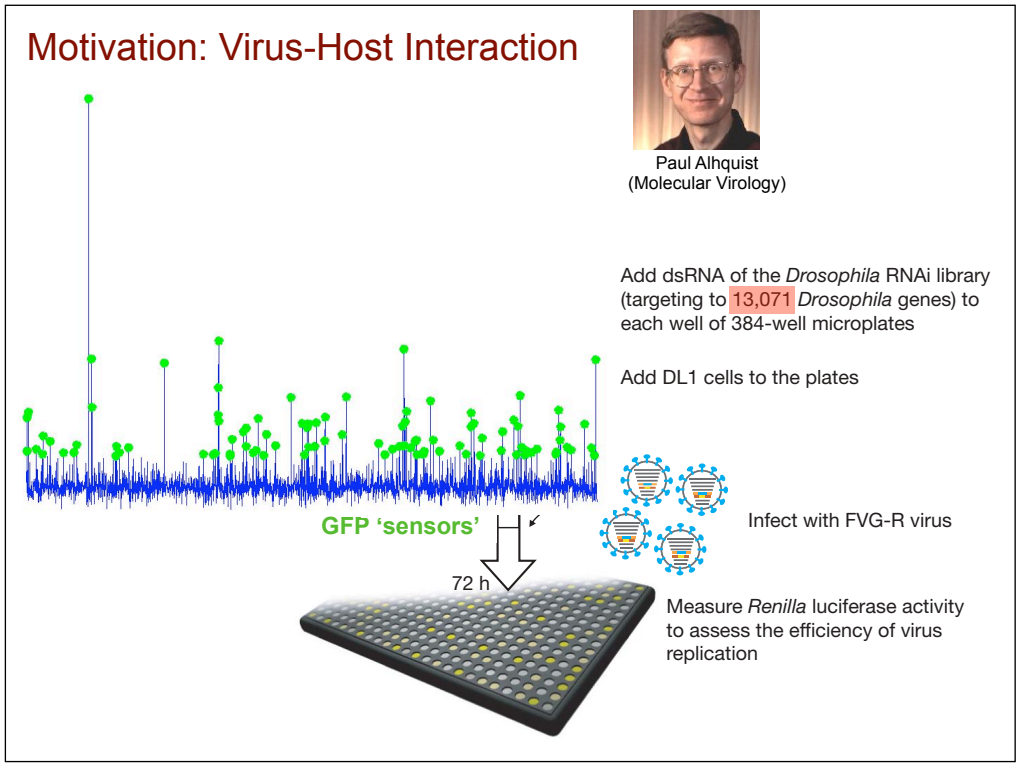
Single test: $\mathbb{P}(y > \tau) \leq \exp(-\tau^2/2)$

n tests: $\mathbb{P}(\max_{i=1}^{n} y_i > \tau) \sim n \exp(-\tau^2/2) = \exp(-\tau^2/2 + \log n)$

$$\Rightarrow \quad \tau > \sqrt{2 \log n}$$



Laplace defined science as a predicting tool.



Laplace defined science as a predicting tool.

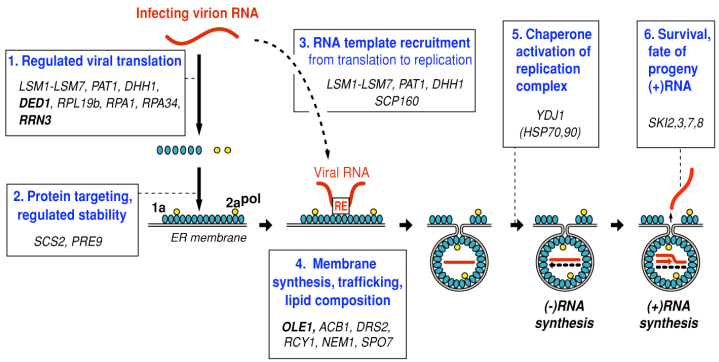
Adaptive Experimentation

How do they confidently determine the ~100 out of 13K genes hijacked for virus replication from extremely noisy data?

Multistage Adaptive Experiments:

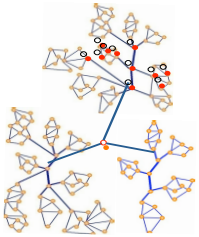
Stage 1: assay all 13K genes, twice; keep all with significant fluorescence in one or both assays for 2nd stage (13K → 1K)

Stage 2: assay remaining 1K genes, 6-12 times; retain only those with statistically significant fluorescence (1K → 100)

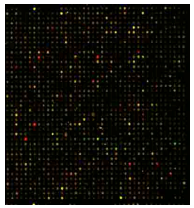


Laplace defined science as a predicting tool.

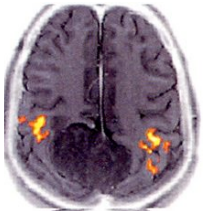
Detection/Estimation of Sparse Signals



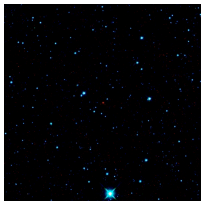
network monitoring



systems biology



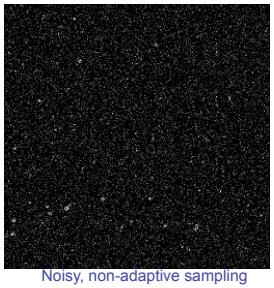
brain mapping



astronomy

We observe sparse signals in noise and wish to reliably detect the signal components.

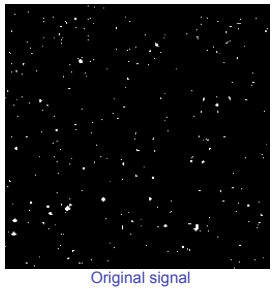
Astronomical Surveying “On a Budget”



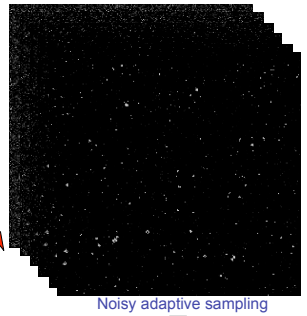
Noisy, non-adaptive sampling

Recovery from non-adaptive samples
(5% of “discoveries” are errors)

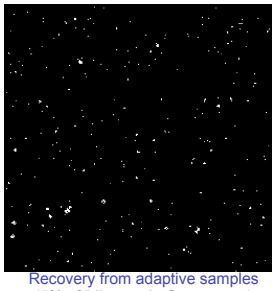
keep only “brightest” 50%
and re-measure each
for ~~two~~ four times as long



Original signal
(~0.8% non-zero components)



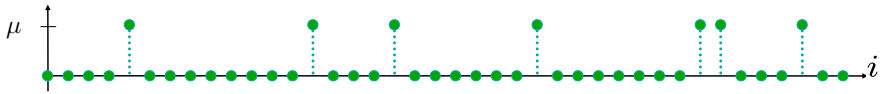
Noisy adaptive sampling



Recovery from adaptive samples
(5% of “discoveries” are errors)

Sparse Signal Model

Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ be an unknown sparse vector;
most (or all) of its components x_i are equal to zero.



$$x_i \;=\; \begin{cases} \mu > 0\,, & i \in \mathcal{S} \\ 0\,, & i \notin \mathcal{S} \end{cases} \;, \text{ where } |\mathcal{S}| \ll n$$

signal **support set**

deterministic
but unknown

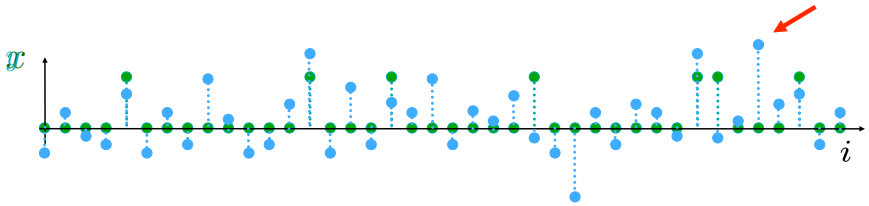
Assume sublinear sparsity level: $|\mathcal{S}| \;=\; n^{1-\beta}\,, 0 < \beta < 1$

number of signal
components

Noisy Observation Model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

$$z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$$



Suppose we want to locate just **one** signal component: $\widehat{i} \;=\; \arg\max_i y_i$

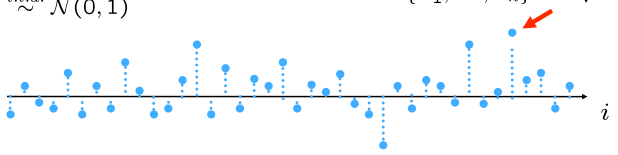
Even if no signal is present, $\max_i y_i \;\sim\; \sqrt{2\log n}$

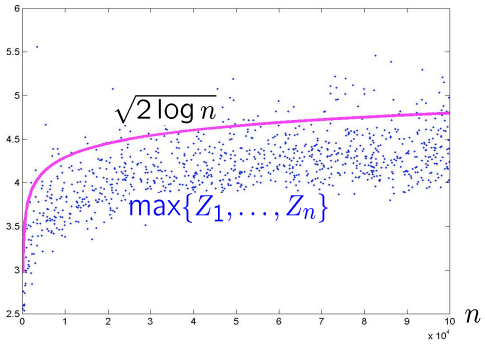
It is *impossible* to reliably detect signal components weaker than $O(\sqrt{\log n})$

Too much noise !

$Z_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$

$\max\{Z_1, \ldots, Z_n\} \sim \sqrt{2 \log n}$

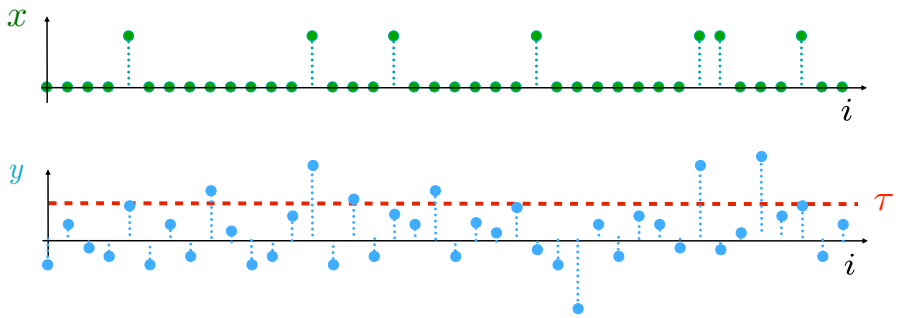




10,000 random trials
at various n

Threshold Tests

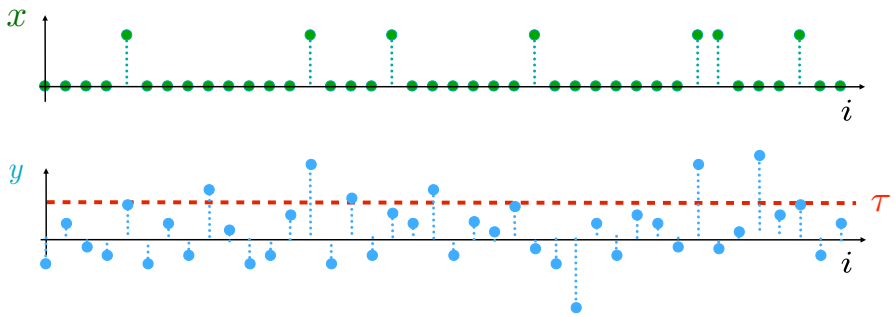
Our goal is to estimate the set of non-zero components: $\mathcal{S} := \{i : x_i \neq 0\}$



Definition 1 *A* threshold test *is an estimator of the form:*

$$\widehat{\mathcal{S}}_\tau(y) \;:=\; \{i \in \{1,\dots,n\} : y_i \geq \tau > 0\}$$

Family-wise Error Rate



Definition 2 *The family-wise error rate (FWER) is the probability of making one or more* false discoveries

$$\mathrm{FWER}(\widehat{\mathcal{S}}) \;:=\; \mathbb{P}\left(|\widehat{\mathcal{S}} \backslash \mathcal{S}| \geq 1\right) \;=\; \mathbb{P}\,(\#\ \textit{falsely discovered components} \geq 1)$$

Bonferroni Correction: To keep the FWER small (e.g., less than 5%) the threshold must be on the order of $\sqrt{\log n}$.

False Discovery Rate

FWER control: $\mathbb{P}(\text{one or more false discoveries}) \leq \alpha$ too conservative

Alternative: $\#$ false discoveries $\sim \alpha\,|\widehat{\mathcal{S}}|$

number of false discoveries should be a small fraction of number of total discoveries

Definition 3 *Let $\mathcal{S} := \{i : x_i \neq 0\}$ and let $\widehat{\mathcal{S}}(y)$ denote an estimator of $\mathcal{S}$. The false-discovery proportion is*

$$\text{FDP}(\widehat{\mathcal{S}}) := \frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|} = \frac{\#\ \textit{falsely discovered components}}{\textit{total}\ \#\ \textit{discovered components}}$$

Definition 4 *The false-discovery rate is the expected value of the false-discovery proportion*

$$\text{FDR}(\widehat{\mathcal{S}}) := \mathbb{E}\left[\frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|}\right] = \mathbb{E}\left[\frac{\#\ \textit{falsely discovered components}}{\textit{total}\ \#\ \textit{discovered components}}\right]$$

False Discovery Rate (FDR) Control

To guarantee FDR $\leq \alpha$

1. Compute p-values

   $p_i := \mathbb{P}(\mathcal{N}(0,1) > y_i),\ i = 1, \ldots, n$

2. Sort p-values

   $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(n)}$

3. Set threshold

   $i^* := \max\{i : p_{(i)} < \frac{i\alpha}{n}\}$
   $\tau = p_{(i^*)}$

This is the so-called Benjamini-Hochberg (BH) threshold

The BH threshold is more lenient than Bonferroni when the signal is not too sparse

Theoretical Performance in High-Dimensional Limit

$$\mathrm{FDP}(\widehat{\mathcal{S}}) := \frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|} = \frac{\#\text{ falsely discovered components}}{\text{total }\#\text{ discovered components}}$$

Definition 5 *The non-discovery proportion is*

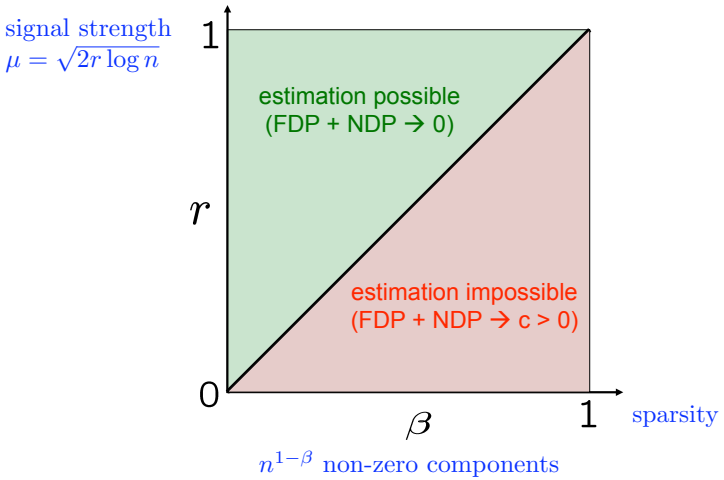
$$\mathrm{NDP}(\widehat{\mathcal{S}}) := \frac{|S \backslash \widehat{\mathcal{S}}|}{|S|} = \frac{\#\ \textit{missed components}}{\#\ \textit{true non-zero components}}$$

Theorem 1 *Assume x has $n^{1-\beta}$, $\beta \in (0,1)$, non-zero components of amplitude $\sqrt{2r\log n}$, $r > 0$. There exists a threshold test that yields an estimator $\widehat{\mathcal{S}} = \widehat{\mathcal{S}}(y)$ such that if $r > \beta$, then as $n \to \infty$,*

$$\mathrm{FDP}(\widehat{\mathcal{S}}) \xrightarrow{P} 0\,, \quad \mathrm{NDP}(\widehat{\mathcal{S}}) \xrightarrow{P} 0\,,$$

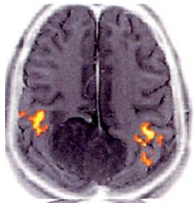
where $\xrightarrow{P}$ denotes convergence in probability. Moreover, if $r < \beta$, then there does not exist a coordinate-wise thresholding procedure that can guarantee that both quantities above tend to 0 as $n \to \infty$.

Limit of Detectability (Ingster '97, Jin & Donoho '03)



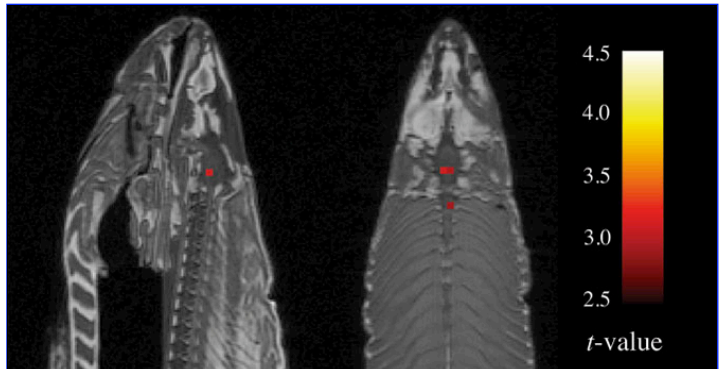
signal amplitude must be $O(\sqrt{\log n})$!

Is there really a problem ?



Scanning Dead Salmon in fMRI Machine Highlights Risk of Red Herrings

By Alexis Madrigal ✉ September 18, 2009 | 5:37 pm | Categories: Brains and Behavior



Science and Data Analysis

Often there are no components that exceed $\sqrt{\log n}$ in magnitude!

Nothing is statistically significant at FWER=5% or FDR=5%.

What's a scientist to do?

Take a closer look at the most significant components.

What do scientists need?

perhaps **not** an FWER/FDR certification

Principled guidance on how to design the next experiment
(and the next, and the next...)

* how many components should be re-tested
* guarantees that important effects will not be overlooked or discarded prematurely

Sequential Experimental Design

Instead of the usual non-adaptive observation model

$$y_i \;=\; x_i \;+\; z_i\,,\; i = 1, \ldots, n$$

suppose we are able to sequentially collect multiple noisy measurements of each component of x, according to

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

where

- j indexes the measurement steps
- k denotes the total number of steps
- $z_{i,j} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$
- $\gamma_{i,j} \geq 0$ controls the precision of each measurement

Total precision budget is constrained, but the choice of $\gamma_{i,j}$ can depend on past observations $\{y_{i,\ell}\}_{\ell<j}$.

Experimental (Precision) Budget

sequential measurement model

$$y_{i,j} \;=\; x_i \;+\; \gamma_{i,j}^{-1/2}\, z_{i,j}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, k$$

The precision parameters $\{\gamma_{i,j}\}$ are required to satisfy

$$\sum_{j=1}^{k} \sum_{i=1}^{n} \gamma_{i,j} \;\leq\; n$$

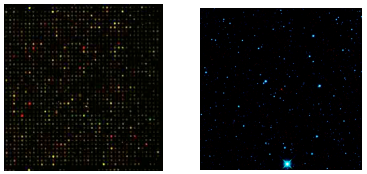
For example, the usual non-adaptive, single measurement model corresponds to taking $k = 1$, and $\gamma_{i,1} = 1$, $i = 1, \ldots, n$. This baseline can be compared with adaptive procedures by allowing $k > 1$ and variable $\{\gamma_{i,j}\}$ satisfying budget.

Precision parameters control the SNR per component.

SNR is increased/decreased by

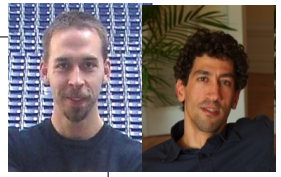
— more/fewer repeated samples or

— longer/shorter observation times



Distilled Sensing

Jarvis Haupt Rui Castro



Distilled Sensing

initialize: $j = 0$, $\mathcal{S}_0 = \{1, \ldots, n\}$

number of measurement steps: k

precision per step: $R_j = \sum_{i=1}^{n} \gamma_{i,j}$, $j = 1, \ldots, k$

total precision: $R = \sum_{j=1}^{k} R_j = n$

for $j = 1, \ldots, k$

1) Allocate precision budget R_j uniformly over indices in $\mathcal{S}_{j-1}$

2) Measure components in $\mathcal{S}_{j-1}$ accordingly to obtain $\{y_{i,j}\}$

3) Set $\mathcal{S}_j = \{i : y_{i,j} \geq 0\}$

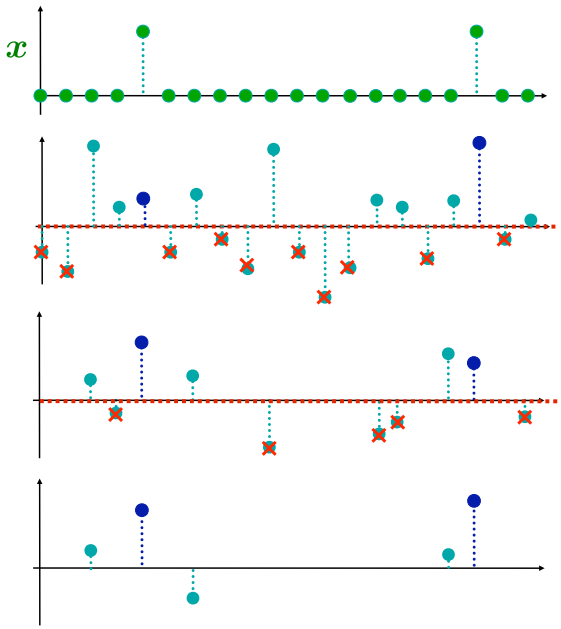
end

output: $\{y_{i,k}\}$

Key idea: At each step DS eliminates about 1/2 of the components from further consideration. If $R_j/R_{j-1} = \rho > 1/2$, then SNR in remaining components is *increased* at next step by a factor $\approx 2\rho$.

total number of measurements $\approx 2n$

Idealized Example



Let $k = 3$ and
$R_j = n/3$, $j = 1, 2, 3$

$$y_{i,1} = x_i + \mathcal{N}(0, 3)$$

$$y_{i,2} = x_i + \mathcal{N}\left(0, \frac{3}{2}\right)$$

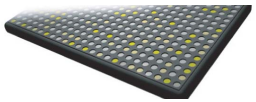
$$y_{i,3} = x_i + \mathcal{N}\left(0, \frac{3}{4}\right)$$

SysBio Example

virus

fruit fly

assay



How to find genes involved in virus replication ?

Distilled Sensing Idea

Budget: **k assays**, n tests/assay

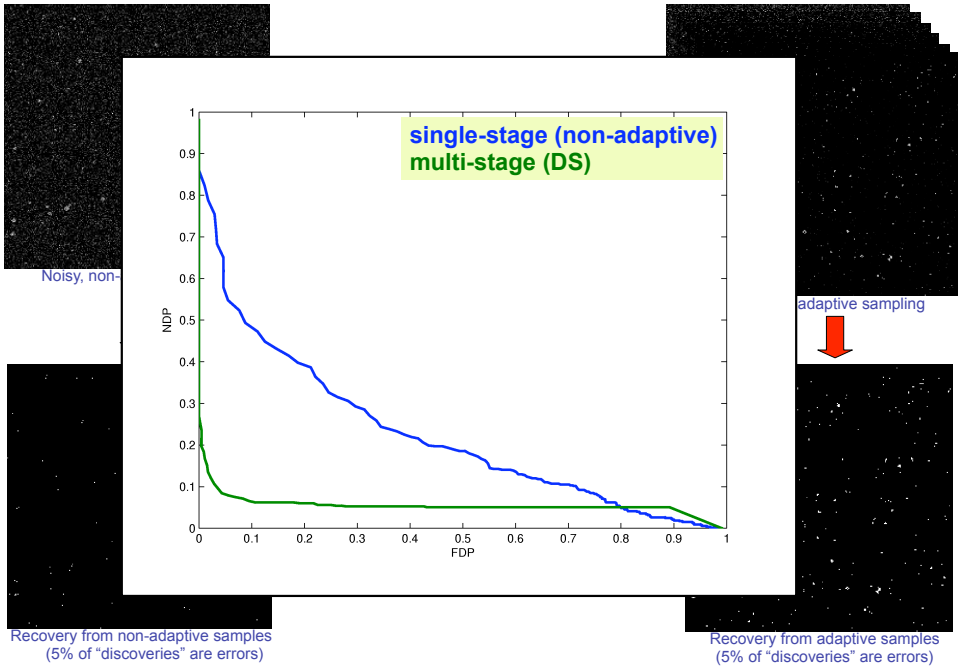
Assay 1: measure fluorescence of all n genes; discard n/2 genes with weakest fluorescence.

Assay 2: measure fluorescence for remaining n/2 genes, each tested twice (double SNR); discard n/4 genes with weakest fluorescence.

Assay 3: measure fluorescence for remaining n/4 genes, each tested four times (quadruple SNR); discard n/8 genes with weakest fluorescence.

continue distilling....

Astronomical Surveying “On a Budget”



Noisy, non-adaptive sampling

Recovery from non-adaptive samples
(5% of “discoveries” are errors)

Recovery from adaptive samples
(5% of “discoveries” are errors)

Distilled Sensing Theorem

Theorem 2 *Assume* $x \geq 0$ *with* $n^{1-\beta}$, $\beta \in (0,1)$, *non-zero components of amplitude* $\mu(n)$, *and sequential measurement model DS with* $k = \lceil \log_2 \log n \rceil + 2$, *and precision budget distributed over the measurement steps so that* $\sum_{j=1}^{k} R_j \leq n$, $R_{j+1}/R_j \geq \delta > 1/2$, *and* $R_1 = c_1 n$ *and* $R_k = c_k\, n$ *for some* $c_1, c_k \in (0,1)$. *Consider the thresholding estimator based on the output of the DS procedure:*

$$\widehat{\mathcal{S}}_{\mathrm{DS}} \quad := \quad \{i \in I_k : y_{i,k} > \sqrt{2/c_k}\}$$

If $\mu(n)$ *is any positive diverging sequence in* n, *then as* $n \to \infty$

$$\mathrm{FDP}(\widehat{\mathcal{S}}_{\mathrm{DS}}) \xrightarrow{P} 0\,, \quad \mathrm{NDP}(\widehat{\mathcal{S}}_{\mathrm{DS}}) \xrightarrow{P} 0\,.$$

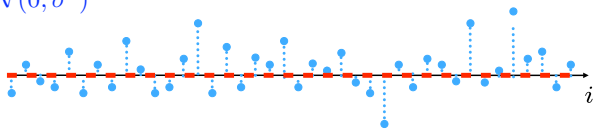
- Take $k \sim \log \log n$ steps, with geometric (decreasing) allocation of budget
- Discard the weakest half of measured components at each step
- Re-measure only surviving components

With very high probability will detect almost all components with amplitude $\geq$ than some constant, independent of n

Key Elements of Proof:

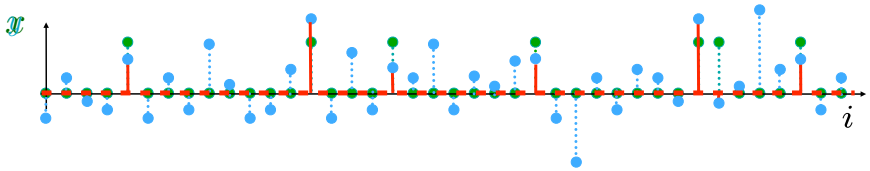
Each step *probably* removes about 1/2 of noise

$y_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$



If we threshold an iid zero-mean Gaussian process at 0,
then only about 1/2 of the components will be retained

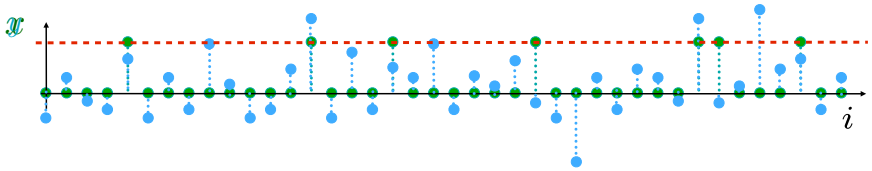
Each step *probably* retains almost all signal components



Thresholding at 0 retains almost all of the non-zero signal components

Summary of DS Performance

$\widehat{\mathcal{S}}(y) := \{\, i : y_i > \text{threshold} \,\}$



Let $\mathcal{S} := \{i = 1, \ldots, n : x_i \neq 0\}$ and let $\widehat{\mathcal{S}}(y)$ denote an estimator of $\mathcal{S}$.

$$\text{FDP}(\widehat{\mathcal{S}}) := \frac{|\widehat{\mathcal{S}} \backslash \mathcal{S}|}{|\widehat{\mathcal{S}}|} = \frac{\#\text{ falsely discovered components}}{\text{total }\#\text{ discovered components}}$$

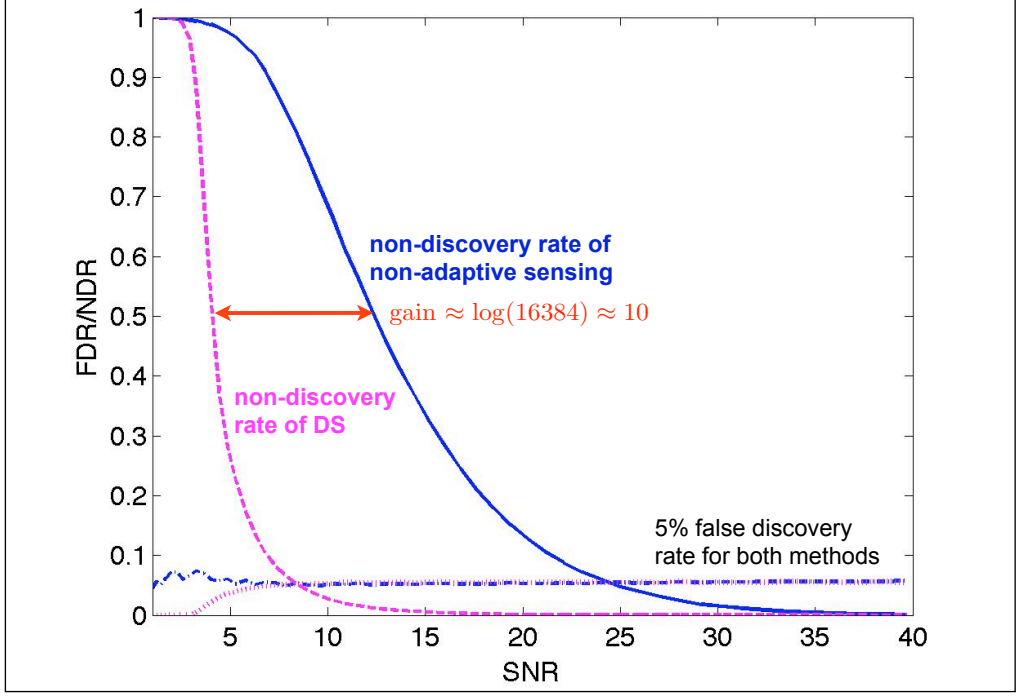
$$\text{NDP}(\widehat{\mathcal{S}}) := \frac{|S \backslash \widehat{\mathcal{S}}|}{|S|} = \frac{\#\text{ missed components}}{\#\text{ true non-zero components}}$$

To guarantee $\text{FDP}(\widehat{\mathcal{S}}) \xrightarrow{P} 0$, $\text{NDP}(\widehat{\mathcal{S}}) \xrightarrow{P} 0$ as $n \to \infty$, we require

non-adaptive	$\text{SNR} \sim \log n$
adaptive	$\text{SNR} \sim$ arbitrarily slowly growing function of n

Adaptivity buys about $\log n$ in SNR; e.g., if $n = 13071$, then $\log n \approx 10$

Example $n = 2^{14},\ \|x\|_0 = \sqrt{n} = 128$



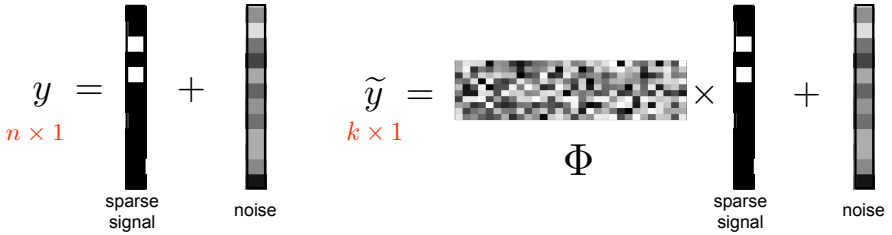
Experimental Budgets

Distilled Sensing acquires approximately $2n$ measurements:
n in the first stage, $\sim n/2$ in the second, $\sim n/4$ in the third, etc.

But there are far fewer, $|\mathcal{S}| \ll n$, relevant components to detect.

Can we find these components with less than n measurements?

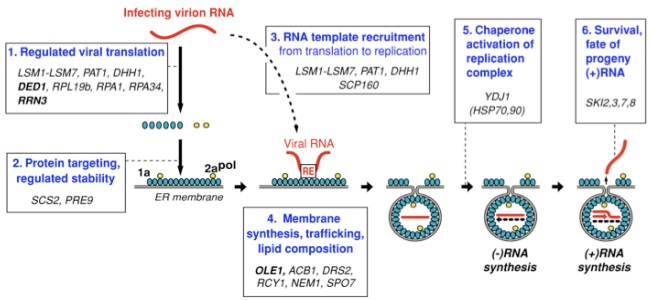
Yes, if we measure randomized combinations of the components.



direct measurement

indirect (randomized) measurement

Motivation: Systems Biology



Linear sparsity model isn't always right! Two or more genes may produce proteins will similar functionality or interact to perform a certain function. Need to study high-dimensional systems with sparse *interactions*.

$$\textbf{Linear model:} \quad y(x) \;=\; \langle x, \phi \rangle \;=\; \sum_i x_i\, \phi_i \qquad \phi \in \{-1, +1\}^n$$

$$\textbf{Nonlinear model:} \quad y(x) \;=\; \sum_i x_i^{(1)}\, \phi_i \;+\; \sum_{i<j} x_{i,j}^{(2)}\, \phi_i \phi_j$$

$\binom{13000}{2} \approx 85,000,000$ possible two-fold gene deletion strains !

The fluorescence data are very noisy, and so biologists employ a multistage adaptive experimental procedure to home-in on the important genes. The results of the first stage inform the design of the second stage. This is an example of feedback from data analysis to data collection. This is what scientists are doing all the time. The figure above shows the inferred network pathway exploited by the virus for the purposes of replication. Unfortunately, single gene deletions may not tell the whole story, since certain genes act together in concert. Such effects can only be detected through multiple gene deletions. Multiple deletion studies become daunting or impossible quite rapidly; e.g., there are 85 million pairwise deletions that can be considered in the case of Drosophila. Biologists have the technology (gene knockdowns) to study multiple deletions, but it is difficult to decide where to start and how to proceed.

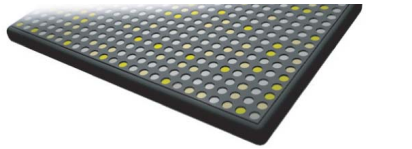
Randomized Experiments

$$\underset{k \times 1}{\widetilde{y}} = \times + +$$

sparse signal

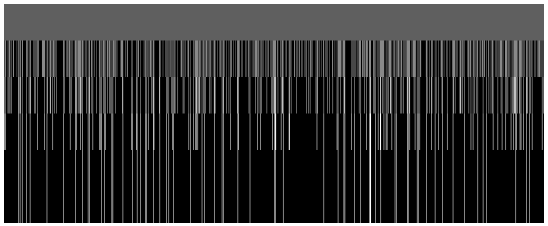
noise

indirect (randomized) measurement



put randomly modulated “knock-down” strains into each well

Randomized Distilled Sensing



stage 1: random combinations of all components

stage 2: random combos top 1/2 components

stage 3: random combos top 1/4 components

stage 4: random combos top 1/8 components

stage 5: random combos top 1/16 components

stage 6: random combos top 1/32 components

support of random sensing matrix

Theorem 4 *Assume $x \geq 0$ with $|\mathcal{S}|$ non-zero components of amplitude μ, and take $O(|\mathcal{S}| \log n)$ randomized DS measurements. If $\mu \succeq \log\log\log n$, then there is a threshold test guaranteeing that*

$$\mathrm{FDP}(\widehat{\mathcal{S}}_{\mathrm{RDS}}) \xrightarrow{P} 0\,, \quad \mathrm{NDP}(\widehat{\mathcal{S}}_{\mathrm{RDS}}) \xrightarrow{P} 0\,.$$

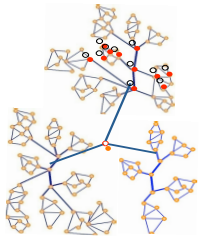
In essence, this implies that we can identify most of the $|\mathcal{S}|$ relevant components using only slightly more than $|\mathcal{S}|$ measurements.

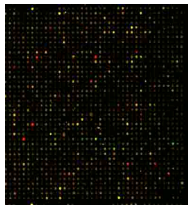
Scientific Discovery is a Closed-Loop Process

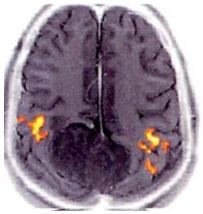
To guarantee that the proportions of false-discoveries and missed components tends to zero as $n \to \infty$, we require

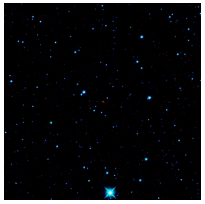
non-adaptive	SNR	$\sim \log n$
adaptive	SNR	$\sim$ constant

Adaptivity effectively eliminates the fundamental statistical challenge in high-dimensional multiple testing.


network monitoring


systems biology


brain mapping


astronomy

more information: www.ece.wisc.edu/~nowak

http://arxiv.org/abs/1001.5311

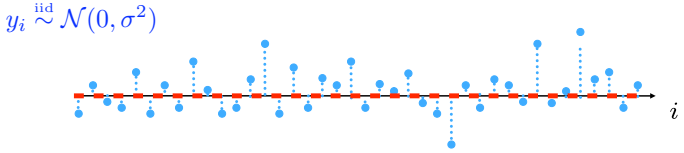
Laplace defined science as a predicting tool.

Key Elements of DS Theorem: Remove 1/2 the Noise

Lemma 1 *If* $\{y_i\}_{i=1}^m \overset{\text{iid}}{\sim} \mathcal{N}(0,\sigma^2)$, $\sigma > 0$, *then for any* $0 < \varepsilon < 1/2$,

$$\left(\frac{1}{2} - \varepsilon\right) m \leq \Big|\{i \in \{1,\ldots,m\} : y_i > 0\}\Big| \leq \left(\frac{1}{2} + \varepsilon\right) m,$$

with probability at least $1 - 2\exp\left(-2m\varepsilon^2\right)$.



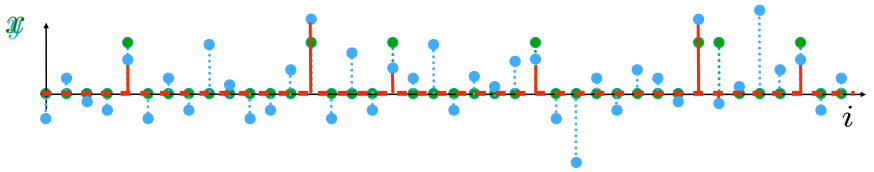
If we threshold an iid zero-mean Gaussian process at 0, then only about 1/2 of the components will be retained

Key Elements of DS Theorem: Keep the Signal

Lemma 2 *Let* $\{y_i\}_{i=1}^m \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$, *with* $\sigma > 0$ *and* $\mu \geq 2\sigma$. *Define* $\epsilon = \frac{\sigma}{\mu\sqrt{2\pi}} < 1$. *Then*

$$(1-\epsilon)m \leq \left|\{i \in \{1, 2, \ldots, m\} : y_i > 0\}\right| \leq m,$$

with probability at least $1 - \exp\left(-\frac{\mu m}{4\sigma\sqrt{2\pi}}\right)$.



Thresholding at 0 retains almost all of the non-zero signal components

Key Elements of DS Theorem: Iteration

At each step of DS, $j = 1, \ldots, k$, define

$$\begin{aligned} s_j &:= \text{number of non-zero components} \\ z_j &:= \text{number of zero components} \end{aligned}$$

Lemma 3 *Let* $0 < \varepsilon < 1/6$ *and assume that* $R_j > \frac{8(s_1+z_1)}{\mu^2}(1/2+\varepsilon)^{j-1}$, $j = 1, \ldots, k-1$. *Then*

$$(1-\varepsilon)^{j-1} s_1 \leq s_j \leq s_1$$

and

$$\left(\frac{1}{2}-\varepsilon\right)^{j-1} z_1 \leq z_j \leq \left(\frac{1}{2}+\varepsilon\right)^{j-1} z_1$$

for $j = 2, \ldots, k$ *with probability at least*

$$1 - \sum_{j=1}^{k-1} \exp\left(\frac{-s_1(1-\varepsilon)^{j-1}}{2\sqrt{2\pi}}\right) - 2\sum_{j=1}^{k-1} \exp\left(-2z_1(1/2-\varepsilon)^{j-1}\varepsilon^2\right) ,$$

With high probability each distillation keeps almost all the non-zero components and rejects about half of the non-signal components.

SNR *increases* at each step