Data Analysis: Modeling FMRI Signals and Noise

- Signal = Measurable response to stimulus

  Noise = Components of measurement that interfere with detection of signal

- Statistical detection theory:
  - ◇ Must understand relationship between stimulus and signal
  - ◇ Must characterize noise statistically
  - ◇ Can then devise methods to distinguish noise-only measurements from signal+noise measurements, and assess their reliability

- FMRI signals and noise:
  - ◇ Stimulus→signal and noise statistics are both poorly characterized
  - ◇ Result is that there is no "best" way to analyze FMRI time series data: there are only "reasonable" analysis methods
  - ◇ To deal with data, must make some assumptions about the signal and noise
  - ◇ These assumptions will be wrong, but have to do something
  - ◇ Different kinds of experiments require different kinds of analyses, since signal models and questions you ask about the signal will vary

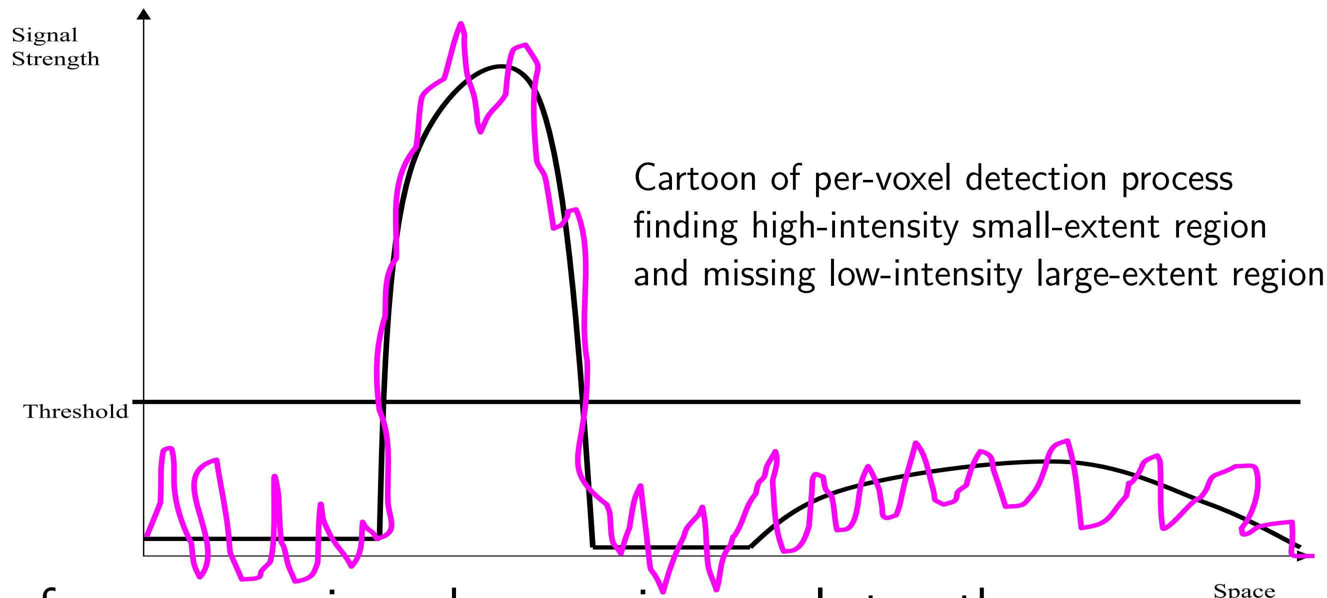
  ↪ Therefore it is important to understand what is going on, so you can select and evaluate "reasonable" analysis approaches

- Meta-method for generating analysis methods:
  - ◇ Write down a mathematical model connecting stimulus to signal
  - ◇ Write down a statistical model for the noise
  - ◇ Combine them to produce an equation for measurements given signal+noise (signal may have zero strength)
  - ◇ Use statistical detection theory to produce an algorithm for processing the measurements to assess signal presence and characteristics
- Fundamental difficulty with neuroimaging data: don't have enough measurements
  - ◇ Sounds crazy: typically get $\frac{1}{2}$ Gbyte of data per scanning sesssion
  - ◇ But most of this is not relevant to neural activity (BOLD signal is weak)
  - ◇ Must make <u>many</u> decisions to make a brain map: at least one per voxel
    - ↪ Typically have $10^4 \dots 10^5$ voxels in the brain
    - ↪ If chance of making a mistake in any one voxel is 1% ($p = 0.01$), then expect $100 \dots 1000$ errors in every brain map
    - ↪ This may be as big as the number of truly active voxels in the brain $\Rightarrow$ results are garbage

◇ There are two ways out of this multiple comparisons problem:

↪ Make the p-value per voxel much more stringent (smaller), so that the number of expected errors goes way down

▷ Problem: Low-intensity large-extent activations will be tossed out



↪ Make fewer comparisons by grouping voxels together

▷ Spatial smoothing of the data prior to detection (à la PET analyses)

▷ Analyze data only after averaging over regions-of-interest (ROIs)

▷ Two-step detection (spatial clustering):

- Provisionally accept as active voxels above some threshold signal level
- Finally accept only voxels contiguous with other supra-threshold voxels
- Isolated "salt and pepper noise" activations are excised

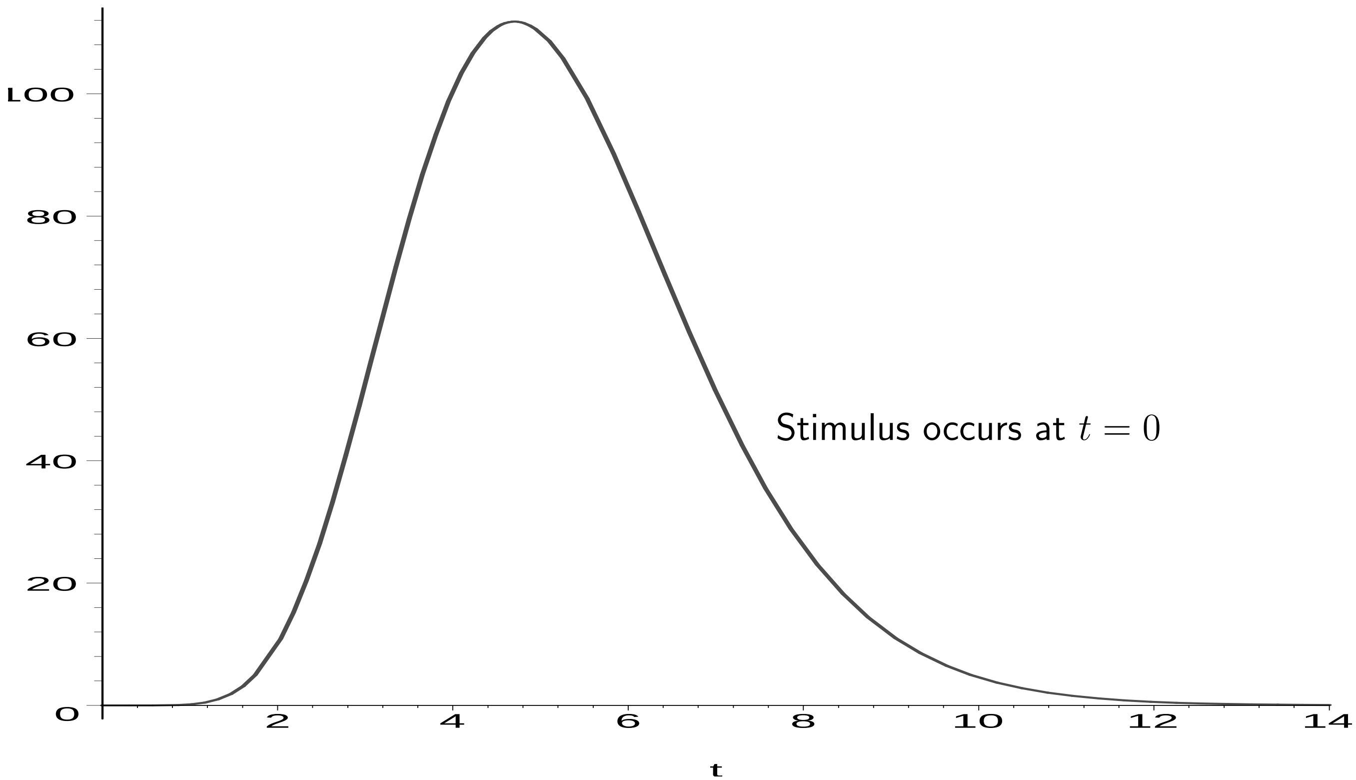
▷ Problem: Loss of spatial resolution

- Mathematical theory follows:
  - ◇ Will try to follow notation Doug Ward used in his manual for `3dDeconvolve`
  - ◇ Time occurs continuously in reality, but in steps in data acquisition
    - ↪ Functions of continuous time are expressed like $f(t)$
    - ↪ Functions of discrete time are expressed like $f(n\Delta t)$, where $n = 0, 1, 2, \ldots$ and $\Delta t =$ time step (also called TR in MRI)
    - ↪ May also use subscript notation f_n to mean same thing as $f(n\Delta t)$
    - ↪ A collection of numbers assembled in a column is a vector and is printed in boldface, as in

$$\begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ \vdots \\ f_M \end{bmatrix} = \mathbf{f} \qquad (\underline{\mathrm{f}} \text{ in handwriting})$$

- Will begin with a simple case, and work upward to models that are realistic for complex neuropsychological experiments

- Model for signal: Known and fixed response to each stimulus
  - ◇ Example: response function $\boldsymbol{h(t) = t^{8.6} e^{-t/0.547}}$ (Mark Cohen)



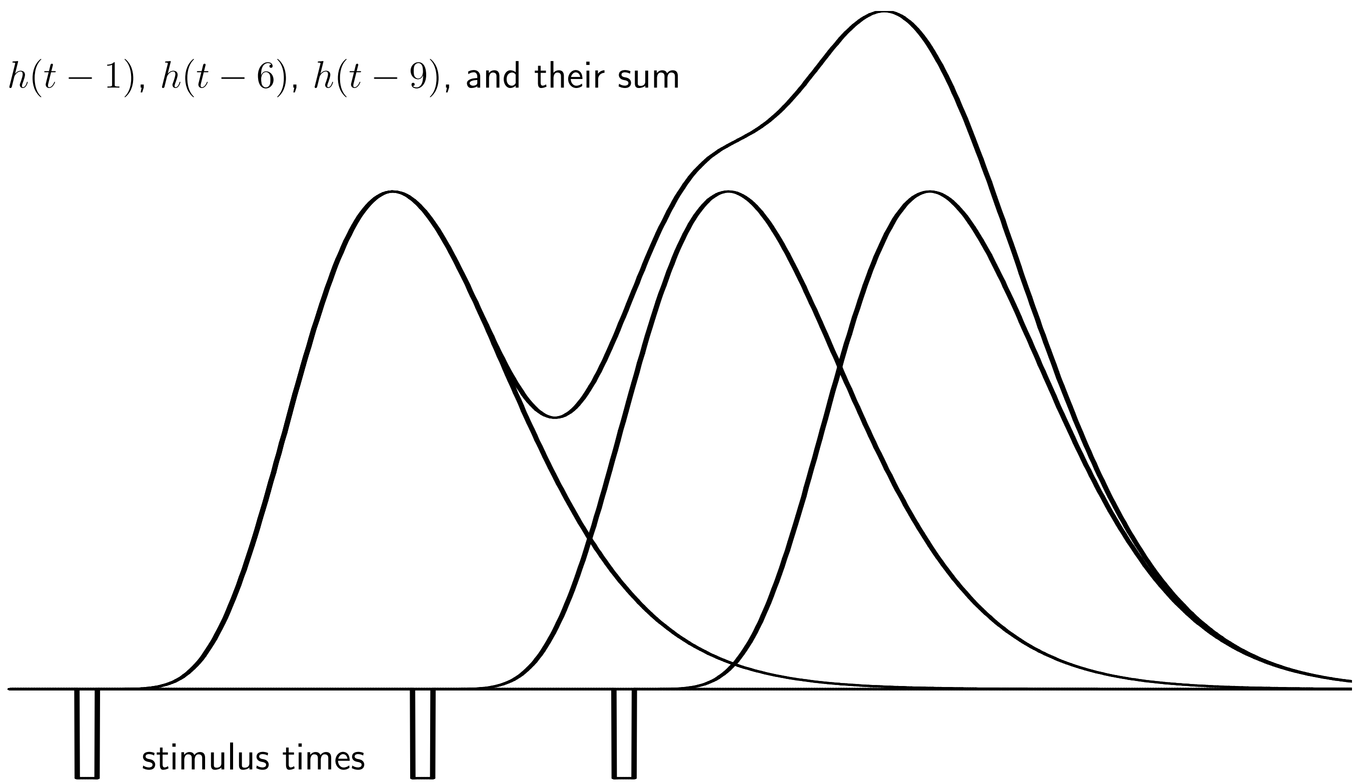
  - ◇ Response model starts at and returns to 0, so final measurement model must additionally include a baseline for each voxel
  - ◇ Stimuli are encoded by a time series that is either 0 or 1 at each point:

$$f_n = \frac{\quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1 \quad 1 \quad \ldots}{t: \; 0 \quad \Delta t \quad 2\Delta t \quad 3\Delta t \quad 4\Delta t \quad 5\Delta t \quad 6\Delta t \quad 7\Delta t \quad 8\Delta t \quad 9\Delta t \quad 10\Delta t \quad 11\Delta t \quad 12\Delta t \quad \ldots}$$

◇ Idealized response to this sequence of stimuli is convolution of individual response function $h(t)$ with stimulus function $f(t)$:

$$r_n = f_n \cdot h_0 + f_{n-1} \cdot h_1 + f_{n-2} \cdot h_2 + \cdots + f_{n-p} \cdot h_p = \sum_{m=0}^{p} f_{n-m} \cdot h_p$$

↪ The response at time $n\Delta t$ is the sum of responses from previous stimuli



↪ We assume that BOLD-derived MRI signals are strictly additive

▷ Only approximately true, but this assumption is widely used

▷ AFNI program `waver` can compute r_n timeseries, given f_n

- $\hookrightarrow$ The maximum convolution lag p determines how far into the future an individual stimulus reaches: $p\Delta t \approx (10 + D)$s for neural activity lasting D seconds

◇ A simple experiment: only two types of conditions (e.g., rest and stimulus)

- $\hookrightarrow$ AFNI interactive `FIM` and `FIM+` handle this case (in graph window)
- $\hookrightarrow$ Measurement model for each voxel separately:

$$
\begin{aligned}
Z_n &= \beta_0 + \beta_1 \cdot n + \alpha \cdot r_n + \varepsilon_n \\
\text{where} \quad Z_n &= \text{measured value at time } t = n\Delta t \\
\beta_0 &= \text{baseline} \\
\beta_1 &= \text{baseline drift rate (or trend)} \\
\alpha &= \text{amplitude of response} \\
\varepsilon_n &= \text{measurement noise (zero mean) in } n^{\text{th}} \text{ sample} \\
n &= 0, 1, 2, \ldots, N-1 \quad \text{(there are } N \text{ measurements in time)}
\end{aligned}
$$

- $\hookrightarrow$ In this model, there are 4 parameters we don't know (in each voxel): β_0, β_1, α, and σ^2 = noise variance
  - ▷ We know $h(t)$ = individual response function and $f(t)$ = stimulus time series, so we know $r(t)$ as well
  - ▷ β_0 and β_1 are "nuisance parameters" in the measurement model: we don't usually care what they are, but must include them to be realistic

- $\hookrightarrow$ This model is called linear since the unknown parameters β_0, β_1, α appear only by multiplying quantities we know (1, n, r_n, respectively)
  - ▷ Linear models are nice, since the algorithms for calculating results from them are relatively straightforward
- $\hookrightarrow$ We can calculate estimates of β_0, β_1, α using the method of linear least squares: minimize the sum E

$$E = \sum_{n=0}^{N-1} \left[Z_n - (\beta_0 + \beta_1 \cdot n + \alpha \cdot r_n) \right]^2$$

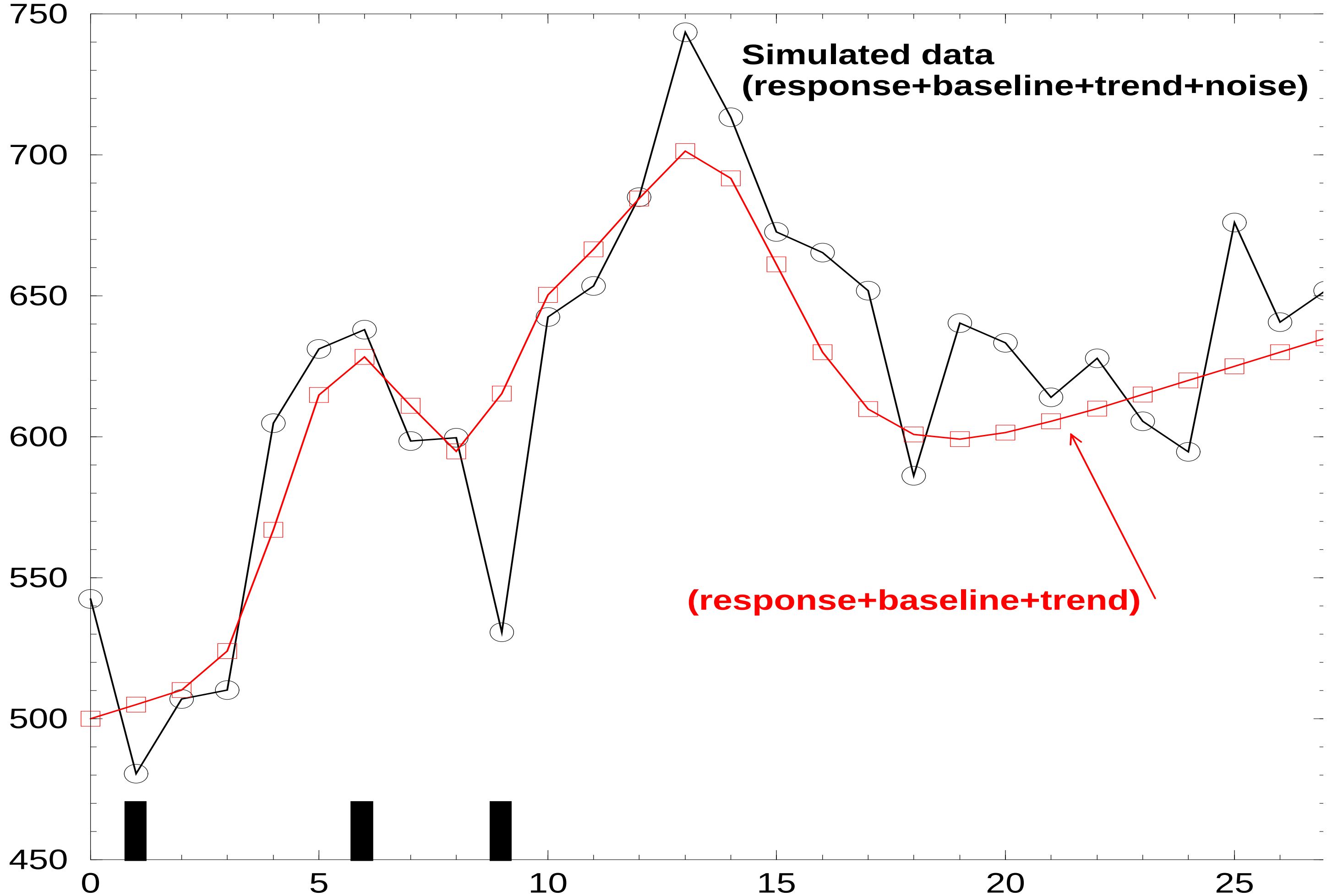
  over all possible values of β_0, β_1, α
  - ▷ We do this because we expect $Z_n \approx \beta_0 + \beta_1 \cdot n + \alpha \cdot r_n$
  - ▷ Estimates of the true (unknowable) parameters are denoted with a "hat", as in $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\alpha}$
  - ▷ Statistical estimation theory lets us also estimate $\hat{\sigma}^2$ from the minimized value of E
  - ▷ Then also can estimate variance of the parameter estimates
- $\hookrightarrow$ Activation detection: is $\alpha \neq 0$? (or, is there any response to stimulation?)
  - ▷ Essentially, we determine if $\hat{\alpha}$ is "far enough" away from 0 to make it very unlikely that the true $\alpha = 0$
  - ▷ "Far enough" is determined by the estimation accuracy

- ▷ Must make some assumptions about the statistical distribution of the noise
  - Gaussian distribution is almost always assumed
  - Stationary in time is almost always assumed
  - Can assume noise is "white" (uncorrelated in time), <u>or</u>
  - Can assume noise is temporally correlated, and try to estimate that too
- ▷ In practice, activation decision is computed by first computing an intermediate statistic, which is then thresholded to make the final decision
  - F-, t-, and correlation coefficient statistics are all used
  - In this simple case (only 1 activation parameter α), these are equivalent
  - Significance and power of test depend on distribution of noise
- ▷ AFNI interactive `FIM` and `FIM+` compute correlation coefficient statistic
  - $\hat{\rho}^2$ is the fraction of the variance in the detrended signal

$$Z_n - \hat{\beta}_0 - \hat{\beta}_1 \cdot n$$

  that is explained by the component $\hat{\alpha} \cdot r_n$
  - If true $\alpha = 0$, and if noise is white, $\hat{\rho}^2$ follows Beta distribution
  - If noise is correlated in time, $\hat{\rho}^2$ distribution is very complex
  - Can approximate with a Beta distribution, but must estimate parameters
  - After have a statistic with a known distribution, can threshold it at a given p-value

$\hookrightarrow$ Graphical view: curve fitting data time series to a sum of model curves



▷ In this case, model curves are $x_0(t) = 1$, $x_1(t) = t$, and $x_2(t) = r(t)$

▷ What's left after subtracting fit are called the residuals:

$$e_n = Z_n - \hat{\beta}_0 - \hat{\beta}_1 \cdot n - \hat{\alpha} \cdot r_n$$

These are used to calculate $\hat{\sigma}^2$, and the fraction of the variance that is not explained by $\hat{\alpha} \cdot r_n$

$\hookrightarrow$ Mathematical view: fitting data vector with a sum of model vectors:

$$\begin{bmatrix} Z_0 \\ Z_1 \\ Z_2 \\ \vdots \\ Z_{N-1} \end{bmatrix} \approx \beta_0 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} + \beta_1 \begin{bmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ N-1 \end{bmatrix} + \alpha \begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_{N-1} \end{bmatrix}$$

<u>or</u>

$$\begin{bmatrix} Z_0 \\ Z_1 \\ Z_2 \\ \vdots \\ Z_{N-1} \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & r_0 \\ 1 & 1 & r_1 \\ 1 & 2 & r_2 \\ \vdots & \vdots & \vdots \\ 1 & N-1 & r_{N-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \alpha \end{bmatrix}$$

<u>or</u>

$$\mathbf{Z} \approx \mathbf{X}\boldsymbol{\beta}$$

- ▷ All notations above are equivalent, but the last 2 (matrix-vector notation) are more convenient
- ▷ They let us tap into the entire mathematics and statistics of linear equation modeling developed in the last 50+ years
- ▷ They also provide a framework for expanding our model to allow for more conditions and for more complicated response functions