BMES Workshop on Parameter Estimation Methods in Physiological Modeling
Hollywood, CA
2007 September 26

Statistical Considerations and Techniques for Understanding Physiological Models, Data, and Treatment

Ben G. Fitzpatrick
Department of Mathematics
Loyola Marymount University
and
Tempest Technologies

http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Grandfatherly Advice in One Slide

- If your model is an attempt to understand a “slow” phenomenon, relative to the data rate and “unmodeled” variations in data, then nonlinear least squares comparison is perfectly reasonable.
- If you are modeling fast phenomena with fluctuations that are inherent in the process, then a stochastic dynamic model may be required.
- Statistical theory for nonlinear least squares, which is closely related to standard linear model theory, can be applied and works reasonably well.
- Do not forget to do some basic pre and post data analysis
  - Covariance and empirical model fitting can help detect dynamics
  - Analysis of the residuals from the fit provides an important post-mortem tool
  - Cross validation can also be helpful

Modeling the planets

- God’s creation must follow the perfect path: circles
- Circles don’t fit the data: epicycles
- Kepler: ellipses are simpler than lots of epicycles
- Newton: law of gravitation “explains” ellipses
- Einstein: relativity “explains” gravity
- Quantum mechanics: does God play dice in perfect circles? now what?

Modeling atmospheric turbulence

- Navier-Stokes equations of fluid flow
  - Conservation of momentum, energy, mass
  - Complicated partial differential equations
  - Exhibit complex output behavior

$$\frac{\partial u}{\partial t} + (u \bullet \nabla) u = \mu \Delta u - \nabla p$$

$$\nabla \bullet u = 0$$

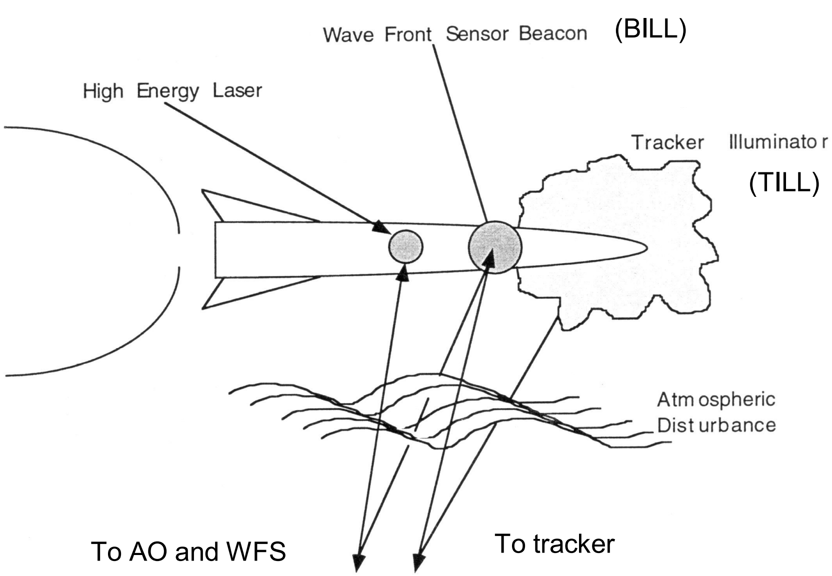
- Kolmogorov's 2/3 law
  - Scaling analysis of flow equations
  - Mean plus fluctuations in the flow
  - Covariance of the fluctuations

$$u = u_0 + \eta$$

$$E\left|\eta(x+h) - \eta(x)\right|^2 = C\left|h\right|^{2/3}$$

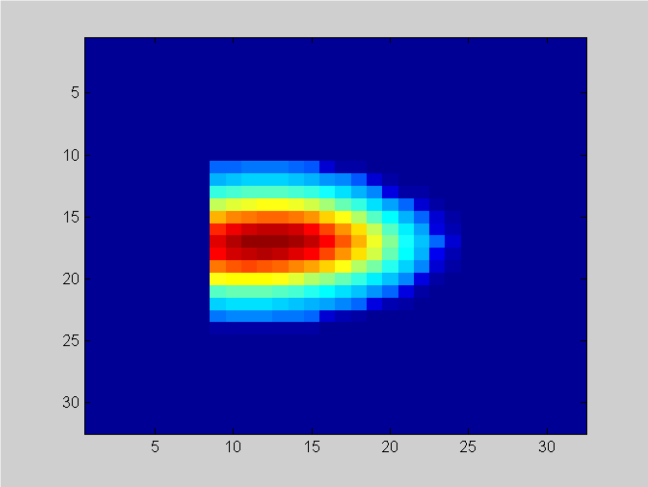
- Is the model the differential equation or the covariance?

Problem: Tracking a missile

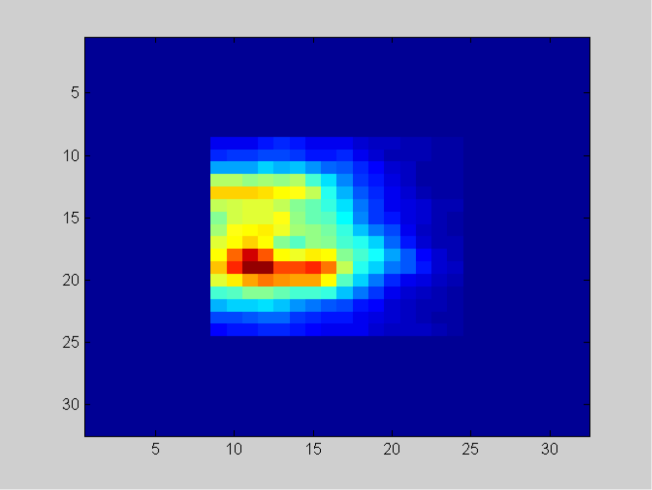


- Aircraft sends out laser beacon signal
- Atmosphere distorts the light
- Signal reflects from missile, propagates back to aircraft
- Return path also distorts laser signal
- First signal used to steer
- Second signal used to correct high order aberrations
- Third beam is the weapon

Tracking via images



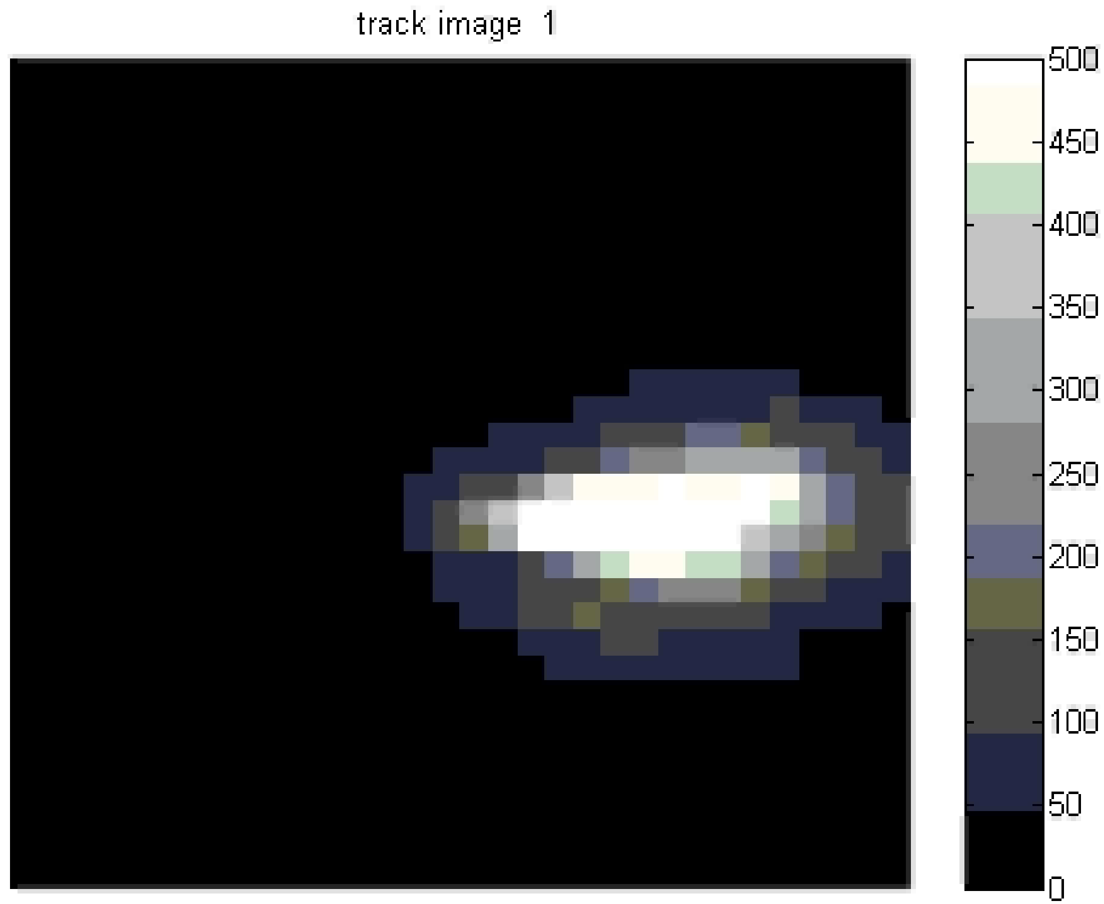
reference image



actual data

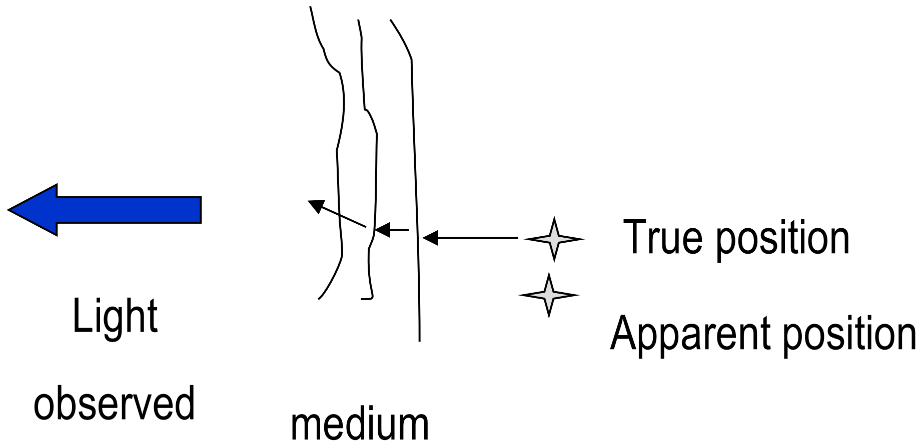
- Images must be compared quantitatively
- Goal: keep the tip in the center of the image plane
- Difference in desired and observed tip position determines line-of-sight adjustment

Tracking via images



The funny part...

- The goal is ***not*** to remove the noise: it's to estimate it.



Firing bullets:

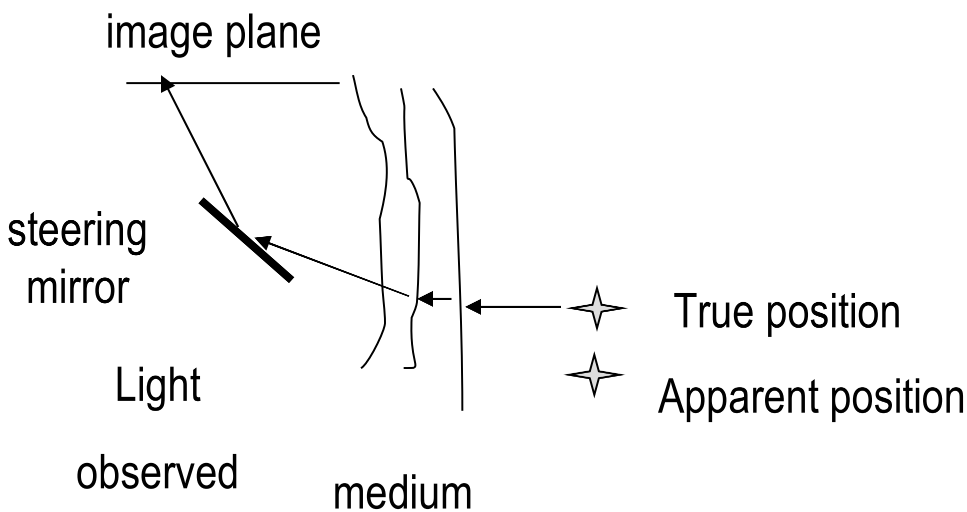
shoot at true position

Firing light:

shoot at apparent position

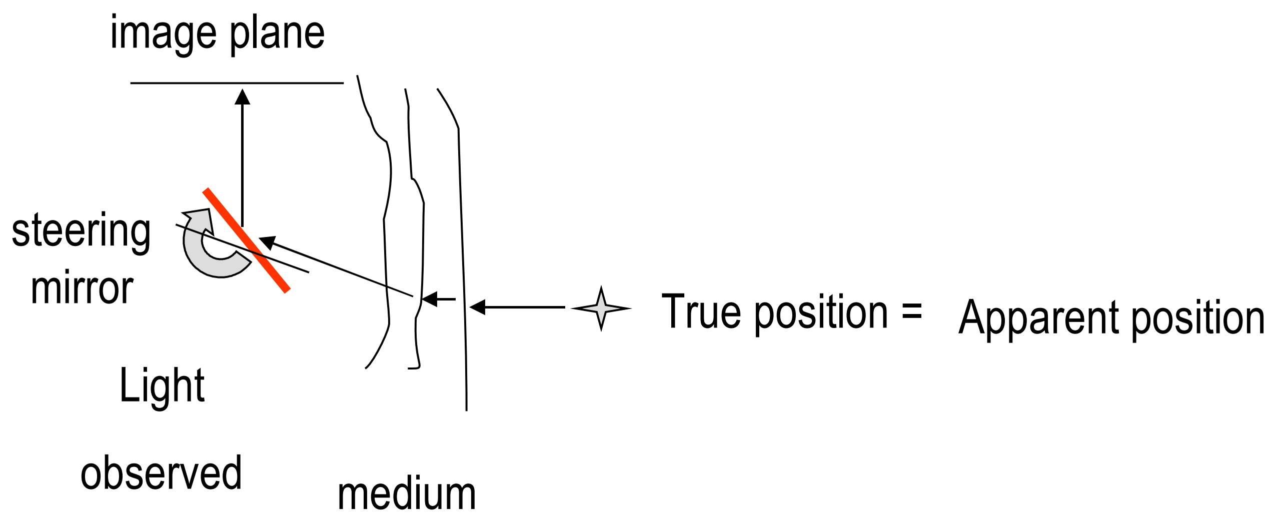
More on the funny part...

- The goal is ***not*** to remove the noise: it's to estimate it.

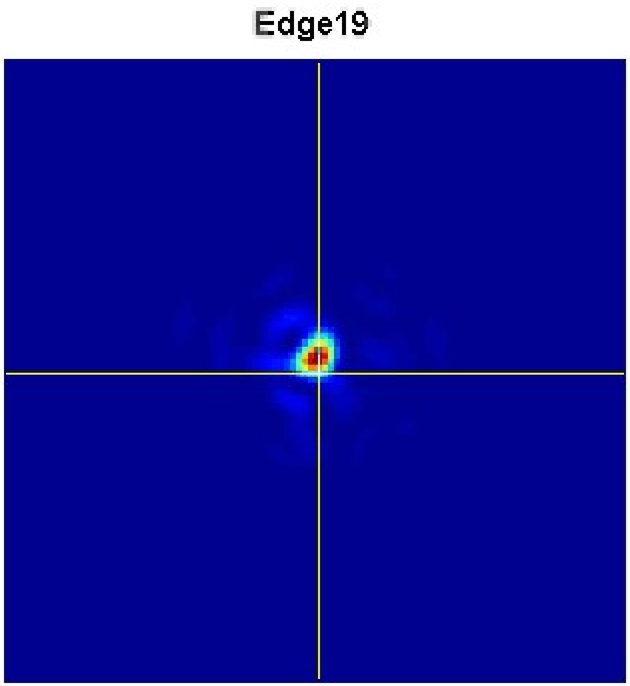


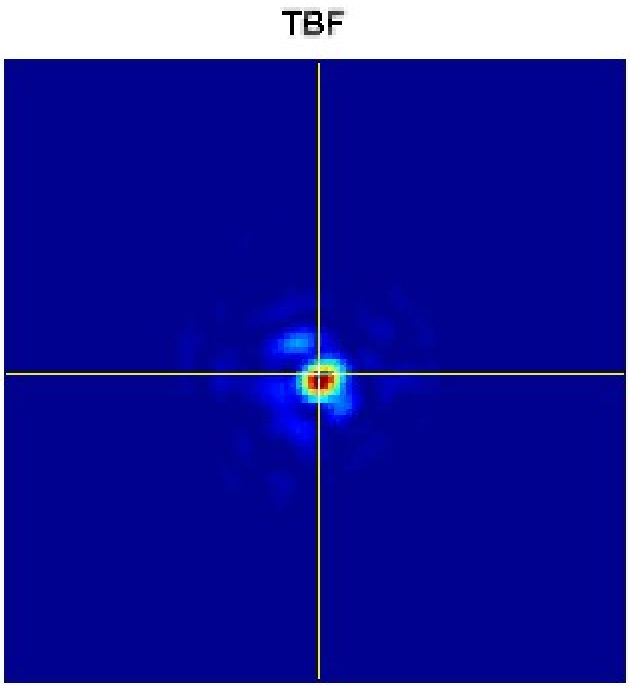
Fixing the funny part...

- The goal is ***not*** to remove the noise: it's to estimate it.



Controllers Based on Two Models





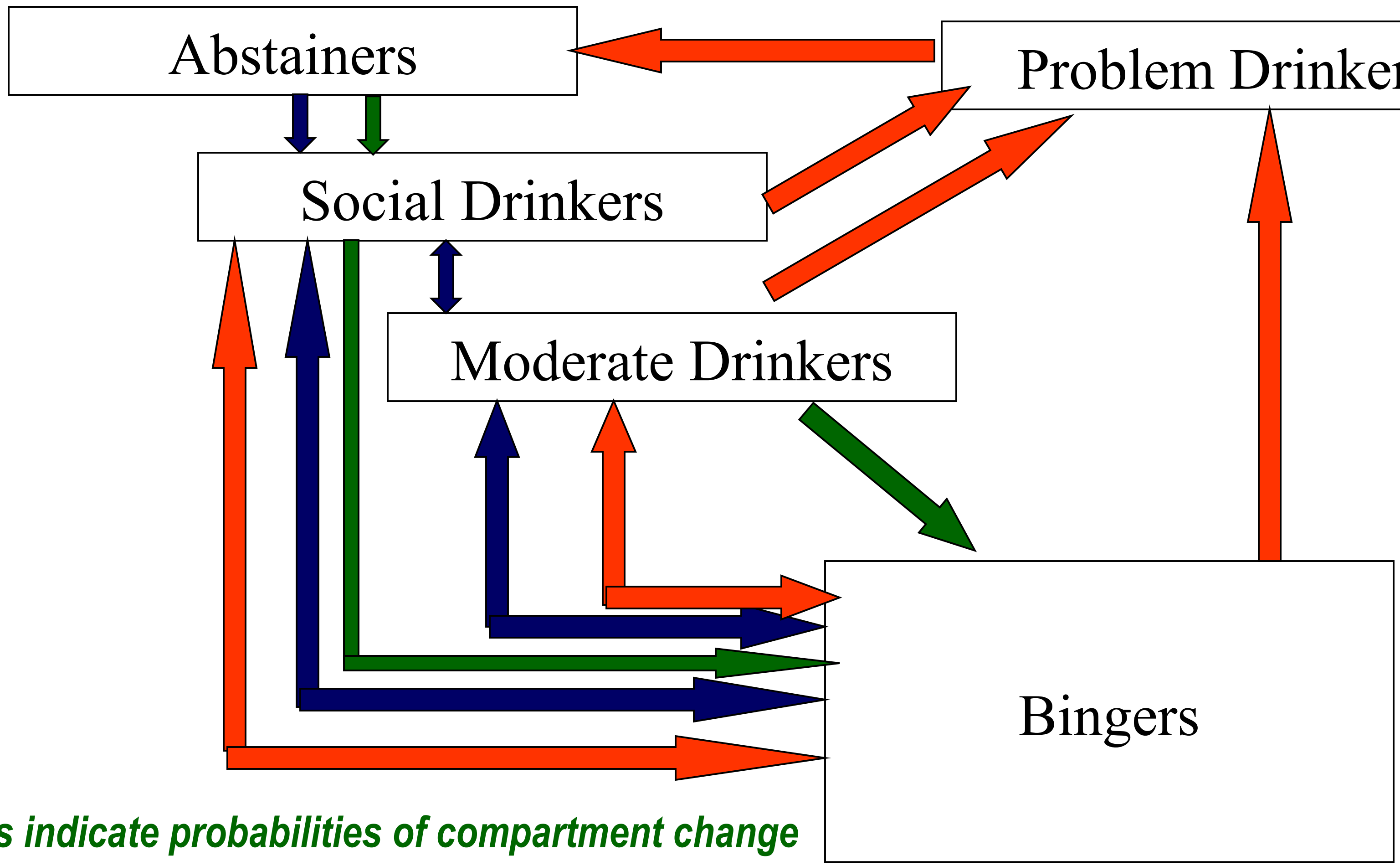
Modeling College Drinking

- Individual interventions have (at best) limited help
- Alcohol is a factor in
  - 40% of academic problems
  - 28% of dropouts
- 42% of students binge
- Binging correlates significantly with rape
- 1400 deaths per year related to alcohol
- Can administrators and politicians develop policy mechanisms to help?
  - Curfew, zoning, marketing restrictions, enforcement, …?

Modeling of populations

- Compartments
  - Lump together “like” individuals
  - Interactions in terms of compartments, not individuals
  - External influences in terms of compartments, not individuals
  - Gives a sense of “average” behavior
- Individuals (Agents)
  - Everyone is (potentially) unique
  - Interactions and external influences work on individuals
  - Aggregate information can be determined by averaging
  - Price of flexibility, variability is complexity

Compartments connected



Arrows indicate probabilities of compartment change

http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

ODEs

$$\frac{dN_1}{dt}=-d_1N_1-n_{12}\frac{\sum_{i=2}^{5}N_i}{\sum_{i=1}^{5}N_i}N_1-s_{12}N_1N_2+r_{41}N_4+r_{21}N_2$$

$$\frac{dN_2}{dt}=-d_2N_2+n_{12}\frac{\sum_{i=2}^{5}N_i}{\sum_{i=1}^{5}N_i}N_1-n_{25}\frac{N_5}{\sum_{i=1}^{5}N_i}N_2+s_{12}N_1N_2+(s_{52}-s_{25})N_2N_5$$
$$+(s_{32}-s_{23})N_2N_3+r_{52}N_5-r_{25}N_2-r_{24}N_2-r_{21}N_2$$

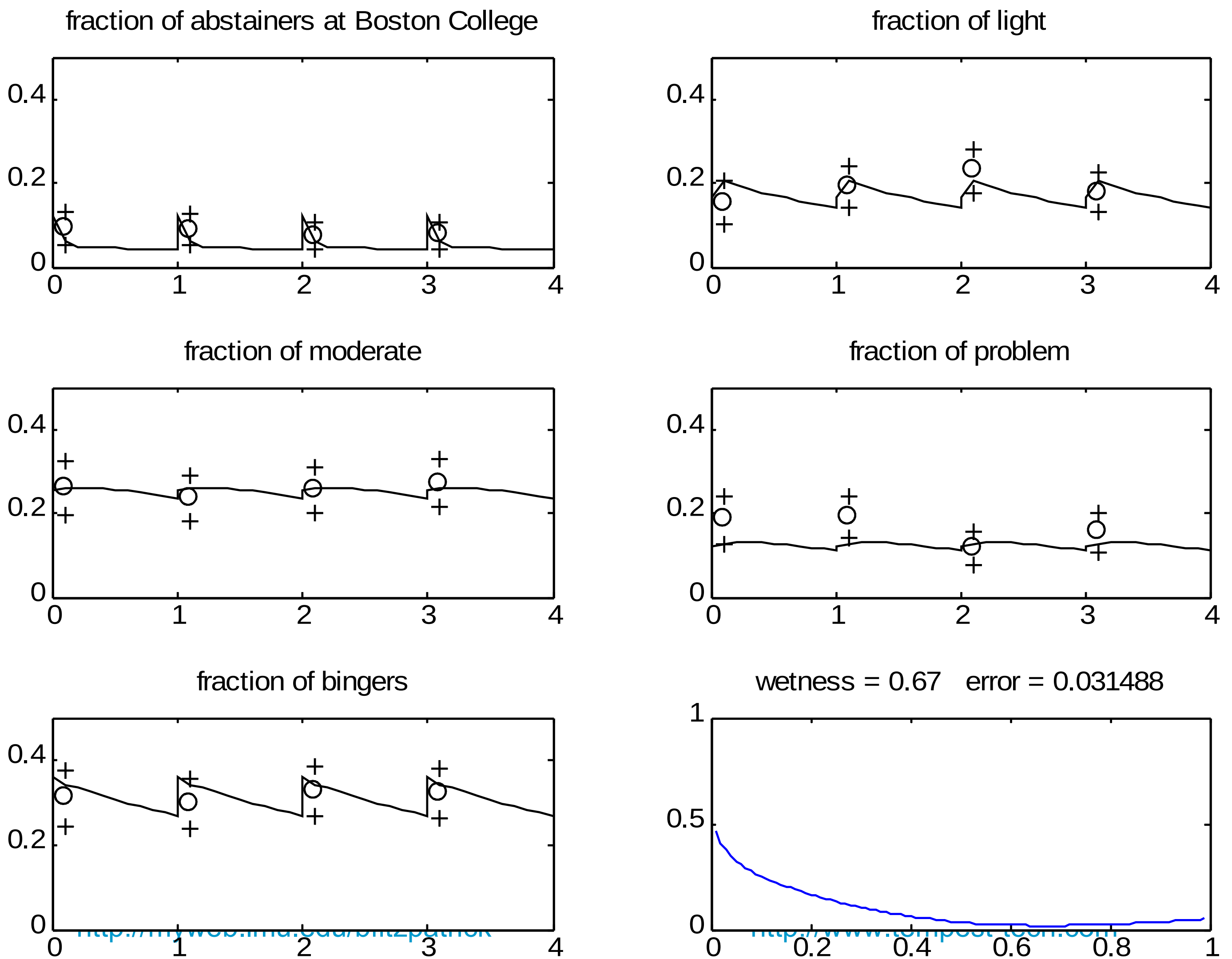
$$\frac{dN_3}{dt}=-d_3N_3-n_{35}\frac{N_5}{\sum_{i=1}^{5}N_i}N_3-(s_{32}-s_{23})N_2N_3+(s_{53}-s_{35})N_3N_5-(r_{34}+r_{35})N_3+r_{53}N_5$$

$$\frac{dN_4}{dt}=-d_4N_4+r_{34}N_3+r_{54}N_5-r_{41}N_4+r_{24}N_2$$

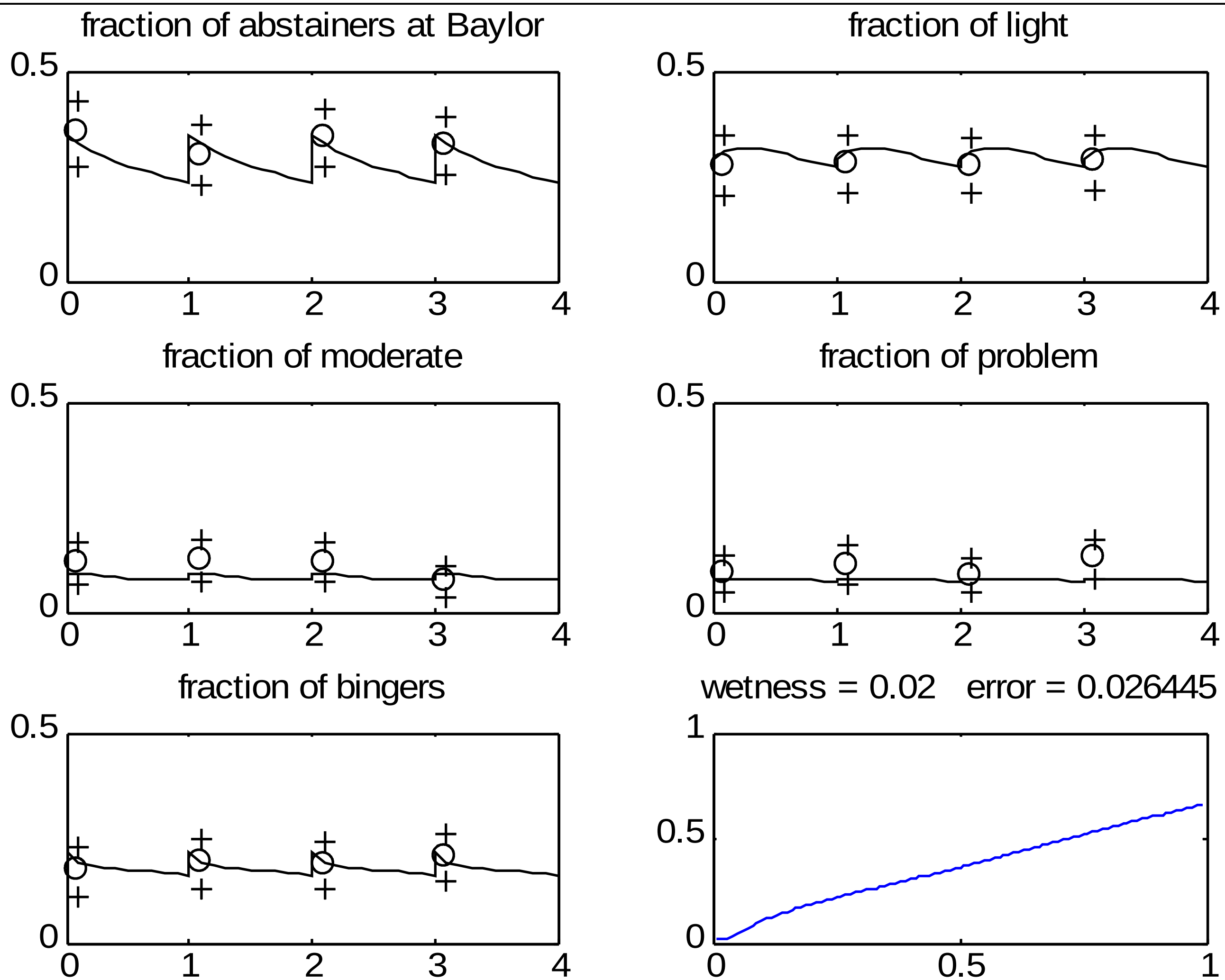
$$\frac{dN_5}{dt}=-d_5N_5+n_{25}\frac{N_5}{\sum_{i=1}^{5}N_i}N_2+n_{35}\frac{N_5}{\sum_{i=1}^{5}N_i}N_3-(s_{52}-s_{25})N_2N_5+(s_{53}-s_{35})N_3N_5-r_{52}N_5$$
$$+r_{25}N_2-r_{54}N_5-r_{53}N_5+r_{35}N_3$$

$$\Delta N_j(t_k)=N(t_k^+)-N(t_k^-)=c_j\sum_{i=1}^{5}[N_i(t_{k-1}^+)-N_i(t_k^-)],\quad j=1,2,3,4,5,\quad t_k=1,2,3,\cdots$$

Data and Model for Boston College

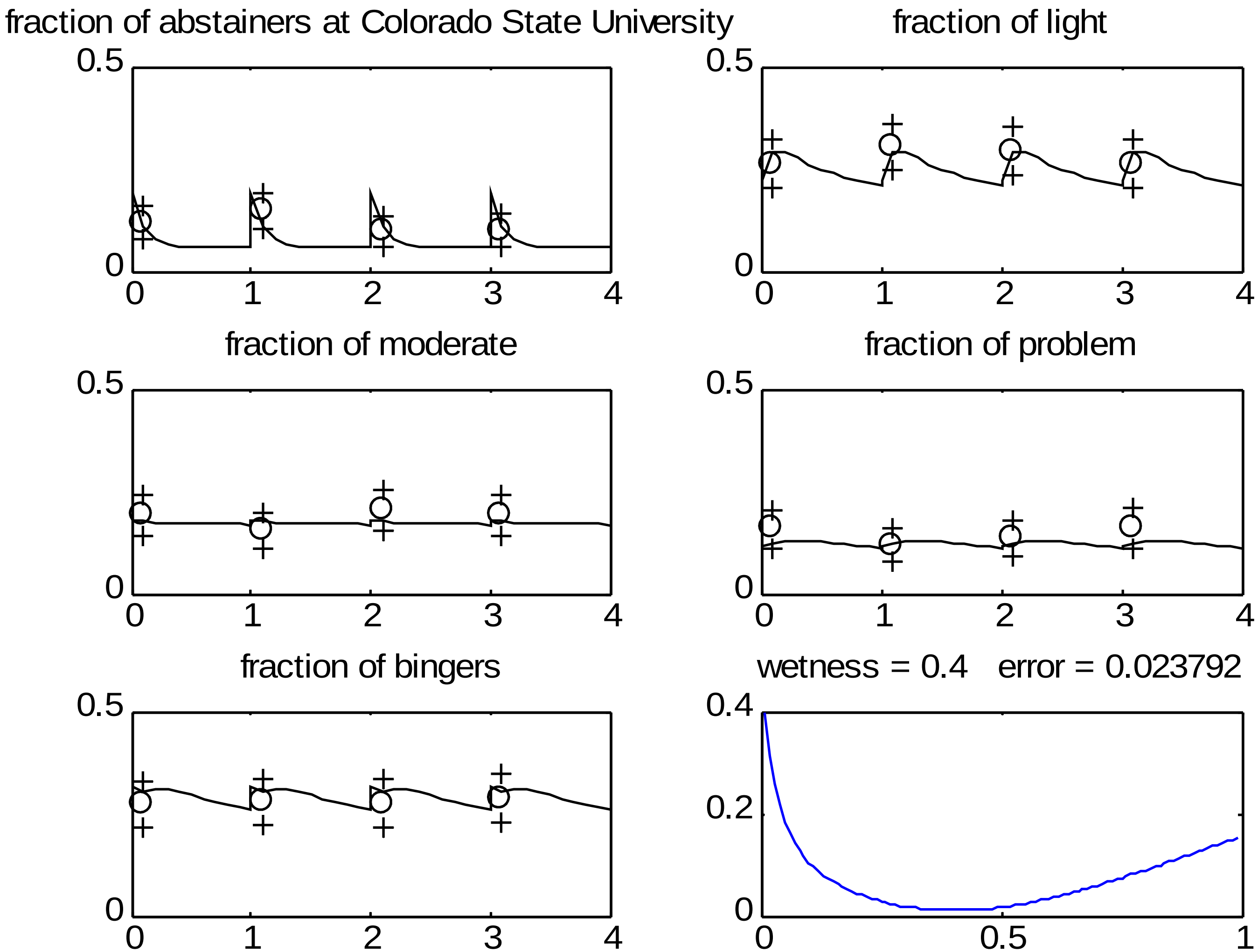


Data and Model for Baylor



http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

Data and Model for Colorado State



http://myweb.lmu.edu/bfitzpatrick

http://www.tempest-tech.com

Modeling: what it is

- Modeling: an attempt to understand something by creating a likeness
  - Architectural scale model in cardboard of a skyscraper
  - Map of Palo Alto
  - Mathematical equations describing gravity
- A model is a simplified representation of reality
- Models can be mental, verbal, graphical, physical, or mathematical
- Models can contain various levels of detail and provide various levels of accuracy
- Models may contain hypotheses about phenomena
- Models can only be demonstrated to be false
  - Validity is the absence of evidence to reject

Modeling basics: why and how

- The goal governs the approach
  - Qualitative understanding
  - Quantitative understanding
  - Quantitative prediction
  - Design/decision making

- The phenomena govern the approach
  - Repeatability (at what level of accuracy)
  - Individual differences
  - Environmental effects
  - Dynamics
  - External/exogenous influences
  - Scales (in time, space, size)

Occam's razor: keep it simple!

Model requirements

- Simple enough to be “study-able”
- Complex enough to capture phenomena of interest
- Parameters estimable and relate-able to problem domain
- Predictions to be quantitatively accurate
- Policy decision impacts to be quantitatively well predicted
- ***A model is at best as good as the assumptions on which it is based***

Possible Approaches

- Empirical statistical approach
  - Collect data
  - Fit a statistical model
    - Linear model, heirarchical linear model, multivariate/covariance, etc
  - Test hypotheses
  - Often motivated by below but tuned for individual situation
- Theoretical mathematical approach
  - Quantify behaviors
  - Develop equations
  - Solve/simulate
  - Test hypotheses
  - Can be simple or complex, depending on goals

Modeling and Uncertainty

- Individual patient uncertainty
  - Parametric? Bayesian approaches are ideal
- Patient activity uncertainty
  - Exogenous? Plant noise and stochastic differential equations
- Measurement sensor noise
  - The noise for which many standard statistical methods have been designed

Inductive and Deductive Science

- Collect data
- Get idea
- Submit proposal
- Manipulate idea to fit data
- Repeat until funding dries up

- Get idea
- Submit proposal
- Collect data
- Manipulate data to fit idea
- Repeat until funding dries up

Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Statistics: the science of data

- Statistics is the science of
  - Collecting,
  - Organizing,
  - Analyzing, and
  - Interpreting data

- Branches of statistics
  - Descriptive
  - Inferential

Questions statistics attempts to answer

- What patterns exist in my data?
- How good are my parameter values?
- Do these samples have the same structure?
- What data should I collect?

Descriptive statistics

- Basic parametrics
  - Means, variances
  - Covariances
- Basic nonparametrics
  - Histograms
  - Emprical CDFs
- Parametric nonparametrics
  - Autocovariance functions
  - PSDs
  - Nonparametric regression

Inferential Statistics

- Three basic problems
  - Parameter estimation
  - Interval estimation (confidence intervals)
  - Hypothesis testing

- Requirements
  - Parameter and noise to data model
  - Data probability distribution function

- Except in the simplest settings, a lot of computation is required

Caution: what statistics does not do

- No definitive decisions!
  - Assumptions are often unverifiable
  - Decisions may not be robust to “decision parameters”
- Two populations: X & Y
- Test that they have the same mean

$$H_0 : \mu_X = \mu_Y$$
$$H_A : \mu_X \neq \mu_Y$$

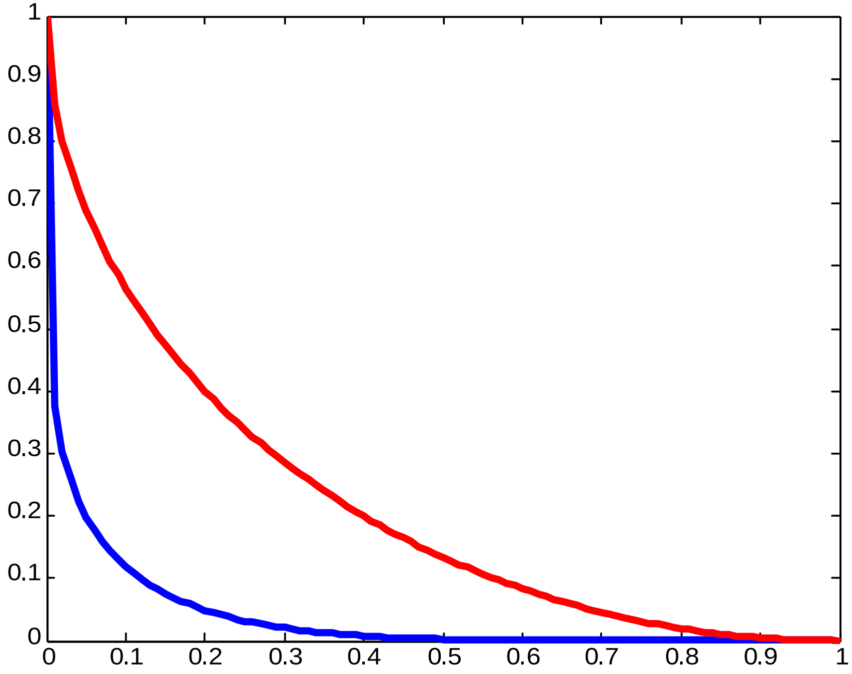
$$\frac{\bar{X} - \bar{Y}}{S_p} \sim T$$

- Stat = 2.3, is it significant? Yes at the 0.05 level, but no at the 0.01 level.
- Phrasing: We either reject H_0 or we fail to reject H_0

More on (Moron?) Hypothesis Testing

- Two types of errors
  - Reject H0 when it is true α
  - Fail to reject H0 when it is false β
  - Typically the probabilities of these erors are
    - Coupled
    - Move in competing directions when test criteria changes

- ROC (receiver operating characteristic) compares the two erros



http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

Techniques of Inferential Statistics

- Frequentist or classical statistics
  - Maximum likelihood
  - Least squares
  - Likelihood ratio tests
  - Asymptotics justify application
- Bayesian statistics
  - Maximum a posteriori
  - Posterior mean
  - Posterior hypothesis probability
  - Computationally intensive

Things We Want from Estimators

- Unbiasedness
  - Is the theoretical mean value of the estimator equal to the parameter it's estimating?
- Consistency
  - Does the estimator converge to the parameter's true value as the number of observations increases?
- Minimal variance
  - Does this estimator have less variability than other possible estimators?
- Asymptotic normality
  - Is the probability distribution of the estimator converging to a normal distribution as the number of observations increases?

Bayesian statistical methods

- Use probability distributions to model uncertainty
  - Unknown parameters to be estimated
  - Measurements and observations
- Compute conditional probabilities for inference

$$\theta = \text{unknown parameter}$$
$$\pi(\theta) = \text{prior probability density}$$
$$y = \text{measurement vector}$$
$$p(y|\theta) = \text{measurement density (conditional)}$$
$$\pi(\theta|y) = \text{posterior}$$

- Requires a lot of thought/investigation to obtain p.

Bayesian inference

- Bayes' theorem:

$$\pi(\theta|y) = \frac{p(y|\theta)\pi(\theta)}{\int_{\Theta} p(y|\theta')\pi(\theta')d\theta'}$$

- The posterior is used for estimation and other inference

$$\theta_{MAP} = \arg\max \pi(\theta|y)$$

$$\theta_{PM} = \int_{\Theta} \theta\pi(\theta|y)d\theta$$

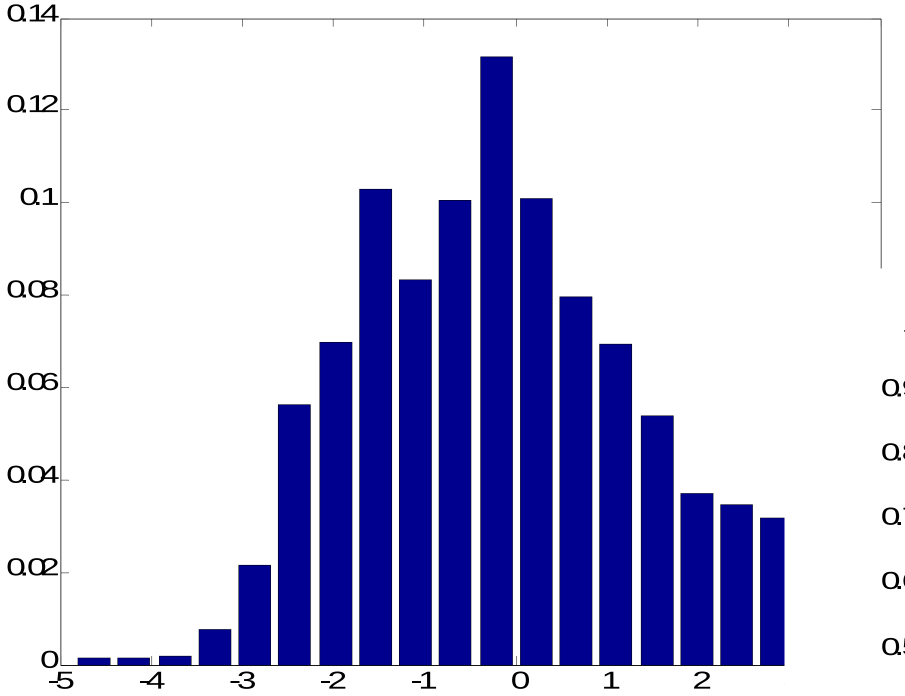
Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Looking for patterns and structure

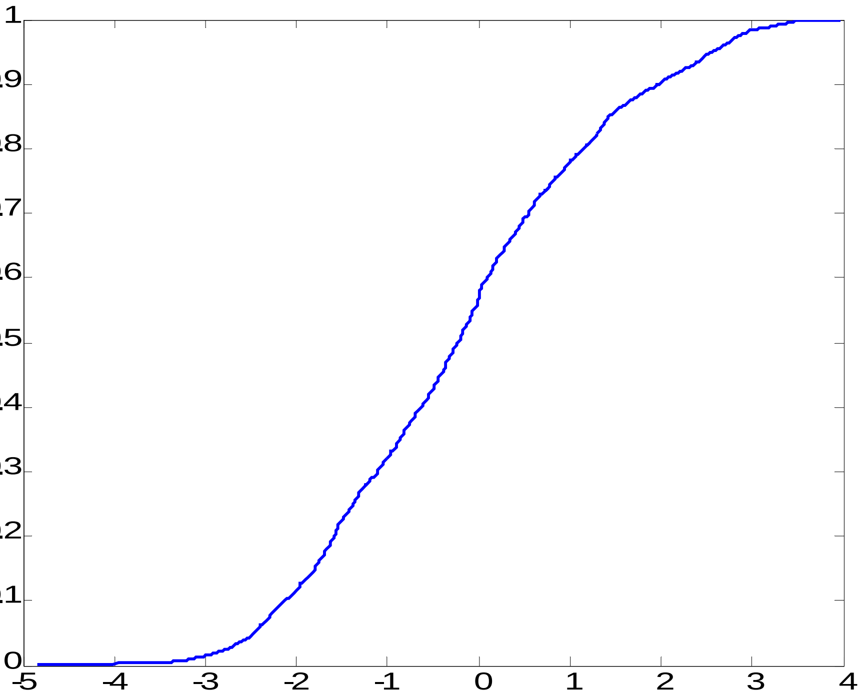
- Standard practice: means and variances
- Frequency analysis: basic variability
  - Histogram
  - CDF
- Correlation analysis: related variables
  - Dot-plots
  - Correlation coefficients and r^2
- Temporal analysis
  - Time domain
  - Frequency domain

Basics



histogram

cumulative

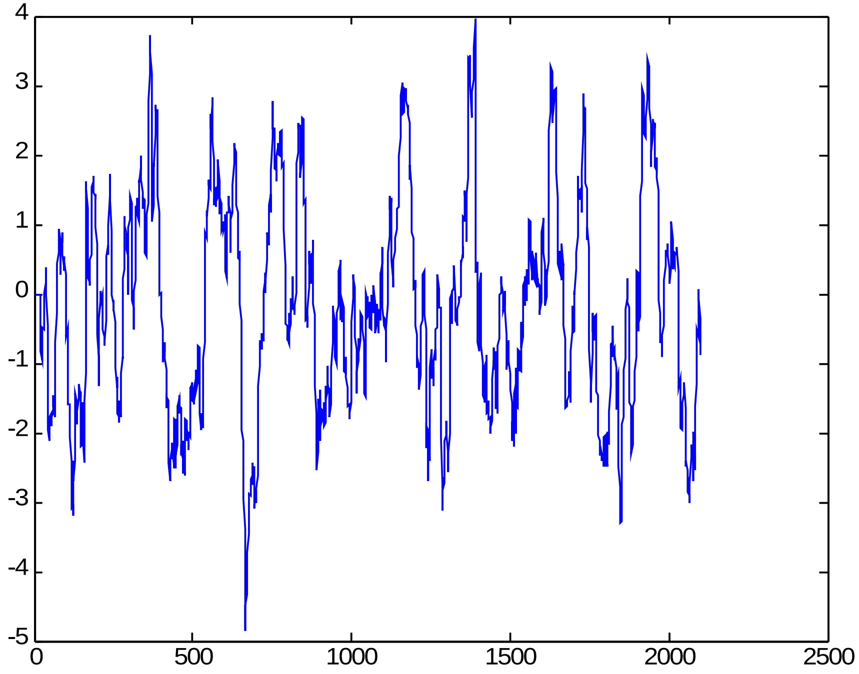


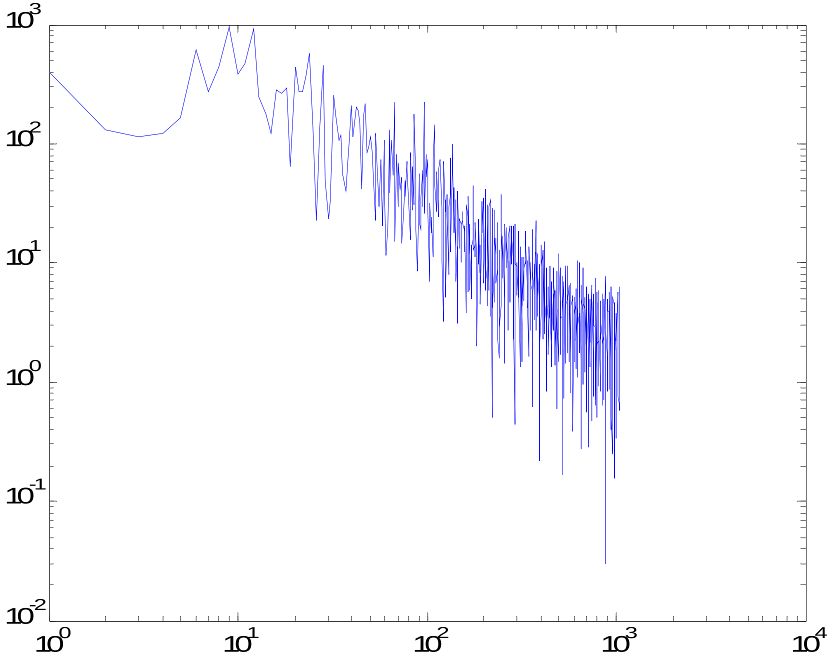
http://myweb.lmu.edu/bfitzpatrick

Time Series Analysis

- Traditional tools
  - Autocovariance
  - Variogram
  - Power spectrum (PSD)
  - Autoregression (AR)
  - AR moving average (ARMA)
  - FARIMA (fractional..integrated…)
- More “modern” tools
  - Nonlinear models
  - Wavelet analysis
  - Local covariance
  - Change point

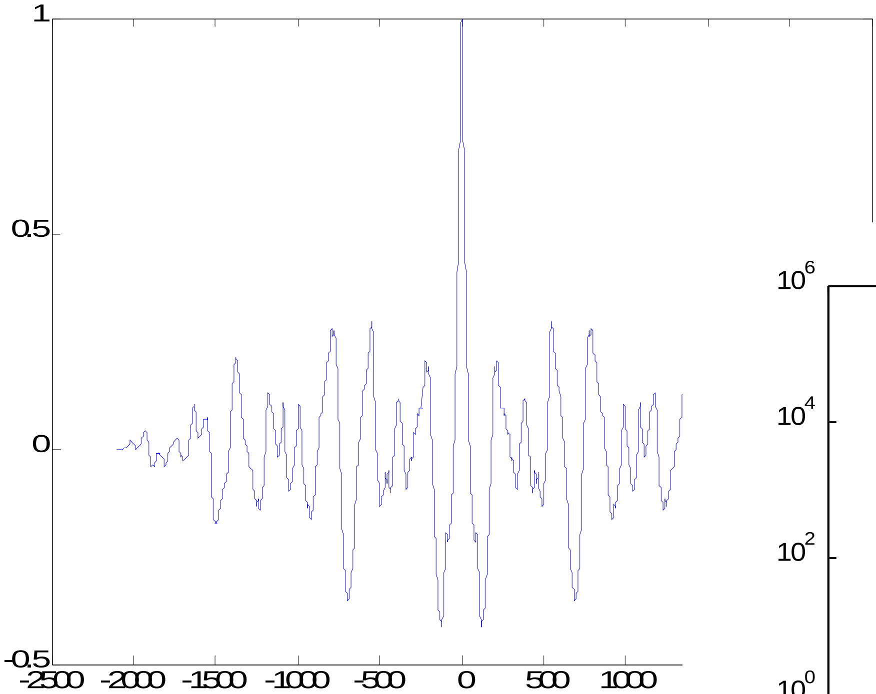
Time and Frequency



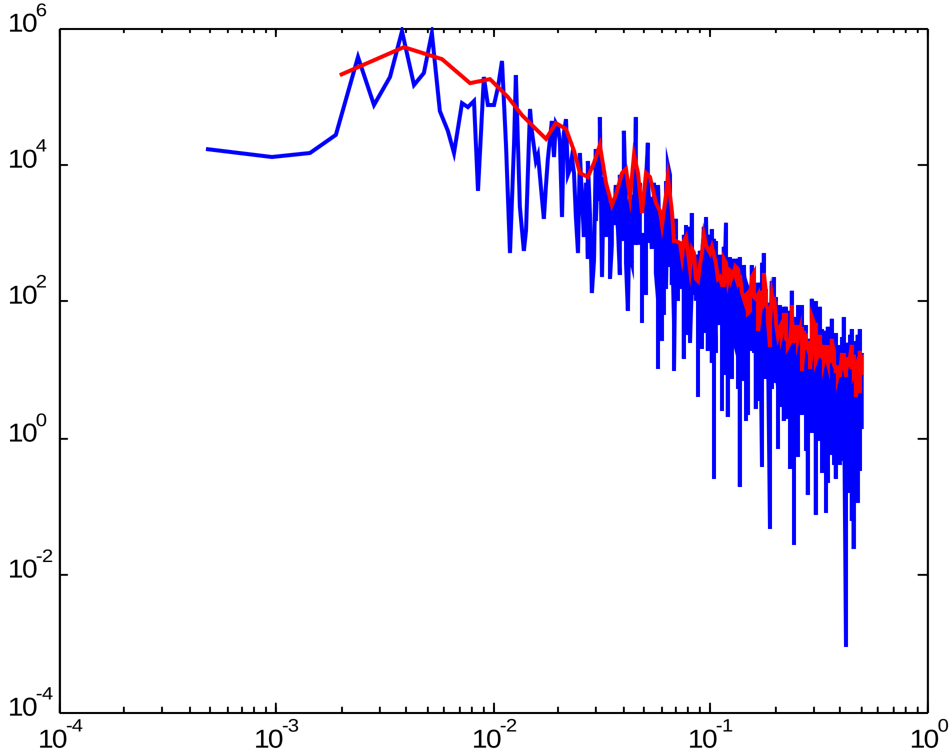


http://myweb.lmu.edu/bfitzpatrick

Second Moments



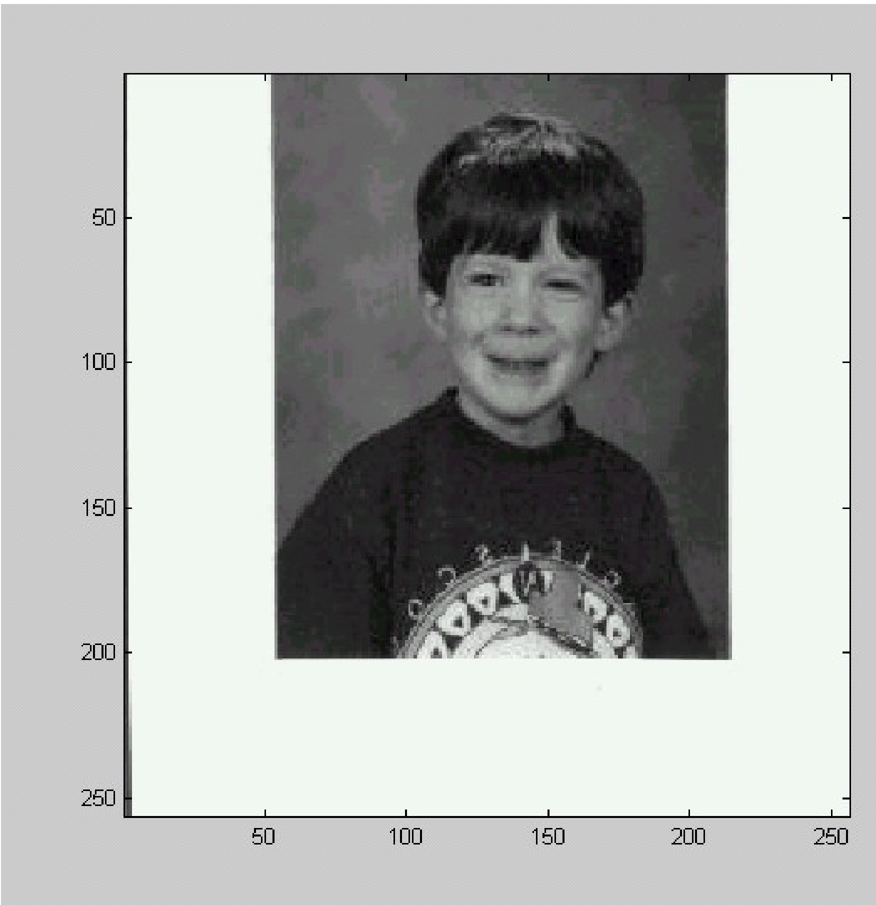
Standard PSD is NOT consistent



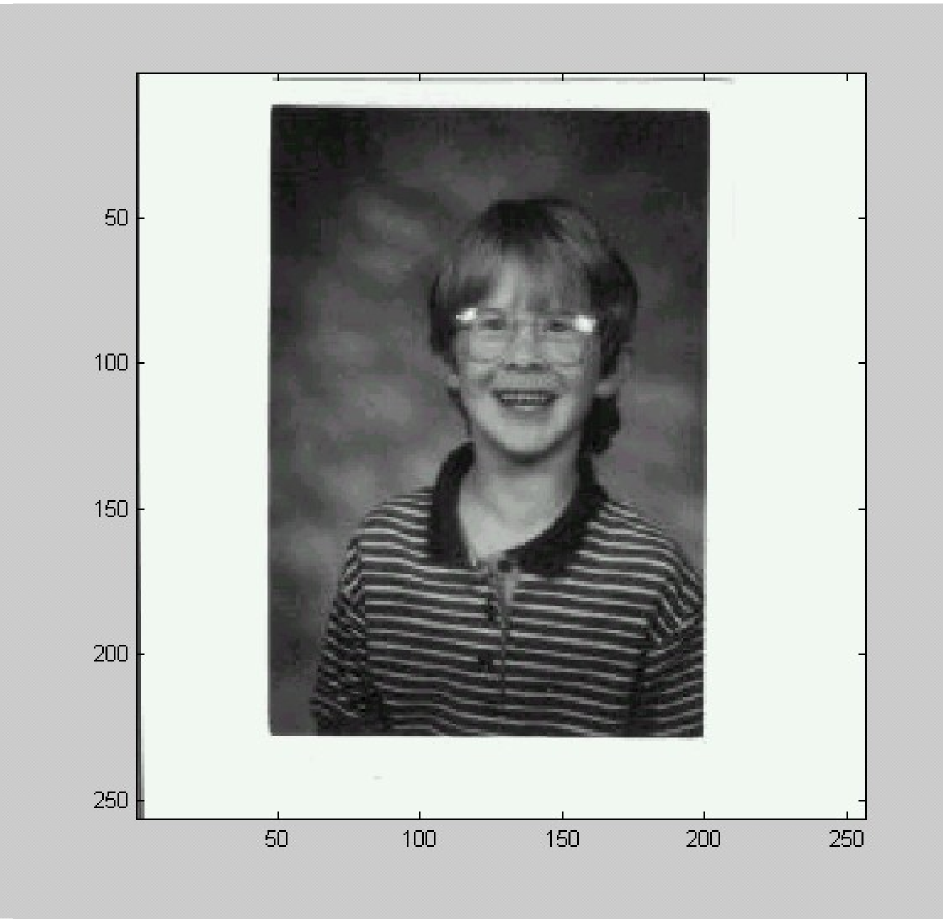
Autocovariance is consistent

http://myweb.lmu.edu/bfitzpat

Two photographs: don’t forget the phase

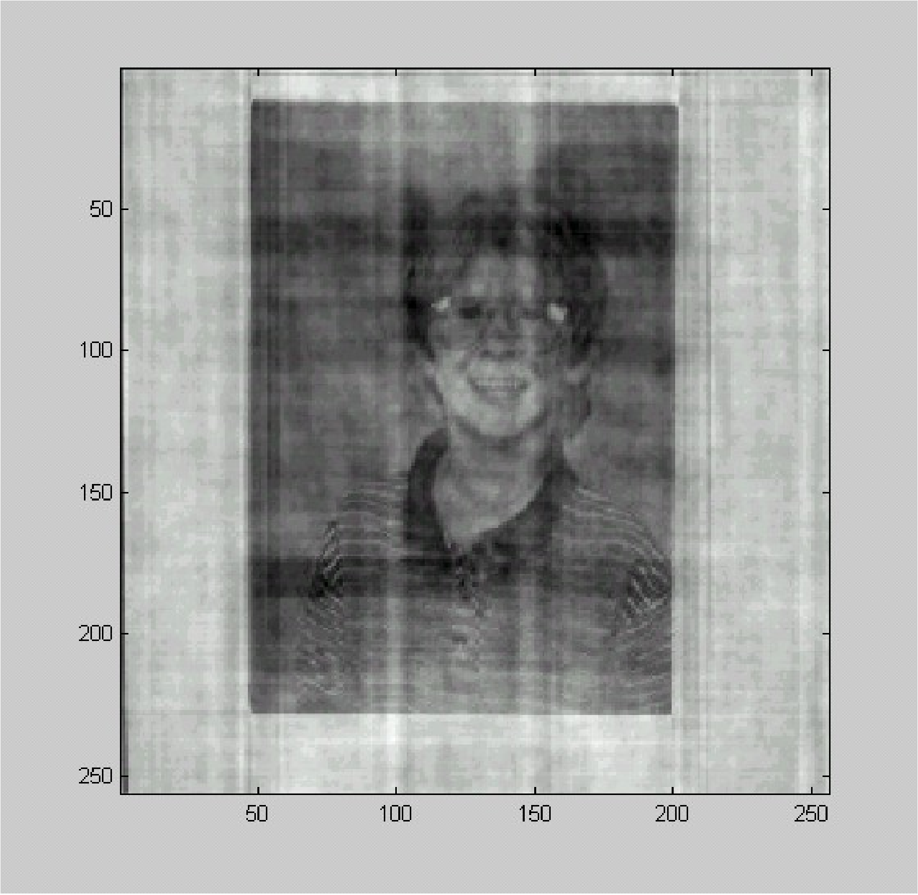


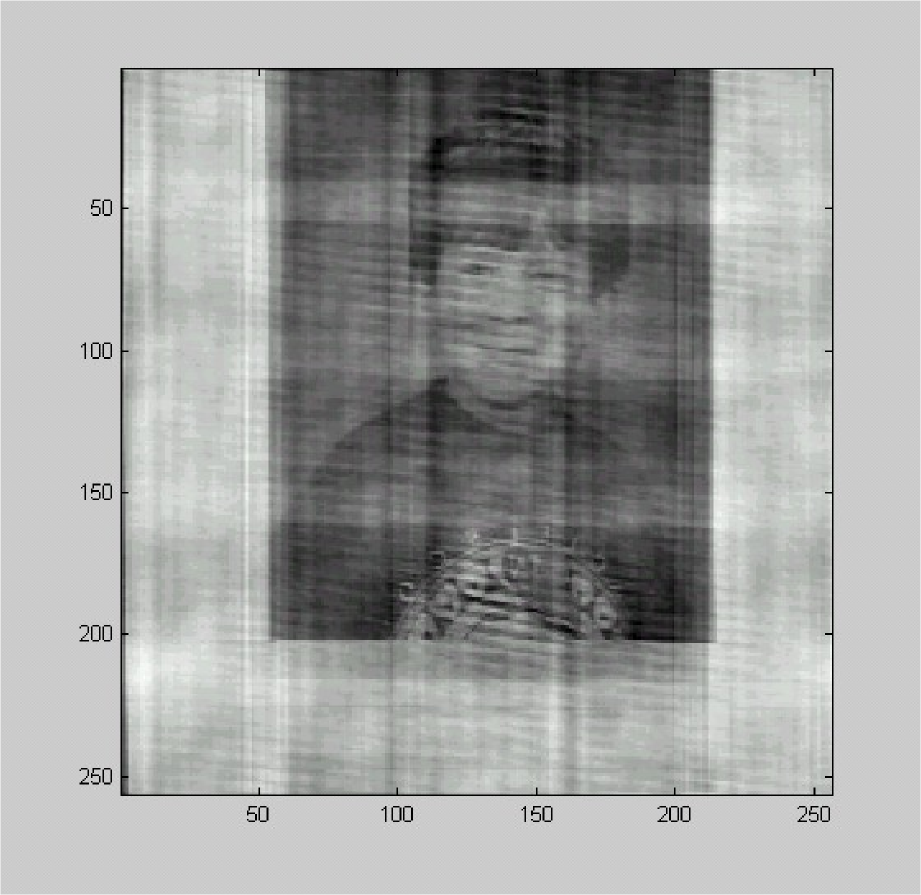
http://myweb.lmu.edu/bfitzpatrick



http://www.tempest-tech.com

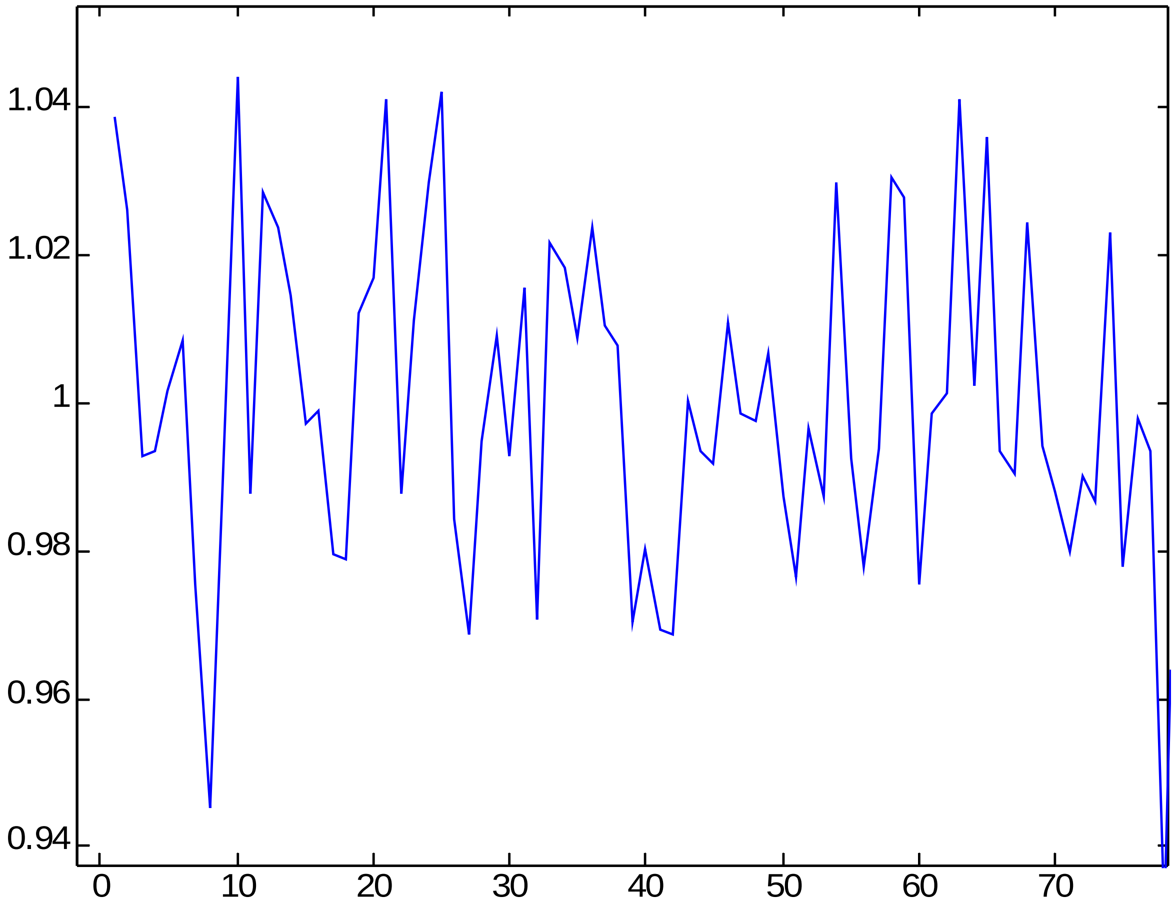
Fourier transform, swap phases, inverse F.T.



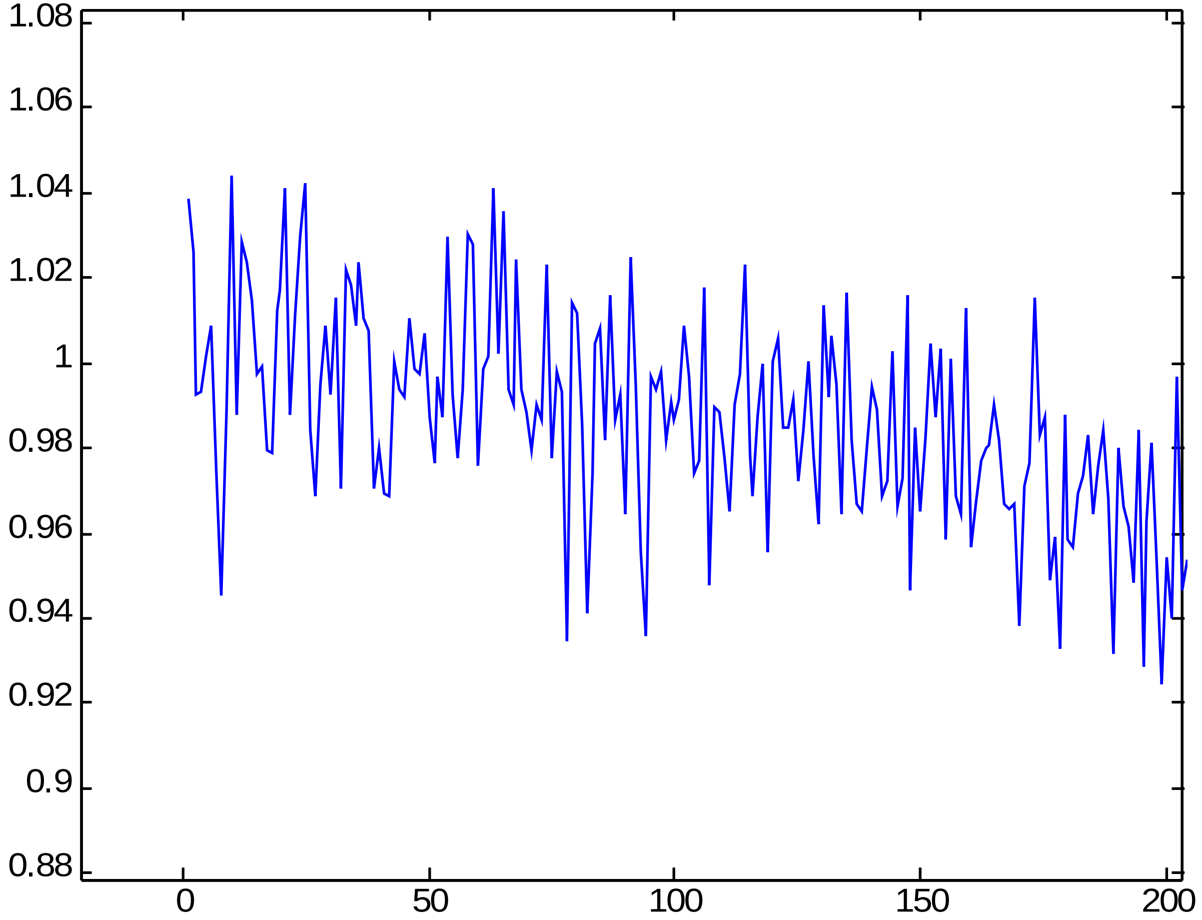


http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

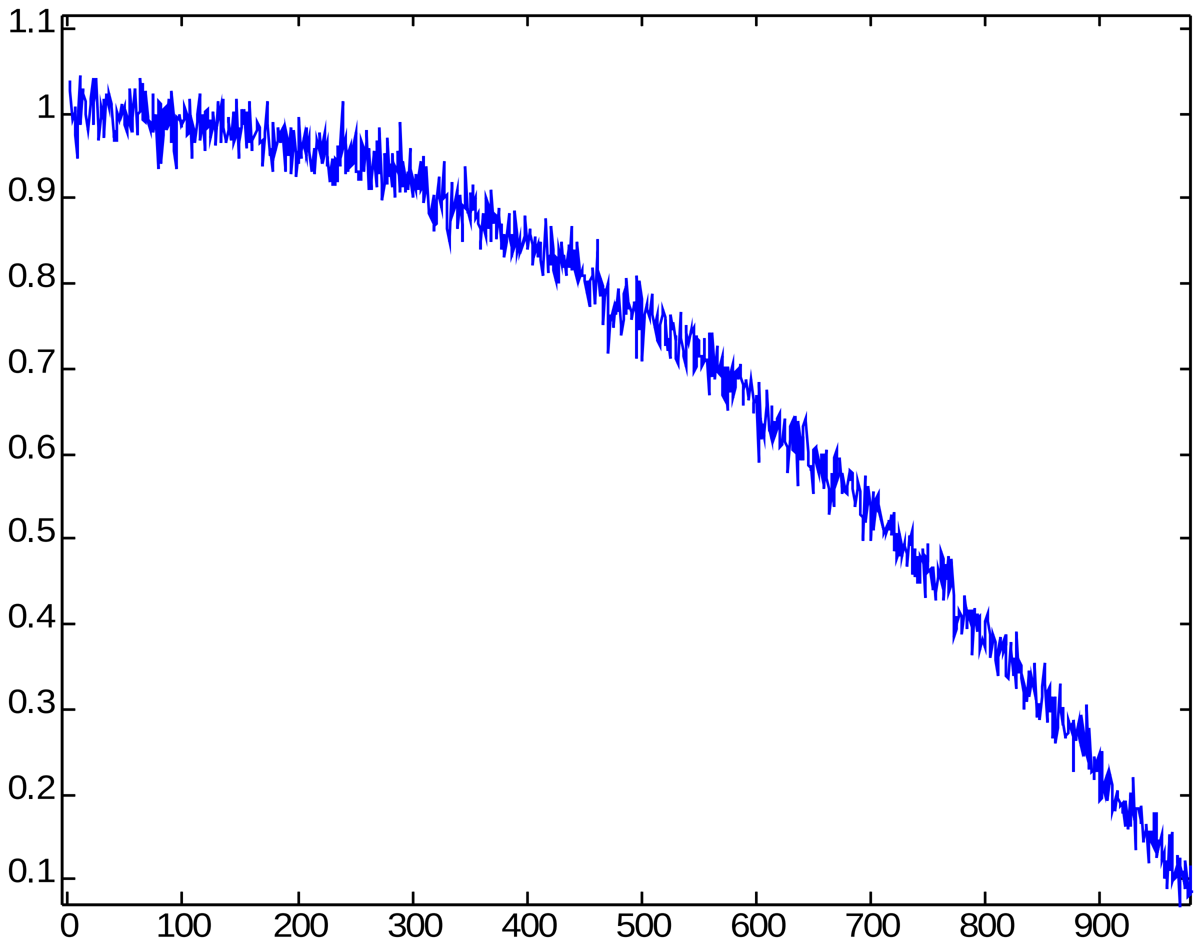
Stationarity



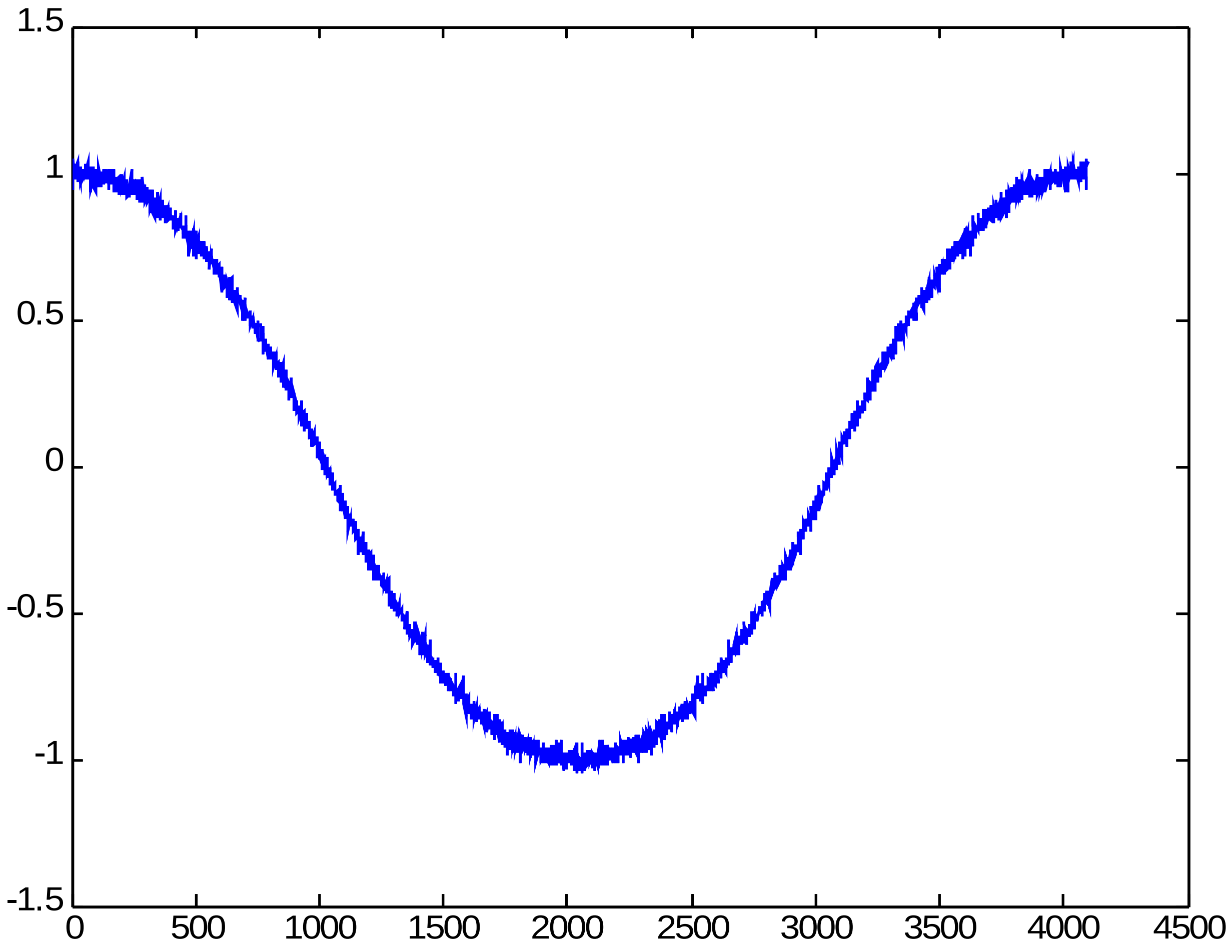
Stationarity



Stationarity

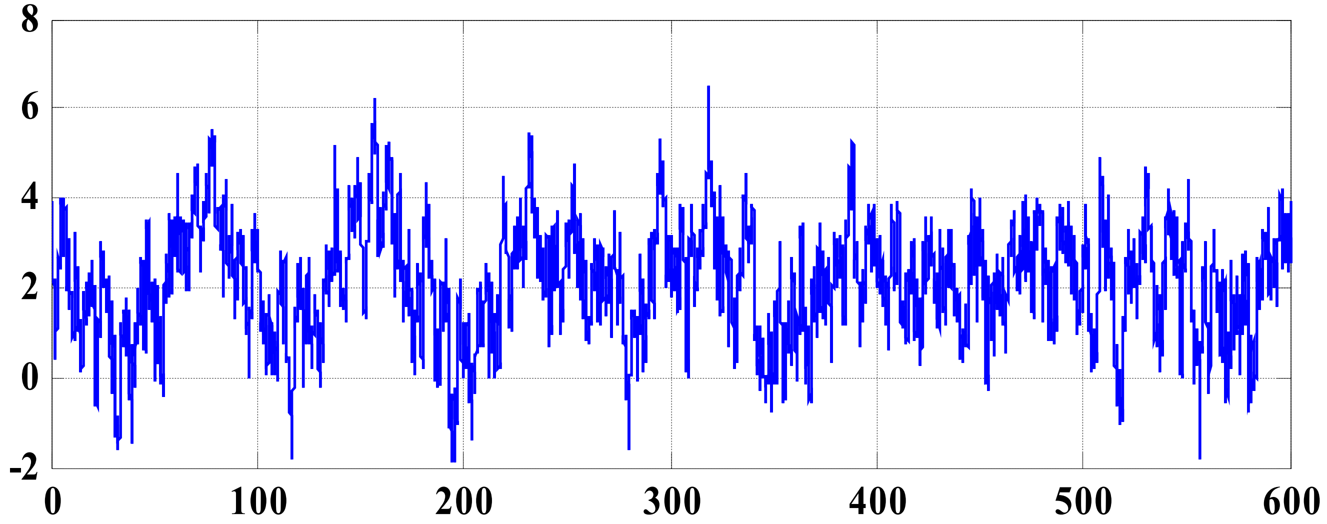


Stationarity

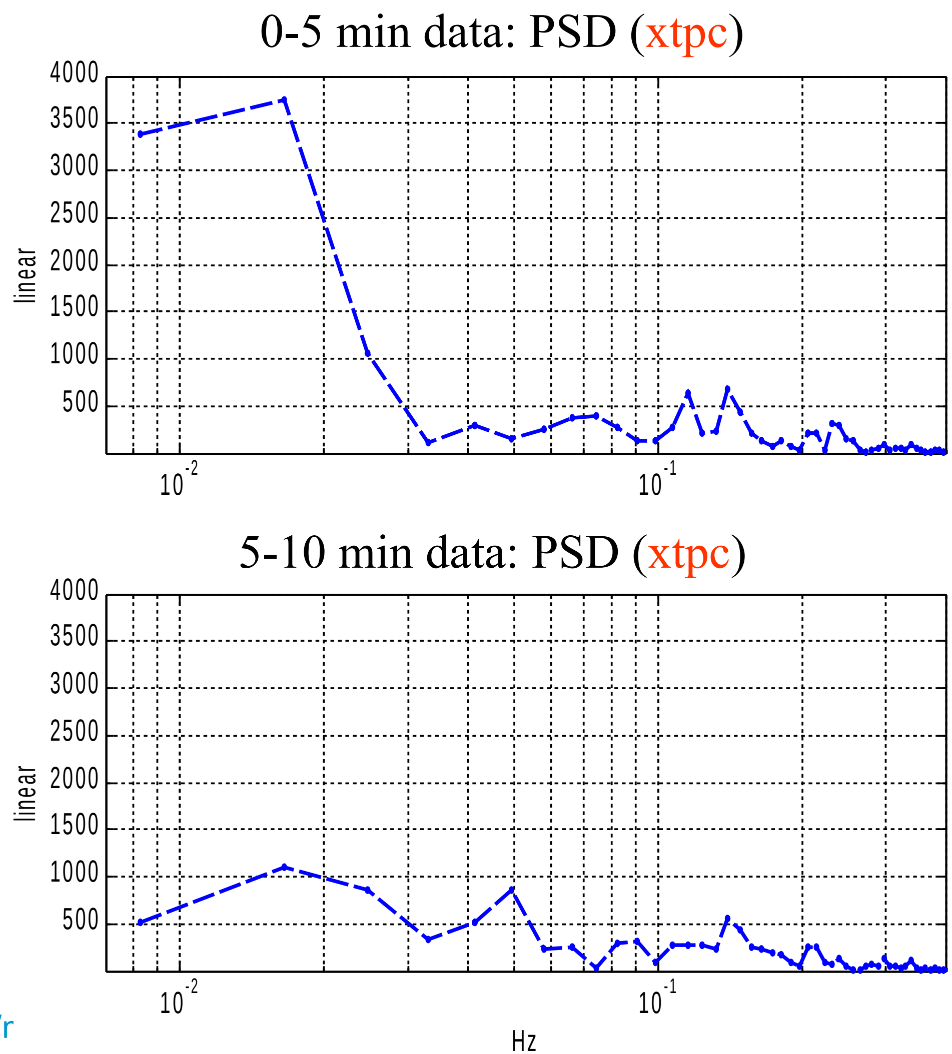


Modeling Atmospheric Turbulence

- 10 minutes of data
- Is it stationary?



PSD's 0-5 minute and 5-10 minute segments.



Nonparametric parametrics?

- Parameter is the covariance
- Estimate the covariance function from data

- Assume stationarity: autocovariance

- Smooth changes: local autocovariance
- Abrupt changes: change-point methods
- Dunno: try wavelets

Nonstationarity: stationary increments

- Brownian motion or random walk: not stationary, but the "derivative" is
- Variogram (or structure function) is a helpful quantity

$$\Gamma(s) = E\left|X(t+s) - X(t)\right|^2$$

- Fractional Brownian motions (long memory processes, 1/f noise, etc) have the form

$$\Gamma(s) = E\left|X(t+s) - X(t)\right|^2 = C\left|s\right|^{2H}$$

  - H is the Hurst parameter: H=1/2 is Brownian motion

Estimating the Variogram

- Standard estimator

$$\hat{\Gamma}(h) = \frac{1}{N_h} \sum_{i=1}^{N_h} \left| X(t_i) - X(t_i + h) \right|^2$$

- This estimator can be fit to functional forms
  - Log-linear regression can estimate the Hurst parameter

- A similar technique, DFA or Detrended Fluctuation Analysis, is used in certain cardiovascular problems

Local Covariance Approach

- Use a sliding window to construct multiple stationarily-modeled covariance functions
- These covariances are examined in terms of the local Karhunen-Loeve expansions (approximately cosine bases)
- In each window, the expansion is compared to the covariance by how well its Hilbert-Schmidt norm compares: how well it captures the total variance in the basis vector directions
- Best window is the one that provides the maximal total variance

Change-Point Approach

- Time series modeled as conditional $g_\theta(y_t \mid y_{t-1}, \cdots, y_1, y_0)$

- Parameter changes/is constant

$$H_0 : \theta = \theta_0$$

$$H_A : \text{there is an } r \text{ with } 1 \le r \le n \text{ such that}$$

$$\theta = \theta_1, 1 \le t \le r$$

$$\theta = \theta_2, r+1 \le t \le n$$

Wavelet Approaches

- Decompose by scale (as in Fourier) and time-locality

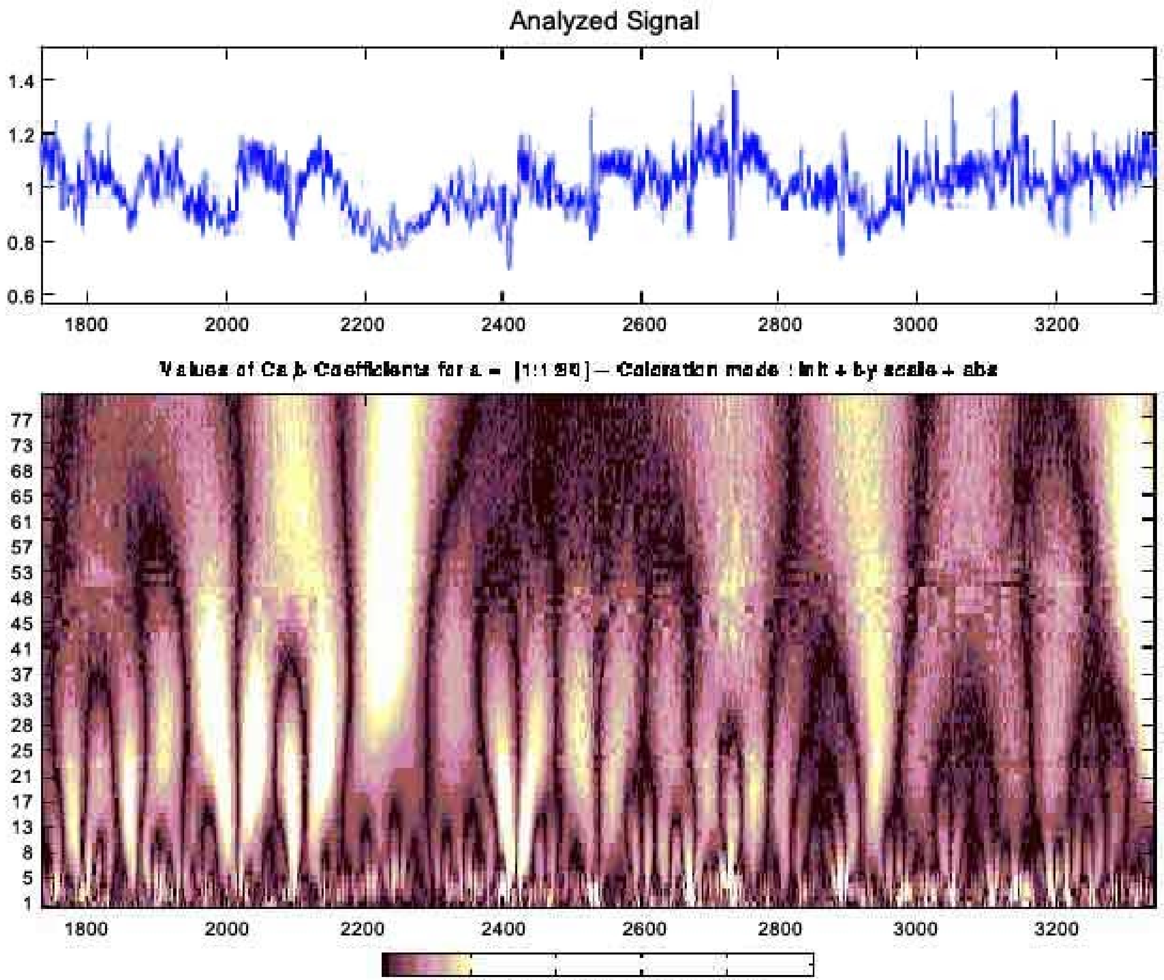
$$W_f(s,a) = |s|^{-\frac{1}{2}} \int_{-\infty}^{\infty} f(x)\psi^*\left(\frac{x-a}{s}\right)dx$$

- Reconstruction has a similar form

$$f = C_\psi^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\langle f, \psi^{s,a} \rangle}{s^2} \psi^{s,a} \, ds\, da$$

- Choice of the wavelet function is important

Wavelet decomposition (from physionet)



http://myweb.lmu.edu/bfitzpatrick http://www.tempest-tech.com

Inferential Wavelets

- Statistical theory for the decomposition exists (in some form)
  - Dependence on the underlying mother wavelet
  - Dependence on the underlying data structure (eg power law)
- Tools for segmentation of signals exist
  - Compare scale estimate to theoretical scale stats
- Parameter estimation tools exist
  - The power in the power law

Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Model-Based Statistics

- Data = Model plus Noise
  - Is plus really “+”?
- How does noise enter the data?
- What is the probabilistic structure of the noise?
- What parameters are unknown?

Simplest Parametric Model

- Data: $Y_k \approx y(t_k)$
- Model: $f(t_k, u_k, \theta)$
  - u is an exogenous or control input variable
  - θ is the unknown parameter
- Statistical observation model: $Y_k = g(t_k) + \varepsilon_k$
  - ε_k are independent identically distributed random variables
- Fitting criteria

$$J_n(\theta) = \sum_{k=1}^{n} \left| Y_k - f(t_k, u_k, \theta) \right|^2$$

What is known

- Consistency

$$\hat{\theta}_n = \arg\min J_n \to \theta^*$$

$$\theta^* = !\arg\min \int_{TxU} \left| f(t,u,\theta) - g(t) \right|^2 d\mu(t,u)$$

$$\frac{1}{n} J_n(\hat{\theta}_n) \to \sigma^2 + \int_{TxU} \left| f(t,u,\theta^*) - g(t) \right|^2 d\mu(t,u)$$

- Asymptotic Normality

$$\sqrt{n}\left(\hat{\theta}_n - \theta^*\right) \to N(0,V)$$

$$V = \sigma^2 \left(\int_{TxU} \frac{\partial f}{\partial \theta}(t,u,\theta) \frac{\partial f^T}{\partial \theta}(t,u,\theta) d\mu(t,u) \right)^{-1}$$

Model selection through hypothesis testing

$$H_0 : \theta \in \Theta_0 = \{\theta \in \Theta : H\theta = h\}$$

$$H_A : \theta \notin \Theta_0$$

$$\tilde{\theta}_n = \arg\min\{J_n(\theta) : \theta \in \Theta_0\}$$

$$\frac{J_n(\tilde{\theta}_n) - J_n(\hat{\theta}_n)}{J_n(\hat{\theta}_n)} \to \chi^2(p-q)$$

$p =$ dimension of Θ

$q =$ dimension of Θ_0

Problems with This Approach

- Is the data really model plus noise?
- Is the parameter really identifiable?
  - Common assumption is $g(t) = f(t, u, \theta^*)$
- Is the noise really iid random noise?
- Do we have enough data for the asymptotics?
- Test becomes less powerful as dimensions increase

- ***This hypothesis test is not validation!***
  - ***Compares a simpler model to a more complicated one!***

Analysis of Residuals

- Are the residuals (Gaussian) white noise?
  - Are there unmodeled trends remaining?
    - If so, we need to revisit the model
  - Are the residuals correlated?
    - If so, we may want to estimate the covariance and re-estimate with weighted least squares.
  - Are the residuals non-Gaussian?
    - If so, we may want to apply maximum likelihood
- Do the residuals indicate outliers?
  - If so, we need to consider robust estimation techniques, such as MAD

Adjusting the Fitting Approach

- Weighted Least Squares

$$J_n(\theta) = \sum_{k=1}^{n} \frac{\left| Y_k - f(t_k, u_k, \theta) \right|^2}{\sigma_k^2}$$

- Minimum Variance

$$J_n(\theta) = \left(Y - f(\theta) \right)^T V^{-1} \left(Y - f(\theta) \right)$$

- Maximum Likelihood

$$J_n(\theta) = p\left(Y - f(\theta) \right), \quad p \text{ is the pdf of the residual vector}$$

- Minimum Median Absolute Deviation

$$J_n(\theta) = \text{Median}\left(Y - f(\theta) \right)$$

Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Bayesian Nonparametric Regression

- Data: $Y_k \approx y(t_k)$
- Model: $f(t_k, u_k, \theta)$
  - u is an exogenous or control input variable
  - θ is the unknown parameter
- Statistical observation model: $Y_k = g(t_k) + \varepsilon_k$
  - ε_k are independent identically distributed random variables
- Prior probability measure on the space of continuous functions

Bayesian Nonparametric Regression

- Prior variance $E_\pi\left(g^2\right) = \tau^2 \int_T \left|\nabla g\right|^2 d\mu$
- Maximum a posteriori

$$J_n^\pi(g) = \sum_{k=1}^{n} \left|Y_k - g(t_k)\right|^2 + \tau^2 \int_T \left|\nabla g\right|^2 d\mu$$

- Model testing? Compute the posterior probability…

Model Test of Fit

- Compute the posterior probability

$$\int_S d\pi(g|y)$$

$$S = \{ f(\bullet, u_{(\bullet)}, \theta) : \theta \in \Theta \}$$

$$S = \{ |g - f(\bullet, u_{(\bullet)}, \theta)| < \delta, \theta \in \Theta \}$$

- Computing with Monte Carlo simulation
  - Sample from the prior
  - Average

- ***Could be viewed as validation***
  - ***But within the assumed framework of the measurement model***

Hierarchical Bayesian Approach

- Start with a simple model

$$Y_k = f(t_k, u_k, \theta) + \varepsilon_k$$

- Determine the residuals

$$\hat{\varepsilon}_k = Y_k - f\left(t_k, u_k, \hat{\theta}_n\right)$$

- Estimate the distribution of the residuals (emprical CDF, covariance)
- Use this estimate as the conditional
- Back to Bayes

Outline

- Modeling discussion
- What statistics actually does
- Exploratory data analysis
- Parametric fitting
- Nonparametric fitting
- Stochastic modeling

Stochastics in Multiple Points in the Model

$$\dot{x}(t) = h(t, x(t), \theta, \eta(t))$$

$\eta(t) =$ plant noise

$$y(t) = L(x(t), \varepsilon(t))$$

$\varepsilon(t) =$ measurement noise

$$\dot{x}(t) = h(t, x(t), \theta) + \sigma(t, x(t), \theta)\dot{\eta}(t)$$

$$y(t) = L(x(t), \theta) + \tau(x(t), \theta)\varepsilon(t)$$

$=$ about the most complicated thing we can do

Nonlinear filtering: particle filters

- Discrete time approximation
- Bayes rule determines the posterior at each step
- Posterior probabilities can be difficult to compute

$$x_{t+1} = x_t + \Delta t h(t, x_t, \theta) + \Delta t \sigma(t, x_t, \theta) \dot{\eta}(t)$$
$$y_t = L(x_t, \theta) + \tau(x_t, \theta) \varepsilon(t)$$

- Generate a large number of sample paths
- Follow them over time
- Approximate the posterior by Monte Carlo

Starting the conversation

- How much of the signal needs to be modeled?
- How do we iterate?
  - Approximate model fit
  - Look at the residual for unmodeled dynamics
  - Estimate statistics of the residual
  - Return to the model fit
- Multiple patients/multiple observations of single patient

References of Interest

- **[A]** Allen, L. . *An Introduction to Stochastic Processes with Biology Applications,* Prentice-Hall, Englewood Cliffs (2003).
- **[BF]** Banks, H. T., and B. G. Fitzpatrick. "Statistical Tests for Model Comparison in Parameter Estimation Problems for Distributed Parameter Systems," *J. Math. Bio* **28**, pp. 501-527 (1990)
- **[Be]** Berger, J. *Statistical Decision Theory and Bayesian Analysis*. Springer, New York (1999).
- **[BD]** Bickel, P., and K. Doksum. *Mathematical Statistics*, Holden-Day, Oakland, 1977.
- **[CaR]** Crato, N. and Ray, B. "Model Selection and Forecasting for Long Range Dependent Processes." *Journal of Forecasting,* **15,** 107-125 (1996).
- **[Cr]** Cressie, N. *Statistics for Spatial Data*, Wiley, New York, 1991.

References of Interest

- **[Ga]** Gallant, A. R. *Nonlinear Statistical Models*, Wiley, New York, 1987.
- **[Fi1]** Fitzpatrick, B. G. "Bayesian analysis in Inverse Problems," *Inverse Problems*, **7**(5) (1991) 675-702.
- **[Fi1]** Fitzpatrick, B. G "Statistical Tests of Fit in Estimation Problems for Structured Population Modeling," *Quart. Appl. Math*., **53**, no. 1, 1995, pp. 105-128.
- **[FY]** Ben G. Fitzpatrick and G. Yin "Empirical Distributions in Least Squares Estimation for Distributed Parameter Systems," *J. Math. Systems, Estimation, and Control*, **5**, no. 1, 1995, pp. 37-57.
- **[HBPJ]** Huet, S., A. Bouvier, M. Poursat, and E. Jolivet. *Statistical Tools for Nonlinear Regression*, 2nd Ed. Springer, New York, 2004.

References of Interest

- **[MPZ]** Mallat, S., G. Papanicolaou, and Z. Zhang, “Adaptive Covariance Estimation of Locally Stationary Processes” preprint (http://georgep.stanford.edu)
- **[PS]** Papanicolaou, G., and K. Solna. “Estimation of Local Power Law Processes,” preprint (http://georgep.stanford.edu)
- **[WJ]** Wand, M. P., and M. C. Jones, *Kernel Smoothing*, Chapman and Hall, London, 1995.
- **[Y]** Yaglom, A. *Correlation Theory of Stationary and Related Random Functions I*, Springer, New York, 1987.