Optimal statistical approaches to detection of exoplanets and estimation of their parameters

Harrison H. Barrett
College of Optical Sciences
and Dept. of Radiology
University of Arizona

Outline

- Pure detection tasks
- Pure estimation tasks
- Combined estimation and detection tasks
- Application to exoplanets -- a game plan

Pure detection

- Given a data set $\mathbf{g}$ (one or more images)
  - Size of $\mathbf{g}$ = # pixels/image × # images
- Two possible hypotheses about the object:
  - Signal absent, S_-
  - Signal present, S_+
- Make a statistically-based determination of whether the signal is present or absent
- Many potential sources of randomness
  - Detector readout noise
  - Photon (Poisson) noise
  - Random system PSF (atmosphere, AO, laser guidestar, ...)
  - Random signal (e.g., location, magnitude)
  - Random background (host star, other celestial objects, dust, ...)

Formulation of the detection strategy

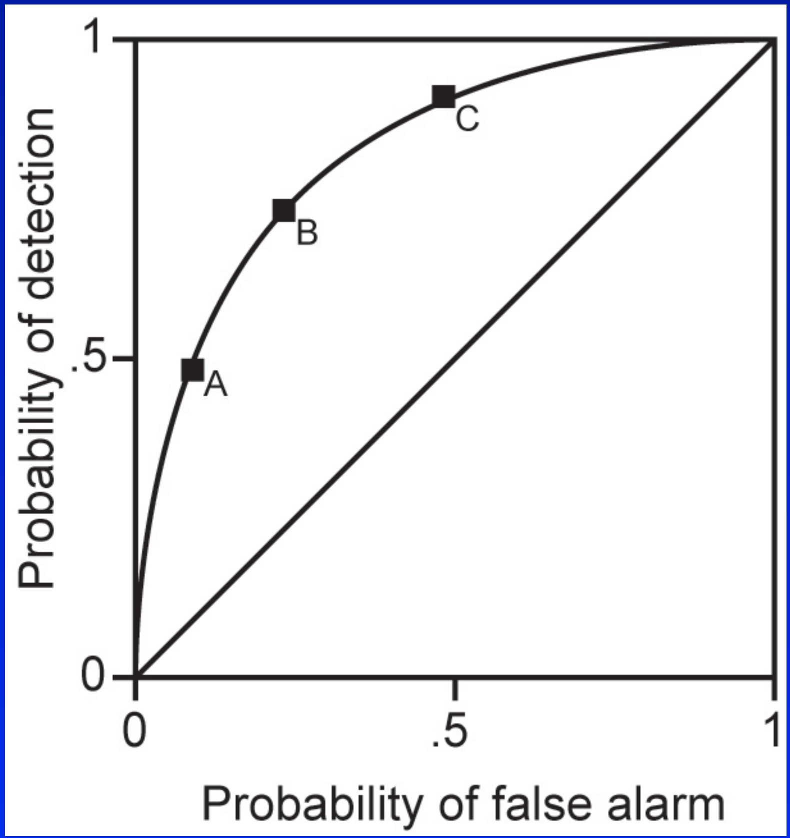
Assume that a binary decision (signal-present or signal-absent) must be made for every image and that there is no randomness in the decision (repeated decisions on the same data vector $\mathbf{g}$ always lead to the same result)

With these assumptions, the decision strategy *always* has the form:

$$t(\mathbf{g}) \underset{D_-}{\overset{D_+}{\gtrless}} \text{threshold}$$

This inequality is to be read:
"decide signal present when the greater-than sign holds;
decide hypothesis signal absent when the less-than sign holds."

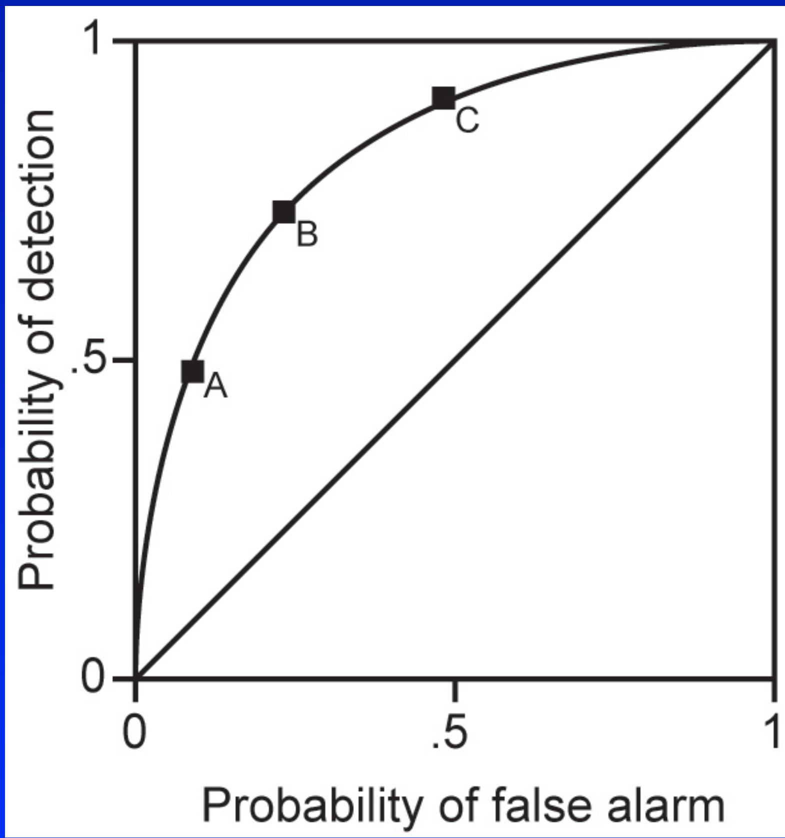
Each signal-present decision can be either a correct detection or a false alarm. Decreasing the threshold increases the probability of detection *and* the probability of a false alarm



Receiver Operating Characteristic (ROC)

Points A, B and C correspond to different decision thresholds

Area under the ROC curve, denoted AUC, is a common figure of merit for detection performance

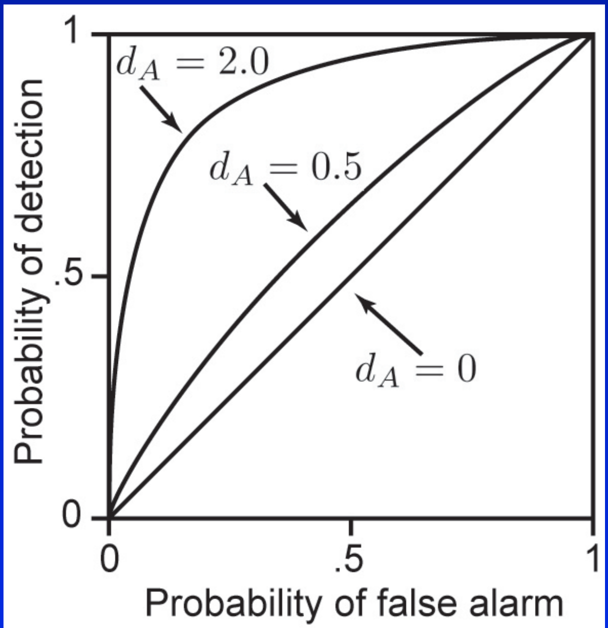


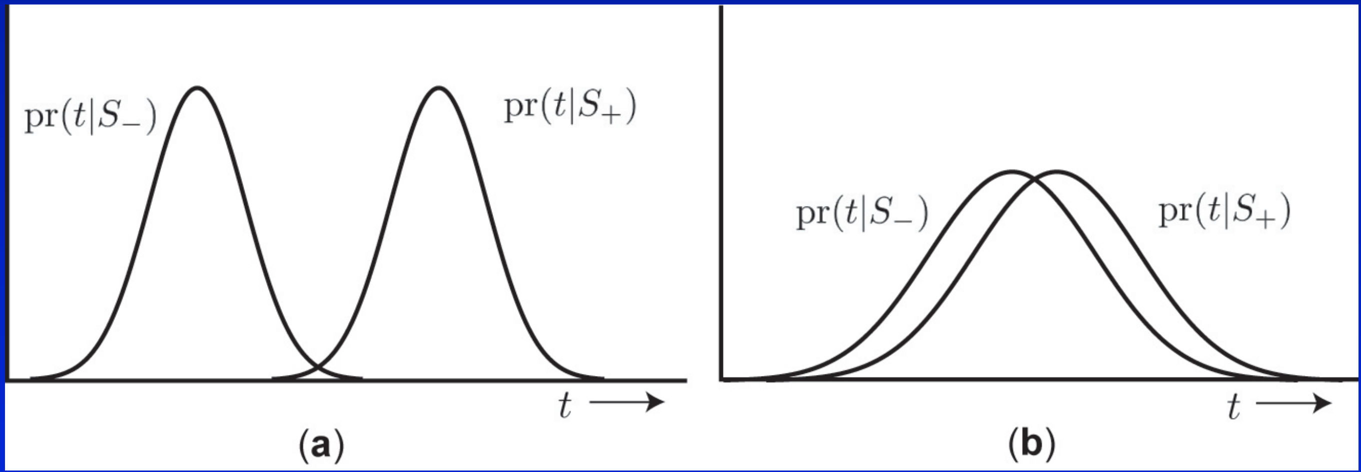
- AUC varies from 0.5 to 1.0
  1.0 ⇔ perfect detection system
  0.5 ⇔ worthless system
- AUC is independent of decision threshold
- AUC is independent of prevalence of signal (probability of signal being present)
- Optimum threshold (operating point) can be determined if one assigns costs to correct and incorrect decisions and has prior knowledge of signal prevalence

Can also define detectability d_A related to AUC

$$d_A = 2\,\mathrm{erf}^{-1}[2(\mathrm{AUC}) - 1]$$

$$(\mathrm{erf} = \text{error function})$$





(a) PDF of test statistic for large d_A

(b) PDF of test statistic for small d_A

If $t(\mathbf{g})$ is normally distributed for S_+ and S_-,

$$d_A = \mathrm{SNR}_t = \frac{\langle t\rangle_+ - \langle t\rangle_-}{\sqrt{\frac{1}{2}\sigma_+^2 + \frac{1}{2}\sigma_-^2}}$$

The ideal observer for signal detection

- Maximizes probability of detection for any probability of false alarm
- Maximizes area under ROC curve
- Maximizes d_A
- Minimizes Bayes risk (if costs and prevalence are specified)
- Requires knowledge of full multivariate PDF of image data under both hypotheses
- Generally nonlinear
- Difficult to calculate

$$\Lambda(\mathbf{g}) = \frac{\mathrm{pr}(\mathbf{g}|S_+)}{\mathrm{pr}(\mathbf{g}|S_-)}$$

$$\lambda(\mathbf{g}) \equiv \ln \Lambda(\mathbf{g}) = \ln \left[\frac{\mathrm{pr}(\mathbf{g}|S_+)}{\mathrm{pr}(\mathbf{g}|S_-)} \right]$$

The test statistic for the ideal observer is the likelihood ratio or its log

The decision is made by comparing the test statistic to a threshold

Varying the threshold generates the ROC curve

1520 J. Opt. Soc. Am. A/Vol. 15, No. 6/June 1998

Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating functions

Harrison H. Barrett, Craig K. Abbey, and Eric Clarkson

- Many mathematical and statistical properties of ideal observers and their ROC curves
- All statistical properties of ideal observer can be derived from likelihood-generating function
- Many simplifications possible if $\lambda(\mathbf{g})$ is normally distributed; true if $\mathbf{g}$ is a sequence of indep. images

The Hotelling observer

Harold Hotelling

- Based on 1931 paper by Harold Hotelling
- Performs only linear operations on data
- Optimum in several senses
- Requires knowledge of ensemble mean and covariance of the images
- Computational difficulty: inversion of large covariance matrix (but many tricks available)
- Equivalent to ideal observer for Gaussian data

Linear discriminants all have the form of a scalar product:

$$t(\mathbf{g}) = \mathbf{w}^t \mathbf{g} \, ,$$

where $\mathbf{w}$ is a template vector the same size as data $\mathbf{g}$,

The Hotelling template is the one that maximizes the SNR:

$$\mathrm{SNR}_t = \frac{\langle t \rangle_+ - \langle t \rangle_-}{\sqrt{\frac{1}{2}\sigma_+^2 + \frac{1}{2}\sigma_-^2}}$$

For weak signals, the data have equal covariance $\mathbf{K_g}$ under the two hypotheses, and the optimal template is

$$\mathbf{w}_{Hot} = \mathbf{K_g}^{-1} \Delta \overline{\mathbf{g}} \, ,$$

where $\Delta \overline{\mathbf{g}}$ is the average difference in the data under the two hypotheses, averaged over all sources of variability.

The resulting SNR is then given by

$$\mathrm{SNR}_{Hot}^2 = \Delta \overline{\mathbf{g}}^t \mathbf{K_g}^{-1} \Delta \overline{\mathbf{g}}$$

Comments on the covariance matrix $\mathbf{K_g}$

- It's huge! (# pixels × # pixels for one image)
- It's an ensemble covariance matrix, not a sample matrix
- It must include all sources of randomness, at least in the signal-absent images
- It must be inverted to get the Hotelling template
- Lecture by Luca Caucci will explain how to deal with these issues

Pure estimation tasks

- Signal known to be present
- Want to estimate some set of parameters $\boldsymbol{\theta}$
- Have a statistical model (likelihood) $\mathrm{pr}(\mathbf{g} \mid \boldsymbol{\theta})$
- Distinguish Bayesian and classical estimation
  - Bayesian: Prior probability model, $\mathrm{pr}(\boldsymbol{\theta})$
  - Classical: No prior knowledge of $\boldsymbol{\theta}$

Estimation figures of merit

Classical: Bias, variance and mean-squared error (MSE)

Bias = systematic error, accuracy
Variance = random error, precision
MSE = bias^2 + variance

Bayesian: Ensemble MSE

$$\mathrm{EMSE} \equiv \left\langle ||\hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta}||^2 \right\rangle$$

Average is over all sources of randomness in the data *and* over an ensemble of parameters

The ideal Bayesian observer

- Minimizes the EMSE

$$\hat{\boldsymbol{\theta}}(\mathbf{g}) = \int d\boldsymbol{\theta} \;\; \mathrm{pr}(\boldsymbol{\theta}|\mathbf{g})\,\boldsymbol{\theta}$$

$$= \frac{\int d\boldsymbol{\theta} \;\; \mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})\,\mathrm{pr}(\boldsymbol{\theta})\,\boldsymbol{\theta}}{\int d\boldsymbol{\theta} \;\; \mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})\,\mathrm{pr}(\boldsymbol{\theta})}$$

- Requires knowledge of PDF of image data *and* a prior distribution on the parameter

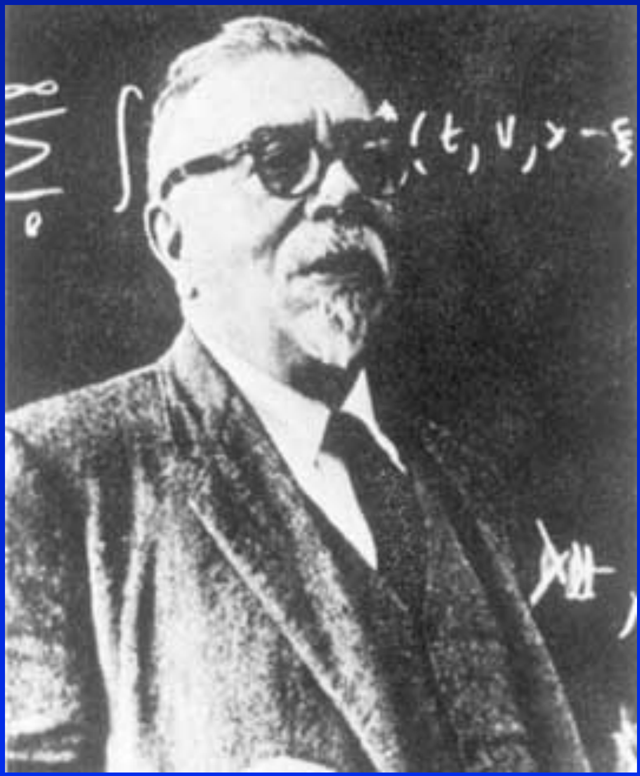
The ideal estimator computes the posterior mean of the parameter

Equivalently, it computes the mean with respect to the likelihood weighted by the prior

- Generally nonlinear

- Difficult to calculate

Generalized Wiener estimator



Norbert Wiener

- Performs only linear operations on data
- Optimizes ensemble mean-square error
- Requires prior knowledge of mean and covariance of data and parameter
- Computational difficulty: inversion of large covariance matrix (but many tricks available)
- Ideal if posterior PDF is Gaussian
- Reduces to Wiener *filter* for stationary,
- Gaussian noise (not a good model for exoplanets)
- Performs poorly with location uncertainty

Maximum A Posteriori (MAP) estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \; \operatorname{pr}(\boldsymbol{\theta}|\mathbf{g}) = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \; \operatorname{pr}(\mathbf{g}|\boldsymbol{\theta})\operatorname{pr}(\boldsymbol{\theta})$$

Find the maximum of the likelihood weighted by the prior

Maximum-likelihood estimation

- MLE maximizes the probability of the data given the parameter :

$$\hat{\theta}_{\mathrm{ML}} \equiv \underset{\theta}{\operatorname{argmax}} \; \mathrm{pr}(\mathrm{g}|\theta)$$

- Equivalently, maximizes the logarithm of this conditional probability:

$$\hat{\theta}_{\mathrm{ML}} = \underset{\theta}{\operatorname{argmax}} \; \ln[\mathrm{pr}(\mathrm{g}|\theta)]$$

Fisher information matrix

Definition

$$F_{jk} = \left\langle \left[\frac{\partial}{\partial \theta_j} \ln \mathrm{pr}(\mathrm{g}|\boldsymbol{\theta}) \right] \left[\frac{\partial}{\partial \theta_k} \ln \mathrm{pr}(\mathrm{g}|\boldsymbol{\theta}) \right] \right\rangle_{\mathrm{g}|\boldsymbol{\theta}}$$

Cramer-Rao lower bound (for unbiased estimator)

$$\mathrm{Var}\{\hat{\theta}_n\} \geq \left[\mathbf{F}^{-1} \right]_{nn}$$

Off-diagonal elements of inverse
relate to covariances of estimates

An *efficient estimator* is one that is unbiased and for which the CR bound become an equality

In any problem, the ML estimator is efficient
if an efficient estimator exists

The ML estimator is always asymptotically efficient ...

... as you get more or better data

....thereby increasing the (Fisher) information content

Maximum-likelihood methods in wavefront sensing: stochastic models and likelihood functions

Harrison H. Barrett

College of Optical Sciences and Department of Radiology, University of Arizona, Tucson, Arizona 85724

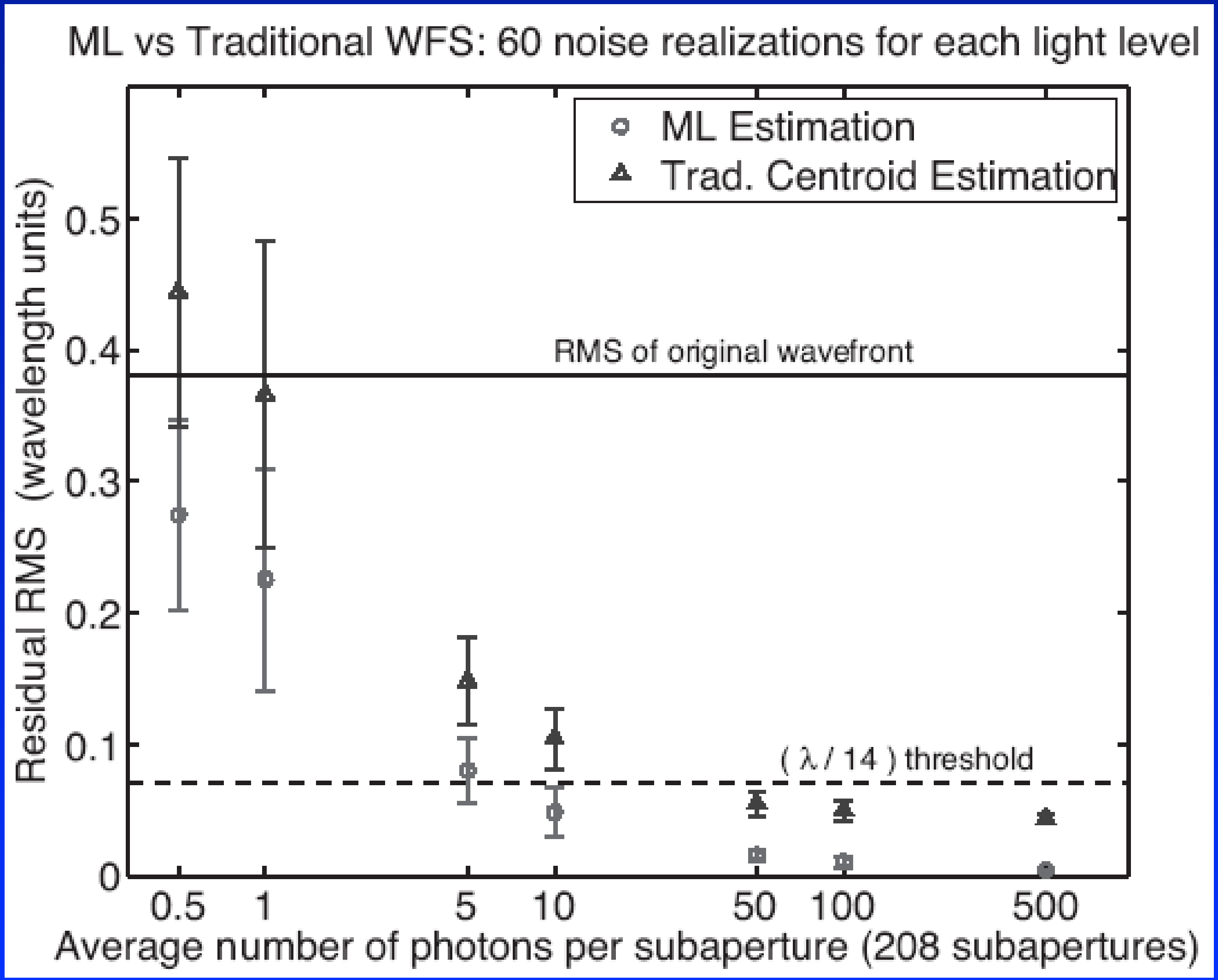
Christopher Dainty

Department of Physics, National University of Ireland, Galway, Ireland

David Lara

Department of Physics, National University of Ireland, Galway, Ireland

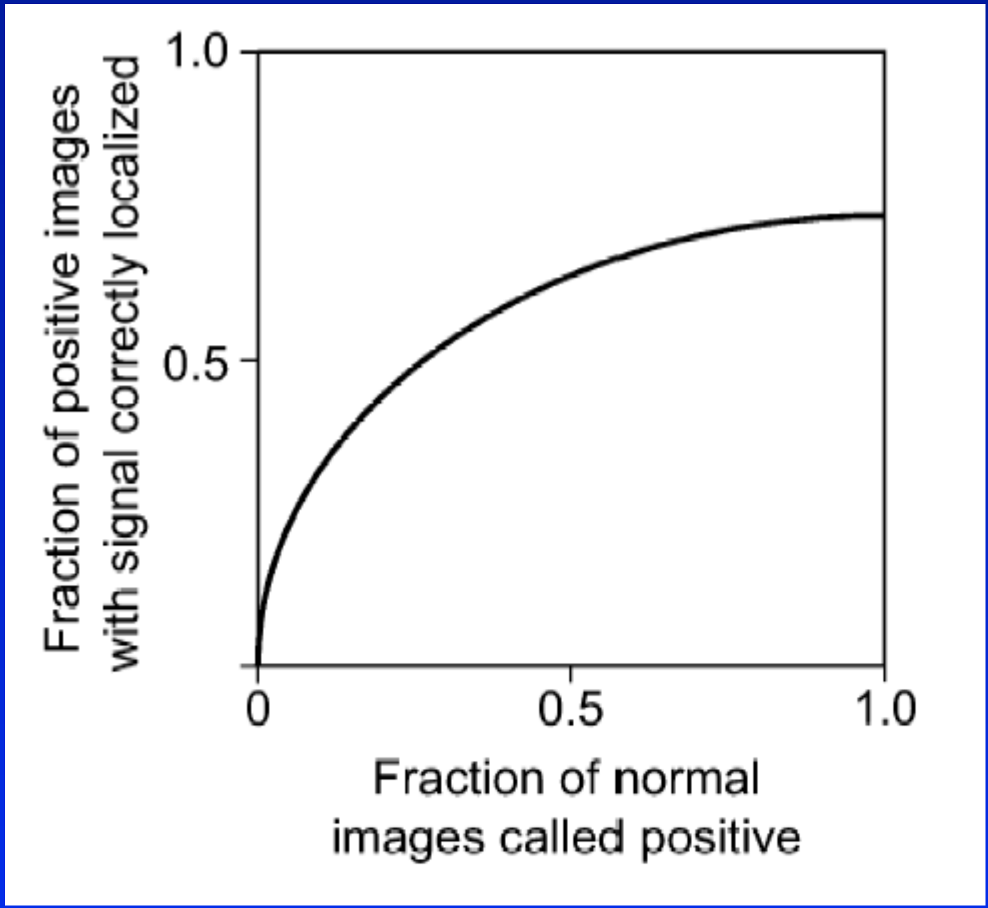
Advantage of MLE over centroids
(work of David Lara, Galway)



Joint detection and estimation

- Problem statement
  - Decide whether a signal is present
  - If it is, estimate location, magnitude, other parameters
- Figures of merit: Area under LROC and EROC
- Ideal observers
- Linear observers
  - Scanning Hotelling template
  - Scanning Linear Estimator (SLE)

Localization tasks: the LROC curve



P. Khurd and G. Gindi, "Decision strategies maximizing the area under the LROC curve," in *Medical Imaging 2005: Image Perception, Observer Performance, and Technology Assessment*, **5749**, pp. 150–161, SPIE, 2005.

General estimation tasks: the EROC curve

EROC: Plot of expected utility of a true-positive detection against false-alarm rate as the detection threshold is varied

Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks

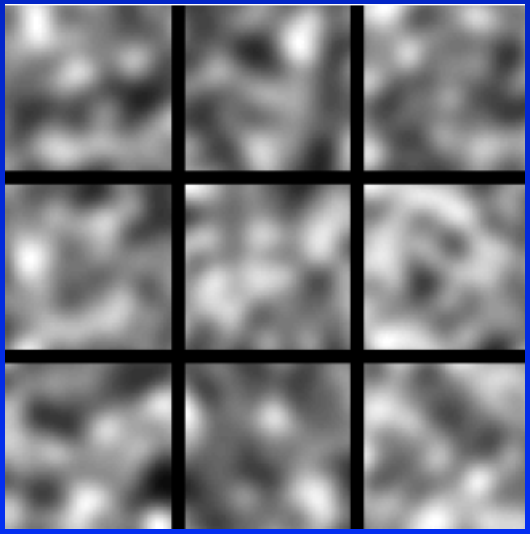
Ideal observers to maximize AEROC must know full PDFs of data under signal-absent and signal-present hypotheses as well as full PDFs of parameters to be estimated

Under Gaussian assumptions, the ideal AEROC observer is equivalent to a scanning Hotelling observer

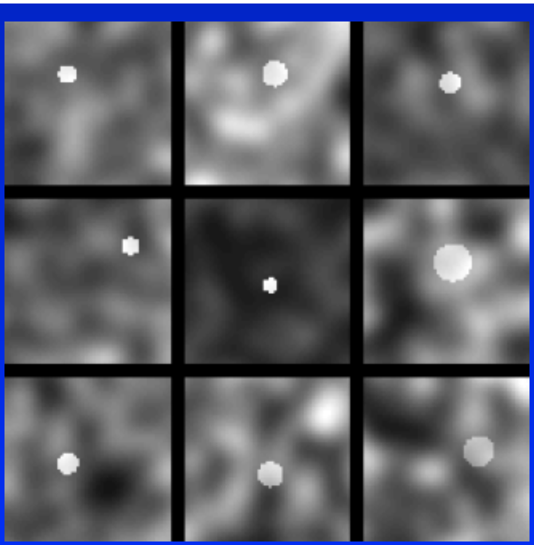
26 May 2008 / Vol. 16, No. 11 / OPTICS EXPRESS 8150

Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods

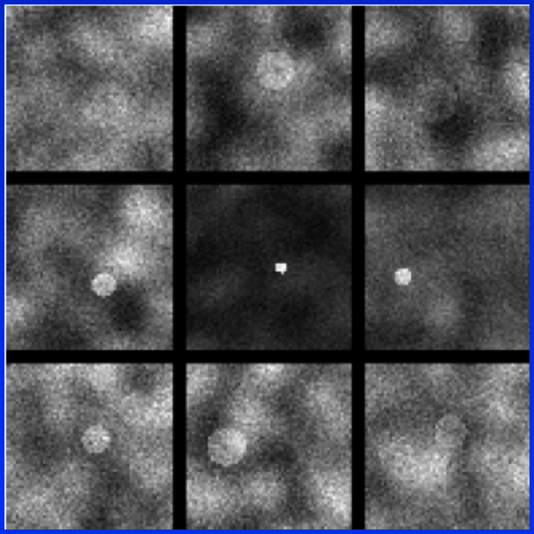
Meredith Kathryn Whitaker, Eric Clarkson, and Harrison H. Barrett
College of Optical Sciences and Department of Radiology, University of Arizona,



Random backgrounds (nuisance parameters)

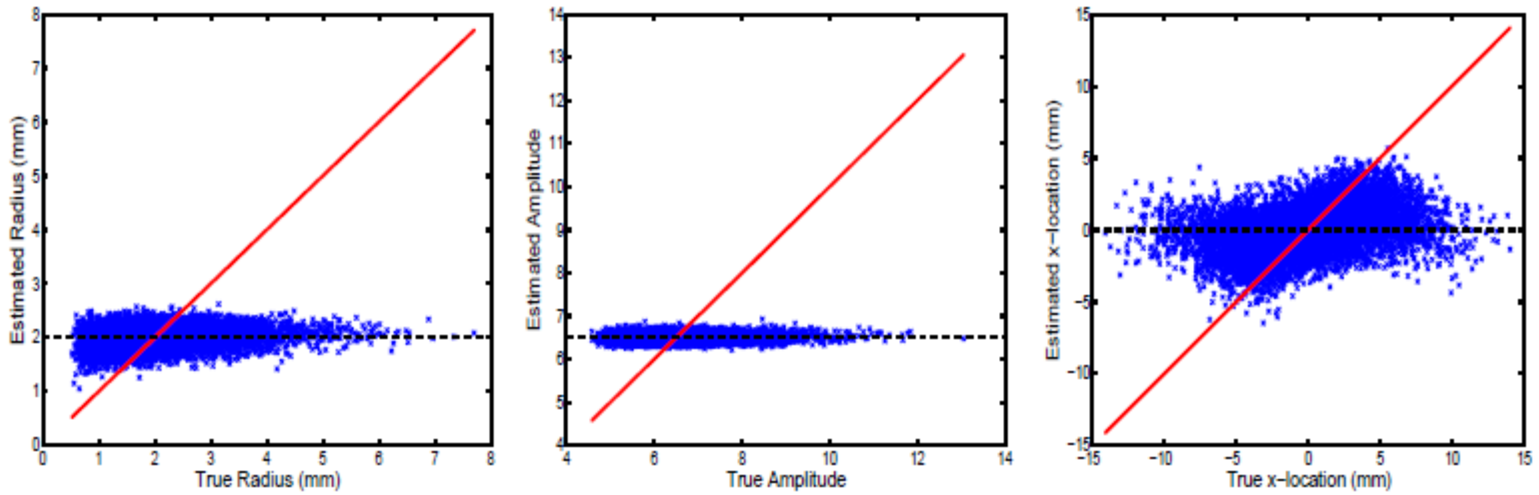


Signals (random size, location and amplitude)



Images for analysis, with Poisson noise

Scatter plots of Wiener estimates of signal radius, amplitude and location
(red line indicates perfect performance)

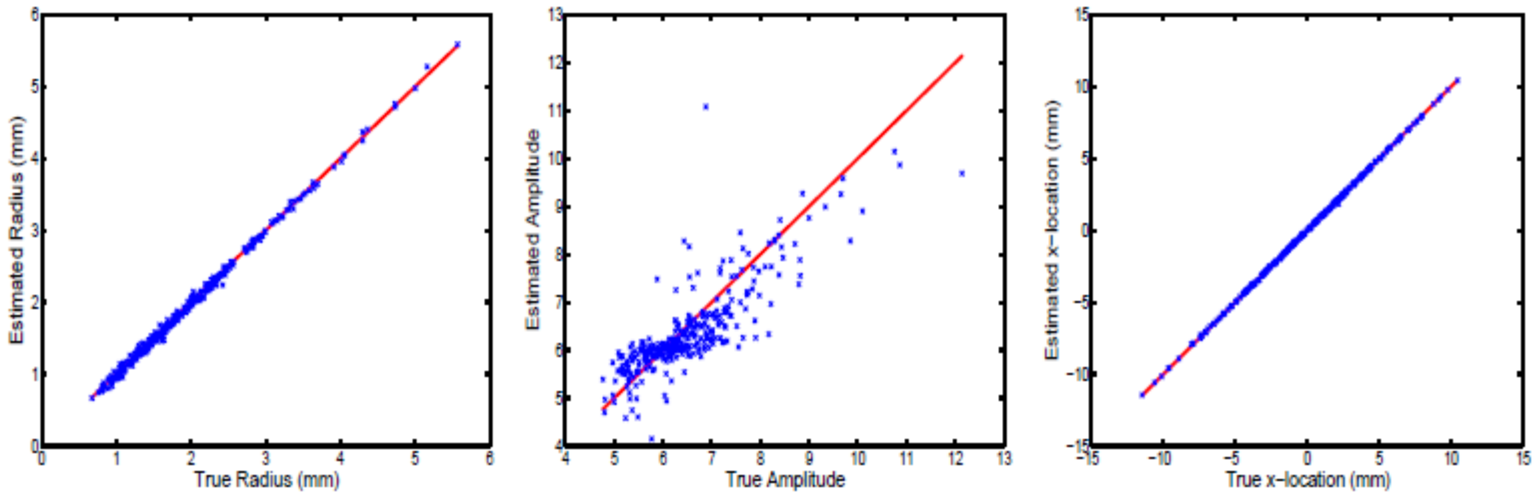


(a) Signal Radius (b) Signal Amplitude (c) y-location

Conclusion: optimum linear estimator does not work at all with location uncertainty

Wiener estimator simply returns the mean of the prior ensemble (dashed line)

Scatter plots of scanning linear estimates of signal radius, amplitude and location (red line indicates perfect performance)



(a) Signal Radius (b) Signal Amplitude (c) y-location

This method was applied to binary stars by Burke, Devaney, Whitaker et al. Proc. of SPIE Vol. 7015, 70152J, (2008)

- What is the SLE?
  - An approximation to a MAP estimator ...

    ... that assumes a Gaussian likelihood ...

    ... and simplifies the covariance by neglecting parameter randomness
  - An estimator that is linear in the data g, ...

    ... but scans a linear template in parameter space

    ... and then performs a nonlinear (argmax) operation
- Why does it work?
  - It knows about the system matrix
  - It knows about the object covariance

Application to exoplanets

- Tasks
- Types of data
- Sources of randomness
- A game plan

Tasks

- Detection
- Photometry (estimation of location of planet)
- Astrometry (estimation of relative magnitude)
- Spectral estimation or detection of spectral signature

Types of data

- Single long-exposure image
- Multiple short-exposure images
  - Temporal sequences
  - Angular sequences
- Images plus spectra
- Images plus AO actuator signals
- Phase-diversity images

Key point: Optimal statistical methods are, in principle, applicable to all of these tasks and all data types

Figures of merit related to task performance can be used to compare data types as well as analysis methods

Major need: Comprehensive, task-specific statistical models

3080 J. Opt. Soc. Am. A/Vol. 23, No. 12/December 2006

Objective assessment of image quality. IV. Application to adaptive optics

Harrison H. Barrett

College of Optical Sciences and Department of Radiology, University of Arizona, Tucson, Arizona 85724

Kyle J. Myers

NIBIB/CDRH Laboratory for the Assessment of Medical Imaging Systems, Rockville, Maryland 20850

Nicholas Devaney and Christopher Dainty

Department of Physics, National University of Ireland, Galway, Ireland

Tasks considered:

- Detection of a star on a random background
- Detection of a faint companion
- Photometry in a crowded star field
- Simultaneous differential imaging

Random effects considered

- Spatial and temporal correlations in residual speckle
- Background models
- Nuisance parameters
- Poisson noise
- Camera readout noise

Related papers

H. H. Barrett, K. J. Myers, N. Devaney, J. C. Dainty and L. Caucci, "Task performance in astronomical adaptive optics," *Proc. SPIE,* 6272, 62721W, 2006.

L. Caucci, H. H. Barrett, N. Devaney, and J. L. Rodriguez, "Application of the Hotelling and ideal observers to detection and localization of exoplanets," *J. Opt. Soc. Am. A*, 24(12), B13-B24, 2007.

D. Burke, N. Devaney, S. Gladysz, H. H. Barrett, M. K. Whitaker, and L. Caucci, "Optimal linear estimation of binary star parameters," *Proc. SPIE 7015*, 70152J, 2008.

L. Caucci, H. H. Barrett, and J. J. Rodriguez, "Spatio-temporal Hotelling observer for signal detection from image sequences," *Opt. Expr.* 17,10946-10958, 2009.

D. Burke, S. Gladysz, L. Roberts, N. Devaney and C. Dainty, "An improved technique for the photometry and astrometry of faint companions," *Pubs. Astron. Soc. Pac.*,121, 767-777, 2009

D. Burke and N. Devaney, "Enhanced faint companion photometry and astrometry using wavelength diversity," *J. Opt. Soc. Am. A*, 27, A246-A252, 2010.

A proposed game plan for the workshop challenge

- Make it a two-stage challenge
- Stage 1 (pre-workshop)
  - Simple telescope model
  - No coronagraph or pupil apodization
  - Only residual atmospheric speckle
  - Use simulated data
- Stage 2 (post-workshop)
  - State-of-the-art telescope
  - Include coronagraph, electric-field conjugation, apodization,...
  - Include quasistatic speckle
  - Consider use of real data with added planets
- In both stages, perform detection, location and magnitude estimation; analyze results by ROC, LROC and EROC

Suggested information to be supplied for Stage 1

- Telescope
  - Specify primary and secondary diameters, spider, wavelength, ideal PSF
- Atmosphere
  - Kolmogorov, specify r_0
  - Residual atmospheric speckle only (no quasistatic)
- Adaptive optics
  - Single adaptive mirror conjugate to telescope pupil
  - Specify actuator configuration, system bandwidth, WFS type and noise level
- Science camera
  - Specify pixel size, array size, readout noise, exposure time
- Images
  - Supply 100 simulated images, half with a single planet in a random location and with random magnitude (ranges to be determined). Planet and host star should have same PSF, including residual speckle.
  - Specify mean # of detected photons from host star (i.e, Poisson noise)
  - Same host star, independent speckle realizations for each of the 100 images

Still to be decided

- PDFs for positions and magnitudes of simulated planets
- Tolerance to use for LROC
- Utility function to use for EROC
- How to set error bars on AEROC (jackknife?)
- How to determine statistical significance of differences in AEROC
- Should we investigate effects of model mismatch, e.g. inaccurate knowledge of pr(θ) ?

Discussion

Acronyms (in order of introduction)

PSF: Point Spread Function
AO: Adaptive Optics
ROC: Receiver Operating Characteristic
AUC: Area Under Curve (usually an ROC curve)
PDF: probability density function
SNR: Signal-to-Noise Ratio
MSE: Mean-Square Error
EMSE: Ensemble Mean-Square Error
ML: Maximum Likelihood
MAP: Maximum A Posteriori
LROC: Localization ROC
EROC: Estimation ROC
AEROC: Area under EROC curve
SLE: Scanning Linear Estimator

Mathematical symbols (in order of introduction)

$\mathbf{g}$ = data set of one or more images (huge vector)

$t(\mathbf{g})$ = scalar test statistic derived from $\mathbf{g}$

S_+ , S_- = signal-present, signal-absent hypotheses

D_+ , D_- = signal-present, signal-absent decisions

SNR_t= SNR associated with $t(\mathbf{g})$

d_A = detectability measure derived from AUC

$\mathrm{pr}(\mathbf{g})$ = full multivariate probability density function (or probability) for $\mathbf{g}$

$\mathrm{pr}(\mathbf{g}|S_+)$ = conditional multivariate PDF for $\mathbf{g}$ given that S_+ is true

$\Lambda(\mathbf{g})$ = likelihood ratio for data $\mathbf{g}$

$\lambda(\mathbf{g})$ = log of the likelihood ratio for data $\mathbf{g}$

Mathematical symbols (continued)

$\mathbf{w}$ = template vector for linear discriminant (same size as $\mathbf{g}$)

$\mathbf{w}^t\mathbf{g}$ = scalar product of $\mathbf{w}$ and $\mathbf{g}$ (t denotes transpose)

$\mathbf{K_g}$ = covariance matrix of data $\mathbf{g}$ [sizeof($\mathbf{g}$) $\times$ sizeof($\mathbf{g}$)]

$\boldsymbol{\theta}$ = vector of parameters to be estimated

$\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})$ = likelihood of $\boldsymbol{\theta}$ for data $\mathbf{g}$ = conditional PDF of $\mathbf{g}$ given $\boldsymbol{\theta}$

$\hat{\boldsymbol{\theta}}(\mathbf{g})$ = estimate of $\boldsymbol{\theta}$ derived from $\mathbf{g}$