# Image Quality: Signal Detection Theory and Classification Tasks

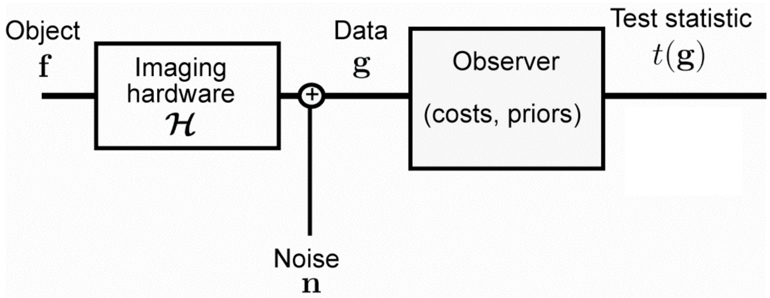
Kyle J. Myers, Ph.D.

NIBIB/CDRH Laboratory for the Assessment of Medical Imaging Systems

# Outline

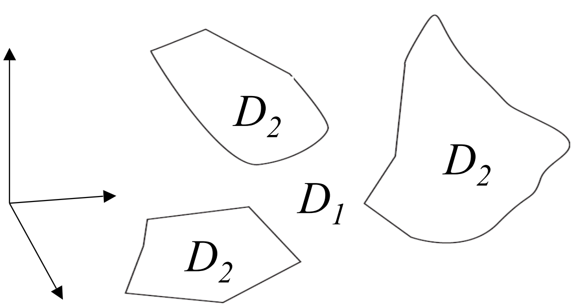
- Statistical Decision Theory: Classification
- The Ideal Observer
- The Hotelling Observer
- Human Observers

# Classification tasks




- The observer computes a scalar test statistic $t(\mathbf{g})$
- Applies a threshold to assign class membership
  - We'll focus on binary tasks here
  - Examples: Tumor detection, Benign vs. malignant
  - Task is to decide in favor of Hypothesis 1 or Hypothesis 2
  - Hypotheses are denoted $H_1$ and $H_2$
  - Decisions are denoted $D_1$ and $D_2$

# Contours of constant $t(\mathbf{g})$ give partitions in data space



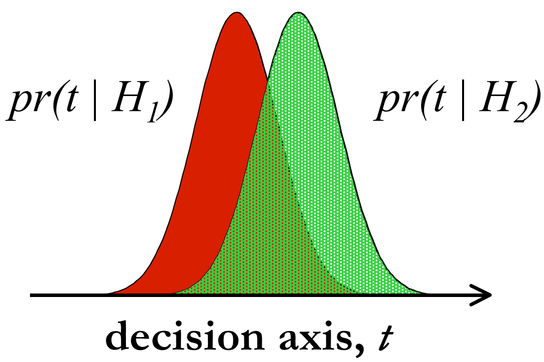

$$t(\mathbf{g}) \underset{D_1}{\overset{D_2}{\gtrless}} \text{constant}$$

Four possible decision outcomes

| | $H_2$: Signal present | $H_1$: Signal absent |
|---|---|---|
| Decide $H_2$<br>$D_2$ | True positive<br>Hit<br>Correct detection | False positive<br>False alarm<br>Type I error |
| Decide $H_1$<br>$D_1$ | False negative<br>Miss<br>Type II error | True negative<br>"good no call" |

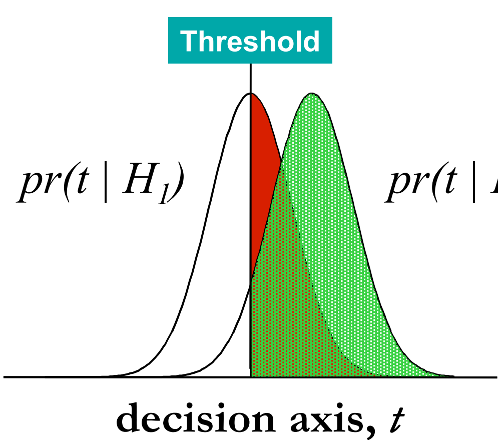
# Test statistic = decision variable

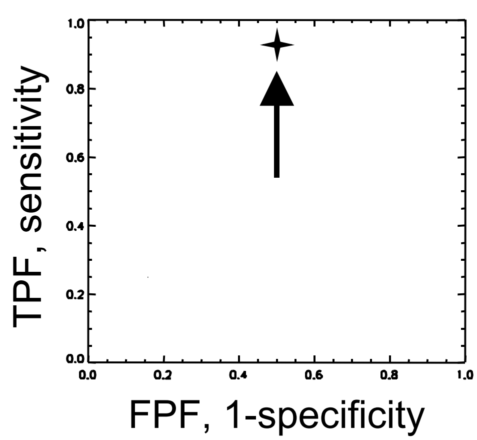
- Distribution over all patients (objects) and noise is denoted $pr(t|H_j)$ for each class $H_j$

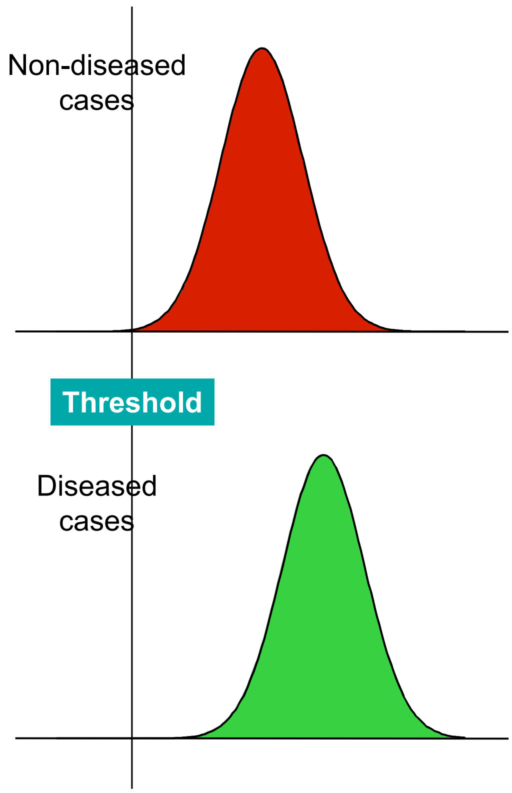



- Analyze statistics of $t$ ($\mathbf{g}$) to determine a figure of merit ➔ measures quality of decisions

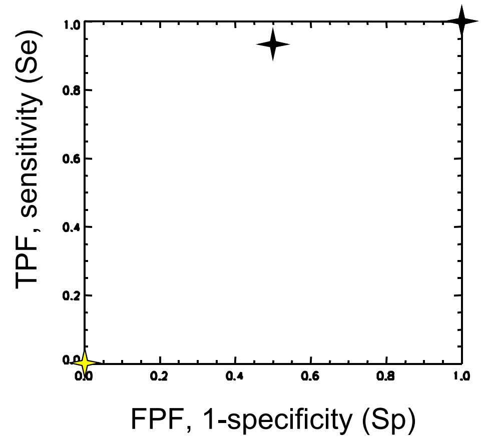
A particular threshold yields an operating point:
(TPF,FPF) = (sensitivity,1-specificity)



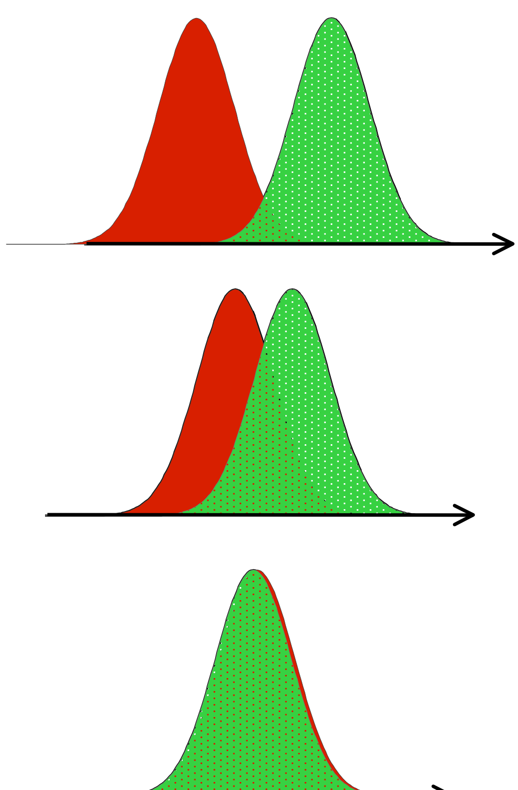





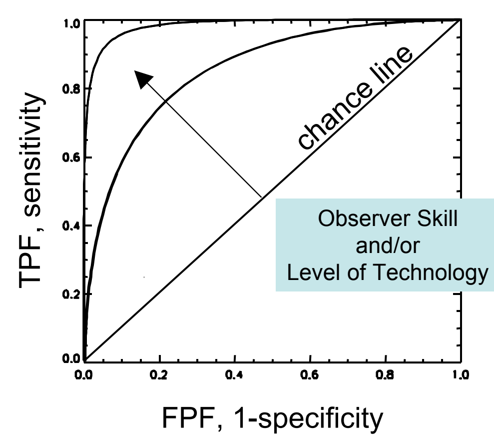

Sweeping across thresholds maps out entire ROC curve




Courtesy Brandon Gallas



Entire ROC curve




Degree of overlap
⇩
Ability to discriminate between 2 conditions

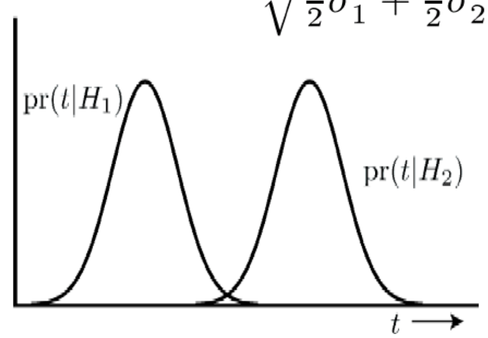
Area under the ROC curve, denoted AUC, is a common figure of merit for detection performance:

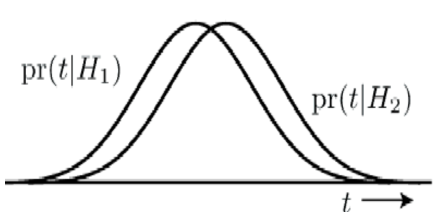
$$\mathrm{AUC} = \int_0^1 d\mathrm{FPF}\ \mathrm{TPF}(\mathrm{FPF}) = \int_0^1 d\mathrm{FPF}(t_c)\ \mathrm{TPF}(t_c)$$

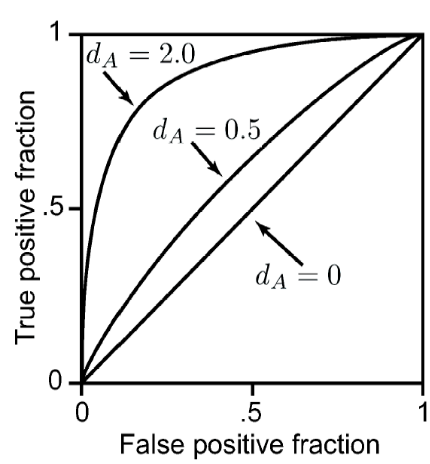
- Average sensitivity over all specificities
- Independent of decision threshold
- Varies from 0.5 to 1.0
  AUC = 1.0 ⇔ perfect system
  AUC = 0.5 ⇔ worthless system
- Independent of prevalence of disease

## Task Signal-to-Noise Ratio: Detectability

$$\mathrm{SNR}_t = \frac{\langle t \rangle_2 - \langle t \rangle_1}{\sqrt{\frac{1}{2}\sigma_1^2 + \frac{1}{2}\sigma_2^2}}$$ = *d´* under conditions of normality



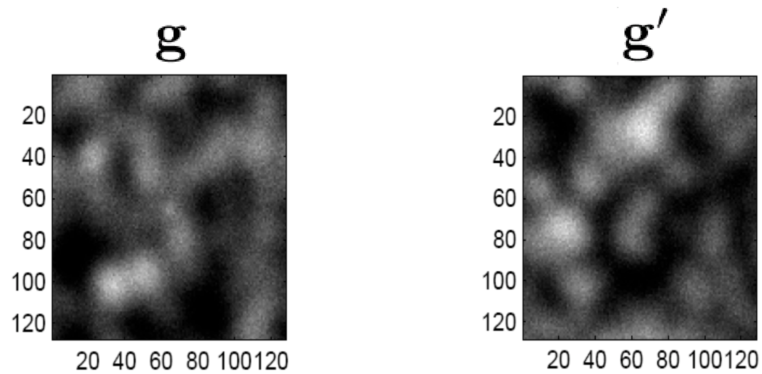




$d_A = 2\ \mathrm{erf}^{-1}\ [2(\mathrm{AUC}) - 1]$

# Another interpretation for AUC

- Percent correct (PC) in a 2-Alternative Forced-Choice experiment (2AFC)
- Present two images, one with tumor and one without
- Observer must pick which image has the tumor

# 2AFC Paradigm




Observer must choose which of two images contains the signal.

$$\Pr(\text{correct}) = \Pr[t(\mathbf{g}') > t(\mathbf{g})]$$

$$= \int_\infty d^M g \int_\infty d^M g' \, q_1(\mathbf{g})\, q_2(\mathbf{g}')\, \text{step}[t(\mathbf{g}') - t(\mathbf{g})] = \text{AUC} \tag{13.39}$$

# What is optimal form for *t(**g**)*?

- We'll consider several "optimality" criteria
- Show that same form for decision variable results

# Finding the Decision Variable of the Ideal Observer

Minimize average cost:

$$\overline{C} = C_{22}\Pr(D_2|H_2)\Pr(H_2) + C_{12}\Pr(D_1|H_2)\Pr(H_2) + C_{21}\Pr(D_2|H_1)\Pr(H_1) + C_{11}\Pr(D_1|H_1)\Pr(H_1) \quad (13.51)$$

Make just two assumptions:

1) Wrong decisions cost more than right ones
2) Data partitions are nonoverlapping and complete:

$$\Pr(D_2|H_2) + \Pr(D_1|H_2) = 1 \quad (13.55)$$

Minimum cost decision rule is:

$$\frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)} > \frac{(C_{21} - C_{11})\,\mathrm{Pr}(H_1)}{(C_{12} - C_{22})\,\mathrm{Pr}(H_2)} \tag{13.58}$$

If instead we maximized the percentage of correct decisions:

$$\frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)} > \frac{\mathrm{Pr}(H_1)}{\mathrm{Pr}(H_2)} \tag{13.61}$$

If we wanted decisions in favor of the most likely class:

$$\frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)} \underset{D_1}{\overset{D_2}{\gtrless}} 1 \tag{13.71}$$

Decisions that maximize TPF for particular FPF have similar form:

$$\frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)} \underset{D_1}{\overset{D_2}{\gtrless}} \gamma \tag{13.69}$$

# Ideal (Bayesian) observer

- Test statistic $t(\mathbf{g})$ is the likelihood ratio: $\Lambda(\mathbf{g}) \equiv \frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)}$
- Optimality:
  - Minimum cost or Bayes risk
  - Minimum probability of error
  - Maximum likelihood
  - Maximum TPF at any FPF
  - Maximum AUC
- Requires full joint PDF on data under each hypothesis (multivariate statistics)
- Performance is determined by statistics of $\Lambda(\mathbf{g}|H_i)$

## Monotonic transformations and the IO

- Monotonic transformations of the decision variable give the same ROC curve.
- The decision axis is rescaled but order is preserved.
- Transform the threshold in the same way to get the same operating point.

➔ We can use either the likelihood ratio or its log as the test statistic for the ideal observer

$$\lambda(\mathbf{g}) \equiv \ln\left[\Lambda(\mathbf{g})\right] = \ln\left[\frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)}\right]$$

## Ideal observer with Gaussian data

When noise is additive and Gaussian and the signals and backgrounds are known exactly:

$$H_1 : \mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}_1 + \mathbf{n} = \mathbf{s}_1 + \mathbf{n} \qquad (13.103)$$

$$H_2 : \mathbf{g} = \boldsymbol{\mathcal{H}}\mathbf{f}_2 + \mathbf{n} = \mathbf{s}_2 + \mathbf{n}$$

with

$$\mathrm{pr}(\mathbf{g}|H_j) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{M}{2}} \prod_{m=1}^{M} \exp\left[-\frac{(g_m - s_{jm})^2}{2\sigma^2}\right] \qquad (13.105)$$

Ideal observer is the Matched Filter:

$$\lambda(\mathbf{g}) = \Delta\mathbf{s}^t\mathbf{g} \underset{D_1}{\overset{D_2}{\gtrless}} \lambda'_c \qquad (13.110)$$

## When the Gaussian noise is correlated

$$\mathrm{pr}(\mathbf{g}|H_j) = \left[(2\pi)^M \det(\mathbf{K_n})\right]^{-\frac{1}{2}} \exp\left[-\tfrac{1}{2}(\mathbf{g}-\mathbf{s}_j)^t\, \mathbf{K_n}^{-1}(\mathbf{g}-\mathbf{s}_j)\right]$$

with $$\mathbf{K}_j = \langle(\mathbf{g}-\mathbf{s}_j)(\mathbf{g}-\mathbf{s}_j)^t|H_j\rangle = \langle\mathbf{n}\mathbf{n}^t\rangle \qquad (13.112)$$

The ideal observer's decision strategy is:

$$\lambda(\mathbf{g}) = \Delta\mathbf{s}^t\mathbf{K_n}^{-1}\mathbf{g} \underset{D_1}{\overset{D_2}{\gtrless}} \lambda'_c \qquad (13.115)$$

Prewhitening Matched Filter:

1) Filter with the inverse of the noise covariance matrix
2) Form scalar product with difference in expected signals

Not a correlation ➔ the observer knows the signal location (no need to scan)

## SNR in Gaussian case

Compute SNR via means and variance of

$$\lambda(\mathbf{g}) = \Delta\mathbf{s}^t\mathbf{K_n}^{-1}\mathbf{g}$$

Means: $$\langle\Delta\mathbf{s}^t\mathbf{K_n}^{-1}\mathbf{g}|H_j\rangle = \Delta\mathbf{s}^t\mathbf{K_n}^{-1}\mathbf{s}_j \qquad (13.116)$$

Variance: $$\sigma_\lambda^2 = \langle\lambda^2(\mathbf{g})|H_j\rangle - \langle\lambda(\mathbf{g})|H_j\rangle^2$$

$$= \Delta\mathbf{s}^t\mathbf{K_n}^{-1}\Delta\mathbf{s} \qquad (13.117)$$

Task SNR:

$$\mathrm{SNR}_\lambda^2 = \Delta\mathbf{s}^t\mathbf{K_n}^{-1}\Delta\mathbf{s} \qquad (13.118)$$

ROC is symmetric; AUC found from inverse error fnct relationship

# Ideal Observer with Poisson data

$$\mathrm{pr}(\mathbf{g}|H_j) = \prod_{m=1}^{M} \frac{e^{-\overline{g}_{jm}}[\overline{g}_{jm}]^{g_m}}{(g_m)!} \qquad (13.130)$$

$$\lambda(\mathbf{g}) = \sum_{m=1}^{M} g_m \ln \frac{\overline{g}_{2m}}{\overline{g}_{1m}} \qquad (13.131)$$

Linear operation on data; filter is nonlinear fnctl of expected data.

Task SNR: $$\mathrm{SNR}^2 = \frac{\left[\sum_{m=1}^{M}(\overline{g}_{2m} - \overline{g}_{1m}) \ln\left(\frac{\overline{g}_{2m}}{\overline{g}_{1m}}\right)\right]^2}{\frac{1}{2}\sum_{m=1}^{M}(\overline{g}_{2m} + \overline{g}_{1m}) \ln^2\left(\frac{\overline{g}_{2m}}{\overline{g}_{1m}}\right)} \qquad (13.132)$$

# Random Signals

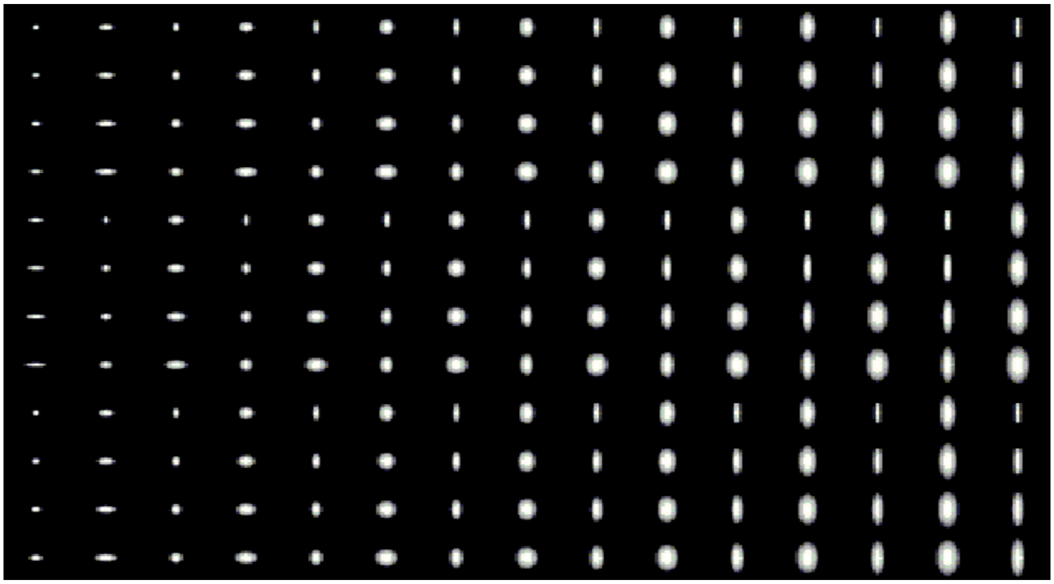
- Signal Known Statistically

$$H_1 : \mathbf{g} = \mathcal{H}\mathbf{f}_1 + \mathbf{n} = \mathbf{b} + \mathbf{n}$$
$$H_2 : \mathbf{g} = \mathbf{s}(\boldsymbol{\theta}) + \mathbf{b} + \mathbf{n}$$

$$\Lambda(\mathbf{g}) = \frac{\mathrm{pr}(\mathbf{g}|H_2)}{\mathrm{pr}(\mathbf{g}|H_1)} = \frac{\langle \mathrm{pr}(\mathbf{g}|H_2, \boldsymbol{\theta})\rangle_{\boldsymbol{\theta}}}{\mathrm{pr}(\mathbf{g}|H_1)}$$

$$= \left\langle \frac{\mathrm{pr}(\mathbf{g}|H_2, \boldsymbol{\theta})}{\mathrm{pr}(\mathbf{g}|H_1)} \right\rangle_{\boldsymbol{\theta}} = \langle \Lambda_{\mathrm{SKE}}(\mathbf{g}, \boldsymbol{\theta})\rangle_{\boldsymbol{\theta}} \qquad (13.148)$$

# Set of possible signals



Zhang, Pham, Eckstein, MIPCX, Sept. 2003

# Example: Location uncertainty

Signal's contribution to *m*<sup>th</sup> data element:

$$s_m(\mathbf{r_s}) = \int_m d^2r \; s(\mathbf{r} - \mathbf{r_s}) \qquad (13.156)$$

For uncorrelated Gaussian noise:

$$\mathrm{pr}(\mathbf{g}|H_2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{M}{2}} \int_\infty d^2r_\mathbf{s} \; \mathrm{pr}_\mathbf{r}(\mathbf{r_s}) \exp\left\{-\frac{1}{2\sigma^2}\sum_{m=1}^{M}[g_m - s_m(\mathbf{r_s}) - b_m]^2\right\} \qquad (13.157)$$

$$\boxed{\Lambda(\mathbf{g}) \propto \int_\infty d^2r_\mathbf{s} \; \mathrm{pr}_\mathbf{r}(\mathbf{r_s}) \exp\left\{\frac{1}{\sigma^2}\sum_{m=1}^{M}[g_m - b_m]s_m(\mathbf{r_s})\right\}} \qquad (13.158)$$

1) Subtract known background at each pixel;
2) Multiply by expected signal for that location;
3) Exponentiate;
4) Repeat for all signal locations; average using prior prob of signal location.

# Random Backgrounds

$$\mathrm{pr}(\mathbf{g}|H_j) = \int_\infty d^M b \; \mathrm{pr}(\mathbf{g}|H_j, \mathbf{b}) \, \mathrm{pr}_\mathbf{b}(\mathbf{b}) \qquad (13.164)$$

$$\Lambda(\mathbf{g}) = \frac{\int_\infty d^M b \; \mathrm{pr}(\mathbf{g}|H_2, \mathbf{b}) \, \mathrm{pr}_\mathbf{b}(\mathbf{b})}{\int_\infty d^M b' \; \mathrm{pr}(\mathbf{g}|H_1, \mathbf{b}') \, \mathrm{pr}_\mathbf{b}(\mathbf{b}')} = \frac{\langle \mathrm{pr}(\mathbf{g}|H_2, \mathbf{b}) \rangle_\mathbf{b}}{\langle \mathrm{pr}(\mathbf{g}|H_1, \mathbf{b}) \rangle_\mathbf{b}} \qquad (13.166)$$

It can be shown that:

$$\boxed{\Lambda(\mathbf{g}) = \langle \Lambda_{\mathrm{BKE}}(\mathbf{g}, \mathbf{b}) \rangle_{\mathbf{b}|\mathbf{g}, H_1}} \qquad (13.169)$$

where: $$\Lambda_{\mathrm{BKE}}(\mathbf{g}, \mathbf{b}) \equiv \frac{\mathrm{pr}(\mathbf{g}|H_2, \mathbf{b})}{\mathrm{pr}(\mathbf{g}|H_1, \mathbf{b})} \qquad (13.170)$$

# Ideal Observer: Summary

- Computes likelihood ratio or its log
- Optimal according to numerous criteria
- Linear for Gaussian data
- (Usually) nonlinear when data are non-Gaussian or there are other sources of variability
  – Random objects
  – Random backgrounds
- Can be difficult to compute
  – But tricks available – see this p.m.'s lecture!

# Linear observers

- Fallback when Ideal Observer is not tractable
- Takes general form $t = \mathbf{w}^t \mathbf{g}$
- What is optimal $\mathbf{w}$?

➔ Found by maximization of $\mathrm{SNR}_t$

# The Hotelling observer

Harold Hotelling

- Optimum linear discriminant

$$\mathbf{w}_{opt\ lin} = \mathbf{K}_{\mathbf{g}}^{-1} \Delta \overline{\mathbf{g}} \qquad (13.177)$$

- Based on 1931 paper by Harold Hotelling
- Requires knowledge of image means and covariance
- Equivalent to ideal observer for Gaussian data

$$\mathrm{SNR}^2_{opt\ lin} = \Delta \overline{\mathbf{g}}^t \mathbf{K}_{\mathbf{g}}^{-1} \Delta \overline{\mathbf{g}} \qquad (13.178)$$

## Random Signals and Backgrounds

$$\mathbf{w} = \mathbf{K_g}^{-1} \langle \Delta \mathbf{g} \rangle_{\mathbf{n,b,s}} \;,$$

where $\mathbf{K_g}$ describes all sources of variability in the data.

$$\mathrm{SNR}^2 = \langle \Delta \mathbf{g} \rangle^t \, \mathbf{K_g}^{-1} \langle \Delta \mathbf{g} \rangle$$

Detectability map: signal known exactly, but variable

$$\mathbf{w}(\theta) = \mathbf{K_g}^{-1}(\theta)\, \mathbf{s}(\theta)$$

$$\mathrm{SNR}^2(\theta) = \mathbf{w}(\theta)^t\, \mathbf{s}(\theta)$$

Computational difficulty: inversion of large covariance matrix (but many tricks available – see this p.m.)

## Observer Efficiency:

$$\eta_{Hot} = \frac{d^2_{A,Hot}}{d^2_{A,ideal}} \qquad (13.273)$$

- Measures efficiency of Hotelling observer relative to ideal observer for same task
- 1 for SKE/BKE tasks in Gaussian noise

## Device Efficiency:

$$\eta_{system} = \frac{d^2_{A,system}}{d^2_{A,ideal}} \qquad (13.246)$$

- Measures information transfer through device or processing algorithm

## Why use linear (suboptimal) observer?

- Ideal observer may be intractable
- To evaluate algorithms
  - Image processing, image enhancement
  - Image reconstruction algorithms
  - Ideal observer is invariant to these!
- To predict human performance

## Task-based assessment via human observers

- Requires psychophysical studies
  - Expensive (observers and cases)
  - Time consuming
    - How to evaluate the many “knobs” of an imaging technology under development?
    - What about enhancements to existing technology?
- Models that predict human performance would enable system design without (or with fewer) such studies

## Models for prediction of human performance

- Humans have been shown to do nonlinear processing poorly; therefore restrict models to linear discriminants
- The visual cortex receives data processed through "channels" or receptive fields
- Random neuronal firing and fluctuating decision thresholds modeled as "internal noise"
- Some include optical aberrations of the eye or contrast sensitivity function ("eye filter")

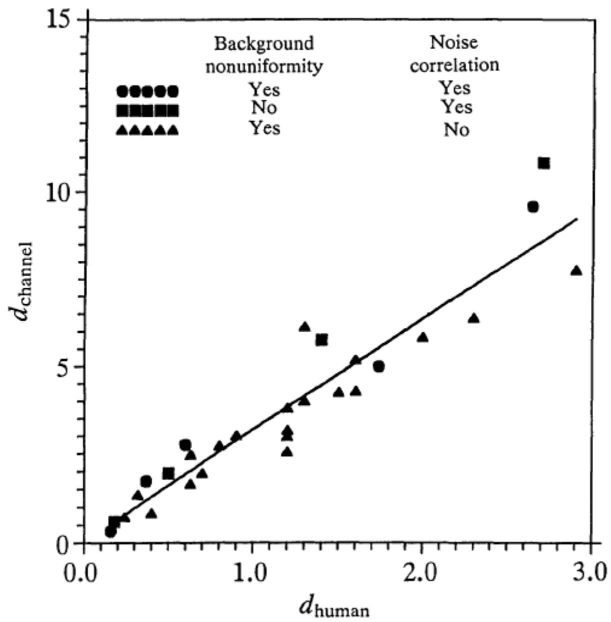
## Channelized linear observer

- Processes data through $P$ channels $\mathbf{u}_p$
- Channel outputs: $v_p = \mathbf{u}_p{}^t \, \mathbf{g}$ = a scalar
- Vector of channel outputs: $\mathbf{v} = \mathbf{U}^t \mathbf{g}$

$$\mathrm{SNR}_{\mathbf{v}}^2 = \Delta\overline{\mathbf{v}}^t \mathbf{K}_{\mathbf{v}}^{-1} \Delta\overline{\mathbf{v}}$$

- $\mathbf{K}_\mathbf{v}$ is $P$ x $P$, where $P$ is small (less than 10)

## Human vs. Channelized Hotelling Observer Detectability

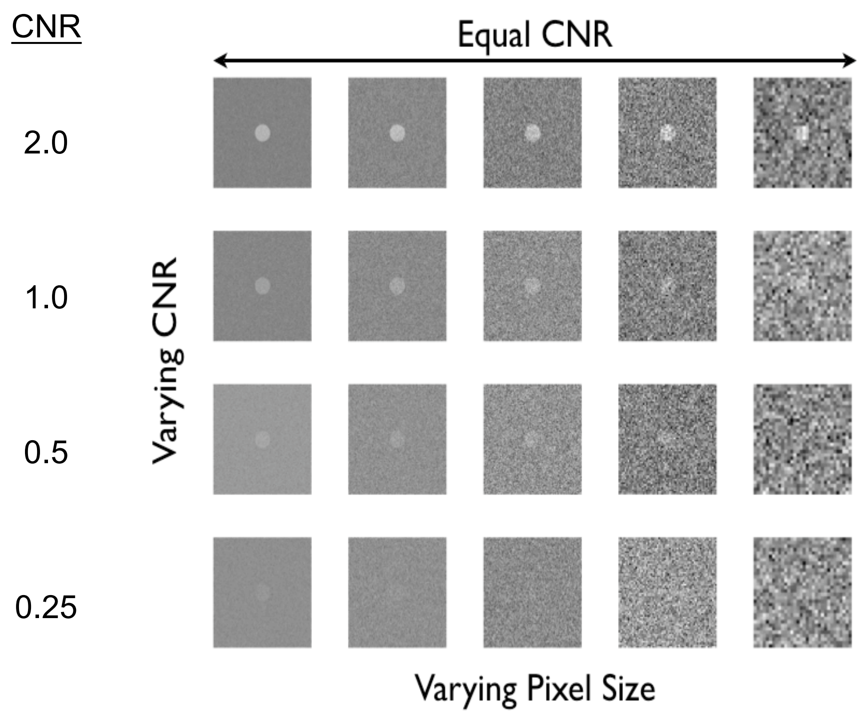



Humans track Channelized Hotelling Observer across range of experiments with different background and noise conditions

H. H. Barrett *et al.*,
Proc. Nat. Acad. Sci., 1993

# Why not Contrast-to-Noise Ratio?

- Contrast-to-noise ratio depends on
  - Contrast
  - Noise
- Detectability depends on:
  - Lesion size and tracer distribution
  - Variance and covariance of measurement noise
  - Background level, variability and correlation
  - Intrinsic spatial resolution, shape of PSF
  - Pixel size
  - Detection threshold (desired false-alarm rate)
  - Detection strategy used (optimal vs. suboptimal)




Courtesy Matt Kupinski

# Conclusions: Model observers for assessment of image quality for classification tasks

- Ideal Observer if possible
  - Captures all information in the data
  - Adds no noise or uncertainty
- Hotelling (Ideal Linear) Observer as next best thing
- Optimal linear discriminant after processing data through channels (channelized Hotelling observer)
  - Useful predictor of human performance
  - Facilitates computation (more this afternoon…)