SAG19: Signal Detection Theory and Rigorous Performance Metrics for Exoplanet Imaging

Chairs: Dimitri Mawet and Rebecca Jensen-Clem (Caltech)

Team members: Olivier Absil (ULg), Ruslan Belikov (NASA AMES), Steve Bryson (NASA AMES), Elodie Choquet (JPL), Brendan Crill (JPL), Thayne Currie (Subaru), Tiffany Glassman (Northrop), Carlos Gomez (ULg), M. Kenworthy (Leiden), John Krist (JPL), Christian Marois (NRC), Johan Mazoyer (STScI), Tiffany Meshkat (JPL), T.J. Rodigas (Carnegie DTM), Garreth Ruane (Caltech), Jean-Baptiste Ruffio (Stanford), Angelle Tanner (MSU), John Trauger (JPL), Maggie Turnbull (SETI)

Keck Institute for Space Studies: Exoplanet Imaging and Characterization Workshop

SAG19: exoplanet imaging signal detection theory and rigorous contrast metrics

Chairs: Dimitri Mawet (Caltech) and Rebecca Jensen-Clem (Caltech)

Team members: John Krist (JPL), Ruslan Belikov (NASA AMES), Steve Bryson (NASA AMES), Jean-Baptiste Ruffio (Stanford), Johan Mazoyer (STScI), Brendan Crill (JPL), Tiffany Glassman (Northrop), Maggie Turnbull (SETI), Angelle Tanner (MSU), T.J. Rodigas (Carnegie DTM), M. Kenworthy (Leiden), Thayne Currie (Subaru), Garreth Ruane (CIT), John Trauger (JPL), Carlos Gomez (ULg), Elodie Choquet (JPL)

As planning for the next generation of high contrast imaging instruments (e.g. WFIRST, HabEx, and LUVOIR, TMT-PFI, EELT-EPICS) matures, and second-generation ground-based extreme adaptive optics facilities (e.g. VLT-SPHERE, Gemini-GPI) are halfway through their large main surveys, it is imperative that the performance of different designs, post-processing routines, observing strategies, and survey results be compared in a consistent, statistically robust framework. SAG19, exoplanet imaging signal detection theory and rigorous contrast metrics, is overarching to all direct imaging instrument, strategies, and methods. The scope of SAG19 is:

1- To go back to the basics of Bayesian Signal Detection Theory (SDT).
Bayesian SDT implies H0:signal absent / H1:signal present hypothesis testing, and invokes well-known concepts such as: the confusion/contingency matrix, false positive (type I error), false negative (type II error), true positive, and true negative fractions, and useful combinations of these quantities such as sensitivity (or completeness) and specificity.

2- To rebuild a solid set of usual definitions used for or in lieu of "contrast" in different contexts, such as astrophysical contrast or ground truth, instrumental contrast used for coronagraph/instrument designs, and the measured on-sky data-driven contrast.

Bayesian, hypothesis testing SDT will automatically force our community to be inclusive of all possible aspects of exoplanet detection, and signal-to-noise ratio (SNR) metrics, including signal-related parameters: planet-star contrast, SED, polarization, variability; instrument parameters: throughput, bandwidth, Strehl ratio/encircled energy, background (sky/thermal, or astrophysical), detector characteristics; noise characteristics as affected by the starlight suppression technique (in a very broad sense): mean intensity, RMS pixel intensities, RMS resolution element (resel, of characteristic size wavelength/telescope diameter) intensities, the probability density function (PDF) computed over pixels, the same PDF computed over resels, their nature and higher order moments, the sample zone and size, outlier management, etc.

3- To identify what we can learn and apply from communities outside our field (e.g. medical imaging). A good example is the widespread use of receiver operating characteristic curve (ROC) and area under the curve (AUC). ROC plots the true positive fraction against the false positive fraction, and is useful to capture the true performance of a given high contrast imaging

SAG19 Motivation and Goals

- The term “contrast” falls short as a general purpose performance metric
- SAG19 Goal #1: Create a unifying figure of merit that can represent the performance of direct imaging testbeds, ground-based observations, space-based observations, post-processing algorithms, internal and external occultors, and surveys
- SAG19 Goal #2: Provide user-friendly code to the community for generating this new figure of merit
- SAG19 Goal #3: Provide a standard dataset for the consistent comparison of new post-processing algorithms

What is the definition of contrast?

- We can all agree that “contrast” has something to do with a planet-to-star brightness ratio
- However, we disagree on the computation of the contrast. Specifically, whether and how to include:
  - Telescope throughput
  - Coronagraph throughput
  - Post processing algorithm throughput
  - Speckle noise
  - Small sample statistics

What’s wrong with our contrast definitions?

- Example #1: a low-throughput testbed
  - We define a raw contrast that’s related to the variance of the counts inside of a half dark hole and the throughput of the coronagraph
  - If we only report the contrast values, and not the details of the calculation, readers may not understand which throughput issues were taken into account (e.g., field and time dependent corrections)
  - Dividing the noise by the throughput artificially increases the number of false positives

What's wrong with our contrast definitions?

- Example #2: an optimistic survey
  - We define a post-processed contrast that's based on the variance of counts at a given separation and the throughput losses from KLIP.
  - We decide to follow up any blobs that exceed a detection threshold of 3σ, to set the false positive fraction to 0.1% assuming white noise
  - Because we didn't have complete knowledge of the noise distribution, we spend much more than 0.1% of the survey time on false positive follow-up

What do we want from a high contrast imaging performance metric?

- We want a noise measurement that is sensitive to both poisson and speckle noise.
- This is challenging due to:
  - Limited sampling
  - Unknown statistics (not Gaussian!)
- Such a noise measurement would enable:
  - Follow-up decision/risk mitigation
  - Management of telescope time/resources

What do we want from a high contrast imaging performance metric?

- We want to measure the fraction of real planets that a system will correctly identify (i.e. the completeness).
- Such a measurement will allow us to:
  - Set more meaningful detection thresholds
  - Place statistically significant constraints on exoplanet population demographics

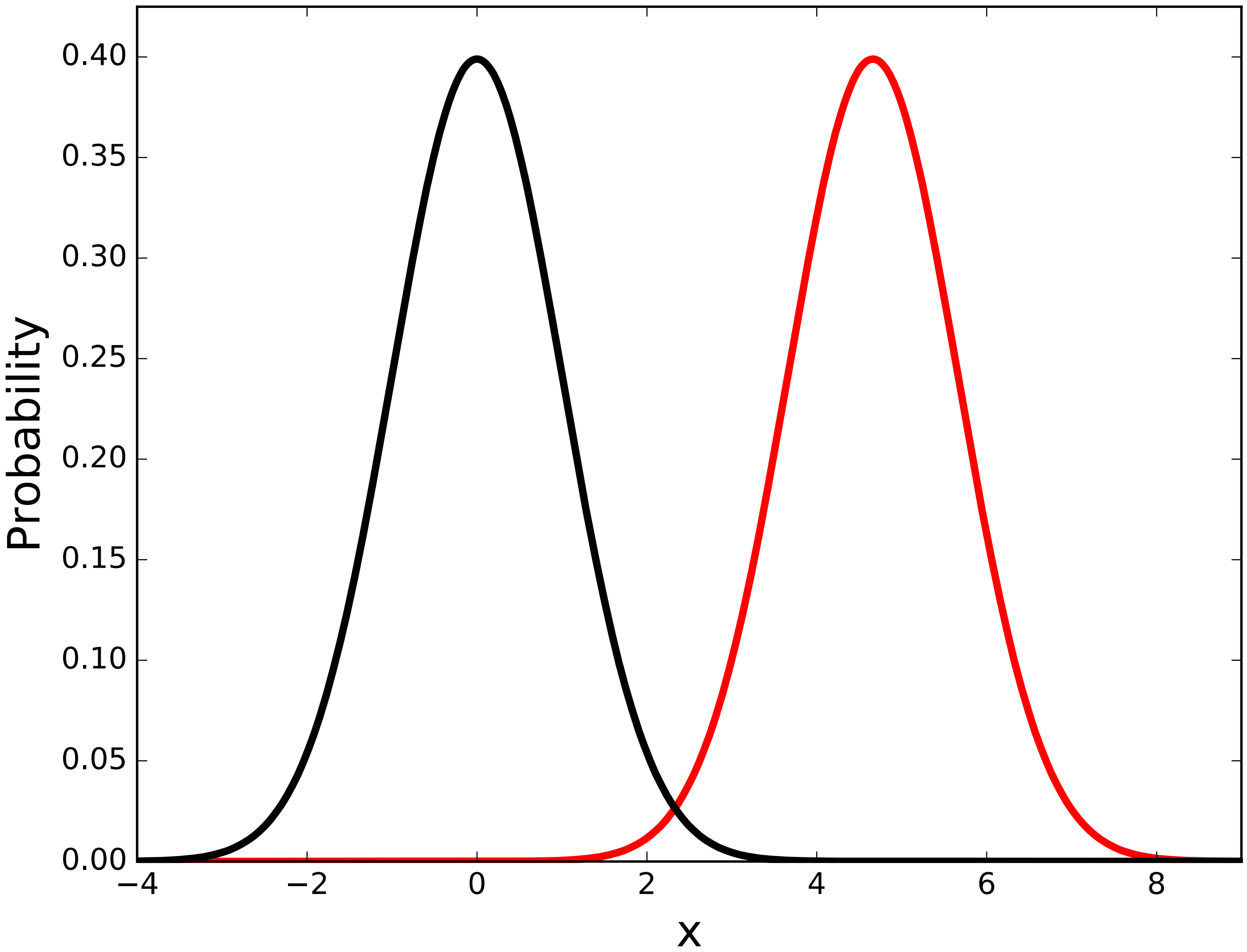
The definition of *contrast* is corrupted

- It currently has too many different meanings in different contexts (simulation, lab tests, telescope data)
- The only true, unambiguous definition of *contrast* is astrophysical planet/star brightness ratio
- *Raw irradiance (flux) and noise variance do not fully describe relevant signal detection metrics such as FPF and TPF*
- The goal of SAG19 is to identify and supply code for an alternative performance metric

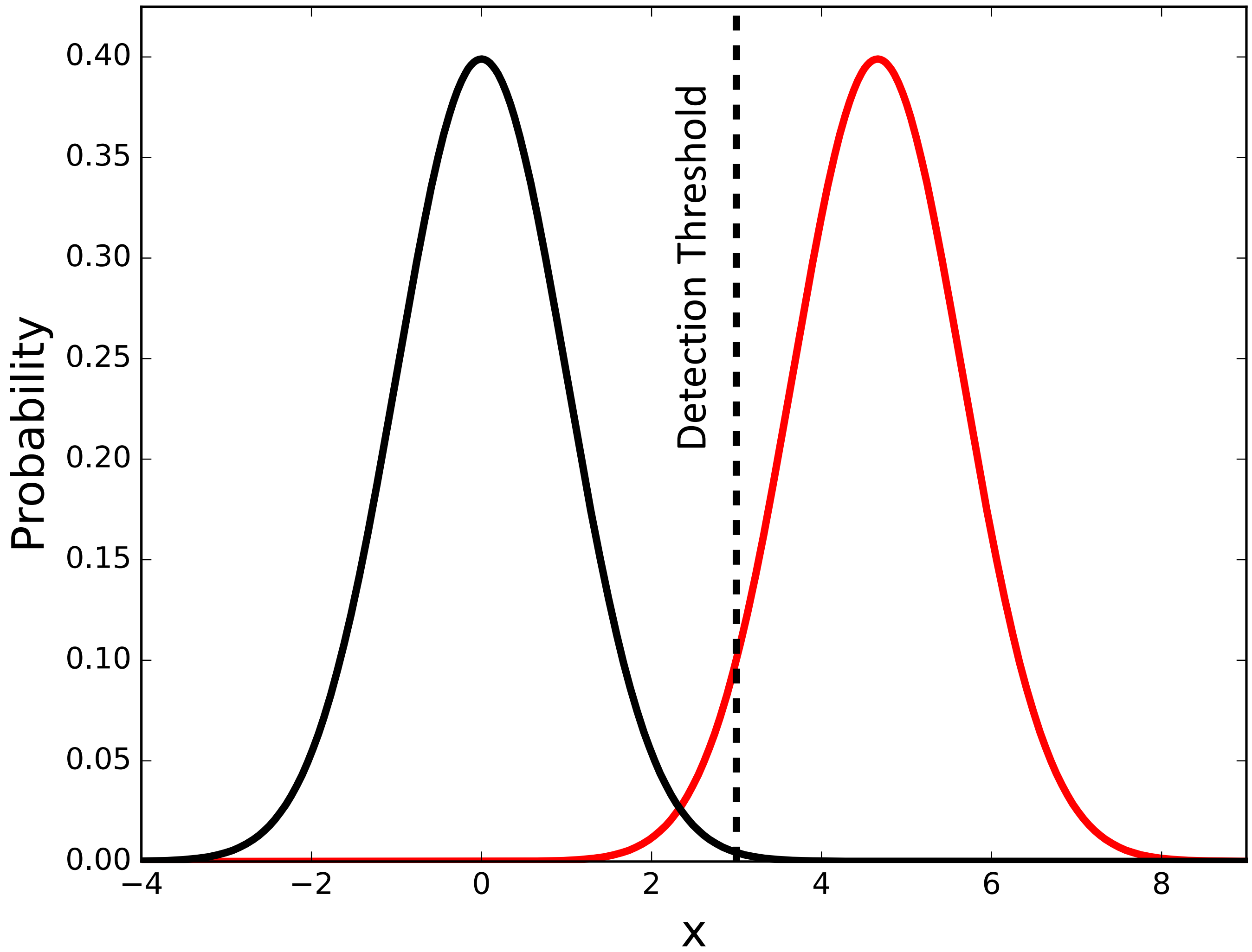
The Confusion Matrix

	H_1: Signal Present	H_0: Signal Absent
Detection	True Positive	False Positive
Null Result	False Negative	True Negative
	True Positive Fraction = TP/(TP+FN)	False Positive Fraction = FP/(FP+TN)

The Noise and the Signal



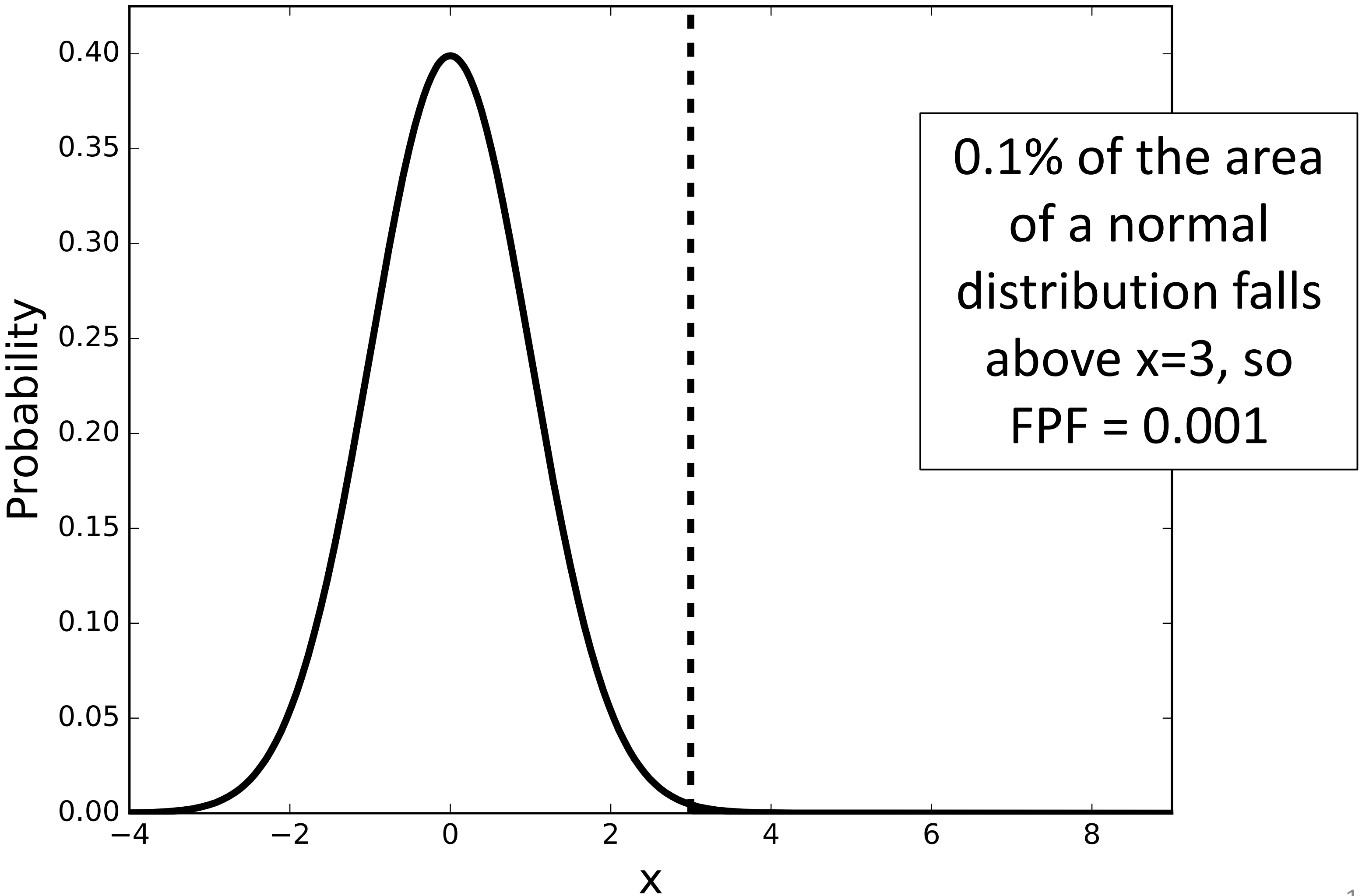
The Noise and the Signal



The False Positive Fraction (FPF)

		H_0: Signal Absent
Detection		False Positive
Null Result		True Negative
		FPF = FP/(FP+TN)

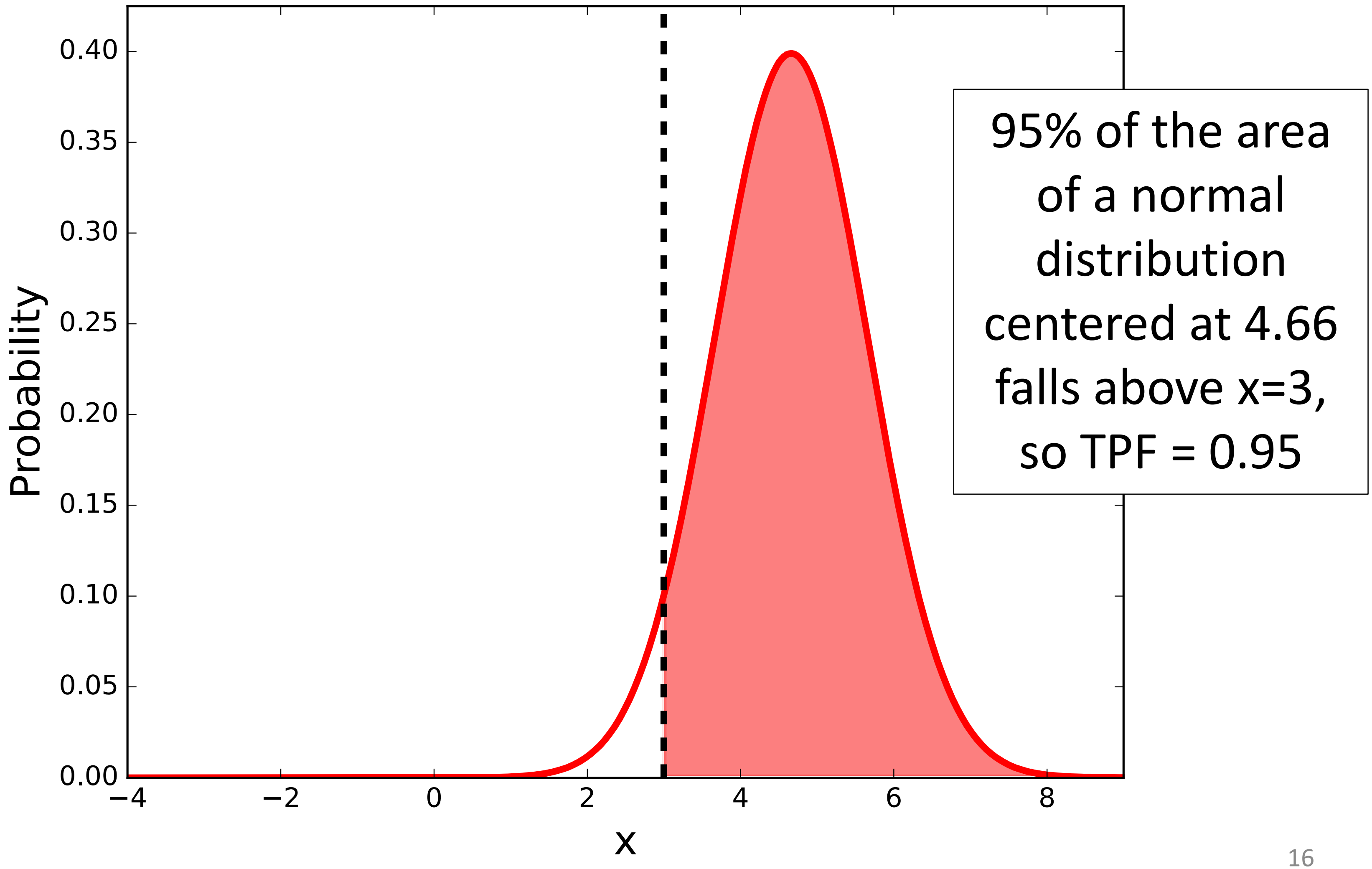
The False Positive Fraction (FPF)



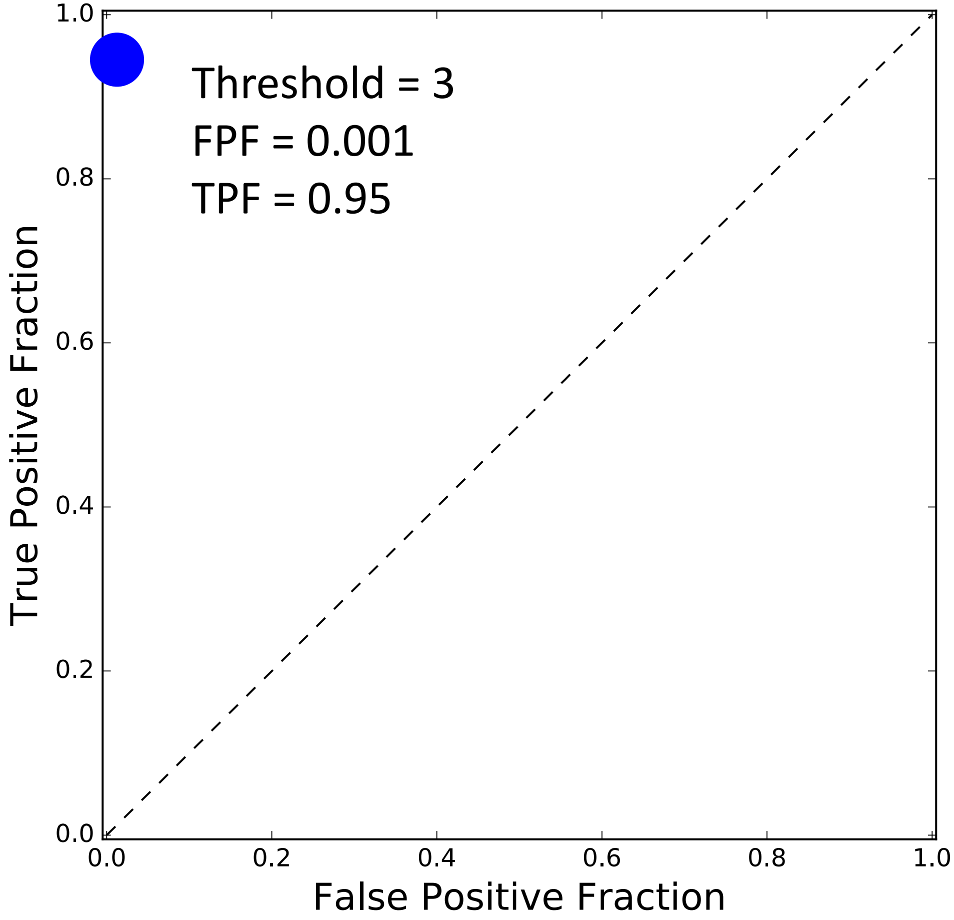
The True Positive Fraction (TPF)

	H_1: Signal Present	
Detection	True Positive	
Null Result	False Negative	
	TPF = TP/(TP+FN)	

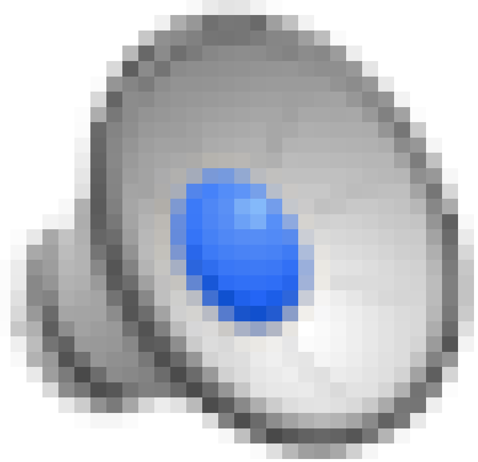
The True Positive Fraction (TPF)



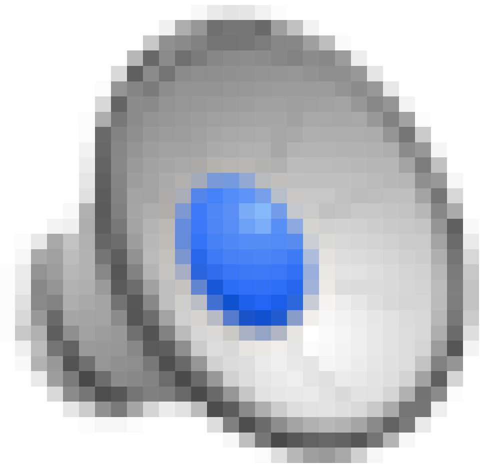
The Receiver Operator Characteristic (ROC) Curve



The Receiver Operator Characteristic (ROC) Curve

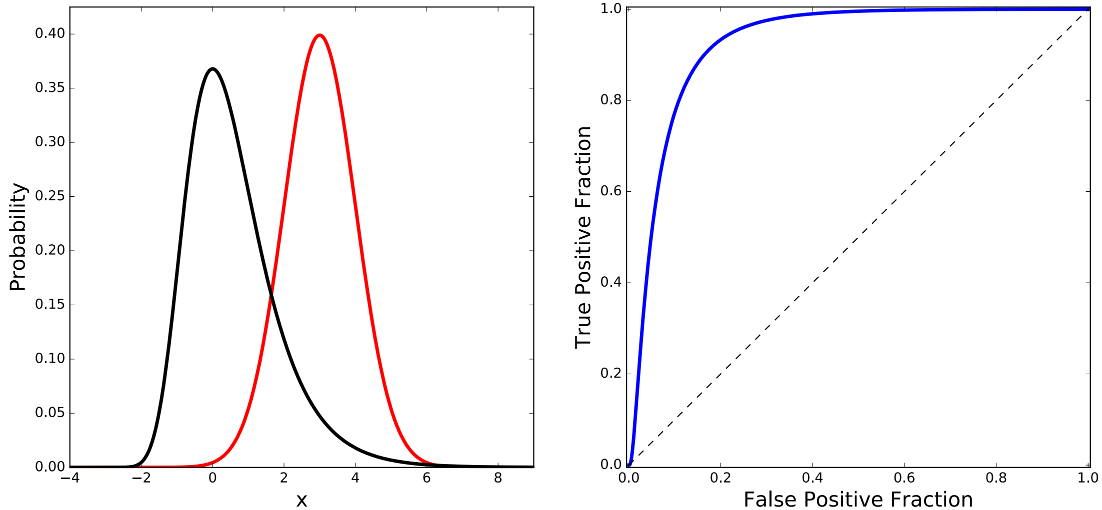


The Receiver Operator Characteristic (ROC) Curve

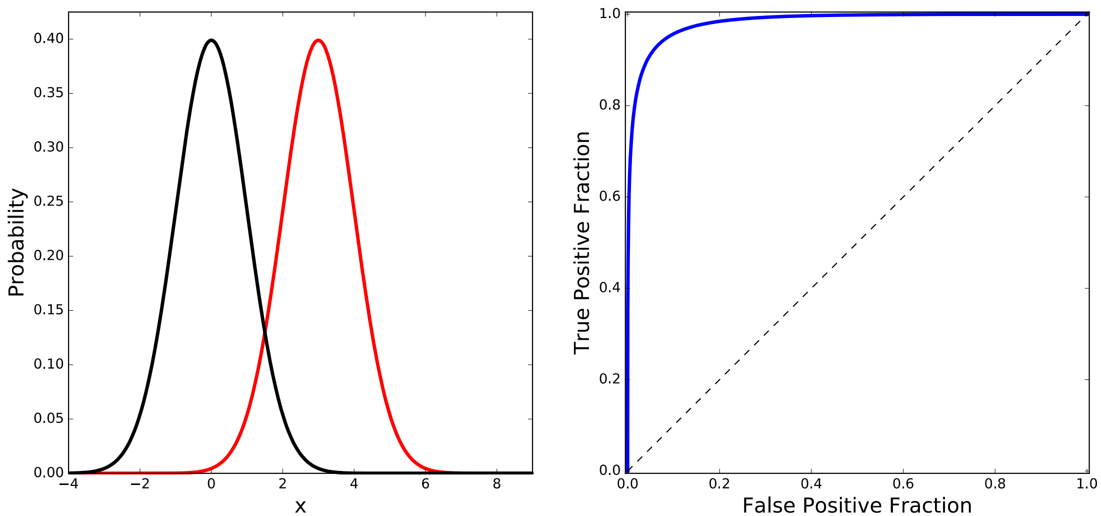


The Receiver Operator Characteristic (ROC) Curve

Fat-tailed noise distribution



Whitened noise distribution

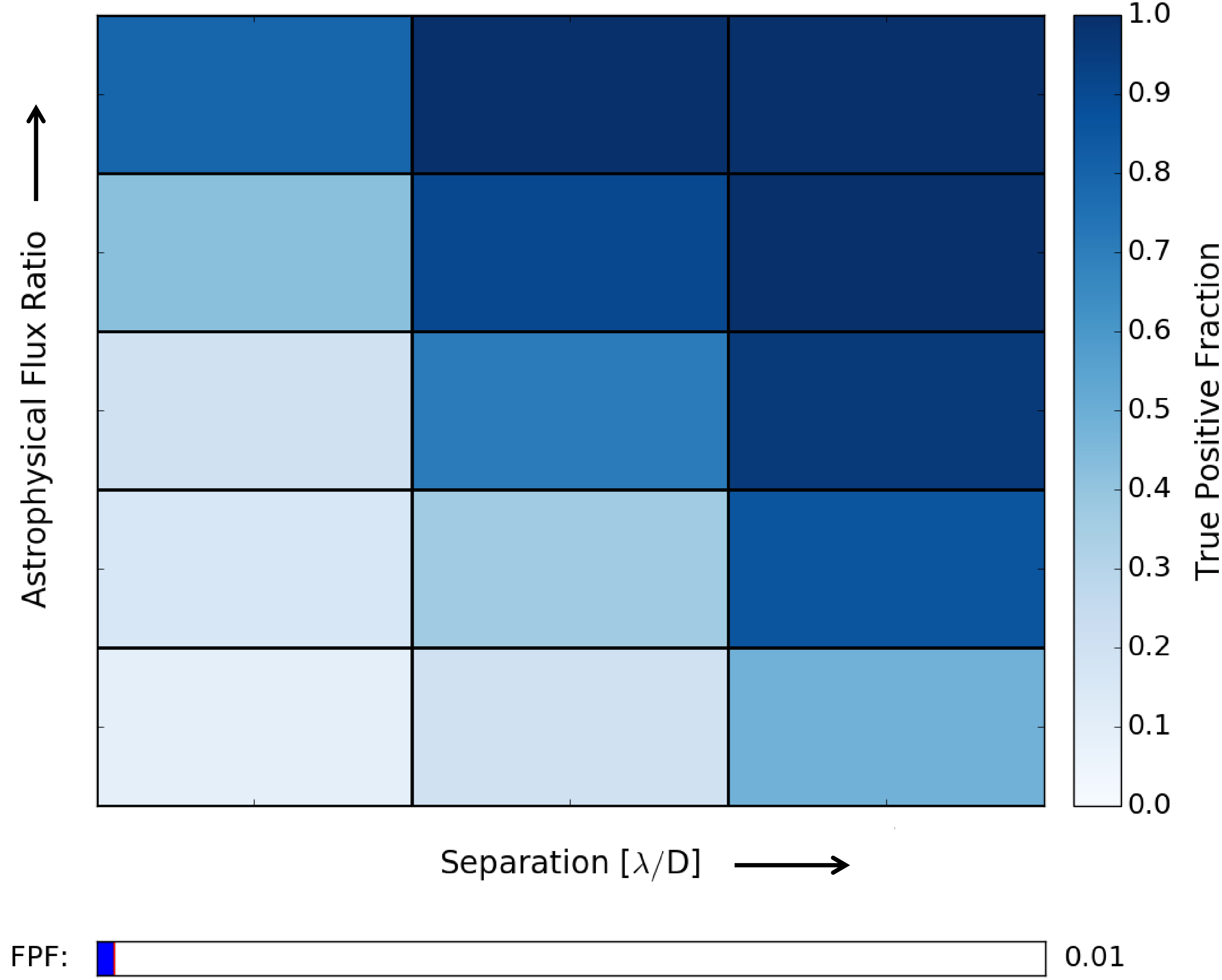


How can we fully capture a system's performance with ROC Curves?

- Regions near the star have different noise properties than regions far from the star
  - If the PSF is symmetric, each annulus requires a new FPF computation (otherwise, more complicated)

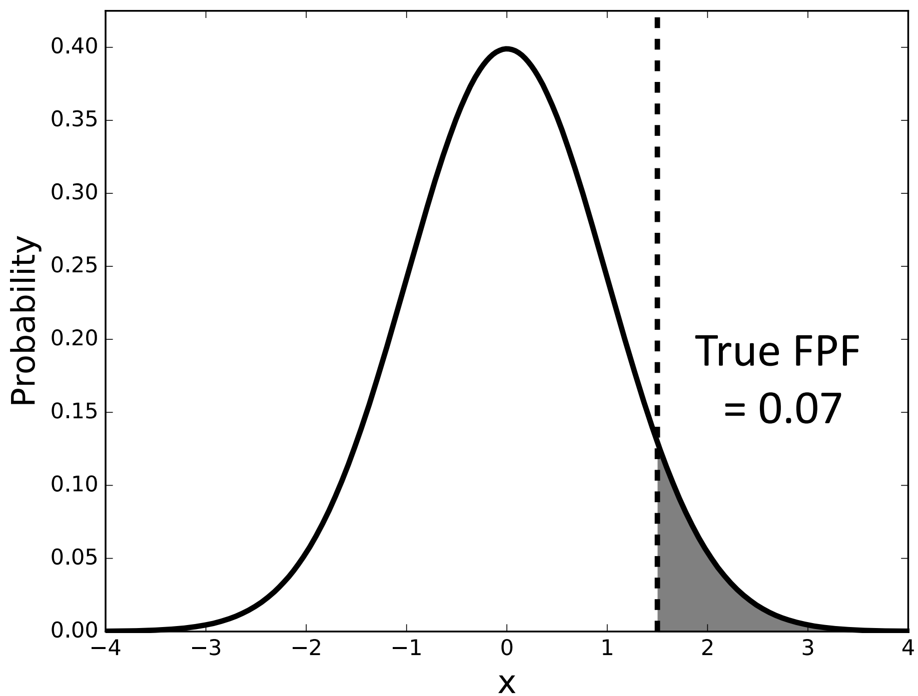
- Planets with different astrophysical flux ratios and separations require new TPF computations

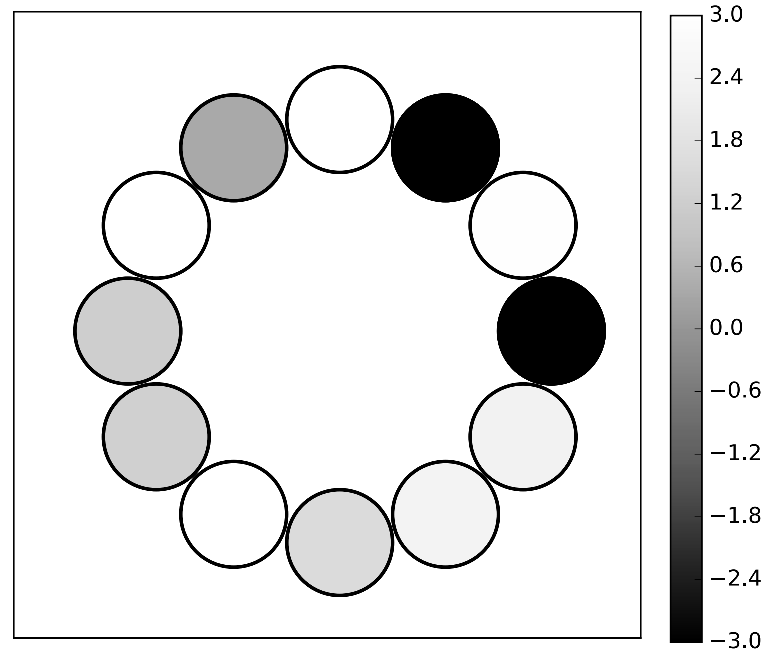
- So, we need to make one ROC curve per separation per injected planet flux ratio. How do we represent all of that information in one figure?


Astrophysical Flux Ratio
Separation [λ/D]
True Positive Fraction
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
FPF:
0.01

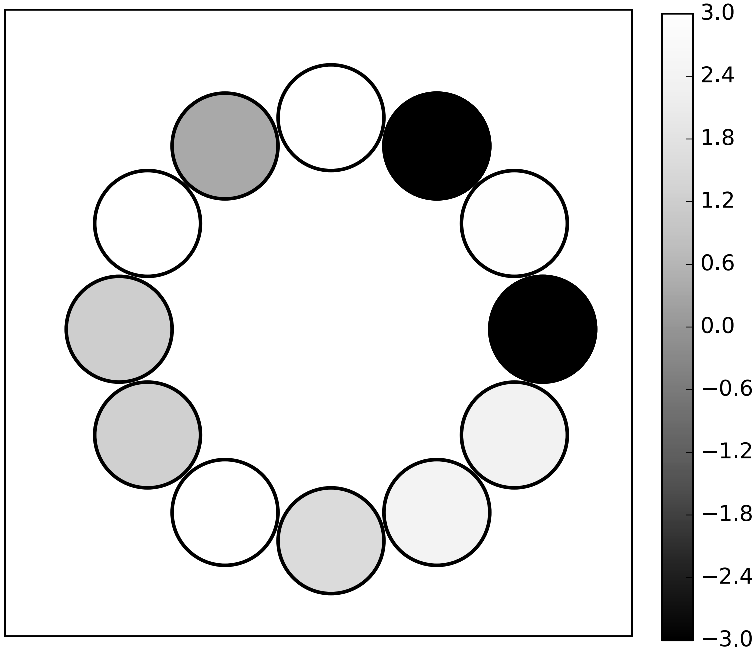
How to Build a ROC Curve: the False Positive Fraction

- Pick a separation (e.g. 2λ/D) and a detection threshold (e.g. 1.5)
- Under the assumption that there is no signal, the fraction of resolution elements falling above the threshold is the FPF
- Try a range of thresholds such that the FPF varies between 0 and 1





How to Build a ROC Curve: the False Positive Fraction



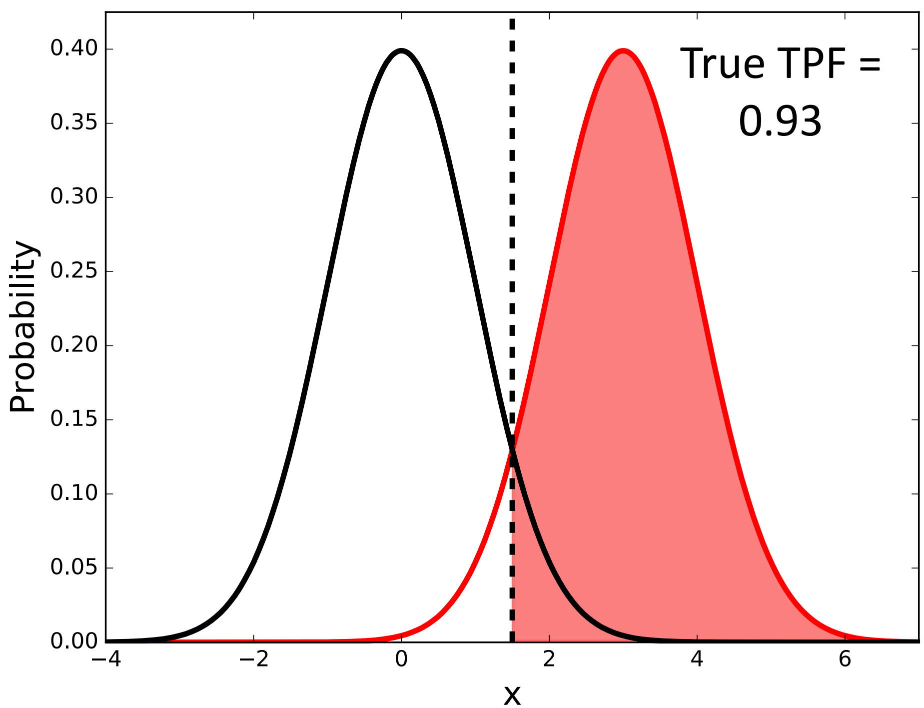
True FPF = 0.07

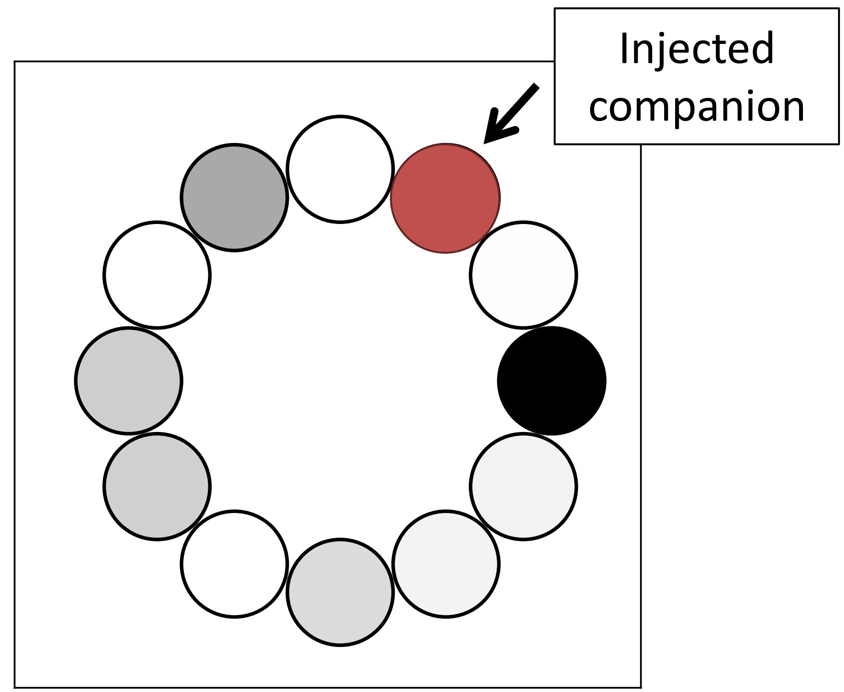
Empirical FPF = 1/12 = 0.08

Res. Elem.	Counts	Meet threshold of counts ≥ 1.5?
1	-0.11	N
2	0.44	N
3	-0.35	N
4	0.3	N
5	0.29	N
6	-2.68	N
7	0.24	N
8	0.1	N
9	0.11	N
10	1.31	N
11	0.01	N
12	1.61	Y

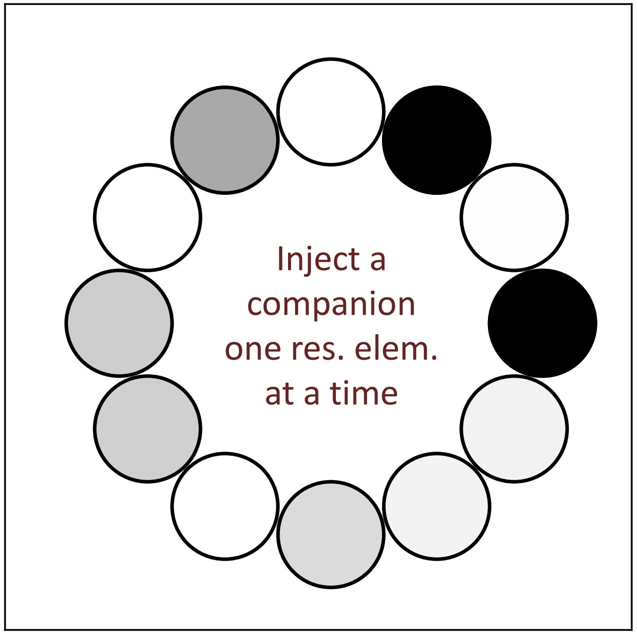
How to Build a ROC Curve: the True Positive Fraction

- Consider the same separation and set of thresholds as in the FPF computations
- Put a fake planet onto each resolution element
- The fraction of resolution elements exceeding the threshold is the true positive fraction





How to Build a ROC Curve: the True Positive Fraction

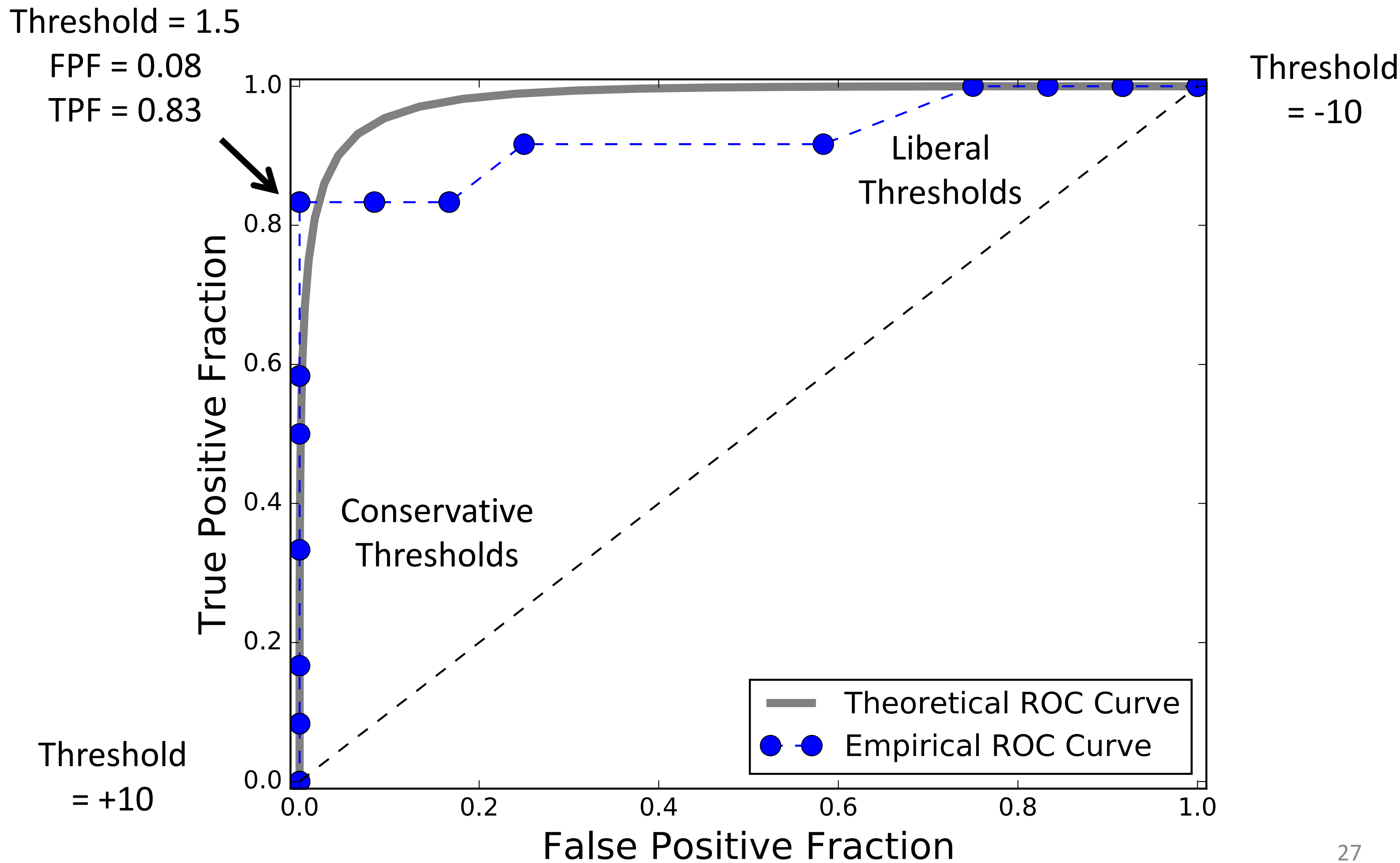


True TPF = 0.93

Empirical TPF = 10/12 = 0.83

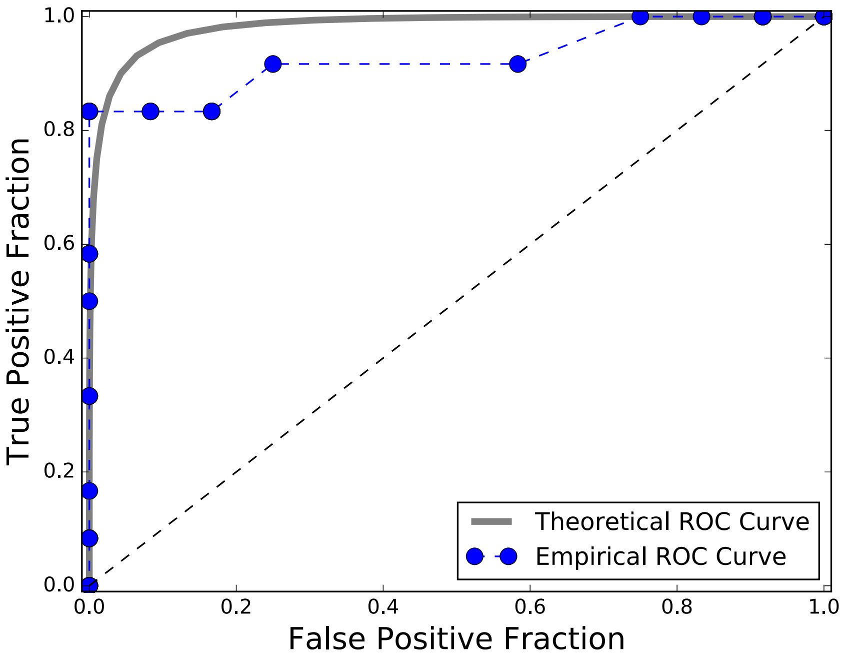
Res. Elem.	Counts	Meet threshold of counts ≥ 1.5?
1	2.76	Y
2	3.39	Y
3	2.74	Y
4	2.8	Y
5	-0.1	N
6	0.48	N
7	3.6	Y
8	4.19	Y
9	3.95	Y
10	4.02	Y
11	3.63	Y
12	5.79	Y

How to Build a ROC Curve: FPF and TPF pairs for a range of thresholds



The Effects of Small Sample Statistics

- Our empirical ROC curve is different from the theoretical ROC curve because we only have a small number of realizations of the noise
- This is especially problematic at small separations:
  - 6 res. elems. at 1 λ/D
  - 12 res. elems. at 2 λ/D



- Here, we can't access the traditional gaussian false positive fractions corresponding to 3σ (FPF = 0.001) and 5σ (FPF = 3×10^{-7})
- However, our performance metrics are now based on the data
- We may not require such small FPF for such small samples

Possibilities for increasing the number of samples at small separations

- Different resampling methods may allow us to access smaller FPFs
  - Angular Differential Imaging (ADI) datasets:
    - Create more “final images” by randomly combining shorter exposures of different parallactic angles within a set of constraints
  - Non-ADI datasets (e.g. WFIRST):
    - Create more “final images” by randomly combining sets of shorter exposures
- Ultimately, however, we may not be able to access all FPFs for all separations

Data challenge

- Ground-based telescopes:
  - SPHERE/GPI/Keck data sets? Injection of false planets.
- Space-based telescopes:
  - WFIRST simulations and lab tests
- The data must be carefully labeled (with ground-truth) and cover a wide range of scenarios (observing techniques, instruments, observation parameters, noise levels, etc)
- New proposed algorithms should use the standard data sets and the metrics accepted by the community when assessing their performance

Continuing Work

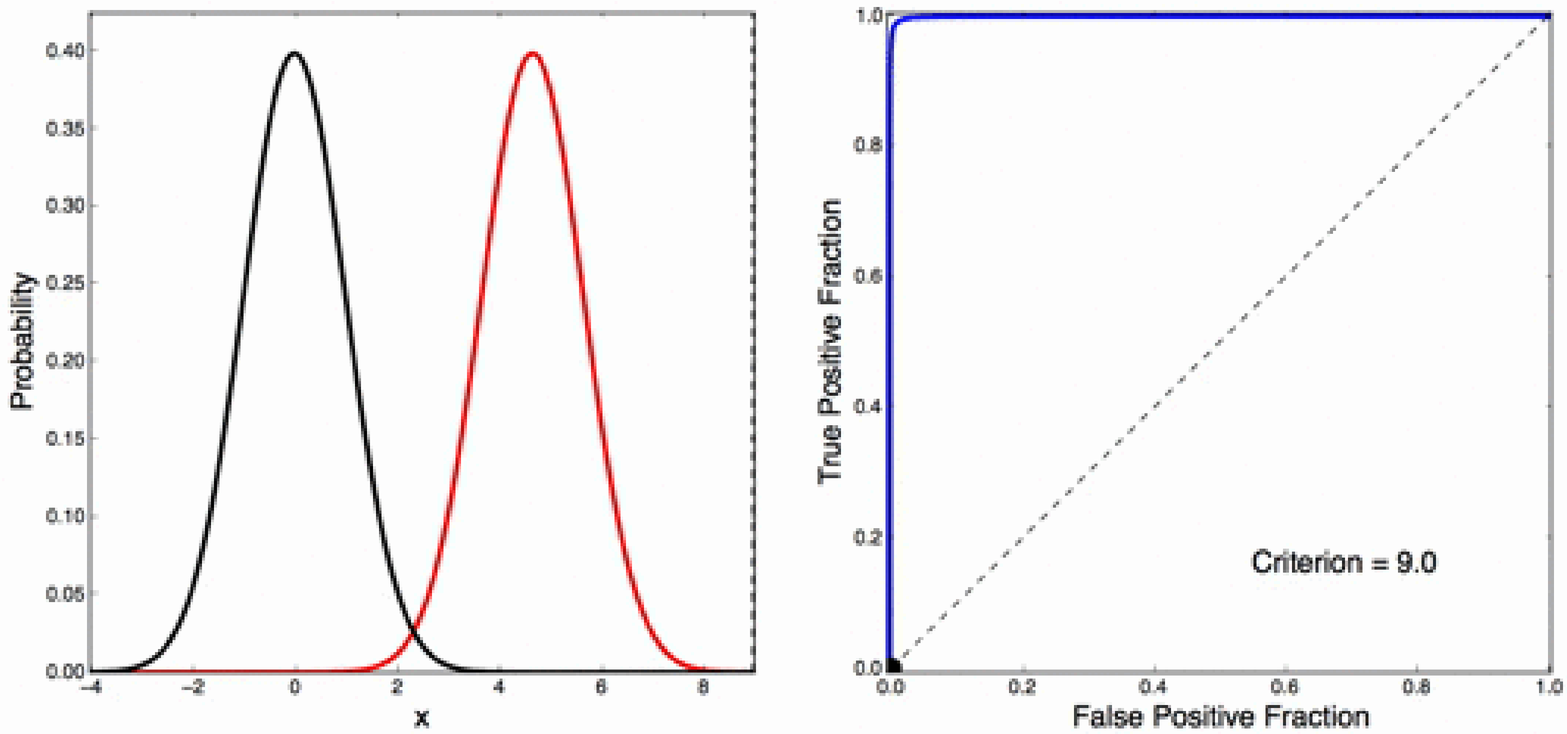
- Simulate bootstrapping methods for increasing the number of samples at small separations
- Investigate the prospects for modeling the noise distributions at small separations in post-processed images.
- Understand the procedures for calculating the false positive fraction in the presence of a suspected planet
- Provide a codebase for generating ROC curves in different scenarios: lab testbeds, ground-based ADI datasets vs. space-based static datasets, single observations vs. multi-target surveys).

Summary

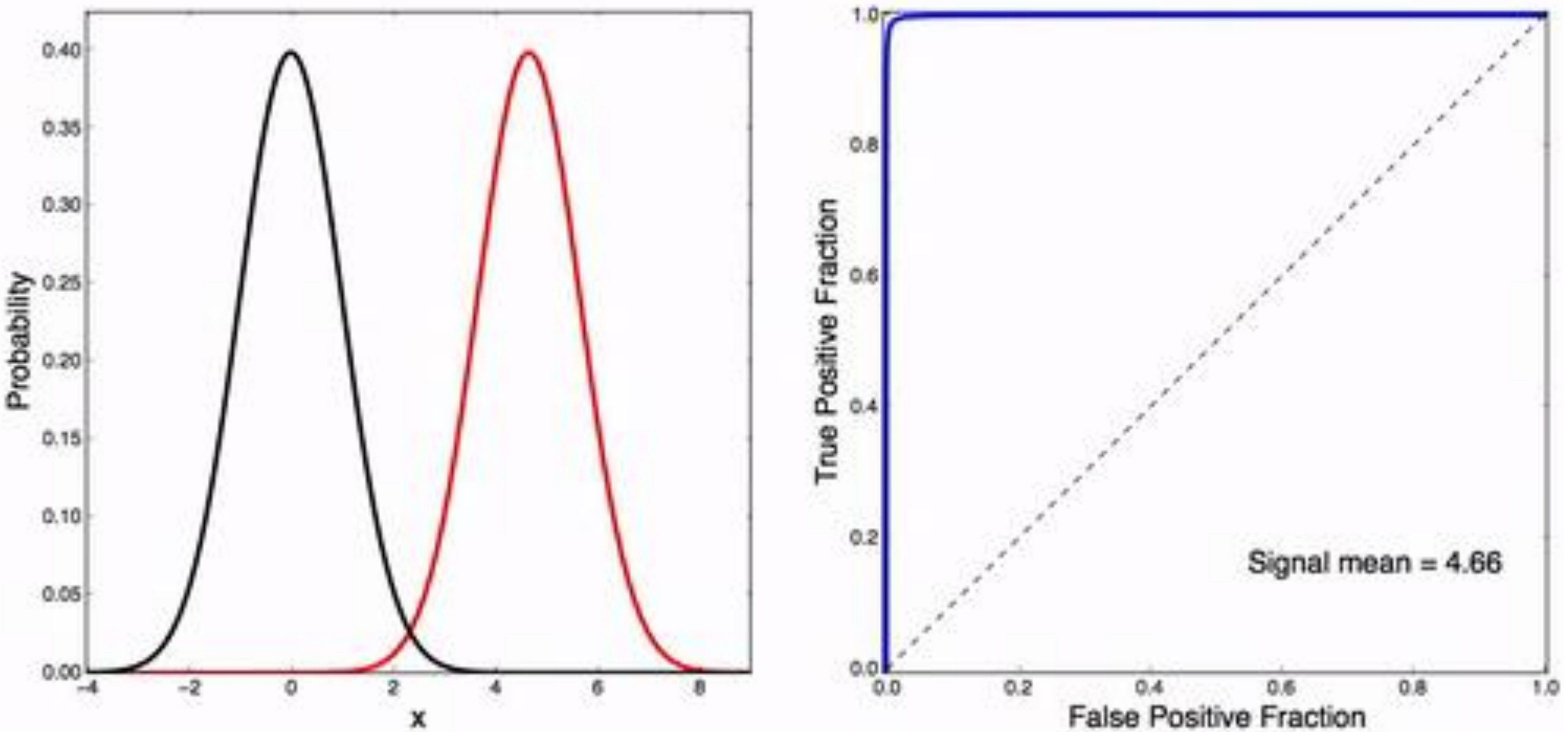
- "Contrast" has too many different meanings in different contexts (simulation, lab tests, telescope data)
- The only true, unambiguous definition of *contrast* is astrophysical planet/star brightness ratio
- We propose ROC curves as a standard performance metric to rigorously account for:
  - False positive fractions
  - True positive fractions

Extra Slides

The Receiver Operator Characteristic (ROC) Curve



The Receiver Operator Characteristic (ROC) Curve



Ambiguous contrast definitions, case I

- Assumption: SNR is solely a function of photon counts
- Aperture photometry: count photons in photometric aperture (or matched filter)
- η_{tel} = telescope throughput, without coronagraph, without apodizer, Lyot stop, etc., independent of photometric aperture size
- η_{cor} = coronagraph throughput or EE, fraction of the energy within photometric aperture (or matched filter)
- ε = astrophysical contrast, planet to star brightness ratio
  - independent of telescope and photometric aperture
- F = photon rate from star per unit time, at telescope input
- C = starlight suppression ratio = ratio of residual starlight photon counts in photometric aperture (or matched filter) to $\eta_{tel}\ \eta_{cor}\ F\ t$ (off-axis PSF with all coronagraph optics in, including final DM settings)

- Planet signal = $\eta_{tel}\ \eta_{cor}\ \varepsilon\ F\ t$
- Photon noise at planet location = sqrt $(C\ \eta_{tel}\ \eta_{cor}\ F\ t)$
- $SNR^2 = (\eta_{tel}\ \eta_{cor}\ \varepsilon\ F\ t)^2 / (C\ \eta_{tel}\ \eta_{cor}\ F\ t) = \eta_{tel}\ \eta_{cor}\ F\ t\ (\varepsilon)^2 / C$ tilda $\eta_{cor}\ t / C$

Ambiguous contrast definitions, case II

- Assumption: SNR is solely a function of photon counts
- Aperture photometry: count photons in photometric aperture (or matched filter)
- η_{tel} = telescope throughput, without coronagraph, without apodizer, Lyot stop, etc., independent of photometric aperture size
- η_{cor} = coronagraph throughput or EE, fraction of the energy within photometric aperture (or matched filter)
- ε = astrophysical contrast, planet to star brightness ratio
  - independent of telescope and photometric aperture
- F = photon rate from star per unit time at telescope input
- C = starlight suppression ratio = ratio of residual starlight photon counts in photometric aperture (or matched filter) to $\eta_{tel}\, F\, t$ (off-axis PSF with coronagraph out, apodizer out, DM flat = iteration 0 of WFC)

- Planet signal = $\eta_{tel}\ \eta_{cor}\, \varepsilon\, F\, t$
- Noise at planet location = sqrt $(C\, \eta_{tel}\, F\, t)$
- $SNR^2 = (\eta_{tel}\ \eta_{cor}\, \varepsilon\, F\, t)^2 / (C\, \eta_{tel}\, F\, t) = \eta_{tel}\ \eta_{cor}\, F\, t\, (\eta_{cor}\, \varepsilon)^2 / C\ \#\ \eta_{cor}^2\, t / C$