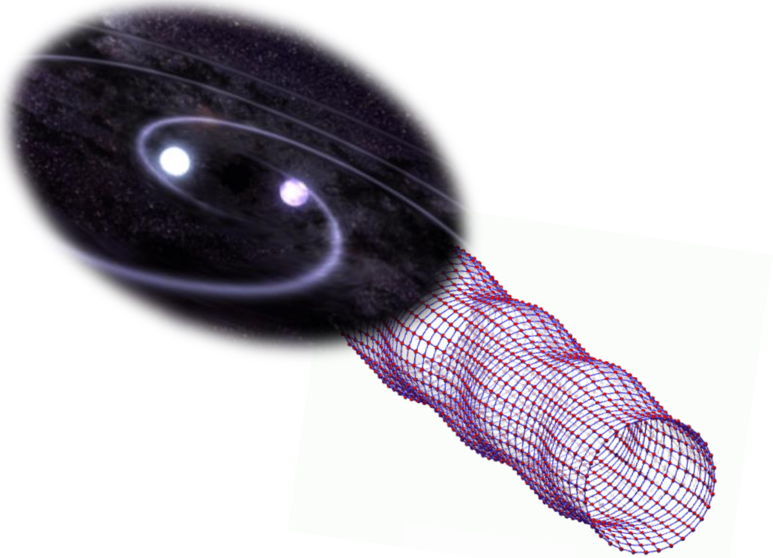
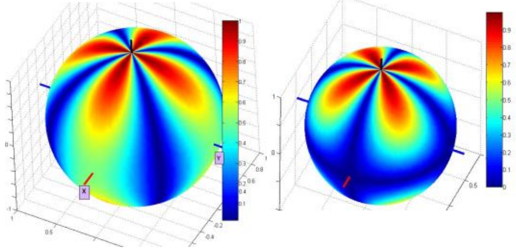
Signal detection and estimation

Gravitational Wave Data Analysis School @CAS, China

Soumya D. Mohanty

UTRGV

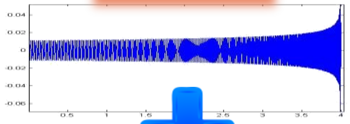


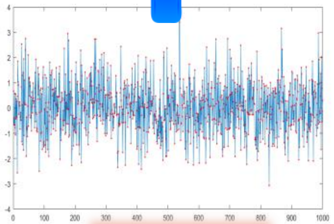


What is the signal shape?

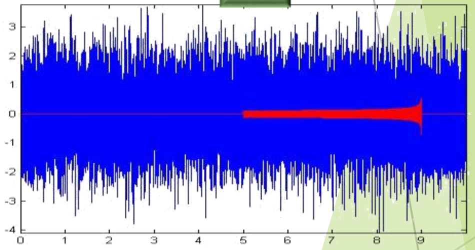
Signal present?

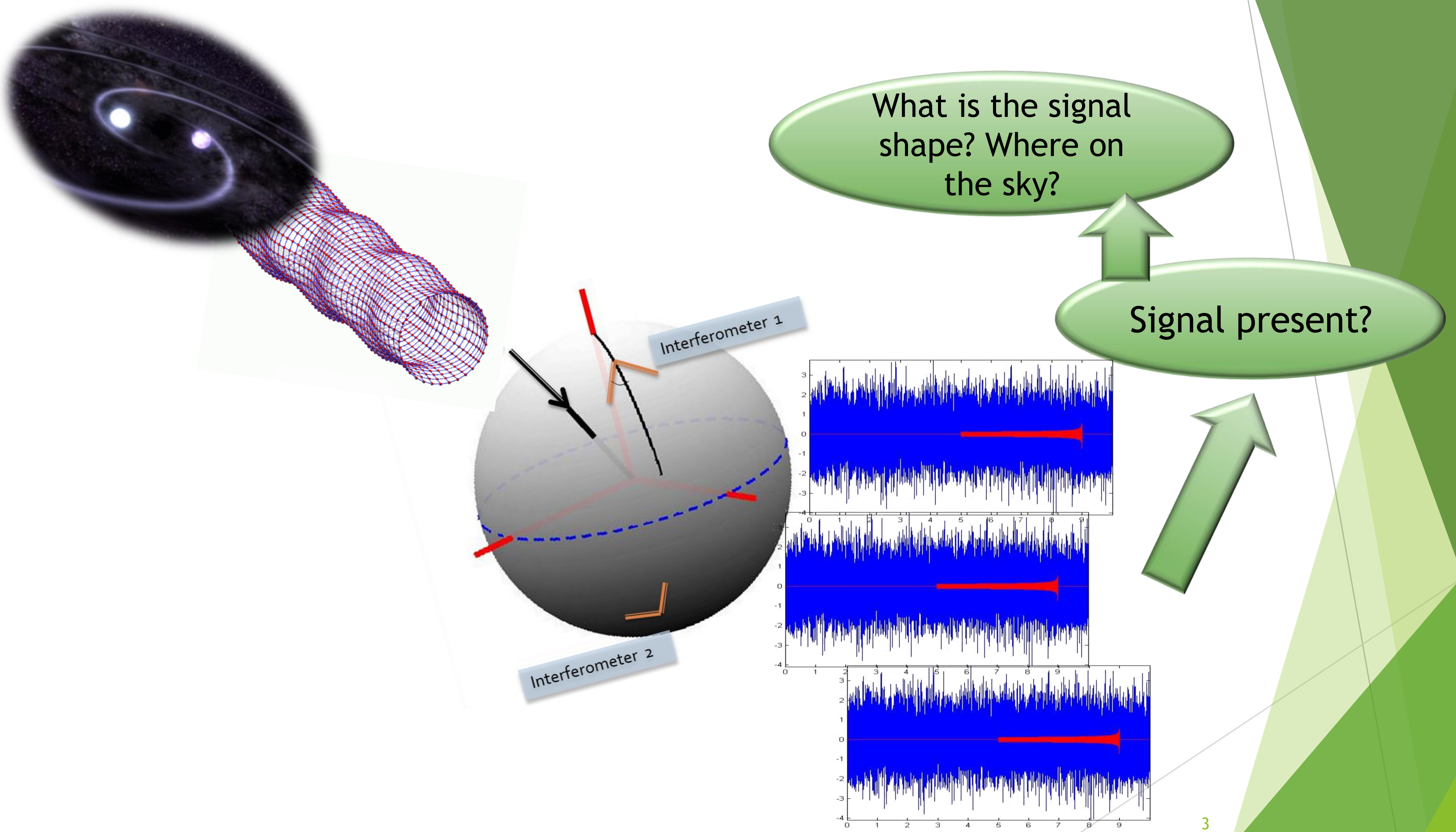
Rare

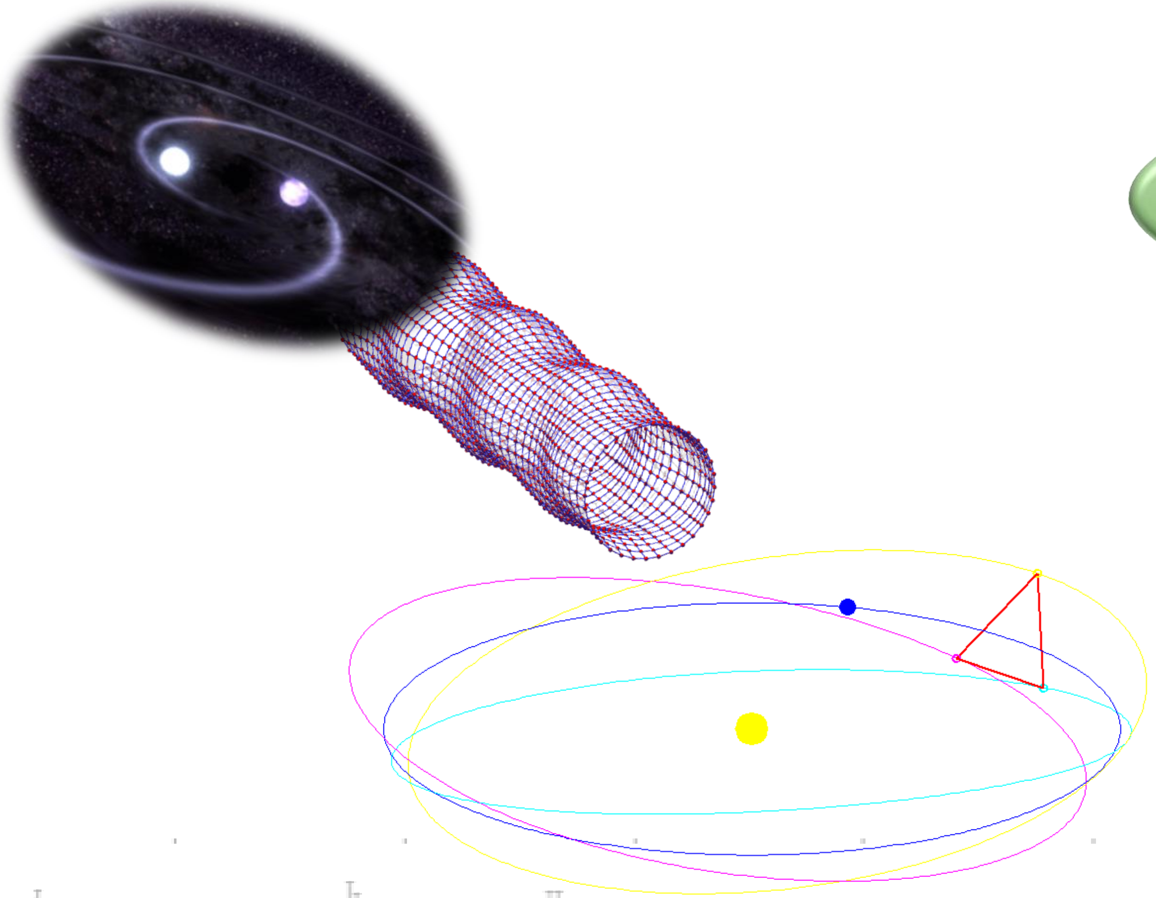




Always

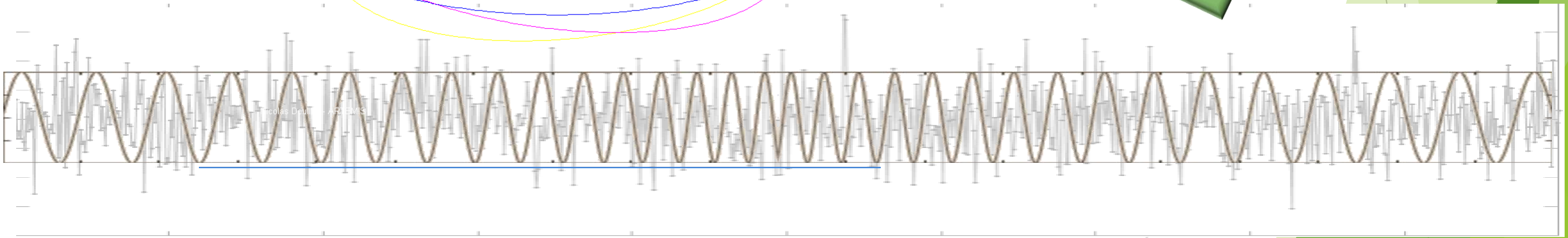



What is the signal shape? Where on the sky?
Signal present?
Interferometer 1
Interferometer 2



What is the signal shape? Where on the sky?

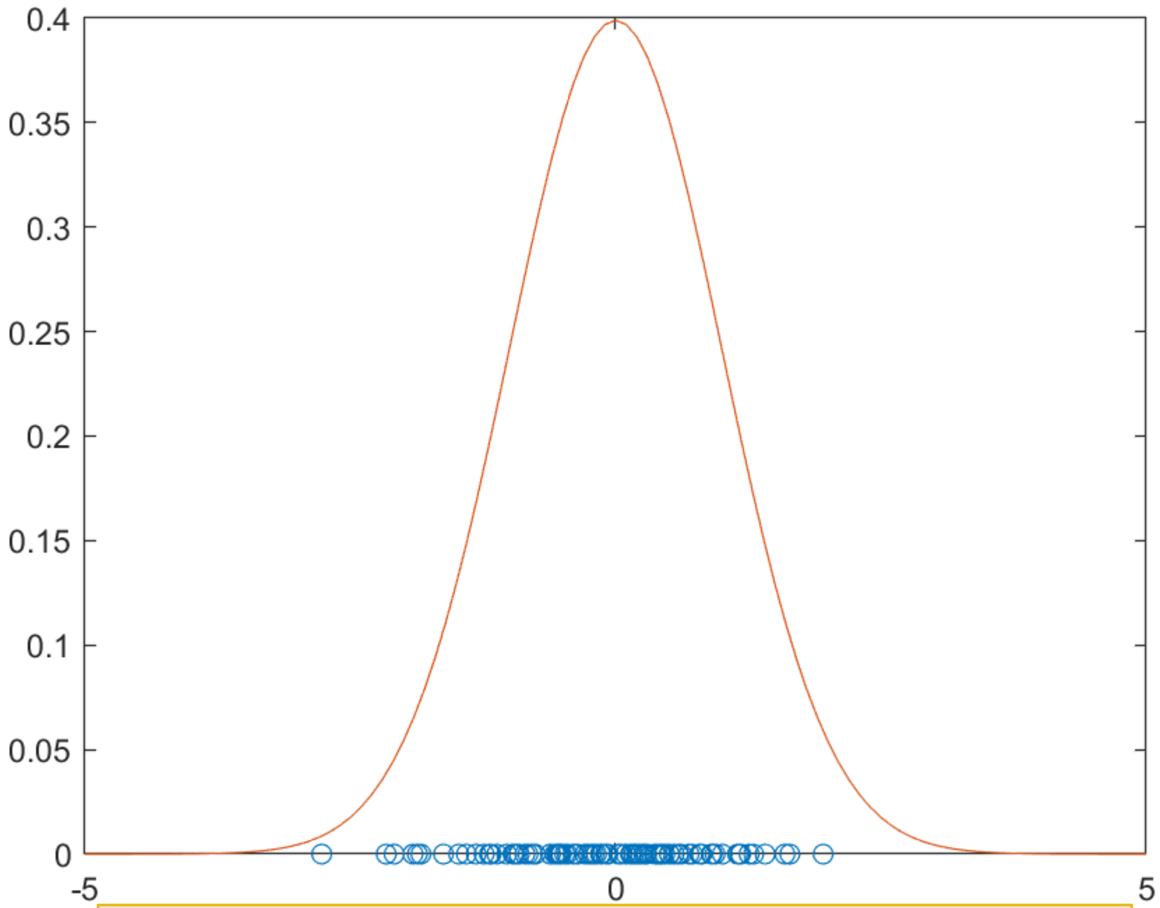
Signal present?



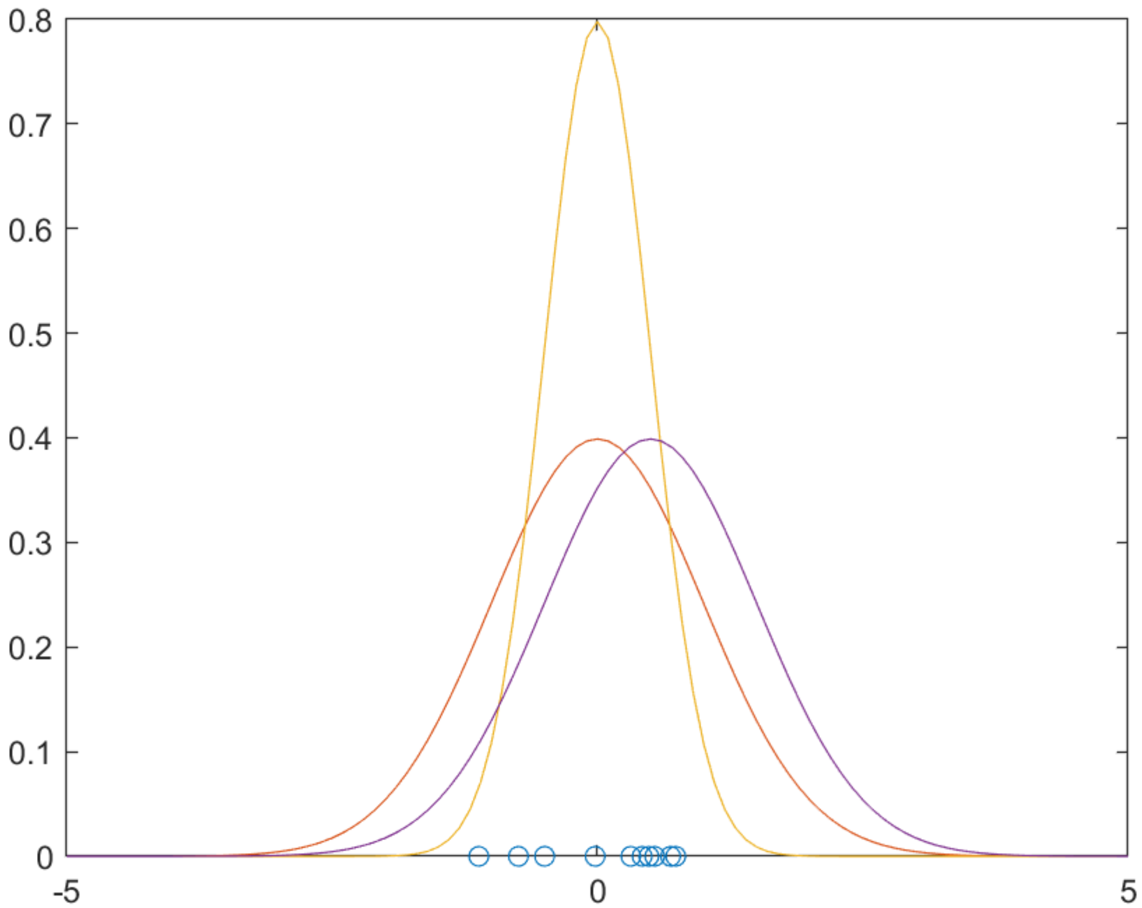
Estimation problem

Motivation for Likelihood

Simple estimation problem



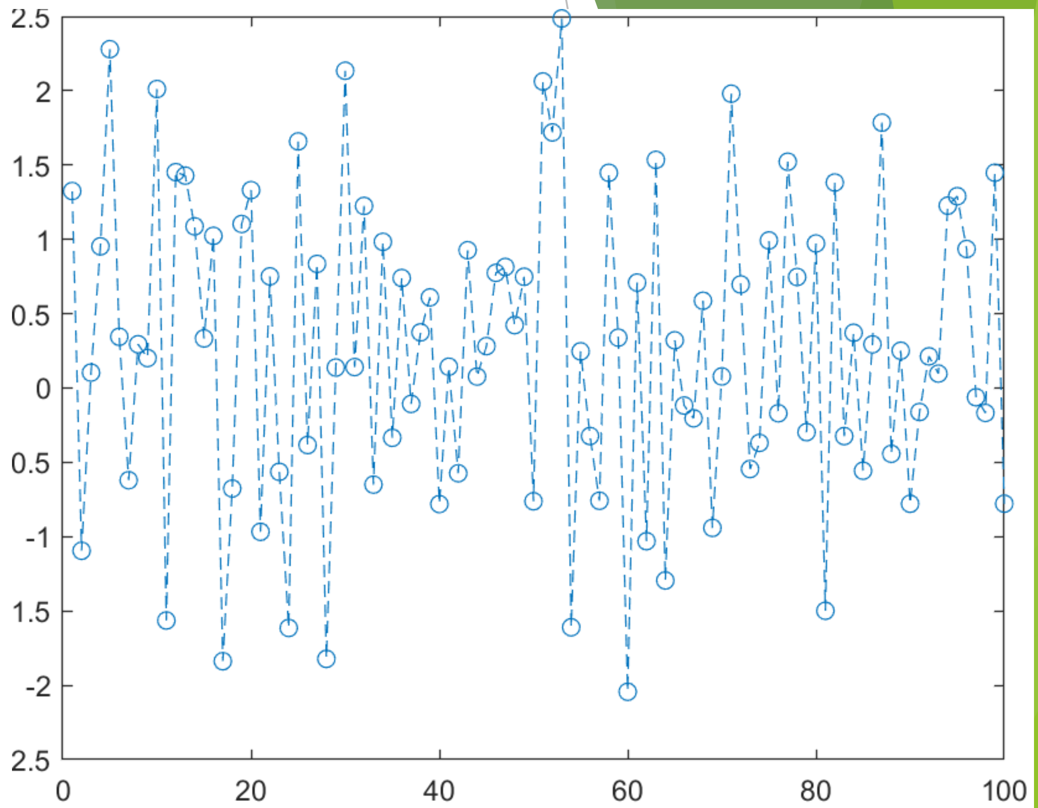
Data values drawn from a given pdf



What pdf were the data values drawn from?

General estimation problem

- **Given**: data $\bar{y} \in \mathbb{R}^N$
  - realization of a stochastic process $\bar{Y} = (Y_1, Y_2, \dots, Y_N)$
- **Given**: set of possible joint pdf's describing the stochastic process : $p_{\bar{Y}}(\bar{y}\,;\bar{\theta})$
  - $\bar{\theta}$: a set of parameters
  - $\bar{y}$ is drawn form one of these pdf's with parameters $\bar{\theta}_{true}$
  - The value of $\bar{\theta}_{true}$ is unknown
- **Task:** Estimate $\bar{\theta}_{true}$



Given: data from WGN with unknown mean μ

The joint pdf of the data is

$$p_{\bar{Y}}(\bar{y};\mu) = \prod_{i=1}^{100} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i-\mu)^2\right)$$

Estimation methods

Likelihood

- In the example, our judgment was based on which pdf appeared to be more "likely" as the correct one

- One way to make this idea mathematically precise is the **likelihood function**
  - Set of parameters: $\bar{\theta}$
  - Joint pdf of the data: $p_{\bar{Y}}(\bar{y}; \bar{\theta})$
  - Likelihood function: consider $p_{\bar{Y}}(\bar{y}; \bar{\theta})$ as a function of $\bar{\theta}$ for fixed $\bar{y}$ (data)
  - Alternative notation: $\mathrm{L}(\bar{y}; \bar{\theta})$

- A high likelihood value means the corresponding pdf gives higher probability of occurrence for the given data

WGN with unknown mean μ

- Set of parameter $\bar{\theta}$ is just μ
- $\bar{y} = (y_1, y_2, \dots, y_N)$
- The joint pdf of the data is

$$p_{\bar{Y}}(\bar{y}; \mu) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(y_i - \mu)^2\right)$$

- **Likelihood function**

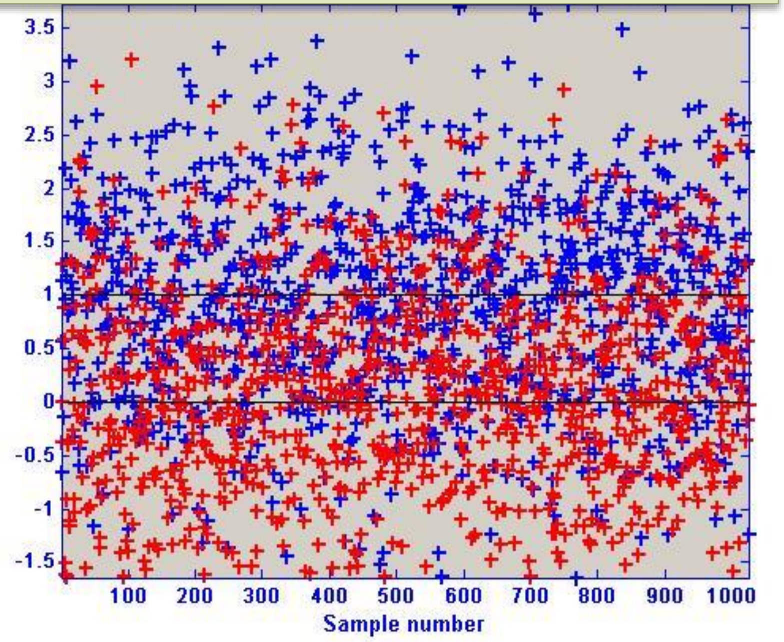
$$\mathrm{L}(\bar{y}; \mu) = p_{\bar{Y}}(y; \mu)$$

Likelihood function: WGN with unknown mean

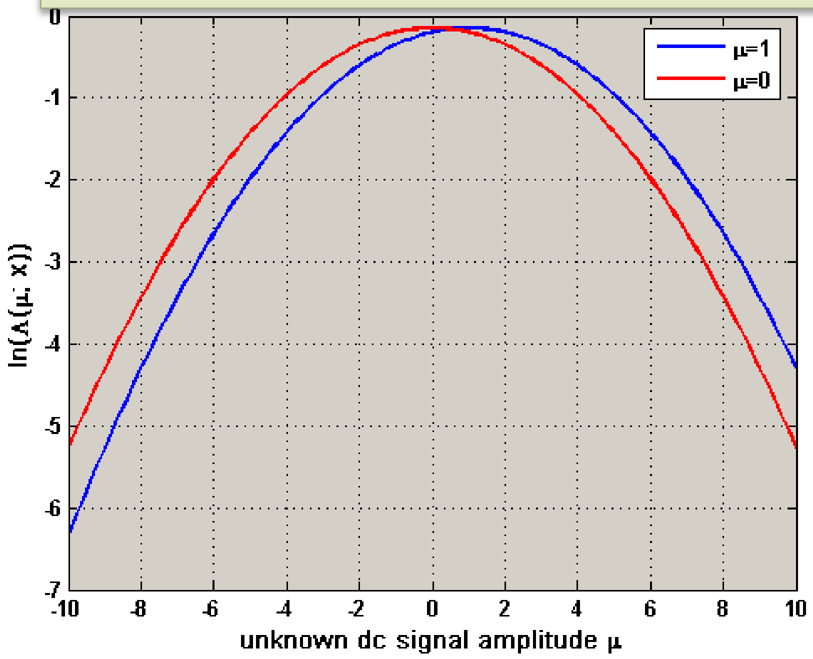
Likelihood function $\mathrm{L}(\bar{y};\mu)$:

$$\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}(y_i-\mu)^2\right)$$

Data realizations for $\mu = 0$ and $\mu = 1$



$\ln(\Lambda(\mu;\ \bar{x})$ for $\mu = 0$ and $\mu = 1$



Maximum Likelihood Estimation

- Find $\bar{\theta}$ at which the Likelihood, $L(\bar{y};\bar{\theta})$, has maximum value
  - This value of $\bar{\theta}$ is called the Maximum Likelihood Estimate (MLE)

$$\bar{\theta}_{MLE} = \arg\max_{\bar{\theta}} L(\bar{y};\bar{\theta})$$

- Instead of maximizing the likelihood, we can use any monotonic function of the likelihood
  - e.g., $\ln(L(\bar{y};\bar{\theta}))$ (**log-likelihood**)
- MLE is just one possible way to get an estimated value; other estimators are possible

MLE for signal in general Gaussian noise

- Data model:

$$\underbrace{\bar{y}}_{\substack{Data \\ realization}} = \underbrace{\bar{s}(\bar{\theta})}_{\substack{Signal\,with \\ parameters \\ \bar{\theta}}} + \underbrace{\bar{n}}_{\substack{Noise \\ realization}}$$

- Where the noise is a realization of $\bar{X}$ with

$$p_{\bar{X}}(\bar{x}\ ) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\|\bar{x}\|^2\right)$$

$$\|x\|^2 = \bar{x}\mathbf{C}^{-1}\ \bar{x}^2$$

MLE for signal in Gaussian noise

- Joint pdf of data

$$p_{\bar{Y}}(\bar{y};\bar{\theta}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\left\|\bar{y}-\bar{s}(\bar{\theta})\right\|^2\right)$$

- Log-likelihood function

$$L(\bar{y};\bar{\theta}) = const. - \frac{1}{2}\left\|\bar{y}-\bar{s}(\bar{\theta})\right\|^2$$

$$\max_{\bar{\theta}} L(\bar{y};\bar{\theta}) \text{ is equivalent to } \min_{\bar{\theta}} \left\|\bar{y}-\bar{s}(\bar{\theta})\right\|^2$$

- This is nothing but the method of **least-squares**

Least-squares for iid noise

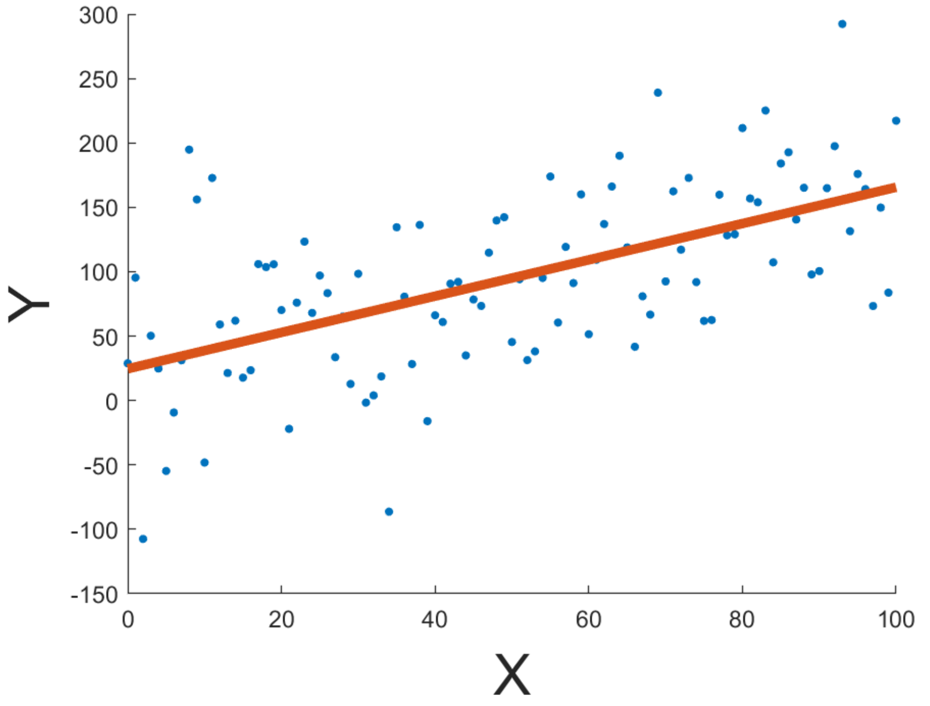
- For iid noise: $\mathbf{C} = \sigma^2 \mathbb{I}$ where $\mathbb{I}$ is the identity matrix

$$\min_{\bar{\theta}} \left\| \bar{y} - \bar{s}(\bar{\theta}) \right\|^2 = \min_{\bar{\theta}} \frac{1}{\sigma^2} \sum_{i=0}^{N-1} \left(y_i - s_i(\bar{\theta}) \right)^2$$

- Example: Fitting a straight line to pairs of points $(y_i, x_i), i = 0, 1, \ldots, N-1$

$$s_i = a x_i + b$$

$$\min_{a,b} \frac{1}{\sigma^2} \sum_{i=0}^{N-1} (y_i - (a x_i + b))^2$$



LINEAR AND NON-LINEAR MODELS

MLE for Gaussian Noise

$$\min_{\bar{\theta}} \left\| \bar{y} - \bar{s}(\bar{\theta}) \right\|^2$$

Linear models

$$\bar{s}(\bar{\theta}) = \sum_{i=0}^{p-1} \theta_i \underbrace{\bar{b}_i}_{Known}$$

Straight line fit:

$$\theta_0 = a, \theta_1 = b, \bar{b}_0 = \bar{x}, b_1 = (1,1, \dots 1)$$

The solution to the optimization problem can be expressed algebraically

Non-linear models

(e.g., Binary inspiral signal in GW data analysis)

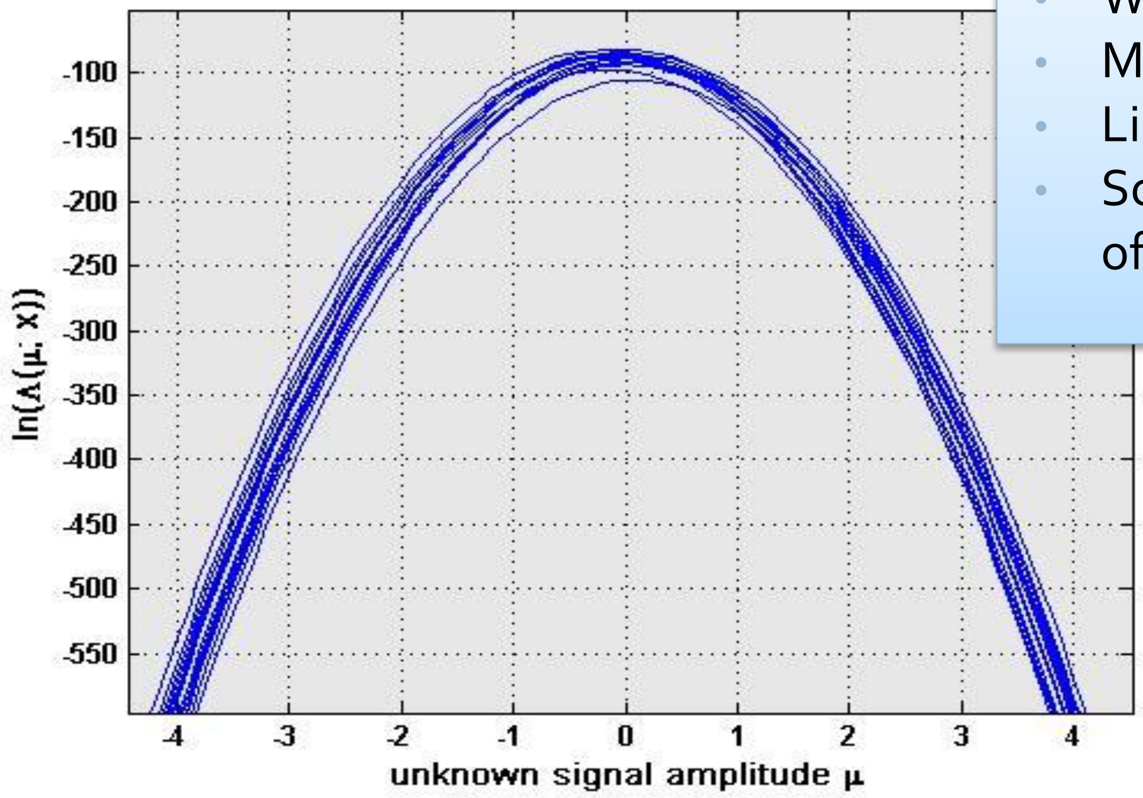
Estimation error

Minor change in notation

- Parameters: $\bar{\theta} = (\theta_0, \theta_1, \ldots, \theta_{M-1}) \rightarrow \Theta$
- Saves effort in producing equations in powerpoint!
- Note that Θ denotes a set of parameters and an element of Θ will continue to be denoted as θ_i

Estimation error

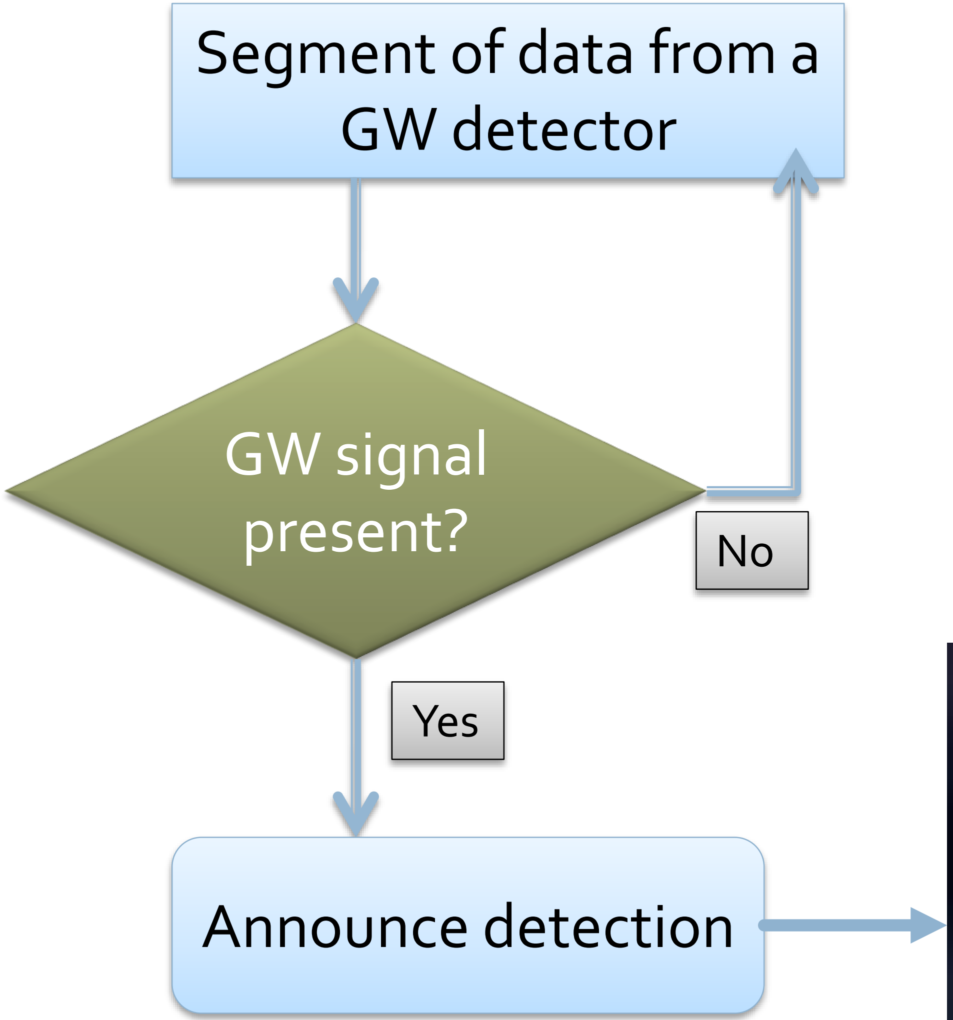
The estimated value of the parameters will not match the true value due to the presence of noise

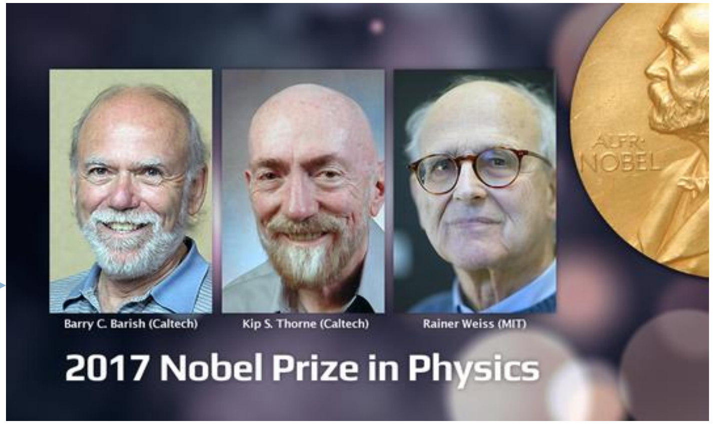


- WGN with non-zero mean
- Mean fixed, different noise realizations
- Likelihood function for each realization
- Scatter in the location of the maximum of the log-likelihood

DETECTION THEORY

Detection problem





Hypothesis testing

Given GW data we need to decide between two possibilities

Null hypothesis (H_0)

- $p_{\bar{Y}}(\bar{y}|H_0)$: Joint pdf of the data is that of pure noise and **no signal**
- $\bar{y} = \bar{n}$
- If H_0: Discard the data and get new data

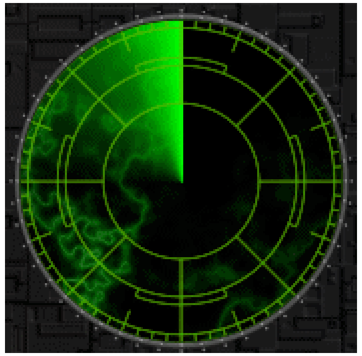
Alternative hypothesis (H_1)

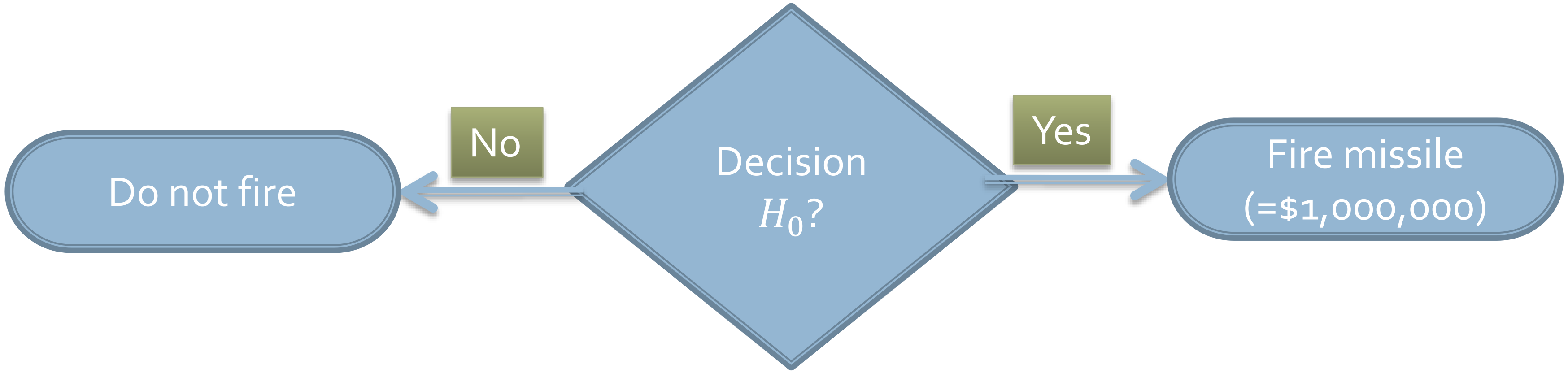
- $p_{\bar{Y}}(\bar{y}\,|H_1)$: Joint pdf of the data is that of pure noise **plus signal**
- $\bar{y} = \bar{s}(\Theta) + \bar{n}$ such that $\bar{s}(\Theta) \neq 0$
- If H_1: Estimate Θ

Hypothesis testing

- Example: radar detection with noise in receiver and clutter

Null (H_0)	Alternative (H_1)
Enemy plane	Not enemy plane (e.g., Friendly plane/flock of birds)





Decision errors

Possible outcomes in the previous example

		Reality	
		Enemy plane (H_0)	Not Enemy plane (H_1)
Decision	Enemy plane (H_0)	✓ Plane destroyed Cost $1,000,000	✗ $1,000,000 wasted
	Not Enemy plane (H_1)	✗ City destroyed Cost $1,000,000,000	✓ No cost

False alarm and False dismissal probabilities

Decision Rule: Chooses $\boldsymbol{H_0}$ or $\boldsymbol{H_1}$ based on given data

Noise in data $\Rightarrow$ Decision outcome is a **random variable**

False dismissal probability ($\boldsymbol{Q_1}$)

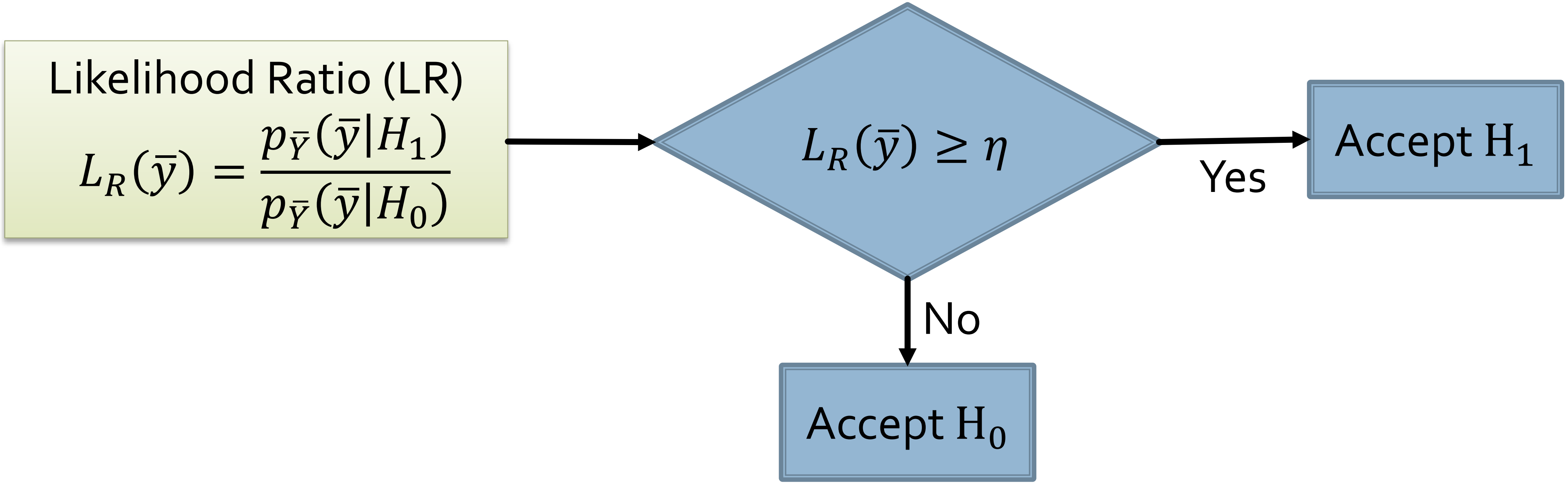
Probability of deciding H_0 when H_1 is true

False alarm probability ($\boldsymbol{Q_0}$)

Probability of deciding H_1 when H_0 is true

Likelihood Ratio Test

- **Binary hypothesis:** No free parameters under H_0 or H_1
- **Neyman-Pearson criterion**: Minimize Q_1 for fixed Q_0 $\rightarrow$ Optimal decision rule exists
- The decision rule is called the Likelihood Ratio (LR) test
- η: **Detection threshold**



Likelihood Ratio Test

- L_R is a random variable since it depends on noisy data
- pdf of L_R under H_0: $p_{L_R}(x|H_0)$
- pdf of L_R under H_1: $p_{L_R}(x|H_1)$

- Accept H_1 when $L_R(\bar{y}) \geq \eta \Rightarrow$ False alarm probability for LR test:

$$\mathrm{Q}_0 = \Pr(L_R \geq \eta \text{ when } H_0 \text{ is true}) = \int_\eta^\infty p_{L_R}(x|H_0)dx$$

- The detection threshold η is fixed by the given false alarm probability Q_0

$$Q_0 = \int_\eta^\infty dx \;\; p_{L_R}(x|H_0)$$

Additive dc signal in zero mean, stationary, white Gaussian noise

data: $\bar{x}$

Null hypothesis pdf:

$$p(\bar{x} \mid H_0) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=0}^{N-1} x_i{}^2 \right)$$

Alternative hypothesis pdf:

$$p(\bar{x} \mid H_1) = \frac{1}{(2\pi)^{N/2}\sigma^N} \exp\left(\frac{-1}{2\sigma^2} \sum_{i=0}^{N-1} (x_i - \mu)^2 \right)$$

Likelihood Ratio :

$$\Lambda(\mathbf{x}) = \frac{p(\bar{x} \mid H_1)}{p(\bar{x} \mid H_0)} = \exp\left(\frac{1}{\sigma^2} \left(\mu \sum_{i=0}^{N-1} x_i - \frac{1}{2} N \mu^2 \right) \right)$$

$$\Lambda(\mathbf{x}) \geq \eta \quad \Rightarrow \quad \frac{1}{N} \sum_{i=0}^{N-1} x_i \geq \frac{\ln \eta}{\mu} + \mu = \text{const.}(\ \Gamma\ )$$

Decision rule : Compute $\hat{\mu} = \frac{1}{N} \sum_{i=0}^{N-1} x_i$ and compare with a threshold

Use $Q_0 = \int_\Gamma^\infty dz\ p_{\hat{\mu}}(z \mid H_0)$ to fix Γ

Log-Likelihood Ratio

- Decision rule: Accept H_1 when $L_R(\bar{y}) \geq \eta$
- $\Rightarrow$ We can also use any monotonic function of $L_R(\bar{y})$
- Example: Accept H_1 when $2L_R(\bar{y}) \geq 2\eta$
  - $\Rightarrow$Overall constant factor in the definition of L_R can be ignored as it simply means redefining the detection threshold
- Common choice:

$$\text{Accept } \mathrm{H}_1 \text{ when } \ln(\,L_R(\bar{y})) \geq \ln(\eta)$$

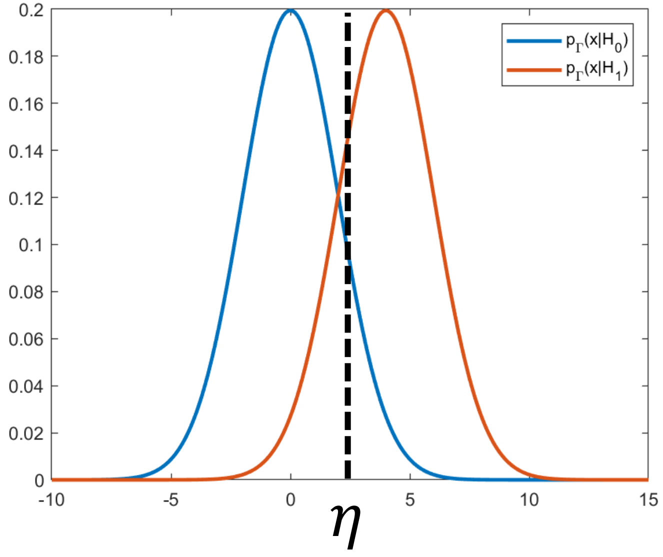
- From now on, we will use log-likelihood ratio LLR (with the same symbol L_R)
  - Overall constant factor in LLR will be ignored
  - Additive constant in LLR will be ignored

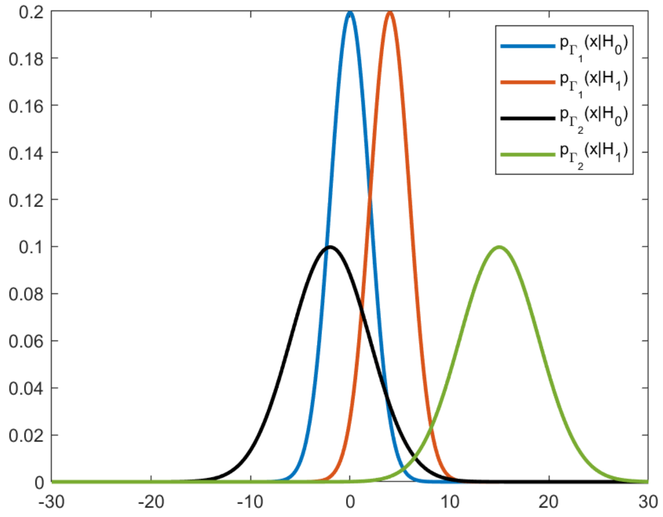
Detection statistic

- L_R is an example of a detection statistic: A function $\Gamma(\bar{y})$ of data $\bar{y}$ that is compared against a threshold η to decide between H_0 and H_1
  - L_R is optimal if the joint pdf of noise matches the one used in deriving the L_R
  - In many cases, we do not know the noise joint pdf accurately $\Rightarrow L_R$ will not be optimal
  - In many cases, the L_R may be too expensive to compute
- For any detection statistic $\Gamma(\bar{y})$, we can define:

$$\text{False alarm probability: } Q_0 = \int_{\eta}^{\infty} p_\Gamma(x|H_0)dx$$

$$\text{Detection probability: } Q_d = \int_{\eta}^{\infty} p_\Gamma(x|H_1)dx$$





- Example:

$$Q_0 = \int_\eta^\infty p_\Gamma(x|H_0)dx$$
$$Q_d = \int_\eta^\infty p_\Gamma(x|H_1)dx$$

- Given two different detection statistics $\Gamma_1(\bar{y})$ and $\Gamma_2(\bar{y})$, the one that has higher Q_d for the same Q_0 is better
- $\Rightarrow$ The two pdfs of Γ_2 are more separated than those of Γ_1

Signal to noise ratio

- How well a statistic $\Gamma(\bar{y})$ performs depends on the separation between the pdfs $p_\Gamma(x|H_0)$, and $p_\Gamma(x|H_1)$
- A convenient measure of their separation is **signal to noise ratio (SNR)**
  - $SNR = E[\Gamma \mid H_1] / [var(\Gamma \mid H_0)]^{1/2}$
  - $var(X) = E[(X - E[X])^2]$
- For the same two hypotheses, the SNR will be higher for a better performing statistic
- SNR: A first, easy step in comparing detection statistics
- Note that SNR is **not a random variable** itself

DC signal in zero mean WGN

LR statistic: $L_R = \frac{1}{N}\sum_{i=1}^{N} x_i$

Zero mean noise $\Rightarrow$ Under H_0: $E[\mathrm{L_R}] = 0$

DC signal amplitude: $\mu \Rightarrow$ Under H_1: $E[L_R] = \mu$

Under H_0: $var(\mathrm{L_R}) = E\left[L_R^2\right]$
Type equation here.

$$= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{j=1}^{N} \mathrm{E}\left[x_k x_j\right] = \frac{1}{N^2}\sum_{k=1}^{N}\sum_{j=1}^{N} \sigma^2 \delta_{kj} = \frac{\sigma^2}{N}$$

$$\therefore\ \mathrm{SNR} = \frac{\mu\sqrt{N}}{\sigma}$$

SNR of LR test for arbitrary signal and Gaussian noise

- Gaussian stationary noise and arbitrary additive signal

$$\bar{y} = \begin{cases} \bar{n} \ ; H_0 \\ \bar{s} + \bar{n} \ ; \ H_1 \end{cases}$$

- Joint pdf of noise:

$$p_{\bar{E}}(\bar{y}\,) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}\|\bar{y}\|^2\right)$$

$$\|\bar{y}\|^2 = \bar{y}\boldsymbol{C}^{-1}\bar{y}^T$$

- Let $\langle \bar{z}, \bar{y} \rangle = \bar{z}\boldsymbol{C}^{-1}\bar{y}^T$
- LR statistic : $\Lambda = \frac{1}{2}\|\bar{y}\|^2 - \frac{1}{2}\|\bar{y} - \bar{s}\|^2 = \langle \bar{y}, \bar{s} \rangle - \frac{1}{2}\langle \bar{s}, \bar{s} \rangle \rightarrow \Lambda = \langle \bar{y}, \bar{s} \rangle$
- SNR:

$$\frac{E[\mathrm{L_R} \mid H_1]}{[var(\mathrm{L_R}|H_0)]^{\frac{1}{2}}} = \frac{\langle \bar{s}, \bar{s} \rangle}{[E[\bar{s}\mathbf{C}^{-1}\bar{y}^T\bar{y}\mathbf{C}^{-1}\bar{s}^T]]^{1/2}} = \sqrt{\langle \bar{s}, \bar{s} \rangle} = \|\bar{s}\|$$

where $E\left[(\bar{y}^T\bar{y})_{ij}\right] = E\left[y_i y_j\right] = C_{ij} \Rightarrow E[\bar{y}^T\bar{y}] = \mathbf{C}$

Composite hypothesis test

- The detection problem in GW data analysis is **not** a binary hypothesis test
- GW data is the sum of noise and a **signal with unknown parameters**

$$\bar{y} = \bar{s}(\bar{\theta}) + \bar{n}\ ;\ \text{Alternative notation}\ \bar{y} = \bar{s}(\Theta) + \bar{n}$$

- Example: The signal may be from a binary inspiral but we do not know the parameters (masses, distance etc) of the source

- H_0: $\bar{y}$ is only noise
- H_1: Not one but many alternative hypotheses corresponding to each possible value of $\Theta \rightarrow$ **Composite hypotheses test**

- Neyman-Pearson criterion for composite hypotheses: **No solution**, in the general case, for an optimal decision surface

GLRT: Generalized Likelihood Ratio Test

- $p_{\bar{Y}}(\bar{y}|H_1;\Theta)$: pdf of the data for the hypothesis $\bar{y} = \bar{s}(\bar{\theta}) + \bar{n}$
- $p_{\bar{Y}}(\bar{y}|H_0)$: pdf of the data under the null hypothesis $\bar{y} = \bar{n}$
- Likelihood Ratio:

$$L(\bar{y};\Theta) = \frac{p(\bar{y}|H_1;\Theta)}{p(\bar{y}|H_0)}$$

Generalized Likelihood Ratio Test (GLRT): test statistic $L_g(\bar{y}) = \max_{\Theta} L(\bar{y};\Theta)$

- We can also define $L_g(\bar{y})$ as $L_g(\bar{y}) = \max_{\Theta} \ln L(\bar{y};\Theta)$
- As with the LR test, Detection threshold η fixed by false alarm probability Q_0

$$Q_0 = \int_{\eta}^{\infty} d\,z\, p_{L_g}(z|H_0)$$