

Chapter 1

Data analysis

Our data model is of a continuous time stochastic process $X(t)$ over an interval $[0, T]$, sampled at n points $t_1, \dots, t_n \subset T$. We assume an additive noise model, that is

$$X_t = h_t + N_t, \quad t \in t_1, \dots, t_n, \quad (1.1)$$

where $X_t = X(t)$, h_t is a sample from a deterministic signal waveform at time t , and $\{N_t : t \in [0, T]\}$ is the detector noise, a zero mean Gaussian random process with covariance matrix Σ_N .

1.1 Hypothesis Testing

Gravitational wave detection can be cast as a hypothesis testing problem. The *null hypothesis*, H_0 , is that the signal is absent and the data consists entirely of noise with probability distribution P_0 . The *alternative hypothesis*, H_1 , is that a signal is present and the data has probability distribution P_1 . This can be written

$$\begin{aligned} H_0 : X_t &= N_t \sim P_0, & t \in t_1, \dots, t_n \\ H_1 : X_t &= h_t + N_t \sim P_1, & t \in t_1, \dots, t_n. \end{aligned} \quad (1.2)$$

A *hypothesis test* or *decision rule* δ is a partition of the observation set \mathcal{X} into two subsets \mathcal{X}_0 and $\mathcal{X}_1 = \mathcal{X}_0^c$. If the data $x \in \mathcal{X}_0$, we accept the null hypothesis H_0 . If $x \in \mathcal{X}_1$, we claim a detection!

We now need a way to choose a decision rule δ that is optimal in some sense. There are several several ways to do this; here the Neyman-Pearson approach is described. The Bayesian and Minimax approaches are also common, and can be found in [poor].

1.1.1 Neyman-Pearson Approach

There are two types of errors that can occur in testing H_0 against H_1 . A *type I error* or *false alarm* occurs if we choose H_1 when H_0 is true, and the probability of this occurring for a decision rule δ is called the *false alarm probability* denoted $P_F(\delta)$. The second error occurs if we choose H_0 when H_1 is true and is called a *type II error* or *miss*. We can define a *miss probability*, $P_M(\delta)$, however it is more common to talk in terms of the *detection probability* or *power* of the test $P_D(\delta) = 1 - P_M(\delta)$. The Neyman-Pearson optimum test maximises the power of the test, subject to a chosen constraint on the false alarm probability:

$$\delta_{NP} = \operatorname{argmax}_{\delta} P_D(\delta) \text{ subject to } P_F(\delta) \leq \alpha. \quad (1.3)$$

The bound α is called the *significance level* of the test. We seek to devise a function $\phi : \mathcal{X} \rightarrow [0, 1]$ representing the probability of choosing H_1 given some data $x \in \mathcal{X}$. The detection and false alarm probabilities for a decision rule associated with such a function are as follows:

$$P_D(\delta) = E_1[\phi(\mathcal{X})] = \int_{\mathcal{X}} \phi(x) p_1(x) d\mathbf{x} \quad (1.4)$$

$$P_F(\delta) = E_0[\phi(\mathcal{X})] = \int_{\mathcal{X}} \phi(x) p_0(x) d\mathbf{x} \quad (1.5)$$

The existence and uniqueness of such a test function is given by the Neyman-Pearson Lemma.

Lemma 1.1.1. *Consider the hypothesis pair 1.2. Suppose P_0 and P_1 have densities p_0 and p_1 respectively and $\alpha > 0$. Then*

1. *Existence: For every $\alpha \in (0, 1)$, there is a function $\phi(x)$ associated to a test δ_{NP} which is of the form*

$$\phi(x) = \begin{cases} 1 & \text{if } p_1(x) > \lambda p_0(x) \\ \gamma & \text{if } p_1(x) = \lambda p_0(x) \\ 0 & \text{if } p_1(x) < \lambda p_0(x) \end{cases} \quad (1.6)$$

where $\eta \geq 0$, $0 \leq \gamma \leq 1$ and such that $P_F(\delta_{NP}) = \alpha$.

2. *Optimality: Let δ' be any decision rule satisfying $P_F(\delta') \leq \alpha$, and that it is associated with some function ϕ' . Then $P_D(\delta_{NP}) \geq P_D(\delta')$, where δ_{NP} is as defined by the function 1.6.*
3. *Uniqueness: Suppose that δ''_{NP} is any α -level Neyman-Pearson decision rule for H_0 versus H_1 . Then the function $\phi''(x)$ associated with δ''_{NP} must be of the form 1.6, except for possibly on a subset of \mathcal{X} having zero probability under H_0 and H_1 .*

Proof. 1. Let η be the smallest number such that

$$P_0(p_1(X) > \eta p_0(X)) \leq \alpha. \quad (1.7)$$

If $P_0(p_1(X) > \eta p_0(X)) < \alpha$, choose

$$\gamma = \frac{\alpha - P_0(p_1(X) > \eta p_0(X))}{P_0(p_1(X) = \eta p_0(X))}, \quad (1.8)$$

otherwise γ can be chosen arbitrarily. Note that for a test of the form 1.6, the false alarm probability is the expectation of the function ϕ under H_0 :

$$P_F(\delta_{NP}) = E_0[\phi] = 1 \times P_0(p_1(X) > \eta p_0(X)) + \gamma P_0(p_1(X) = \eta p_0(X)) = \alpha. \quad (1.9)$$

2. Note that $\forall x \in \mathcal{X}$

$$[\phi(x) - \phi'(x)][p_1(x) - \lambda p_0(x)] \geq 0,$$

and therefore,

$$\int_{\mathcal{X}} [\phi(x) - \phi'(x)][p_1(x) - \lambda p_0(x)] d\mathbf{x} \geq 0. \quad (1.10)$$

Expanding this and rearranging terms gives

$$\int_{\mathcal{X}} \phi(x) p_1(x) d\mathbf{x} - \int_{\mathcal{X}} \phi'(x) p_1(x) d\mathbf{x} \geq \lambda \left[\int_{\mathcal{X}} \phi(x) p_0(x) d\mathbf{x} - \int_{\mathcal{X}} \phi'(x) p_0(x) d\mathbf{x} \right], \quad (1.11)$$

which, using 1.4 and 1.5, becomes

$$P_D(\delta_{NP}) - P_D(\delta') \geq \lambda \left[\underbrace{P_F(\delta_{NP})}_{=\alpha} - \underbrace{P_F(\delta')}_{\leq \alpha} \right] \geq 0. \quad (1.12)$$

Thus, $P_D(\delta_{NP}) \geq P_D(\delta')$.

3. By definition, we must have $P_D(\delta''_{NP}) = P_D(\delta_{NP})$. Therefore, by 1.12 we have

$$0 \geq \lambda [\alpha - P_F(\delta''_{NP})] \geq 0 \quad (1.13)$$

or, $P_F(\delta''_{NP}) = \alpha$. Reversing the steps in part 2 above, we find

$$\int_{\mathcal{X}} [\phi(x) - \phi''(x)][p_1(x) - \lambda p_0(x)] d\mathbf{x} = 0. \quad (1.14)$$

Since we know the integrand to be non-negative, this implies that it is zero except possibly on a set of zero probability under H_0 and H_1 . Therefore, $\phi(x)$ and $\phi''(x)$ differ only on the set $\{x \in \mathcal{X} | p_1(x) - \lambda p_0(x) = 0\}$, which implies ϕ'' is also of the form 1.6, differing only in γ .

□

We define the quantity

$$L(x) = \frac{p_1(x)}{p_0(x)}, \quad x \in \mathcal{X}, \quad (1.15)$$

to be the *likelihood ratio*. Then, by rearranging the conditions in 1.6, we see that a Neyman-Pearson test involves comparing the likelihood ratio for an observed value of X to a threshold λ .

1.2 Introduction

For a known waveform the most efficient data analysis tool is matched filtering. But often the waveforms of GW are poorly known, limiting the matched filter's applicability.

A linear time frequency transform correlates the signal with a family of waveforms that are well concentrated in time and space. These waveforms are called *time frequency atoms*.

1.3 Fourier Transforms

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt$$

1.4 Windowed Fourier Transforms

Gabor atoms

$$g_{u,\xi}(t) = g(t - u)e^{i\xi t}$$

When g is Gaussian, best localisation and the atoms are called *Gabor functions*. Fourier transform of Gabor atom

$$\hat{g}_{u,\xi}(\omega) = \hat{g}(\omega - \xi)e^{-iu(\omega - \xi)}$$

Windowed Fourier transform:

$$Sf(u, \xi) = \int_{-\infty}^{+\infty} f(t)g_{u,\xi}(t)dt = \int_{-\infty}^{+\infty} f(t)g(t - u)e^{-i\xi t}dt$$

Also can be written as frequency integral

$$Sf(u, \xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{g}_{u,\xi}^*(\omega)d\omega$$

1.5 wavelet transforms

Wavelet transforms are described in terms of families of basis functions $\psi_{a,b}(x)$ which are dilations and translations of a real function $\psi(x)$ of zero average:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$$

A wavelet transform of a function f has a different time frequency resolution to a windowed Fourier transform. It is computed by correlating f with a wavelet atom

$$Wf(a, b) = \int_{-\infty}^{+\infty} f(t) \psi_{a,b}^*(t) dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{\psi}_{a,b}^*(\omega) d\omega$$

1.6 Notes from "method for detection and reconstruction of gravitational wave transients with networks of advanced detectors"

1.6.1 Low latency search

Designed to identify potentially significant candidates in real time (within minutes). Streaming? Promptly share sky location with partner telescopes to search for EM counterparts eg Gamma ray burst and afterglow or kilonova which will fade within seconds to days.

1.6.2 Archive analysis

Establish significance of event and identify progenitors. Slower and more detailed.

1.6.3 Inverse problem for data from a network of detectors

Network of K detectors. Time series data from the detectors is transformed with the Wilson-Debauchies-Meyer (WDM) transform so that each detector k presents data $x_k[i]$ in a general time-frequency (TF) domain, where i is the data sampling (TF pixel) index.

Detector noise assumed to be Gaussian, described by WDM power spectral density $S_k[i]$ estimated for every data sample. The noise-scaled (whitened) data is defined as $\omega_k[i] = x_k[i] / \sqrt{S_k[i]}$

TF series combined to obtain energy TF maps $E[i] = \sum_k \omega_k^2[i]$ where the $E[i]$ are maximised over all possible time-of-flight delays in the network (i think this means it's lining up the signals based on the travel time of GW from one detector to another. Obviously this depends on where

in the sky the GW is coming from). The energy maps are used to identify TF areas with excess energy above baseline detector noise using an appropriate clustering algorithm. The inverse problem is to extract the signal waveform, polarisation and sky location. (Note that this slightly differs from my definition of the inverse problem which is to determine the parameters from the waveform. Many steps to the inverse problem).

Data vector $\mathbf{x}[i] = x_1[i], \dots, x_K[i]$ recorded by detector at time of a GW signal $\mathbf{h}[i] = [h_+[i], h_\times[i]]$ coming from sky location θ and ϕ is a superposition of network response $\mathcal{F}\mathbf{h}[i]$ and noise $\mathbf{n}[\mathbf{h}]$:

$$\mathbf{x}[i] = \mathcal{F}\mathbf{h}[i] + \mathbf{n}[\mathbf{h}]$$

where \mathcal{F} is the network antenna pattern matrix:

$$\mathcal{F} = \begin{bmatrix} F_{1+}(\theta, \phi) & F_{1\times}(\theta, \phi) \\ \dots & \dots \\ F_{K+}(\theta, \phi) & F_{K\times}(\theta, \phi) \end{bmatrix}$$

(Antenna patterns often also include a transformation by polarisation angle Ψ , but this is equivalent to a rotation of the wave frame and can be included in the definition of \mathbf{h}).

BX-LBTN-KYW2-9V7B-9K2D

1.6.4 Matched Filtering

from [?] Searches for GWs from CBC events employ matched filtering as the first step.

"whitening": Optimal strategy, weight detector output and template wavefor by inverse of amplitude spectral density of detector noise

discretely sampled time series whitened data: $\vec{s} = s_i$ whitened template waveform: $\vec{h}_\alpha = h_{\alpha i}$ noise: $\vec{n} = n_i$ zero-mean $\langle n_i \rangle = 0$, unit-variance Gaussian random process $\langle n_i, n_j \rangle = \delta_{ij}$ output of matched filter is given by the vector inner product

$$\rho_\alpha = \vec{h}_\alpha^* \cdot \vec{s} \quad (1.16)$$

search over phase by using complex-valued templates where $\Re \vec{h}_\alpha$ contains the cosine like phase and $\Im \vec{h}_\alpha$ contains the sine-like phase.

1.7 Jaranowski et.al

Random Variables

A *probability space* is a triple $(\mathcal{X}, \mathcal{O}, P)$ where

- \mathcal{X} is a sample space, usually an n -dimensional Euclidean space \mathbb{R}^n or a discrete set $\Gamma = \{\gamma_1, \gamma_2, \dots\}$, or for continuous process, a function space.
- \mathcal{O} is a σ -algebra of subsets of \mathcal{X} called observation events, these are the set of values that data can take.

– if $\mathcal{X} = \mathbb{R}^n$, consider open intervals

$$\{\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n : a_1 < y_1 < b_1, \dots, a_n < y_n < b_n\} \quad (1.17)$$

where a_i and b_i are real numbers. Then usually \mathcal{O} is the smallest σ -algebra containing all such intervals. This is known as the *Borel* σ -algebra and is denoted by B^n .

– if \mathcal{X} is a discrete set Γ , the σ -algebra \mathcal{O} is the set of all subsets of Γ . This is called the *power set* of Γ and is denoted by 2^Γ .

- P is a probability distribution (measure) on \mathcal{O} . This is a function assigning a number $P(A) \geq 0$ to any set $A \in \mathcal{O}$ such that $P(\emptyset) = 0$ and $P(\mathcal{X}) = 1$. For probability spaces (\mathbb{R}^n, B^n, P) usually there exists a *probability density function* $p : \mathbb{R}^n \rightarrow [0, \infty)$ such that

$$P(A) = \int_A p(\mathbf{x}) d\mathbf{x} \quad (1.18)$$

where \mathbf{x} are some coordinates on \mathbb{R}^n .

σ -algebra:

1. $\mathcal{X} \in \mathcal{O}$
2. if $A \in \mathcal{O}$, then $A^c := \mathcal{X} - A \in \mathcal{O}$
3. for any countable family A_k of members of \mathcal{O} , their union is in \mathcal{O}

Together these imply that the intersection of a finite or countable family of members of \mathcal{O} belong to \mathcal{O} .

A *random variable* X is a real-valued function $X : \mathcal{X} \rightarrow \mathbb{R}$. The sets $\{\xi \in \mathcal{X} : X(\xi) \leq x\} \in \mathcal{O} \forall x \in \mathbb{R}$.

The *cumulative distribution function* of X is

$$P_X(x) := P(X \leq x). \quad (1.19)$$

It is non-decreasing and right-continuous.

A *continuous* random function is one with a continuous cumulative distribution function. If it is differentiable, then the probability density function is

$$p_X(x) = \frac{dP_X(x)}{dx}. \quad (1.20)$$

This is always non-negative as the cdf is non-decreasing. It is also normalised

$$\int_{-\infty}^{\infty} p_X(x) dx = 1 \quad (1.21)$$

The *expectation value* or *mean* of a continuous random variable is

$$E\{X\} := \int -\infty^{\infty} x p_X(x) dx \quad (1.22)$$

The *variance* of the random variable X is

$$\text{Var}\{X\} \equiv \sigma_X^2 := E\{(X - E\{X\})^2\}. \quad (1.23)$$

The *covariance* between two random variables X and Y defined on the same probability space is

$$\text{Cov}\{XY\} := E\{(X - E\{X\})(Y - E\{Y\})\} = E\{XY\} - E\{X\}E\{Y\}. \quad (1.24)$$

Random variables for which $E\{XY\} = E\{X\}E\{Y\}$ are called *uncorrelated*. $E\{XY\} = 0$ they are called *orthogonal*.

X, Y continuous. $p_{X,Y}$ the *joint* pdf. *independent* if $p_{X,Y} = p_X(x)p_Y(y)$. The *conditional* pdf of X given $Y = y$ is

$$p_{X|Y}(x|Y = y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}. \quad (1.25)$$

A *complex* random variable $Z = X + iY$ has pdf p_Z defined as the joint pdf of the two real random variables X and Y

$$p_Z(z) := p_{X,Y}(x, y). \quad (1.26)$$

Mean

$$E\{Z\} := E\{X\} + iE\{Y\} \quad (1.27)$$

Variance

$$\text{Var}\{Z\} := E\{|Z - E\{Z\}|^2\} = E\{|Z|^2\} - |E\{Z\}|^2. \quad (1.28)$$

Stochastic Processes

Let $T \subset \mathbb{R}$. A *stochastic process* $X(t)$ is a family of random variables labelled by the numbers $t \in T$, all defined on the same probability space $(\mathcal{X}, \mathcal{O}, P)$. For each finite subset $t_1, \dots, t_n \subset T$ the random variables $X(t_1), \dots, X(t_n)$ have a joint n -dimensional cumulative distribution function F defined by

$$F_{t_1, \dots, t_n}(x_1, \dots, x_n) = P(x(t_1) \leq x_1, \dots, x(t_n) \leq x_n). \quad (1.29)$$

When T is a set of discrete points the stochastic process is called a *random sequence*, and if the variable t is time it's called a *time series*.

The stochastic process is *Gaussian* if the cdf F_{t_1, \dots, t_n} is Gaussian for any n and any $t_1, \dots, t_n \in T$. *Stationary* if for all finite dimensional cdfs and all τ

$$F_{t_1+\tau, \dots, t_n+\tau}(x_1, \dots, x_n) = F_{t_1, \dots, t_n}(x_1, \dots, x_n) \quad (1.30)$$

The *moments* of the probability distribution are

$$\mu_{m_1 \dots m_n} := E\{X(t_1)^{m_1} \dots X(t_n)^{m_n}\} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{m_1} \dots x_n^{m_n} dF_{t_1, \dots, t_n}(x_1, \dots, x_n). \quad (1.31)$$

This is an n -fold *Stieltjes integral*. When F is differentiable, it can be written in terms of the joint pdf

$$f_{t_1 \dots t_n}(x_1 \dots x_n) = \frac{\partial^n dF_{t_1, \dots, t_n}(x_1, \dots, x_n)}{\partial x_1 \dots \partial x_n} \quad (1.32)$$

and the above integral can be written

$$\mu_{m_1 \dots m_n} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} x_1^{m_1} \dots x_n^{m_n} f_{t_1 \dots t_n}(x_1 \dots x_n) dx_1 \dots dx_n. \quad (1.33)$$

The first order moment is called the *mean* value of the stochastic process and is denoted by $m(t) := E\{X(t)\} = \mu_1$, if the process is stationary this is constant. The second order moment is called the *autocorrelation function* and is denoted by $K(t, s) := E\{X(t)X(s)\}$, if the process is stationary it depends only on the time difference $K(t, s) = R(t - s)$ for some function R s.t. $R(0) > 0$, $R(-\tau) = R(\tau)$ and $|R(\tau)| < R(0)$.

For a complex stationary process, $R(t) := E\{X(t+\tau)X(t)^*\}$, and $R(-\tau) = R(\tau)^*$. Consider a complex stationary stochastic process

$$X(t) = X_0 U(t), \quad (1.34)$$

where X_0 is a random variable with $E\{X_0\} = 0$, and $U : T \rightarrow \mathbb{C}$ with $\int_{-\infty}^{\infty} U(t) dt = 0$. Stationarity gives $R(\tau) = E\{X(t+\tau)X(t)^*\} = E\{X_0 U(t+\tau)X_0^* U(t)^*\} = E\{|X_0|^2\} U(t+\tau)U(t)^*$ is

independent of t . If we take $\tau = 0$, this means $|U(t)|^2 = r^2$, a constant, thus $U(t) = r \exp(i\phi(t))$. For non-zero τ , $r^2 \exp(i\phi(t+\tau)) \exp(-i\phi(t)) = r^2 \exp(i(\phi(t+\tau) - \phi(t)))$ is independent of t , implying $\phi'(t) = \omega$ for some constant ω , and therefore $\phi = \omega t + \theta$. Therefore, $X(t) = X_0 \exp(i\omega t)$, where the r and $\exp(\theta)$ have been “absorbed” into the X_0 . The autocorrelation function $R(\tau) = E\{|X_0|^2\} \exp(i\omega\tau) = b_0 \exp(i\omega\tau)$, where $b_0 = E\{|X_0|^2\} = E\{|X_0 - E\{X_0\}|^2\}$ is the variance of X_0 .

A more general complex stochastic process can be written as a superposition

$$X(t) = \sum_{k=1}^{\infty} X_k \exp(i\omega_k t), \quad (1.35)$$

where for each complex r.v. X_k , $E\{X_k\} = 0$ and $E\{X_k X_l^*\} = 0$ for $k \neq l$. The autocorrelation function is then

$$R(\tau) = \sum_{k=1}^{\infty} b_k \exp(i\omega_k \tau), \quad (1.36)$$

with each variance $b_k < \infty$. Such a stationary process is called a *process with discrete spectrum*, where the set $\{\omega_k\}$ is the *spectrum*.

Any complex stationary stochastic process can be obtained as the limit of a sequence of processes with discrete spectra.

Cramér Representation Theorem Let $X(t)$, $-\infty < t < \infty$, be a zero-mean stochastically continuous complex stationary process. Then there exists an orthogonal process $Z(\omega)$ such that for all t , $X(t)$ may be written in the form,

$$X(t) = \int_{-\infty}^{\infty} \exp(i\omega t) dZ(\omega), \quad (1.37)$$

the integral being defined in the mean-square sense. The process $Z(\omega)$ has the following properties:

1. $E\{dZ(\omega)\} = 0 \forall \omega$
2. $E\{|dZ(\omega)|^2\} = dF(\omega) \forall \omega$, where $F(\omega)$ is the (non-normalised) integrated spectrum of $X(t)$
3. for any two distinct frequencies ω, ω'

$$\text{Cov}\{dZ(\omega), dZ(\omega')\} = E\{dZ(\omega) dZ(\omega')^*\} = 0 \quad (1.38)$$

Wiener-Khinchin Theorem A necessary and sufficient condition for $\rho(\tau)$ to be the auto-correlation function of a stochastically continuous stationary process $X(t)$ is that there exists a function $F(\omega)$ having the properties of a cdf on $(-\infty, \infty)$ such that for all τ , $\rho(\tau)$ may be expressed in the form

$$\rho(\tau) = \int_{-\infty}^{\infty} \exp(i\omega\tau) dF(\omega) \quad (1.39)$$

If F differentiable everywhere,

$$dF(\omega) = S(\omega)d\omega \quad (1.40)$$

S is called the *two-sided spectral density* of the stationary stochastic process. In this case, the correlation between the Fourier components are

$$E\{\tilde{X}(\omega)\tilde{X}(\omega')^*\} = S(\omega)\delta(\omega' - \omega) \quad (1.41)$$

In the discrete-time case, a time series $X_k = X_0U(k)$, $k \in \mathbb{Z}$, is stationary iff it is of the form

$$X_k = X_0 \exp(i\omega_0 k). \quad (1.42)$$

The autocorrelation function is $R(l) = E\{X_k X_{k+l}\}$. The Cramér representation and Wiener-Khinchin theorems become

$$\begin{aligned} x_k &= \int_{-\pi}^{\pi} \exp(i\omega k) dZ(\omega) \\ R(l) &= \int_{-\pi}^{\pi} \exp(i\omega l) dF(\omega) \end{aligned} \quad (1.43)$$

If F is differentiable with derivative S , we can write

$$R(l) = \int_{-\pi}^{\pi} \exp(i\omega l) S(\omega) d\omega, \quad (1.44)$$

which can be solved to find

$$S(\omega) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \exp(-i\omega l) R(l). \quad (1.45)$$

Hypothesis Testing

Bayesian Approach Assign costs C_{ij} , $i, j = 0, 1$ incurred by choosing hypothesis H_i when H_j is true. We say these costs are *uniform* if $C_{ij} = 0$ when $i = j$ and $C_{ij} = 1$ when $i \neq j$.

Conditional Risk R_j of a decision rule δ is

$$R_j(\delta) = C_{0j}P_j(\mathcal{R}) + C_{1j}P_1(\mathcal{R}') \quad (1.46)$$

where $P_j(\mathcal{R}) = P(X = x \in \mathcal{R} | H_j) = \int_{\mathcal{R}} p_j(x) dx$ is the probability distribution of the data when hypothesis H_j is true.

A priori probabilities/ priors $\pi_0, \pi_1 = 1 - \pi_0$ of hypotheses H_0 and H_1 .

Bayes' Risk average cost of rule δ :

$$\begin{aligned} r(\delta) &= \pi_0 R_0(\delta) + \pi_1 R_1(\delta) \\ &= \pi_0 C_{00} + \pi_1 \end{aligned} \quad (1.47)$$

Matched filter

$$x(t) = s(t) + n(t) \quad (1.48)$$

- x = data
- s = signal, deterministic
- n = noise, additive, continuous Gaussian stochastic process, zero mean.

Consider it to be continuous for theory, obtain results by suitable sampling of continuous expressions.

Observed on interval $[0, T_0]$. Can be shown that log-likelihood is (the Cameron-Martin formula)

$$\ln \Lambda[x] = \int_0^{T_0} q(t)x(t)dt - \frac{1}{2} \int_0^{T_0} q(t)s(t)dt \quad (1.49)$$

where q is the soln to

$$s(t) = \int_0^{T_0} K(t, t')q(t')dt'. \quad (1.50)$$

and K is the autocorrelation function of the noise. This means the likelihood ratio test is equivalent to comparing

$$G := \int_0^{T_0} q(t)x(t)dt \quad (1.51)$$

to a threshold G_0 . Hypothesis test expectations:

$$E_0\{G\} = 0, E_1\{G\} = \int_0^{T_0} q(t)s(t)dt. \quad (1.52)$$

$$\begin{aligned} Var\{G\} &= \int_0^{T_0} \int_0^{T_0} q(t_1)q(t_2)E\{n(t_1)n(t_2)\}dt_1dt_2 \\ &= \int_0^{T_0} \int_0^{T_0} q(t_1)q(t_2)K(t_1, t_2)dt_1dt_2 \\ &= \int_0^{T_0} q(t)s(t)dt = \rho^2 \end{aligned} \quad (1.53)$$

ρ is the *signal to noise ratio*.

Since data x is Gaussian, and G is linear in x , pdf's when signal is absent or present are

$$p_0(G) = \frac{1}{\sqrt{2\pi\rho^2}} \exp\left(-\frac{G^2}{2\rho^2}\right) p_1(G) = \frac{1}{\sqrt{2\pi\rho^2}} \exp\left(-\frac{(G - \rho^2)^2}{2\rho^2}\right) \quad (1.54)$$