

KAGRA データ解析スクール@RESCEU  
重力波天文学で便利な統計学の初歩

伊藤 洋介

平成 25 年 2 月 19 日



# 目次

<b>第 1 章</b>	<b>Fourier 変換とスペクトル理論</b>	<b>5</b>
1.1	Fourier 変換 . . . . .	5
1.2	Sampling frequency, Nyquist theorem and aliasing . . . . .	6
1.3	Frequency resolution . . . . .	8
1.4	Fourier 変換の実際 . . . . .	9
1.5	ナイーブな Powerspectrum density の推定 . . . . .	10
1.6	Powerspectrum と powerspectrum density の違い . . . . .	11
1.7	$\chi^2$ 分布の性質 . . . . .	11
1.8	ナイーブな Powerspectrum 推定の推定誤差 . . . . .	12
1.9	実際の Powerspectrum の推定 . . . . .	12
<b>第 2 章</b>	<b>仮説検定 (Hypothesis testing)</b>	<b>15</b>
2.1	判別機的能力 . . . . .	15
2.1.1	Bayesian approach: error rate . . . . .	15
2.1.2	Bayesian approach: Cost matrix . . . . .	16
2.1.3	False alarm と False dismissal . . . . .	17
2.1.4	Bayes ルール、FAR, FDR の例 . . . . .	18
2.1.5	Odds . . . . .	19
2.2	Minimax approach . . . . .	20
2.3	Neyman-Pearson approach . . . . .	20
2.4	Bayesian と frequentist(頻度主義) . . . . .	21



# 第1章 Fourier変換とスペクトル理論

重力波のデータ解析をする上で、感度曲線がどのような周波数依存性を持ち、どこにどのような線ノイズが現れるかを調べることは基本的である。たとえば連続波解析であれば、

- パルサー解析においては、パルサーのように振る舞ってしまう感度曲線上の輝線ノイズはやっかいな問題である。パルサー信号との弁別には、輝線ノイズの振幅や周波数時間依存性を調べる必要がある。
- ある周波数における感度を調べるには、その周波数における検出器ノイズパワースペクトル密度 (PSD) を正確に求めなければならない。誤差の少ない PSD の方法はどのようなものだろうか？

この章は、パワースペクトルの推定が実際どのようにしておこなわれているかを示す。その過程で、実際のデータ解析にはほとんど不可避な時間離散化の問題（離散的な時間間隔でしかデータを取得できない）ことにより生じる、スペクトル解析における問題に触れる。

この章は、“Spectrum and spectral density estimation by the Discrete Fourier transform (DFT), including a comprehensive list of window functions and some new flat-top windows.” (G. Heinzel A. Rüdiger and R. Schilling, Max-Planck-Institut für Gravitationsphysik (Albert-Einstein-Institut) Teilinstitut Hannover February 15, 2002), 「スペクトル解析」(日野幹夫 著 朝倉書店、1977年) および、「TECHI シミュレーションで学ぶデジタル信号処理」(尾知博 著、CQ 出版社、2001年) にしたがう。

## 1.1 Fourier変換

この文章では、Fourier 変換は以下のように定義する。

$$x(t) = \int_{-\infty}^{\infty} X(\omega)e^{i\omega t} d\omega, \quad (1.1)$$

$$X(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt, \quad (1.2)$$

あるいは、

$$x(t) = \int_{-\infty}^{\infty} X(f) e^{2\pi i f t} df, \quad (1.3)$$

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-2\pi i f t} dt \quad (1.4)$$

離散 Fourier 変換は、

$$x_k = \sum_{m=0}^{N-1} X_m \exp\left(\frac{2\pi i m k}{N}\right), \quad (1.5)$$

$$X_m = \frac{1}{N} \sum_{k=0}^{N-1} x_k \exp\left(\frac{-2\pi i m k}{N}\right) \quad (1.6)$$

FFTW は、規格化していない結果を与える。つまり FFTW を使って Fourier 変換と Fourier 逆変換をおこなうと、結果はもとの  $N$  倍になっている。式で言うと、式 (1.6) でなく、

$$x_k^{FFTW} = \sum_{m=0}^{N-1} X_m \exp\left(\frac{2\pi i m k}{N}\right), \quad (1.7)$$

$$Y_m^{FFTW} = \sum_{k=0}^{N-1} y_k \exp\left(\frac{-2\pi i m k}{N}\right) \quad (1.8)$$

が FFT の出力である。

## 1.2 Sampling frequency, Nyquist theorem and aliasing

本来連続的なデータを  $\Delta t$  ごとにサンプルする。サンプリング周波数は  $f_s = 1/\Delta t$  である。本来連続的なデータであるから、間引きして観測することによって、高周波数の情報を失っていると考えられる。実際、Nyquist の定理より、サンプリングしたデータは、連続波に含まれている周波数成分が  $f_{Ny} \equiv f_s/2$  より小さい周波数を持つ場合のみにもともとの連続的なデータを完全に再現できる。

たとえば、サンプリング周波数が 1kHz のとき、500Hz までの周波数を持つ波のみがデータに含まれているなら、離散化されたデータは元の連続的なデータの情報を完全に再現できる。

今、連続時系列データ信号の離散化と、データを有限時間しか取得できないことの2つの効果を式を使って見てみよう。もともとの連続時系列データが

$$x(t) = \int_{-\infty}^{\infty} X(f) \exp(2\pi i f t) df \quad (1.9)$$

離散化によって

$$x(t_k) = \int_{-\infty}^{\infty} X(f) \exp(2\pi i f t_k) df \quad (1.10)$$

両辺を離散 Fourier 変換をおこない、 $x_k = x(t_k)$  ( $k = 0, \dots, N-1$ ) の Fourier 変換を  $\tilde{X}_m = \tilde{X}(f_m = mf_s)$  と書く。

$$\begin{aligned}\tilde{X}_m &= \frac{1}{N} \sum_{k=0}^{N-1} x(t_k) \exp\left(-\frac{2\pi imk}{N}\right) \\ &= \frac{1}{N} \int_{-\infty}^{\infty} \sum_{k=0}^{N-1} \exp\left(-\frac{2\pi imk}{N}\right) X(f) \exp(2\pi ift_k) df\end{aligned}\quad (1.11)$$

$t_k = k\Delta t$  をつかって

$$= \frac{1}{N} \int_{-\infty}^{\infty} \sum_{k=0}^{N-1} \exp\left(-\frac{2\pi imk}{N}\right) X(f) \exp(2\pi ifk\Delta t) df\quad (1.12)$$

$T \equiv N\Delta t$  として

$$= \frac{1}{N} \int_{-\infty}^{\infty} X(f) \frac{\exp(2\pi i(fT - m)) - 1}{\exp\left(\frac{i2\pi}{N}(fT - m)\right) - 1} df\quad (1.13)$$

$f_{res} = 1/T$  を使って

$$\equiv \int_{-\infty}^{\infty} X(f) W(f - mf_{res}) df\quad (1.14)$$

$f_{res} = 1/T$  ( $f_{res}$  は周波数分解能) より  $Nf_{res} = f_s$  に注意すると、関数  $W(f - mf_{res})$  は  $f = mf_{res} + (\text{整数})f_s$  に振幅 1 のするどいピークを持つ無限個の山を持つ関数である。また、 $fT$  が整数であるような波はある周波数ビンでのみ値を持つが、 $fT$  が整数にならない波はある周波数ビンにピークに周囲の周波数ビンにも振幅を持つ (この現象をスペクトラルリーケージ spectral leakage と呼ぶ。 )。これが有限時間積分の効果である。

無限個の山を持つことから、 $X(f)$  が  $1/T$  に比べてゆるやかな周波数依存性を持つ関数のとき、

$$\begin{aligned}\int_{-\infty}^{\infty} W(f - mf_{res}) &= \frac{1}{N} \sum_{k=0}^{N-1} \int_{-\infty}^{\infty} \exp\left(-\frac{2\pi imk}{N}\right) \exp(2\pi ifk\Delta t) df \\ &= 1\end{aligned}\quad (1.15)$$

に注意すると、以下のように近似できる。

$$\tilde{X}(f_m) \simeq \sum_{n=-\infty}^{\infty} X(f_m + nf_s)\quad (1.16)$$

この式の意味は、離散化時系列データの周波数ビン  $f_m$  の振幅は、もともとの連続時系列データの周波数ビン  $f_m$  にあった振幅のみならず、サンプリング周波数の整数倍の周波数ビンにあった振幅が足し合わされたものであるということである。

逆に離散化時系列データの周波数ビン  $f_m$  の振幅がもともとの連続時系列データの周波数ビン  $f_m$  の振幅と一致するためには、もともとの連続時系列

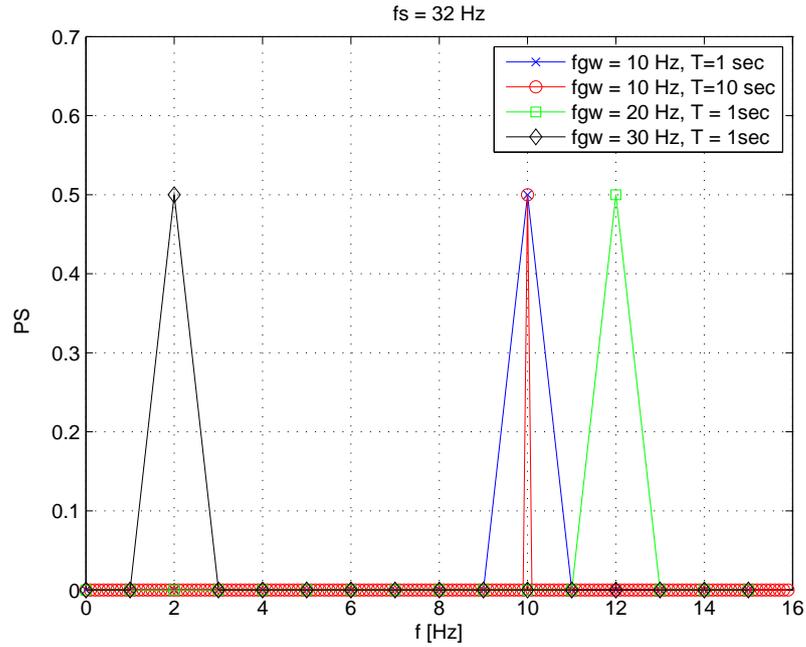


図 1.1:

データは周波数  $f_s/2$  より高周波数の周波数成分を持っていてはいけない。  
 $f_{Ny} = f_s/2$  をナイキスト周波数と呼ぶ。

連続的なデータがナイキスト周波数よりも高周波数成分を持つ場合、ナイキスト周波数よりも高周波数の成分はサンプリングによって失われるわけではなく、サンプリングされたデータの低周波数成分に影響を与える。これを aliasing と呼ぶ。

図 1.1 はエイリアシングの影響を示す。本来 20Hz のサイン波が  $16 - (20-16) = 12$  Hz に、本来 30Hz のサイン波が  $16 - (30-16) = 2$  Hz に現れている。また、同時に積分時間の増大が周波数分解能の増加を与えていることも分かる(赤い線と青い線の比較)。

### 1.3 Frequency resolution

サンプル数が  $N$  のとき、周波数分解能は、

$$f_{res} = \frac{f_s}{N} = \frac{1}{T} \quad (1.17)$$

で与えられる。

<sup>1</sup> $f_s$  でなくて  $f_s/2$ なのは次の理由による。離散化時系列データのフーリエ変換  $\tilde{X}(f)$  は  $[-f_s/2, f_s/2]$  に値を持つが、連続時系列データが持っている可能性のある  $f = f_s/2 + \delta f$  の周波数成分は、離散時系列データの  $f' = f_s/2 + \delta f - f_s = -f_s/2 + \delta f$  の周波数成分に影響を与え得ることができ、 $-f_s/2 < f' < f_s/2$  だから。

## 1.4 Fourier 変換の実際

有限時間のデータしか扱えないので、データ両端の不連続性はスペクトルに影響を与え、求めたい本来のスペクトルが得られない。これを防ぐために、実際のデータ解析では Window 関数を導入する。

窓関数の例として、Hannig window がある。

$$w_j = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi j}{N} \right) \right] \quad (1.18)$$

$J = 0, 1, \dots, N-1$ 。Hanning window をかけたサイン波は図 1.2 のように両端がゼロに近づく。<sup>2</sup>

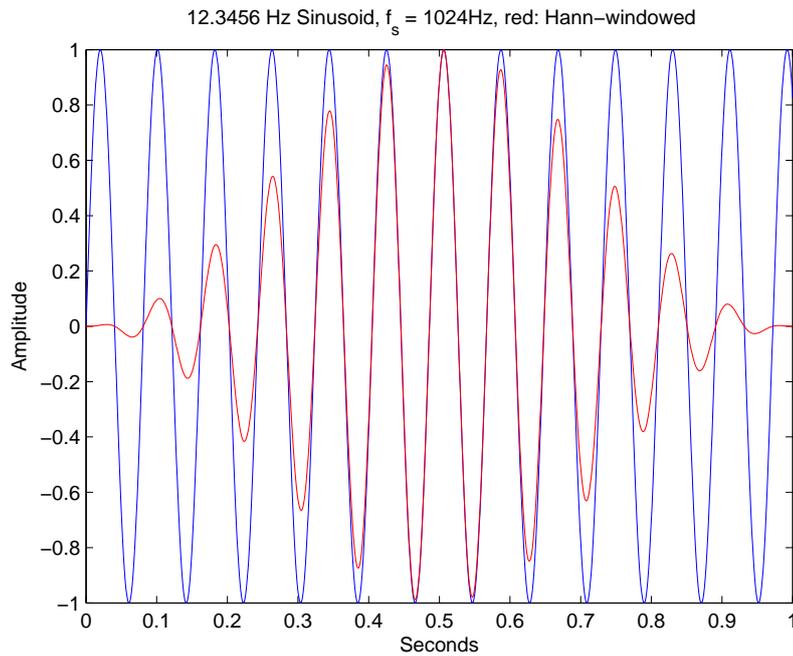


図 1.2:

実際のデータ解析では、以下の Fourier 変換を求める。

$$Y_m = \frac{1}{N} \sum_{k=0}^{N-1} w_k x_k \exp \left( \frac{-2\pi i k m}{N} \right) \quad (1.19)$$

窓関数を導入したことによって、周波数分解能は一般に下がる。

図 1.3 は窓関数を導入した影響を示す。窓関数の導入によって、spectral leakage が抑えられる一方、有効的な周波数分解能は悪くなる。

<sup>2</sup>Window によっては、必ずしも両端をゼロにしない場合もある。

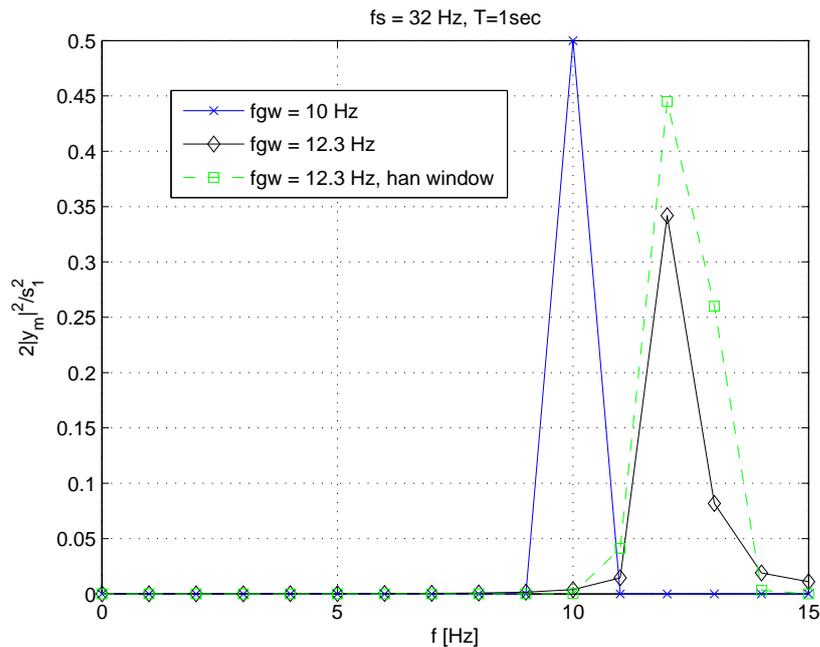


図 1.3:

## 1.5 ナイーブな Powerspectrum density の推定

便利のため以下の 2 つの和を導入する。

$$S_1 = \sum_{j=0}^{N-1} w_j, \quad (1.20)$$

$$S_2 = \sum_{j=0}^{N-1} w_j^2. \quad (1.21)$$

Powerspectrum density は以下のように推定する。

$$PSD(f_m = m f_s) = \frac{2|Y_m|^2}{f_s S_2} \quad (1.22)$$

$m = 0, 1, \dots, N/2$ . <sup>3</sup>本来 Fourier 変換は負の周波数成分を与えるが、 $x_k w_k$  が実数列であるので、 $Y_m^* = Y_{N-m}$  が成り立つため正周波数の成分だけ考えれば良い。2 倍したのは、正周波数成分のみを考えたためである。

PSD は  $[|Y_m|^2]/\text{Hz}$  の次元を持つ。

Powerspectrum は以下のように推定する。

$$PS(f_m = m f_s) = \frac{2|Y_m|^2}{S_1^2} \quad (1.23)$$

<sup>3</sup> $S_2$  で割るのは、 $Y_m$  がノイズのみを含む場合に PSD の振幅が時間によらないようにするためである。すなわち、 $\langle |Y_m|^2 \rangle = \langle \mathcal{F}(wy)\mathcal{F}(wy) \rangle = S_2$  である。

$m = 0, 1, \dots, N/2$ .<sup>4</sup>

PS は  $[|Y_m|^2]$  の次元を持つ。

## 1.6 Powerspectrum と powerspectrum density の違い

長時間積分によって信号対雑音比 (Signal to noise ratio, SNR) は  $\sqrt{T}$  だけ増えると言う。これは積分によって信号が  $T$ 、ノイズが  $\sqrt{T}$  だけ増えた結果である。これらの長時間積分の効果は PSD や PS でどのように現れるのか？

ガウス雑音にサイン波を入れたものを Fourier 変換してみると分かるが、長時間積分によって、PS では、サイン波の振幅が変化せず、ノイズレベルが長時間積分によって  $1/T$  下がる。一方、PSD では、サイン波の振幅が  $T$  増大し、ノイズレベルは一定になる。

一般に Powerspectrum と言ったときには、実は PSD を意味していることが多いので注意が必要である。単位が異なるので注意すると良い。

図 1.4 と図 1.5 はランダムガウシアンノイズ + 10Hz サイン波信号の PSD と PS を示す。PS の場合、ノイズのパワーが積分時間に反比例して減少する一方、信号のパワーは一定である。PSD の場合、ノイズのパワーが積分時間に無関係になる一方、信号のパワーは積分時間に比例して増大する。

## 1.7 $\chi^2$ 分布の性質

少し寄り道して、 $\chi^2$  分布の性質について述べる。

独立に正規分布  $N(0, 1)$  にしたがう  $n$  個の乱数  $x_k$  ( $k = 1, \dots, n$ ) を考えると、

$$\sum_{k=1}^n x_k^2 \quad (1.24)$$

は自由度  $n$  の  $\chi^2$  分布にしたがう。

自由度  $\chi^2$  分布の平均は  $n$ 、分散は  $2n$  であることが知られている。いわゆる reduced  $\chi^2$  が「1 に近いと良い」言われるのは

$$\left\langle \frac{\chi^2}{n} \right\rangle = 1 \quad (1.25)$$

という性質があるからである。reduce  $\chi^2 = 1.01$  でも、自由度が  $2 \times 10^8$  なら

$$\frac{\chi^2/n - 1}{\sqrt{V(\chi^2)/n}} = \frac{0.01}{\sqrt{2/n}} = 100 \quad (1.26)$$

となって (大雑把に言って)  $100\sigma$  で棄却すべきである。

<sup>4</sup> $S_1^2$  で割るのは、 $Y_m$  が輝線 (単一のサイン波) のみを含む場合に PS の振幅が時間によらないようにするためである。

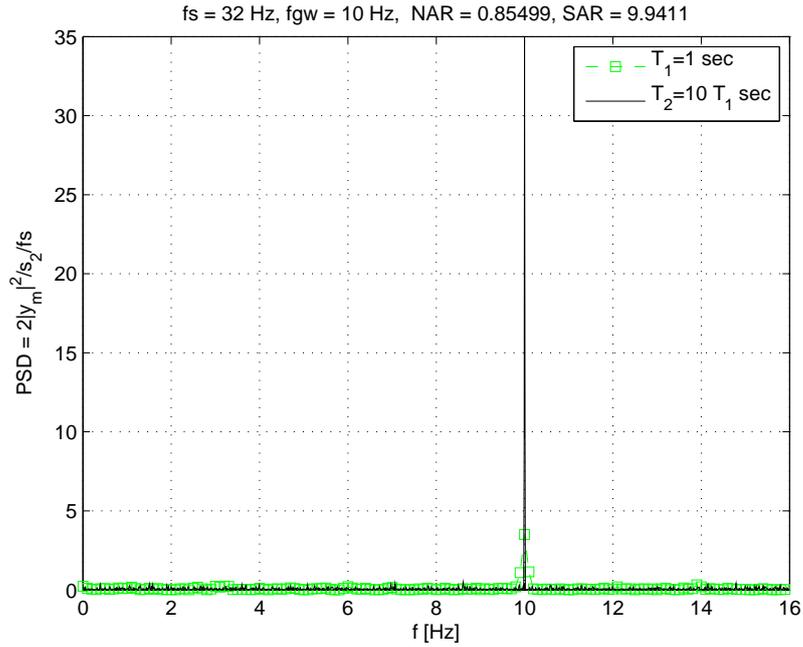


図 1.4: SAR: signal amplitude ratio, NAR: noise amplitude ratio

## 1.8 ナイーブな Powerspectrum 推定の推定誤差

PS、PSD ともに  $|Y_m|^2 = Y_{Rm}^2 + Y_{Im}^2$  に比例する。 $Y_{Rm}$  と  $Y_{Im}$  が独立に正規分布  $N(0, \sigma^2)$  にしたがうなら、PS、PSD を適当にスケールした量は自由度 2 の  $\chi^2$  分布にしたがう。 $PSD = A\chi^2$  としよう。

自由度 2 の  $\chi^2$  分布の分散は、4 なので、期待されるスペクトルの推定誤差は

$$\frac{E(PSD)}{\sqrt{V(PSD)}} = 1 \quad (1.27)$$

よって、100パーセントになる。

このようなナイーブな推定スペクトルの推定誤差は、観測時間を増やしても改善されない。

## 1.9 実際の Powerspectrum の推定

実際のデータ解析においては、スペクトルは averaging と overlapping method などを使用して求める。これは、周波数分解能を犠牲にする代わりに、スペクトルの推定誤差を小さくする方法である。

今、観測時間  $T$  のデータがあったとすると、これを時間幅  $T/M$  の  $M$  個のセグメントにわけると、それぞれのセグメントで PSD を求め、 $M$  個のセグメ

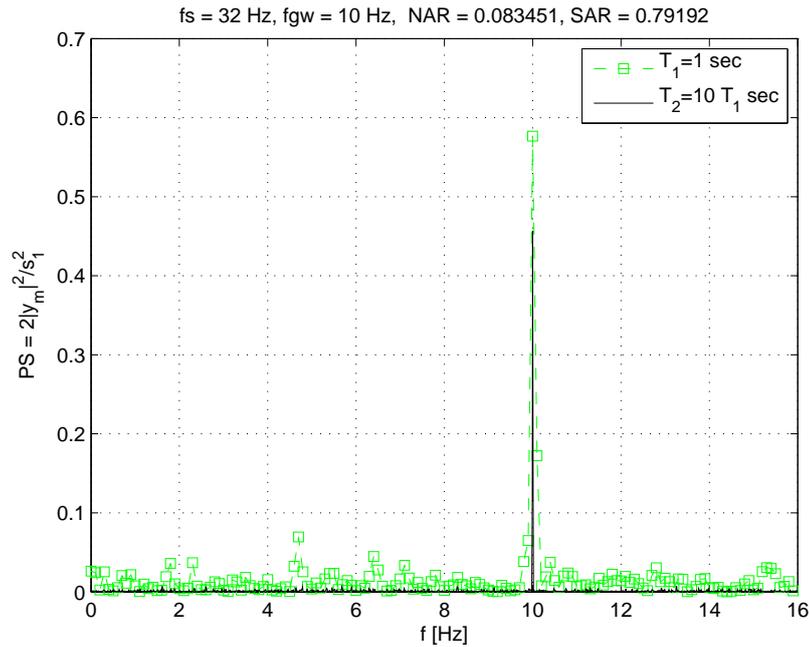


図 1.5:

ントで平均すれば、

$$PSD = \frac{1}{M} \sum_i PSD_i \quad (1.28)$$

周波数分解能は  $f_s/M$  になるが、PSD の推定誤差は  $\sqrt{1/M}$  だけ良くなる。これが averaging である。

ここで、窓関数を導入したことによって、各データセグメントの両端付近のデータがほとんど有効に活用されていないことを思い出そう。このような窓関数の使用によって "捨てられた" データは、各データセグメントを overlap させることによって復活させることができる。これが、overlapping 法である。

図 1.6 は、Welch の overlapping method を用いた PSD と periodgram の方法 (ナイーブな PSD 推定) の結果を比較したものである。Welch の方法では、信号 (10Hz のサイン波) の振幅が power leakage によって小さく評価されている一方で、ノイズの PSD の分散は小さくなっている。計算はともに Matlab を用いておこなっている。

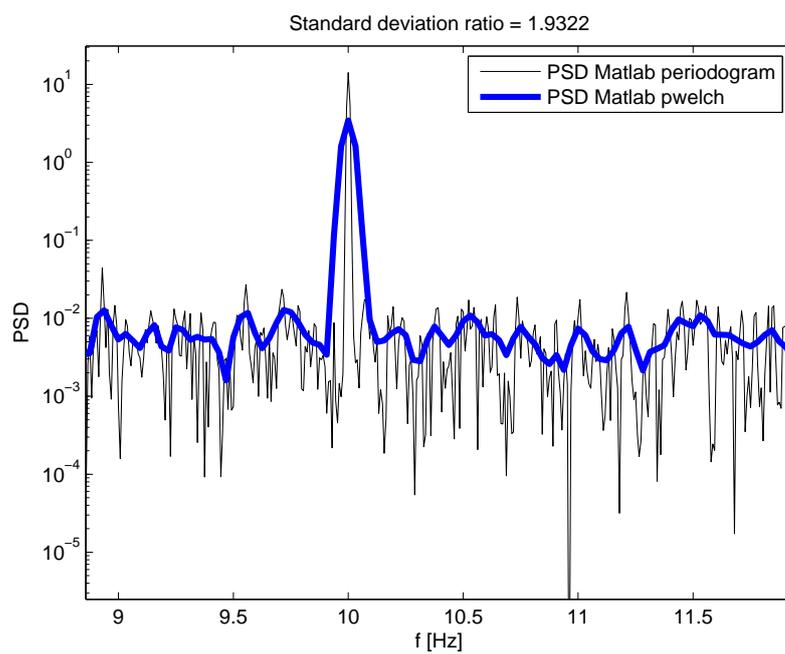


図 1.6:

## 第2章 仮説検定 (Hypothesis testing)

この章は『Rで学ぶデータサイエンス5 パターン認識』(金森敬文、竹之内高志、村田昇 著、共立出版、2009年)、“Gravitational-Wave Physics and Astronomy An Introduction to Theory, Experiment, and Data Analysis” (Jolien D. E. Creighton, Warren G. Anderson, Wiley-VCH, 2011), “Gravitational Wave Data Analysis” (B. F. Schutz ed., P. 73-94 B. F. Schutz ed., Kluwer Academic Publications, 1989), および “Analysis of Gravitational-Wave Data” (Jaranowski and Królak, Cambridge University Press, 2009) にしたがう。

### 2.1 判別機有能力

データから、検出  $H_1$  と非検出  $H_0$  を判断する方法を考える。データ  $x$  が得られたときに、ラベル  $y \in \{0, 1\}$  を返す判別機  $F(x)$  を考える。最適な  $F(x)$  を探す問題を考えよう。データ  $x$  と  $y$  は確率分布  $p(x, y)$  にしたがうとする。 $y$  が確率変数であるのは、たとえば、本来  $y_0$  が正しい統計的性質であるときに、得られた  $x$  から必ずしも  $y_0$  が得られるとは限らないからである。

#### 2.1.1 Bayesian approach: error rate

データ  $x$  の存在する空間 (入力空間) のうち、 $F$  によってある  $y$  に写像される領域を  $R_y(F)$  とする。

$$R_y(F) = \{x | F(x) = y\}. \quad (2.1)$$

このとき error rate は以下の式で定義される。Error rate は、本来性質  $Y$  をもつ母分布から得られたデータ  $X$  を判別機  $F(\cdot)$  にかけてときに、 $F(X) \neq Y$  を得るような確率を、すべての  $Y$  について足したものである。

$$\begin{aligned} Pr(F(X) \neq Y) &= \sum_y Pr(X \in R_y(F), Y \neq y) \\ &= \sum_y \int_{R_y(F)} \sum_{y' \neq y} p(x, y') dx \\ &= \int_{R_0(F)} p(x, 1) dx + \int_{R_1(F)} p(x, 0) dx \end{aligned} \quad (2.2)$$

= (本来  $y=1$  なのに、0 と判断する確率) + (本来  $y=0$  なのに、0 と判断する確率)

$$= 1 - \sum_y \int_{R_y(F)} p(x, y) dx \quad (2.3)$$

= 1 - (正しい判断をする確率)

$$(2.4)$$

ここで、

$$F^*(x) = \max_y p(x, y) = \max_y p(y|x)p(x) \quad (2.5)$$

を考えると、

$$\begin{aligned} & Pr(F(X) \neq Y) - Pr(F^*(x) \neq Y) \\ &= \sum_y \int_{R_y(F^*)} p(x, y) dx - \sum_y \int_{R_y(F)} p(x, y) dx \\ &= \int_{R_0(F^*)} p(x, 0) dx - \int_{R_0(F)} p(x, 0) dx \\ &+ \int_{R_1(F^*)} p(x, 1) dx - \int_{R_1(F)} p(x, 1) dx \\ &= \sum_y \sum_{y'} \int_{R_y(F^*) \cap R_{y'}(F)} (p(x, y) - p(x, y')) dx \geq 0. \end{aligned} \quad (2.6)$$

最後の等号は下から上に逆方向に式変形して見ると示しやすい。

したがって、条件付き確率分布  $P(y|x)$  (尤度関数でもある) を最大にする判別機  $F^*(\cdot)$  が最も error rate を小さくするという意味において最適である。このような  $F^*$  を金森らの本では Bayesian ルールと呼んでいる。

### 2.1.2 Bayesian approach: Cost matrix

非負のコスト行列  $C_{ij}$  を考える。 $C_{ij}$  は  $H_j$  が正しいときに  $H_i$  を選択することによるコストを表す。今、データが  $R$  にあるとき、 $H_0$  を選択し、 $R'$  にあるとき  $H_1$  を選択するような判別機  $F$  を考えよう。また、 $P_j(R)$  を  $H_j$  が正しいときにデータが  $R$  をとるような確率とする。

ここで条件付きリスク関数を

$$R_j(F) = C_{0j}P_j(R) + C_{1j}P_j(R') \quad (2.7)$$

( $j$  について和は取らない) と定義する。たとえば、 $C_{00} = C_{11} = 0$  と置くと、 $R_0(F) = C_{10}P_0(R')$  となって、仮説  $H_0$  が正しいときに、判別機  $F$  が仮説  $H_1$  を選択してしまうような領域  $R'$  をデータが取ってしまう確率  $P_0(R')$  に重み  $C_{10}$  をかけたものが仮説  $H_0$  が正しいとしたときの条件付きリスク関数  $R_0(F)$  である。

全リスク関数を求めるために、仮説に対する事前確率分布を  $\pi_0, \pi_1$  とする。  
全リスク関数は

$$\begin{aligned}
r(F) &= \pi_0 R_0(F) + \pi_1 R_1(F) \\
&= \pi_0 C_{00} P_0(R) + \pi_0 C_{10} P_0(R') + \pi_1 C_{01} P_1(R') + \pi_1 C_{11} P_1(R) \\
&= \pi_0 C_{00} + \pi_1 C_{01} + \pi_0 (C_{10} - C_{00}) P_0(R') + \pi_1 (C_{11} - C_{01}) P_1(R') \\
&= \pi_0 C_{00} + \pi_1 C_{01} \\
&\quad + \int_{R'} (\pi_0 [C_{10} - C_{00}] p_0(x) + \pi_1 [C_{11} - C_{01}] p_1(x)) \\
&= \pi_0 C_{00} + \pi_1 C_{01} \\
&\quad + \int_{R'} p_0(x) \pi_0 (C_{01} - C_{11}) (\lambda - \Lambda(x)) \tag{2.8}
\end{aligned}$$

ここで、

$$\Lambda(x) = \frac{p_1(x)}{p_0(x)}, \tag{2.9}$$

$$\lambda = \frac{\pi_0 C_{10} - C_{00}}{\pi_1 C_{01} - C_{11}}. \tag{2.10}$$

Cost matrix の方法では、全リスクが最小になるように判別機  $F$  を決定する。これは、尤度比  $\Lambda(x)$  が  $\lambda$  より大きい  $x$  において仮説  $H_1$  を選択することに等しい。(つまりデータ  $x$  が領域  $R'$  にあるときには常に  $\Lambda(x) > \lambda$  とするように  $F$  を決める。)

この判別機  $F : \{ \Lambda(x) > \lambda \text{ then take } H_1 \}$  を Jaranowki and Królak, Schutz らは Bayes ルールと呼ぶ。

### 2.1.3 False alarm と False dismissal

False alarm probability と detection probability を

$$P_{FA} = \int_{R'} p_0(x) dx, \tag{2.11}$$

$$P_D = \int_{R'} p_1(x) dx \tag{2.12}$$

と定義する。このとき、

$$r(F) = \pi_0 C_{00} + \pi_1 C_{01} + \pi_0 (C_{10} - C_{00}) P_{FA} + \pi_1 (C_{11} - C_{01}) P_D \tag{2.13}$$

例として、 $\pi_0 = \pi = 1/2$  として、さらに

$$C_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.14}$$

を採用すると、

$$r(F) = \frac{1}{2} (P_{FA} + P_{FD}) \tag{2.15}$$

また、このとき  $\lambda = 1$  となる。つまり、作業仮説が間違っていた時のコストが同じならば、尤度比（所与のデータに対して2つの仮説のもっともらしさ）が等しいときに判断を変えるしきい値になる。

### 2.1.4 Bayes ルール、FAR, FDR の例

データ  $X$  を得たとき、以下の2つの仮説を検定する。

$$H_0 : X = N, \quad (2.16)$$

$$H_1 : X = N + \mu \quad (2.17)$$

$N$  は正規分布  $N(0, \sigma^2)$  にしたがって、 $\mu > 0$  とする。 $d = \mu/\sigma$  を導入すると尤度比は、

$$\begin{aligned} \Lambda(x) &= \frac{p_1(x)}{p_0(x)} = \exp\left(-\frac{(x-\mu)^2}{2\sigma^2} + \frac{x^2}{2\sigma^2}\right) \\ &= \exp\left(\frac{x}{\sigma}d - \frac{d^2}{2}\right) \end{aligned} \quad (2.18)$$

図2に信号があるときと無いときの確率密度関数  $p_0, p_1$  を示す。 $(p_1$  については  $\mu = 4, \sigma = 1, p_0$  については  $\sigma = 1$  としている。)

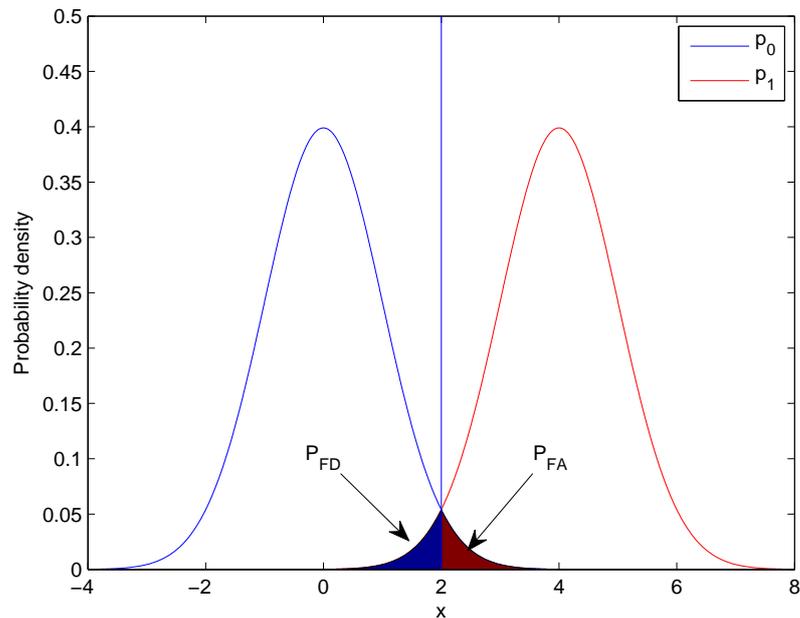


図 2.1:

Bayes ルール  $\Lambda(x) > \lambda$  は

$$x > x_0 \equiv \sigma \left( \frac{\ln \lambda}{d} + \frac{d}{2} \right) \quad (2.19)$$

と等価である。とくに  $C_{01} = C_{10} = 1/2, C_{00} = C_{11} = 0$  のとき、 $\lambda = 1$  より  $x_0 = \mu/2$  である。図 2 には  $\mu = 4, \sigma = 1$  のときの FAR と FDR が図示されている。

### 2.1.5 Odds

Bayesian approach と関係している考え方にオッズがある。A が与えられたときに B である確率を  $P(B|A)$ 、B で無い確率を  $P(NotB|A)$  とする。オッズ  $O(B|A)$  は

$$O(B|A) = \frac{P(B|A)}{P(NotB|A)} \quad (2.20)$$

で与えられる。今、ベイズの定理

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} \quad (2.21)$$

を使うと、

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \\ &= \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|NotB)P(NotB)} \\ &= \frac{\Lambda(B|A)}{\Lambda(B|A) + P(NotB)/P(B)} \end{aligned} \quad (2.22)$$

ただし、

$$\Lambda(B|A) = \frac{P(A|B)}{P(A|NotB)} \quad (2.23)$$

は尤度比である。オッズは

$$O(B|A) = \frac{P(B|A)}{P(NotB|A)} = \frac{P(B)}{P(NotB)} \frac{P(A|B)}{P(A|NotB)} = O(B)\Lambda(B|A) \quad (2.24)$$

と書けることがわかる。

今 A をデータ、B を信号が存在するという仮説  $H_1$  とする。尤度比は、信号がする場合としない場合にデータ A を得る確率の比、オッズはデータを得たときに、信号を存在すると判断する確率としないと判断する確率の比を与えている。データが何らかのパラメータに依存するとき、尤度比を最大にするパラメータはオッズを最大にするパラメータでもある。

## 2.2 Minimax approach

$\pi_0$  が分からない場合、Cost 関数による方法において、リスク関数を

$$r(\pi_L, F) = \max_{0 \leq \pi_0 \leq 1} r(\pi_0, F) \quad (2.25)$$

とする。ただし、 $\pi_L$  は、リスク最悪の  $\pi_0$  (条件  $0 \leq \pi_L \leq 1$  の下で、 $r(\pi, F)$  を最大にする  $\pi$ ) である。すなわち、安全マージンをとって、リスクが最大になる事前確率分布関数を探す。そしてこのリスクを最小にする  $F$  を探す方法を Minimax approach と呼ぶ。

$$r(\pi_L, F_{\pi_L}^*) = \min_F \max_{0 \leq \pi_0 \leq 1} r(\pi_0, F) \quad (2.26)$$

Minimax approach が与えるのは、最悪を想定した場合の最善の判別方法と言える。このように  $\pi_L$  を探索した後で、最適な  $F$  を探すという方法 (同時に最適かしくなくても良いという方法) が解を与えることは Minimax 原理によって保証されている<sup>1</sup>。

## 2.3 Neyman-Pearson approach

Neyman-Pearson approach は所与の False alarm probability に対して、detection probability を最大化する方法である。

$$P_{FA} = \int_{R'} p_0(x) dx, \quad (2.27)$$

$$P_D = \int_{R'} p_1(x) dx. \quad (2.28)$$

仮説検定の問題では、誤検出率 (False Alarm Rate) は type I の誤りの確率または有意度 (Significance)、一方誤非検出率 (False Dismissal Rate) は type II の誤りの確率と呼ばれ、検出率 (Detection Probability) = 1 - FDR は仮説検定テストのパワーと呼ばれる。

最適な FAR と FDR を求める方法として receiver-operating-characteristics (ROC) がある。ROC では以下のような設定を考える。データがそこにあったときに判別機が  $H_0$  と  $H_1$  のどちらが良いかを判断する領域  $R, R'$  を定義するパラメータ (複数可)  $c$  が存在するとする。たとえば、信号対雑音比があるしきい値を越えるときに  $H_1$  であると判断する場合、 $c$  はこの信号対雑音比に対するしきい値である。 $DR$  はパラメータ  $c$  を通して  $FAR$  の関数となるので、これを縦軸・横軸に図示したグラフが Receiver-Operating-Characteristic Curve (受信者動作曲線) である。

なお、以上の設定のもとで、FAR (誤検出率) を指定したときに、検出率  $DR$  が最大になるような  $c$  を探す方法は古典的な Neyman-Pearson の方法と呼ばれる。

<sup>1</sup>ただし、リスク  $r$  の最大値が  $r$  の取りうる境界に位置することがあり、この場合の処理を適切にする必要がある。

## 2.4 Bayesian と frequentist(頻度主義)

Bayesian 流のパラメータ推定は、パラメータ自体が確率分布すると考える。パラメータがある値領域に含まれる確率がこれこれであると言う。

それに対して頻度主義は、パラメータは固定されていると考える。例えば、存在しているある連星の質量は、ある値を取るのであって、確率分布しているわけではないと考える。頻度主義の立場では、パラメータ推定は、以下の言明になる。

頻度主義のパラメータ推定：あるパラメータを持つ分布からデータを得ると、得られたデータをある確率でこれこれの領域に存在する。

たとえば、連続波の検出統計量  $F$  では、以下の言明をする。

頻度主義のパラメータ推定例：重力波の振幅が  $h_a$  であると仮定すると、検出統計量  $F$  が実際得られた検出統計量の値  $\hat{F}$  以下の値を得る確率は、 $xx\%$ (例えば  $xx = 5\%$ ) である。また、重力波の振幅が  $h_b$  であると仮定すると、検出統計量  $F$  が実際得られた検出統計量の値  $\hat{F}$  以上の値を得る確率は、 $yy\%$ (例えば  $yy = 5\%$ ) である。

このとき、 $h_a$  が重力波振幅の  $1 - xx\%$  上限を与え、 $h_b$  が  $1 - yy\%$  の下限を与える。