

**BAB III**  
**MENYELESAIKAN MASALAH REGRESI YANG TIDAK LINIER**  
**DENGAN ANALISIS REGRESI FOURIER**

**3.1 Pengantar**

Model ARIMA digunakan untuk analisis data deret waktu pada kategori data berkala 'tunggal', atau sering dikategorikan model-model *univariate*. Untuk data-data dengan kategori deret berkala berganda (*multiple*), tidak bisa dilakukan analisis menggunakan model ARIMA, oleh karena itu diperlukan model-model *multivariate*. Model-model yang masuk kelompok *multivariate* analisisnya lebih rumit dibandingkan dengan model-model *univariate*. Pada model *multivariate* sendiri bisa dalam bentuk analisis data bivariat (yaitu, hanya data dua deret berkala) dan dalam bentuk data *multivariate* (yaitu, data terdiri lebih dari dua deret berkala). Salah satu model *multivariate* yaitu: Analisis Regresi Fourier. Sebagai ilustrasi, berikut ini merupakan beberapa contoh masalah regresi Fourier:

**Bidang Pertanian**

Misalkan diketahui bahwa terdapat hubungan tidak linier antara jumlah taraf pemupukan Fosfat pada suatu lahan pertanian sebagai variabel bebas  $X$  dengan produksi padi yang dihasilkan suatu lahan pertanian tersebut sebagai variabel terikat  $Y$ . Secara umum produksi padi akan meningkat cepat bila pemberian Fosfat ditingkatkan dari taraf rendah ke taraf sedang. Tetapi ketika pemberian dosis Fosfat diteruskan hingga taraf tinggi, maka tambahan dosis Fosfat tidak lagi diimbangi kenaikan hasil, sebaliknya terjadi penurunan hasil.

### Bidang Pengairan

Misalkan diketahui bahwa terdapat hubungan tidak linier antara jumlah kelarutan oksigen pada suatu sungai sebagai variabel bebas  $X$  dengan kualitas air yang dihasilkan suatu sungai tersebut sebagai variabel terikat  $Y$ . Secara umum kualitas air akan semakin baik bila kelarutan oksigen meningkat dari taraf rendah ke taraf sedang. Tetapi ketika kelarutan oksigen terus meningkat hingga taraf tinggi, maka peningkatan kelarutan oksigen tidak lagi diimbangi kenaikan hasil, sebaliknya terjadi penurunan hasil.

Berdasarkan beberapa ilustrasi di atas, maka masalah regresi Fourier didefinisikan sebagai berikut :

#### Definisi 3.1 :

Misalkan diketahui bahwa  $X$  dan  $Y$  mempunyai hubungan tidak linear (diasumsikan  $x$  diketahui)

$$Y = f(X) = m + \sum_{k=1}^N (\alpha_k \cdot \cos(2k\pi x) + \beta_k \cdot \sin(2k\pi x)) + \varepsilon \quad [3.1]$$

Jika terdapat nilai dari  $Y$ , katakanlah  $y_0$  yang tidak diketahui dan nilai dari  $X$  (atau sampel acak berukuran  $N$  dengan  $k$  titik untuk nilai dari  $X$ ) yang berkorespondensi dengan  $y_0$  dapat diamati, maka masalah untuk menentukan nilai  $y_0$  disebut sebagai masalah regresi Fourier (Thibos, 1993; 2000; 2003).

### 3.2 Asumsi Statistika

Untuk memeriksa pengaruh dari *noise* terhadap koefisien Fourier, harus diketahui sesuatu tentang sifat dasar statistika dari proses *noise*. Dari berbagai

macam jenis *noise* yang terjadi sebenarnya, hal yang paling sederhana dan mudah dikerjakan dari inti matematika ada dua sifat:

1. *Noise* bersifat *additive*. Dengan kata lain, seperti yang diindikasikan pada persamaan [3.2.1], nilai sampel sama dengan jumlah linier dari *signal* yang berupa sampel dan *noise*.

$$\begin{aligned} \mathbf{v} &= (v_1, v_2, v_3, \dots, v_N) \\ &= (f(x_1), f(x_2), f(x_3), \dots, f(x_N)) \end{aligned} \quad [3.2.1]$$

$$\mathbf{v} = (f(x_1) + n_1, f(x_2) + n_2, f(x_3) + n_3, \dots, f(x_N) + n_N) \quad [3.2.2]$$

2. Setiap sampel dari *noise*  $n_j$  *independent* dari proses *noise* (atau populasi) mean nol dan variansi  $\sigma^2$ .

Untuk tujuan dari pengantar pembahasan ini, akan diasumsikan bahwa kedua kondisi tersebut terpenuhi.

Satu implikasi dari kedua asumsi tersebut adalah *noise* yang diketahui bersifat *independent* dari *signal*nya. Dengan kata lain, *noise* tidak akan bertambah besar atau kecil karena hanya *signal* bertambah besar atau kecil. Implikasi lainnya adalah bahwa tiap sampel dari *noise* bersifat *independent* secara statistika dari semua nilai *noise* yang lain dan sebelumnya tergambar dari populasi yang mempunyai sifat statistika seperti semua sampel lainnya. Artinya, pada jargon statistikawan, nilai *noise*  $n_j$  disebut *random variabel* dan kumpulan dari semua *random variabel* ini dikatakan *independent* dan *berdistribusi identik*.

Hal ini berlawanan dengan kasus dimana, dijelaskan, *noise* bertambah besar pada akhir percobaan dari awal percobaan yang melanggar asumsi bahwa *noise* itu

berdistribusi identik. Contoh hal yang tidak *independent* akan terjadi jika nilai *noise* dengan waktu  $t_2$  bergantung pada *noise* di  $t_1$  beberapa waktu sebelumnya.

Karena setiap titik sampel  $v_j$  pada data vektor diasumsikan sebagai jumlah dari *signal*  $f(x_j)$  dan sampel  $n_j$  dari *noise*, ini berarti bahwa  $v_j$  sebenarnya merupakan *random variabel*. Selain itu, karena *noise* mempunyai mean nol dan *noise* bersifat *additive*, itu mengakibatkan bahwa mean dari  $v_j$  sama dengan  $f(x_j)$ . Dengan kata lain, karena *signal* diasumsikan sebagai *noise-free*, variansi  $v_j$  sama dengan variansi  $\sigma^2$  dari *noise*. Kita tuliskan kesimpulan matematikanya sebagai berikut

$$\begin{aligned} v_j &= f(x_j) + n_j && \text{(sampel = signal + noise)} \\ \bar{v}_j &= f(x_j) && \text{(mean sampel = signal)} \\ \text{Var}(v_j) &= \sigma^2 && \text{(variansi sampel = variansi noise)} \end{aligned} \quad [3.2.3]$$

### 3.3 Mean dan Variansi Koefisien Fourier untuk *Signal Noise*

Mengingat kembali dari persamaan [2.3.7.6] dan [2.3.7.7] bahwa koefisien Fourier yang diperoleh untuk data vektor  $\mathbf{v}$  dinyatakan sebagai fungsi basis trigonometri berupa

$$a_k = \frac{2}{D} \cdot \sum_{j=0}^{D-1} v_j \cos k\theta_j \quad \dots \theta_j = \frac{2\pi x_j}{L} \quad [2.3.7.6]$$

$$b_k = \frac{2}{D} \cdot \sum_{j=0}^{D-1} v_j \sin k\theta_j \quad \dots \theta_j = \frac{2\pi x_j}{L} \quad [2.3.7.7]$$

dan untuk fungsi basis kompleks eksponensial oleh

$$c_k = \frac{1}{D} \cdot \sum_{j=0}^{D-1} v_j \exp(ik\theta_j) \quad \dots \theta_j = \frac{2\pi x_j}{L} \quad [2.3.7.7a]$$

Sekarang berdasarkan persamaan [3.2.3] tiap data vektor adalah jumlah *signal* vektor ditambah *noise* vektor. Ini artinya bahwa koefisien yang dihitung oleh persamaan [2.3.7.6, 2.3.7.7] dapat dianggap sebagai estimasi koefisien Fourier sebenarnya dari *signal* tunggal. Untuk melihat hal ini, substitusikan persamaan [3.2.3] ke dalam [2.3.7.6] untuk mendapatkan

$$\begin{aligned}\hat{a}_k &= \frac{2}{D} \cdot \sum_{j=0}^{D-1} (f(x_j) + n_j) \cos k\theta_j \\ &= \frac{2}{D} \cdot \sum_{j=0}^{D-1} f(x_j) \cos k\theta_j + \frac{2}{D} \cdot \sum_{j=0}^{D-1} n_j \cos k\theta_j \\ &= a_k + \varepsilon_k\end{aligned}\quad [3.3.1]$$

Dimana variabel  $\hat{a}$  adalah nilai perhitungan koefisien Fourier. Hasil ini menyatakan bahwa  $\hat{a}_k$  adalah estimasi koefisien  $a_k$  yang sebenarnya dengan errornya dinyatakan oleh  $\varepsilon_k$ .

Hasil yang sama digunakan pada koefisien sinus. Hasil yang bersesuaian untuk koefisien kompleks adalah

$$\begin{aligned}\hat{c}_k &= \frac{1}{D} \cdot \sum_{j=0}^{D-1} (f(x_j) + n_j) \exp(ik\theta_j) \\ &= \frac{1}{D} \cdot \sum_{j=0}^{D-1} f(x_j) \exp(ik\theta_j) + \frac{1}{D} \cdot \sum_{j=0}^{D-1} n_j \exp(ik\theta_j) \\ &= c_k + \varepsilon_k\end{aligned}\quad [3.3.2]$$

Langkah selanjutnya adalah menyelidiki sifat statistika dari estimasi koefisien Fourier. Karena estimasi ini dinyatakan sebagai jumlah kuantitas deterministik  $a_k$  dan *random variabel*  $\varepsilon_k$ , perlu memfokuskan penyelidikan pada bentuk random error. Dari teori probabilitas diketahui bahwa jika  $Y$  merupakan *random variabel* dengan mean  $\mu$  dan variansi  $\sigma^2$ , dan jika  $s$  merupakan konstanta

skalar, maka *random variabel* yang baru  $Z=sY$  dengan mean  $s\mu$  dan variansi  $s^2\sigma^2$ . Ini mengakibatkan bahwa bentuk umum  $(2/D)(\cos k\theta_j)n_j$  pada persamaan [3.3.1] adalah *random variabel* dengan mean 0 dan variansi  $(2/D)^2(\cos k\theta_j)^2\sigma^2$ .

Hasil lainnya dari teori probabilitas adalah jika  $Y$  dan  $Z$  *independent, random variabel* berdistribusi identik dengan masing-masing mean  $\mu, \nu$  dan variansi  $\sigma^2, \tau^2$ , maka *random variabel* yang baru  $W=Y+Z$  dengan mean  $\mu+\nu$  dan variansi  $\sigma^2+\tau^2$ . Singkatnya, mean jumlah dan variansi jumlah. Penggunaan hasil ini pada penyajian akhir kedua persamaan [3.3.1] lihat bahwa  $\varepsilon_k$  adalah jumlah dari *random variabel*  $D$ , masing-masing memiliki mean 0 dan variansi  $(4\sigma^2/D^2)\cos^2 k\theta_j$ , akibatnya variansi dari  $\varepsilon_k$  adalah

$$\begin{aligned} \text{Var}(\varepsilon_k) &= \sum_{j=0}^{D-1} (2/D)^2 (\cos k\theta_j)^2 \sigma^2 = \frac{4\sigma^2}{D^2} \sum_{j=0}^{D-1} \cos^2 k\theta_j \\ &= \frac{4\sigma^2}{D^2} \cdot \frac{D}{2} = \frac{2\sigma^2}{D} \quad \text{untuk } k \neq 0 \\ &= \frac{4\sigma^2}{D^2} \cdot D = \frac{4\sigma^2}{D} \quad \text{untuk } k = 0 \end{aligned} \quad [3.3.3]$$

Penyederhanaan persamaan [3.3.3] berdasarkan faktanya bahwa panjang kuadrat dari sampel fungsi  $\cos$  sama dengan  $D/2$ , kecuali ketika  $k=0$ , pada kasus dimana sama dengan  $D$ . Munculnya  $k=0$  sebagai kasus khusus dirasa agak janggal secara matematika. Hal tersebut dapat dihindari dengan membagi koefisien  $a_0$  oleh  $\sqrt{2}$  yang bertujuan untuk perhitungan variansi seperti yang telah dilakukan pada teorema Parseval (lihat persamaan [3.3.3a]).



$$\begin{aligned}
 \text{energi} &= \int_0^{2\pi} v^2(t) dt = \frac{\pi a_0^2}{2} + \pi \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \\
 \text{power} &= \frac{\text{energi}}{\text{waktu}} = \frac{1}{2\pi} \int_0^{2\pi} v^2(t) dt \\
 &= \left(\frac{a_0}{2}\right)^2 + \frac{1}{2} \sum_{k=1}^{\infty} (a_k^2 + b_k^2) = \left(\frac{a_0}{2}\right)^2 + \sum_{k=1}^{\infty} \frac{m_k^2}{2} \quad [3.3.3a] \\
 &= \frac{1}{2} \left\{ \left(\frac{a_0}{\sqrt{2}}\right)^2 + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) \right\} \\
 &= \frac{1}{2} \{ \text{panjang vektor Fourier}^2 \}
 \end{aligned}$$

Untuk koefisien Fourier kompleks, bentuk umum  $(1/D)\exp(ik\theta_j)n_j$  adalah *random variabel* dengan mean 0 dan variansi  $\left((1/D)\exp(ik\theta_j)\right)^2 \sigma^2$ . Jumlah dari  $D$  sebagai *random variabel*, diberikan rumus berikut untuk variansi *noise*

$$\begin{aligned}
 \text{Var}(\varepsilon_k) &= \frac{\sigma^2}{D^2} \cdot \sum_{j=0}^{D-1} (\exp(ik\theta_j))^2 \\
 &= \frac{\sigma^2}{D^2} \cdot D = \frac{\sigma^2}{D} \quad [3.3.3b]
 \end{aligned}$$

Suatu keuntungan dari koefisien Fourier kompleks adalah bahwa bentuk konstantanya tanpa kasus khusus.

Dari hasil ini, dapat disediakan nilai untuk satu, dua statistik tepat dari estimasi koefisien Fourier. Dari persamaan [3.3.1] diketahui bahwa *random variabel*  $\hat{a}_k$  adalah jumlah koefisien deterministik  $a_k$  dan *random variabel*  $\varepsilon_k$  dengan mean nol dan variansi seperti yang diberikan pada persamaan [3.3.2].

Akibatnya,

$$\begin{aligned}
 \text{Mean}(\hat{a}_k) &= a_k \\
 \text{Var}(\hat{a}_k) &= \frac{2\sigma^2}{D} && \text{untuk } k \neq 0 \\
 &= \frac{4\sigma^2}{D} && \text{untuk } k = 0
 \end{aligned}
 \tag{3.3.4}$$

dan dengan persamaan yang sama untuk estimasi koefisien sin. Persamaan yang sesuai untuk koefisien Fourier kompleks adalah

$$\begin{aligned}
 \text{Mean}(\hat{c}_k) &= c_k \\
 \text{Var}(\hat{c}_k) &= \frac{\sigma^2}{D}
 \end{aligned}
 \tag{3.3.4a}$$

Perhatikan bahwa karena variansi dari  $a_0$  adalah  $4\sigma^2/D$  maka variansi dari  $a_0/2$ , seperti menyatakan variansi dari mean, sama dengan  $\sigma^2/D$  dan juga standar deviasi dari mean adalah  $\sigma/\sqrt{D}$ . (Hasil ini lebih jelas untuk  $c_0$ ) Ini sejenis dengan hasil dari statistika dasar. Pada statistika, standar deviasi dari mean untuk  $D$  nilai data biasanya disebut standar error dari mean dan sama dengan  $\sigma/\sqrt{D}$ , dimana  $\sigma$  adalah standar deviasi populasi dari data yang diketahui.

Ringkasan, di bawah asumsi *additive, independent noise*, variansi dari semua estimasi koefisien Fourier trigonometri (kecuali  $a_0$ ) sama dengan variansi *noise* dikali  $2/D$ . Variansi dari semua estimasi koefisien Fourier kompleks sama dengan variansi *noise* dikali  $1/D$ . Hal ini mengakibatkan bahwa cara untuk memperkecil estimasi variansinya adalah dengan memperbesar  $D$ , jumlah titik sampel. Prosedur tersebut disebut *signal-to-noise ratio* (SNR) sering digunakan untuk menghitung realibilitas dari *signal*. SNR adalah bagian dari koefisien Fourier yang dinyatakan sebagai perbandingan dari mean (sebagai contoh,  $a_k$ )



dengan standar deviasi  $\sigma\sqrt{2}/\sqrt{D}$ . Dari definisi tersebut, SNR dari estimasi koefisien Fourier bertambah senilai  $\sqrt{D}$  dan berkurang senilai  $\sigma$  pada proporsinya, jumlah dari *noise*.

### 3.4 Probabilitas Distribusi dari Koefisien Fourier untuk *Signal Noise*

Mean dan variansi berguna untuk meringkas statistik dari *random variabel*, tetapi karakter yang lebih lengkapnya adalah pada bentuk probabilitas distribusinya. Diberikan signal deterministik, probabilitas distribusi koefisien Fourier dihitung untuk  $D$  sampel dari *waveform noise* yang bergantung pada probabilitas distribusi dari *noise* yang ditambahkan. Seperti kasus yang biasa terjadi pada analisis dasar dari *signal noise*, diasumsikan dari sekarang bahwa *noise* mempunyai kepadatan probabilitas Gauss (atau normal),  $N(\mu, \sigma^2)$ , dari mean  $\mu$  dan variansi  $\sigma^2$ . Di bawah asumsi tersebut, probabilitas  $P$  yang merupakan *signal noise* yang terletak pada range  $(a, b)$  yang diberikan pada daerah di bawah fungsi kepadatan probabilitas Gauss antara batasnya

$$P = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} dx \quad [3.4.1]$$

Beberapa anggapan untuk asumsi Gauss dapat dipertimbangkan. Pertama, banyak penelitian *noise* secara fisika dapat dimodelkan dengan baik oleh bagian fungsi probabilitas ini. Hal itu tidak mengejutkan karena teorema limit pusat dari probabilitas teori probabilitas menyatakan bahwa jumlah dari bilangan yang besar dari variabel *independent* cenderung ke Gauss tanpa memperhatikan probabilitas distribusi dari individu variabelnya, alasan lain adalah bahwa asumsi tersebut

membuat masalah yang ada mudah dihitung. Suatu hasil dari teori probabilitas adalah bahwa distribusi Gauss dekat sekali dengan penjumlahan, artinya bahwa sejumlah nilai dari suatu bilangan *random variabel* Gauss bersisa Gauss juga. Karena error variabel  $\varepsilon_k$  adalah nilai jumlah dari variabel *noise*, jika *noise* adalah Gauss maka sekaligus merupakan estimasi dari koefisien Fourier. Singkatnya, *noise* Gauss menghasilkan koefisien Fourier Gauss.

Distribusi Gauss hanya memiliki dua parameter, mean dan variansi, dimana diketahui dari hasil yang lebih umum dari bagian 3.3 di atas. Sehingga bisa diringkas hasil sebelumnya dengan menyatakan bahwa estimasi koefisien Fourier berdistribusi seperti *random variabel normal* (yaitu Gauss) beserta mean dan variansinya juga (baca “ $N$ ” berdistribusi normal)

$$\begin{aligned}\hat{a}_k &\square N(a_k, 2\sigma^2/D) \\ \hat{b}_k &\square N(b_k, 2\sigma^2/D) \\ \hat{c}_k &\square N(c_k, 2\sigma^2/D)\end{aligned}\quad [3.4.2]$$

*Random variabel* yang menarik lainnya adalah pempoweran harmonik ke- $k$ . seperti yang ditunjukkan pada persamaan [3.3.3a] dimana *power signal* adalah satu per kuadrat amplitudonya. Selanjutnya, estimasi *power signal*  $\rho_k$  di harmonik ke- $k$  adalah

$$\rho_k = \hat{m}_k^2/2 = (\hat{a}_k^2 + \hat{b}_k^2)/2 \quad [3.4.3]$$

Dari teori probabilitas diketahui bahwa jika  $X$  distandarkan *random variabel* Gauss dengan mean nol dan variansi satu, yaitu jika  $X \square (0,1)$ , maka variabel  $Z=X^2$  berdistribusi seperti variabel chi-kuadrat dengan derajat kebebasan 1.

Artinya,  $Z \sim \chi_1^2$ . Hasil ini berguna pada konteks yang ada jika kita standarkan estimasi koefisien Fourier pada persamaan [3.4.2] oleh pengurangan mean dan membaginya dengan standar deviasi, lalu dikuadratkan, bentuk standar koefisien Fourier berdistribusi  $\chi^2$

$$\left( \frac{\hat{a}_k - a_k}{\sqrt{2\sigma^2/D}} \right)^2 \sim \chi_1^2 \quad [3.4.4]$$

dan pernyataan yang sama, tetap dipergunakan untuk koefisien  $b_k$ . sekarang, dari teori probabilitas kita juga mengetahui bahwa jika *random variabel* X dan Y keduanya berdistribusi seperti chi-kuadrat dengan derajat kebebasan 1, maka variabel  $Z=X+Y$  berdistribusi chi-kuadrat dengan derajat kebebasan 2. Sehingga

$$\frac{(\hat{a}_k - a_k)^2 + (\hat{b}_k - b_k)^2}{2\sigma^2/D} \sim \chi_2^2 \quad [3.4.5]$$

Prosedur hasil terakhir ini adalah bahwa hal tersebut memberikan kita cara untuk menguji keberadaan *signal* pada bagian frekuensi harmonik. Pada kasus ini, hipotesis nolnya adalah koefisien Fourier dari harmonik ke- $k$  bernilai nol. Dengan asumsi tersebut, persamaan [3.4.5] menjadi

$$\frac{\hat{a}_k^2 + \hat{b}_k^2}{2\sigma^2/D} \sim \chi_2^2 \quad [3.4.6]$$

Pengombinasian hasil tersebut dengan definisi dari *power signal* pada persamaan [3.3.1] diketahui bahwa

$$\frac{\rho_k}{\sigma^2/D} = \frac{\text{Power harmonik ke } -k}{\text{Rata-rata power noise}} \sim \chi_2^2 \quad [3.4.7]$$

Penggunaan hasil terakhir ini, perhatikan bahwa penyebut pada sisi kanan persamaan [3.4.7] adalah total *power noise* yang diteliti dibagi dengan jumlah koefisien Fourier yang ditentukan. Interpretasi tersebut berasal dari pemahaman teorema Parseval pada persamaan [2.3.10.1] dan faktanya bahwa  $\sigma^2$  adalah nilai ekspektasi dari variansi sampel  $s^2$  yang diperoleh dari suatu bagian data vektor yang terdiri dari titik sampel  $D$  dari *noise* yang diteliti. Sehingga, berdasarkan interpretasi tersebut, penyebut dari persamaan [3.4.7] adalah jumlah ekspektasi dari *power noise* per koefisien, dengan kata lain, rata-rata *power* pada *spectrum power noise*.

Perbandingan pada sebelah kiri selanjutnya merupakan jumlah ukuran dari *power* untuk harmonik ke- $k$ , menormalkan dengan rata-rata *power noise*. Jika disebut hal ini sebagai kuantitas masing-masing *power* relatif dari harmonik ke- $k$ , maka persamaan [3.4.7] merupakan *power* relatif harmonik ke- $k$  yang berdistribusi  $\chi^2$  di bawah hipotesis nol yaitu ada *power signal* bernilai nol pada harmonik ke- $k$ .

Pada sub-bab berikutnya kita akan menggunakan hasil persamaan [3.4.7] tersebut untuk membangun statistik uji dari hipotesis nol. Dalam hal ini, memanfaatkan sebelumnya bahwa mean variabel  $\chi^2$  sama dengan jumlah derajat kebebasan variabelnya, dan variansinya sama dengan dua kali mean. Karena *power signal*  $\rho_k$  adalah skala variabel  $\chi^2$  di bawah hipotesis nol, diketahui bahwa

$$\begin{aligned} \text{Mean}\left(\frac{P_k}{\sigma^2/D}\right) &= 2, & \Rightarrow & \text{Mean}(\rho_k) = 2\sigma^2/D \\ \text{Var}\left(\frac{P_k}{\sigma^2/D}\right) &= 4, & \Rightarrow & \text{Var}(\rho_k) = 4\left(\sigma^2/D^2\right) \end{aligned} \quad [3.4.8]$$

Perhatikan bahwa standar deviasi dari  $\rho_k$ , merupakan akar kuadrat dari variansinya, sama dengan mean, sehingga SNR=1 pada kasus seperti ini. Biasanya SNR rendah tidak diinginkan, disebut metode SNR yang diperbaiki seperti yang dijelaskan di bawah.

### 3.5 Distribusi Koefisien Fourier untuk Random Signal

Kadang-kadang penelitian suatu *signal* di bawah penyelidikan tidak semuanya memiliki komponen deterministik, tetapi sesederhana proses random. Satu contoh seperti *electroencephalogram*, tegangan yang sangat kecil yang terekam oleh elektroda terletak pada tengkorak. Contoh lainnya adalah fluktuasi normal dari diameter pupil, atau “hippus” yang kadang disebut sebagai kondisi alam bawah sadar. *Signal* tersebut disebut *stochastic* karena mereka tidak mudah menyesuaikan ke dalam model persamaan [3.3.2] sebagai jumlah komponen deterministik ditambah komponen random *noise*, kecuali kita sederhanakan semua bentuk *signal*nya bersama-sama.

Analisis Fourier dari stochastic, atau random, *signal* biasanya berlaku pada bentuk polar karena random *signal* sebenarnya menghilangkan fase yang penting, hanya menyisakan porsi jarak *spectrum* yang terpakai. Selanjutnya, mengacu dari plot  $m_k$ , jarak koefisien Fourier, lebih terbiasa memplot  $\rho_k = m_k^2/2$ , yang merupakan *power* komponen harmonik. Berdasarkan hal itu, graf *power* dari tiap komponen Fourier seperti fungsi frekuensi yang disebut *power spectrum*. *Power spectrum* dari prose random yang memenuhi asumsi bahwa tiap sampel *independent*, dan berdistribusi identik, setiap sampel lainnya akan mempunyai

*power spectrum* datar. Hal ini karena, seperti yang ditunjukkan pada persamaan [3.4.7] untuk kasus *signal* bernilai nol, *power* pada tiap harmoniknya sama. Penelitian *noise* yang mempunyai *power spectrum* datar disebut “white” *noise*, analogi dengan *spectrum* cahaya. Hasilnya berakibat bahwa jika penelitian *noise* itu menggunakan cara yang menghasilkan *spectrum* yang tidak datar, dengan kata lain, “colored” *spectrum*, maka sampel *noise* tidak lagi *independent* dan berdistribusi identik. Akibatnya, korelasi antar sampel pun tidak lagi berdistribusi identik.

Pada bagian akhir sub-bab 3.4 observasi tersebut, tentang *signal deterministik*, standar deviasi  $\rho_k$  sama dengan mean, sehingga SNR=1. Arti “*signal*” pada konteks ini adalah nilai estimasi  $\rho_k$ , *power* komponen harmonik ke- $k$  dari random *signal*. Serupa dengan nilai SNR rendah yang tidak diinginkan dan juga mean untuk perbaikan realibilitas yang diperoleh. Satu metode untuk mengulang proses sampling *waveform* dan menghitung *power spectrum*. Jika  $M$  *spectra* ditambahkan secara bersamaan, *power* pada tiap harmonik akan menjadi jumlah dari *random variabel*  $M$ , masing-masing berdistribusi  $\chi^2$  dengan derajat kebebasan 2. Sehingga total *power* akan berdistribusi  $\chi^2$  dengan derajat kebebasan  $2M$ , yang mempunyai mean  $2M$  dan standar deviasi  $2\sqrt{M}$ . Rata-rata *power* adalah *total power* dibagi dengan  $M$ .

$$\bar{P}_k \square \frac{1}{M} \sum_N \chi_2^2 \square \frac{1}{M} \chi_{2M}^2 \quad [3.5.1]$$

mean, variansi, dan SNRnya adalah



$$\begin{aligned}
 \text{mean}(\bar{p}_k) &= \frac{1}{M} \cdot 2M = 2 \\
 \text{variansi}(\bar{p}_k) &= \left(\frac{1}{M}\right)^2 \cdot 4M = \frac{4}{M} \\
 \text{SNR} &= \frac{\text{mean}}{\sqrt{\text{variansi}}} = \frac{2}{2/\sqrt{M}} = \sqrt{M}
 \end{aligned}
 \tag{3.5.2}$$

Selanjutnya disimpulkan bahwa realibilitas dari estimasi *power spectrum* dihasilkan oleh penyetaraan individu *spectra*  $M$  bertambah pada proporsinya sebesar  $\sqrt{M}$ .

Teknik yang ekivalen adalah menyetarakan vektor sampel  $M$  dan selanjutnya menghitung *power spectrum* dari mean data vektor. Karena tiap komponen dari data vektor bertambah realibilitasnya pada proporsi sebesar  $\sqrt{M}$ , jadi, dapat menghitung *power spectrum*.

### 3.6 Analisis Regresi Fourier

Pada teori statistika tentang regresi, metode yang biasa digunakan berupa pendekatan model *goodness of fit* untuk menghitung statistik  $S$  yang didefinisikan oleh perbandingan

$$S = \frac{\text{variansi data model}}{\text{variansi residu}}
 \tag{3.6.1}$$

Dalam hal ini, gagasan dasarnya adalah *waveform* memiliki variansi dikarenakan dua faktor: fungsi deterministik dan random error. Pada regresi linier, sebagai contoh, fungsi deterministik yang mendasarinya diasumsikan sebagai garis lurus, yang mempunyai dua parameter bebas: *slope* dan *intercept*. Begitu model memprediksikan dengan pasti jumlah variansi dari data (pembilang pada [3.6.1]),

tetapi beberapa variansi residu (penyebut [3.6.2] ) tidak didapatkan dari model. Jika  $S$  membesar, akibatnya adalah modelnya bisa digunakan untuk menghitung variansi datanya. Pada konteks Fourier, dua faktor tersebut kita kenal sebagai *signal* dan *noise*. Statistik  $S$  persis seperti SNR yang didefinisikan sebelumnya karena pembilangnya adalah ukuran kekuatan *signal*nya dan penyebutnya bergantung pada jumlah *noise* yang ada.

Penggunaan umum suatu statistik seperti  $S$  biasa digunakan untuk uji hipotesis tentang kecocokan model yang disebut uji parametrik. Agar bisa mengembangkan kegunaan dari bentuk uji yang seperti ini, perlu dipahami distribusi probabilitas  $S$ . Distribusi yang seperti ini mungkin sudah kita kenal sebagai distribusi F-Snedecor (nama kehormatan dari Bangsawan R.A. Fisher) yang digunakan ketika pembilang persamaan [3.6.1] adalah variabel  $\chi^2$  dengan derajat kebebasan  $a$ , dibagi dengan  $a$ , dan penyebutnya adalah variabel  $\chi^2$  dengan derajat kebebasan  $b$ , dibagi dengan  $b$ . seperti berikut,

$$\frac{\frac{\chi_a^2}{a}}{\frac{\chi_b^2}{b}} \square F_{a,b} \quad [3.6.2]$$

Dijelaskan hasil dari sub-bab 3.2 hal tersebut ditunjukkan sebagai *power* harmonik yang berdistribusi  $\chi^2$  ketika *noise* tunggal Gauss diketahui, itu tidak sulit untuk menemukannya, uji F kadang-kadang bisa digunakan untuk uji *goodness of fit* dari model deret Fourier. Hartley (1949) telah mengembangkan uji yang seperti ini dan metodenya dijelaskan di bawah.

Dari versi teorema Parseval yang ditunjukkan persamaan [2.3.9.6] variansi  $D$  titik sampel sama dengan jumlah *power* dari komponen harmonik yang sesuai.

$$\frac{1}{D} \sum_{k=1}^D Y_k^2 - m^2 = \frac{1}{2} \sum_{k=1}^{D/2} a_k^2 + b_k^2 = \sum_{k \neq 0} c_k^2 = \sum_{k=1}^{D/2} \rho_k \quad [2.3.10.1]$$

Selanjutnya, jika model Fourier dalam pembahasannya termasuk semua  $D$  komponen harmonik, maka itu akan menghitung semua variansi datanya, sehingga variansi residunya nol, dan model akan sesuai dengan data yang sebenarnya. Dengan kata lain, jika hanya beberapa harmonik yang termasuk dalam model, maka harmonik yang diabaikan dihitung sebagai variansi residunya. Pada kasus ini, kita dapat membentuk statistik seperti  $S$  untuk memutuskan apakah modelnya cocok atau tidak.

Untuk melihat kegunaanya, kita masukkan ke model Fourier hanya harmonik ke- $k$  saja. Dengan kata lain, asumsikan semua harmonik yang lainnya adalah *noise*. Menurut persamaan [2.3.10.1] di atas, variansi yang dihitung dari model ini adalah  $\rho_k$ , yang didapatkan pada persamaan [3.4.7] bahwa jika  $\rho_k$  dinormalkan yaitu membaginya dengan ekspektasi jumlah *power* di bawah hipotesis nol yang hanya diketahui *noise*-nya saja, maka “*power* relatif” juga berdistribusi  $\chi^2$  dengan derajat kebebasan 2.

$$\frac{\rho_k}{\sigma^2/D} \square \chi_2^2 \quad [3.6.3]$$

Dengan jelas nilai ini akan menunjukkan pembilang dari statistik-F. untuk mendapatkan penyebut yang dibutuhkan, perhatikan kembali bahwa terdapat  $D-3$  harmonik residu pada kasus ini. Jumlah total *power* relatif residunya adalah

jumlah  $R=(D-3)/2$  variabel acak, tiap distribusi  $\chi^2$  dengan derajat kebebasan 2, yang selanjutnya berdistribusi  $\chi^2$  dengan derajat kebebasan  $2R=D-3$

$$\sum_{j=1}^R \frac{\rho_j}{\sigma^2/D} \square \chi_{2R}^2 \quad [3.6.4]$$

Sekarang untuk merumuskan statistik Hartley, bagi tiap variabelnya dengan masing-masing bilangan derajat kebebasan dan bentuk perbandingannya

$$H = \frac{\frac{\rho_k}{2\sigma^2/D}}{\frac{1}{2R} \sum_{j=1}^R \frac{\rho_j}{2\sigma^2/D}} = \frac{\text{power relatif harmonik ke-}k}{\text{rata-rata power relatif residu}} \square F_{2,2R} \quad [3.6.5]$$

Sehingga, nilai  $\sigma$  yang tidak diketahui yang terdapat pada pembilang dan penyebut selanjutnya dapat dihilangkan

$$H = \frac{\rho_k}{\frac{1}{R} \sum_{j=1}^R \rho_j} \square F_{2,2R} \quad [3.6.6]$$

Selanjutnya, uji Hartley dari hipotesis nol yang *power signal* pada harmonik ke- $k$  nya adalah nol ini akan menolak hipotesis nol jika  $H > F_{2,2R}$ . Untuk mengolahnya biasa digunakan tingkat signifikansi 1% atau 5%, bergantung nilai F yang sesuai dari tabel distribusi-F. Jika perhitungan statistik ujinya lebih dari nilai pada tabel, tolak hipotesis nol yang *power signal* harmoniknya adalah nol. Tingkat signifikansi diinterpretasikan sebagai probabilitas penolakan kesalahan hipotesis nol.

### 3.7 Interval Konfidensi

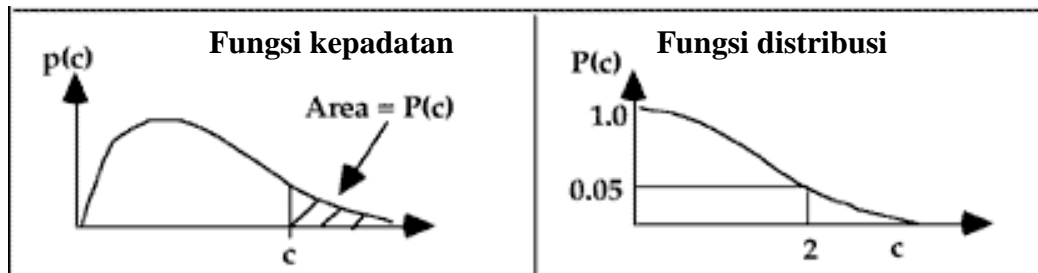
Suatu hasil yang sangat penting dari statistik yang biasa kita gunakan adalah spesifikasi batas konfidensi untuk mean sampel dari populasi. Jika kita melihat kembali hasilnya, itu akan berguna sebagai pendahuluan untuk memperoleh interval konfidensi untuk koefisien Fourier. Sehingga  $\bar{x}$  adalah mean dari  $N$  sampel dan akan menyatakannya dalam konfidensi 95% (yaitu kurang dari 5% kekeliruannya), mean populasi  $\mu$  yang sebenarnya berada pada range

$$\bar{x} - A \leq \mu \leq \bar{x} + A \quad [3.7.1]$$

Pertanyaannya adalah, apakah nilai  $A$  itu? Jawaban yang tepat untuk menjawab pertanyaan tersebut adalah dua kali standar error mean. Untuk melihat kenapa ini benar, lihat kembali bahwa standar mean sampel  $t$ , yang dikenal sebagai statistik- $t$  Student,

$$t = \frac{|\bar{x} - \mu|}{s/\sqrt{N}} \quad [3.7.2]$$

distribusi- $t$  dengan derajat kebebasan  $N-1$ . Pada persamaan ini,  $s$  adalah standar deviasi sampel dan  $s/\sqrt{N}$  adalah standar error dari mean. Distribusi- $t$  Student sudah diketahui sebagai fungsi distribusi yang diparameterkan oleh jumlah derajat kebebasan. Contohnya seperti tampak pada Gambar 3.1. Sebelah kirinya adalah probabilitas fungsi kepadatan dan sebelah kanannya adalah 1 dikurangi distribusi probabilitas kumulatif, yaitu daerah di bawah fungsi kepadatan memenuhi beberapa kriteria  $c$ , sebagai fungsi  $c$ .



**Gambar 3.1 Distribusi-t Student**

Nilai eksak  $c$  yang diperoleh dari  $P(c)$  sampai pada 5% bergantung  $D$ , tetapi untuk sampel  $c$  yang besar nilainya konvergen ke 2. Ini artinya bahwa probabilitas  $t$  yang lebih dari 2 hanya 5%. Berdasarkan persamaan [3.7.2], diperoleh

$$\text{Prob} \left( \frac{|\bar{x} - \mu|}{s(\bar{x})} > 2 \right) = 5\% \quad [3.7.3]$$

pertidaksamaan seperti persamaan di atas dapat menghasilkan bentuk yang serupa dari persamaan [3.7.1] yaitu

$$\text{Prob}(\bar{x} - 2s(\bar{x}) < \mu < \bar{x} + 2s(\bar{x})) = 95\% \quad [3.7.4]$$

Dengan kata lain, batas konfidensi 95% untuk  $\mu$  adalah  $\bar{x} \pm 2s(\bar{x})$ .

Berdasarkan penjelasan tersebut, diketahui dari persamaan [3.6.6] yaitu perbandingan Hartley dari *power* harmonik dengan *power* residunya merupakan distribusi-F di bawah hipotesis nol. Jika diletakkan batas hipotesis nol maka harus kembali lagi pada bentuk pembilang dari persamaan [3.4.4], sehingga

$$H = \frac{(\hat{a}_k - a_k)^2 + (\hat{b}_k - b_k)^2}{\frac{1}{R} \sum_{j=1}^R p_j} \square F_{2,2R} \quad [3.7.5]$$



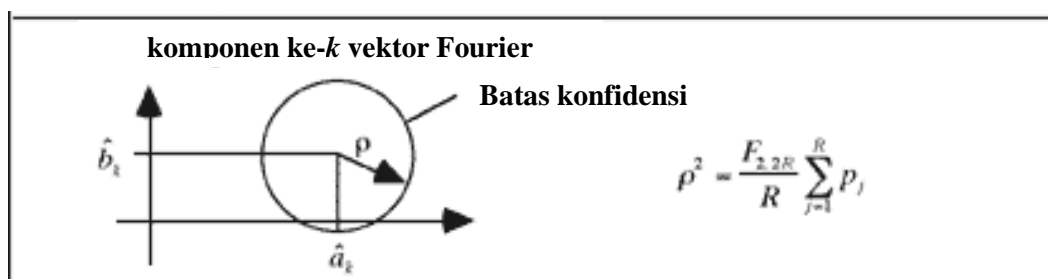
Analogi persamaan [3.7.4] selanjutnya adalah

$$\text{Prob} \left( \frac{(\hat{a}_k - a_k)^2 + (\hat{b}_k - b_k)^2}{\frac{1}{R} \sum_{j=1}^R P_j} > F_{2,2R} \right) = 5\% \quad [3.7.6]$$

Pertidaksamaan yang mendefinisikan batas kefidensi dengan interpretasi geometri sederhana seperti ditunjukkan pada Gambar 3.2. menggambarkan lingkaran berpusat pada titik  $(\hat{a}_k, \hat{b}_k)$  dan dengan jari-jari  $\rho$  yang diberikan oleh

$$\rho^2 = \frac{F_{2,2R}}{R} \sum_{j=1}^R P_j \quad [3.7.7]$$

maka dengan kefidensi 95% kita dapat menyatakan bahwa nilai sebenarnya dari koefisien Fourier  $(a_k, b_k)$  bersesuaian dengan titik di sekitar lingkaran. Jika lingkaran tersebut mengandung koefisien awalnya, maka *power* dari bentuk harmonik ke- $k$  tidak berbeda secara signifikan dengan nol.



**Gambar 3.2 Batas Kefidensi untuk Koefisien Fourier**