Comments on ‘New Approaches in Period Analysis of Astronomical Time Series’ by Pavlos Protopapas (Or: ‘A Pavlosian Response’)

Don Percival

Applied Physics Laboratory
Department of Statistics
University of Washington
Seattle, Washington, USA

`http://faculty.washington.edu/dbp`

Background: I

- let's set the stage by considering some background
- suppose we have observations y_j of brightness of a star at times x_j, to be modeled as

$$y_j = g(x_j) + \epsilon_j, \quad j = 1, 2, \ldots, n,$$

  where

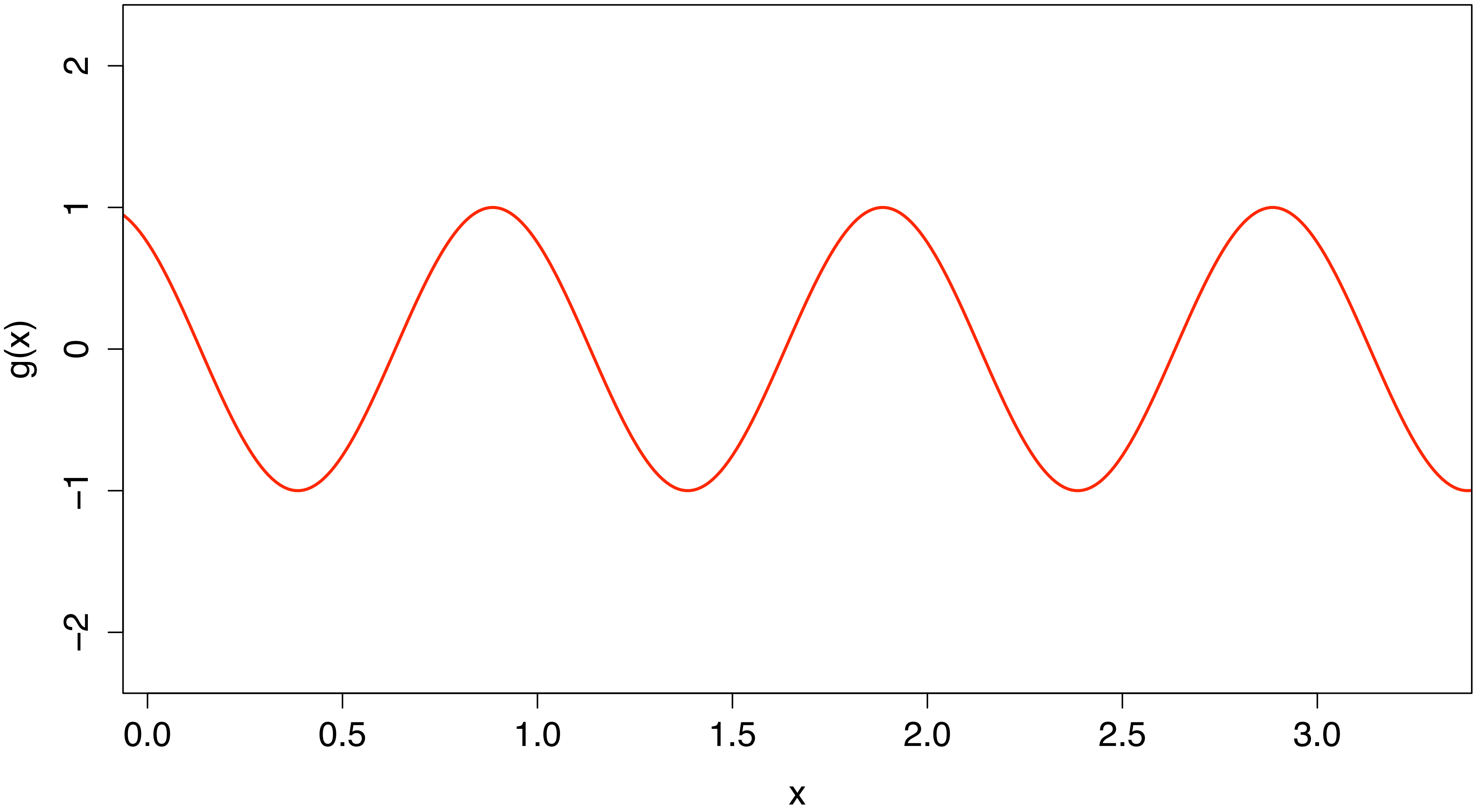
  - $g(x)$ is a periodic function with unknown period p_0
  - ϵ_j represents observational noise

- problem of estimating p_0 and $g(x)$ has received considerable attention in statistical and astronomical literature (Hall, 2008, has good review of statistical approaches)

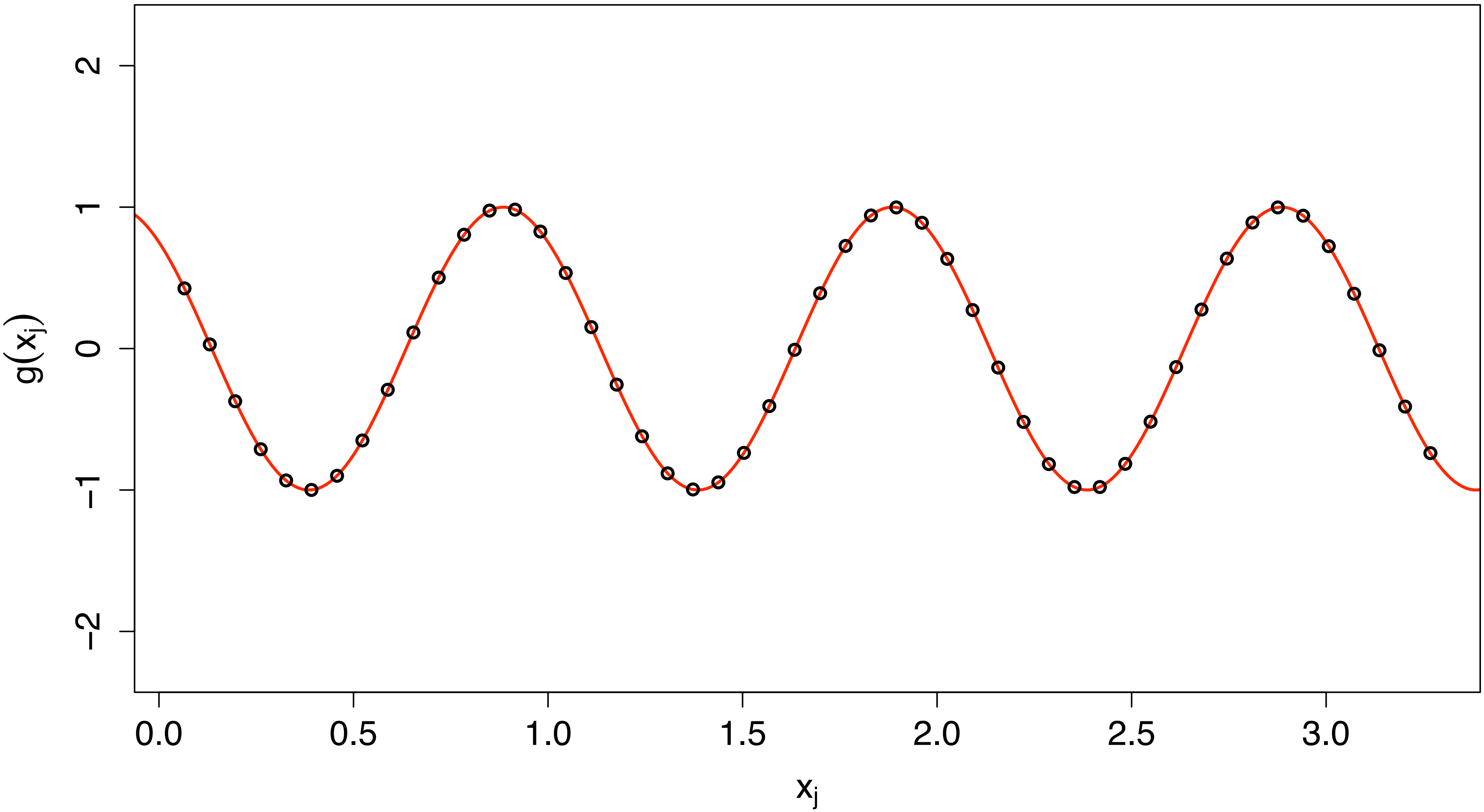
Background: II

- toy parametric version of problem: assume
  - $g(x) = A\cos(2\pi \frac{1}{p_0} x + \phi)$
  - $x_j = j\,\Delta$ (i.e., regular sampling) with $2\,\Delta < p_0$
  - ϵ_j's are realizations of independent and identically distributed (IID) random variables (RVs) with finite variance σ_ϵ^2
- consider example with $A = 1$, $p_0 = 1$, $\phi = 0.72$, $\sigma_\epsilon = 0.5$, $\Delta = 1/15.3$ and $n = 50$

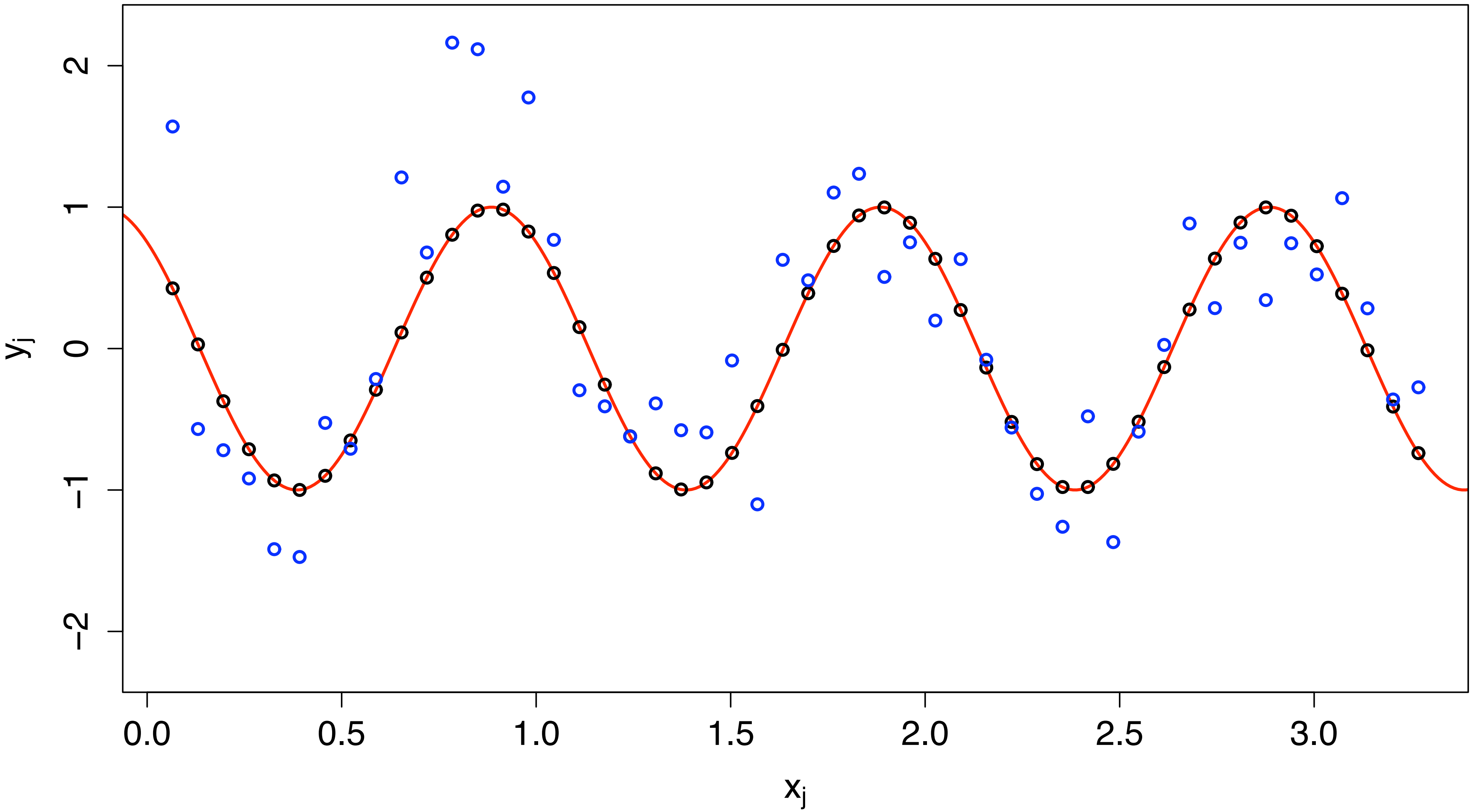
Toy Parametric Model



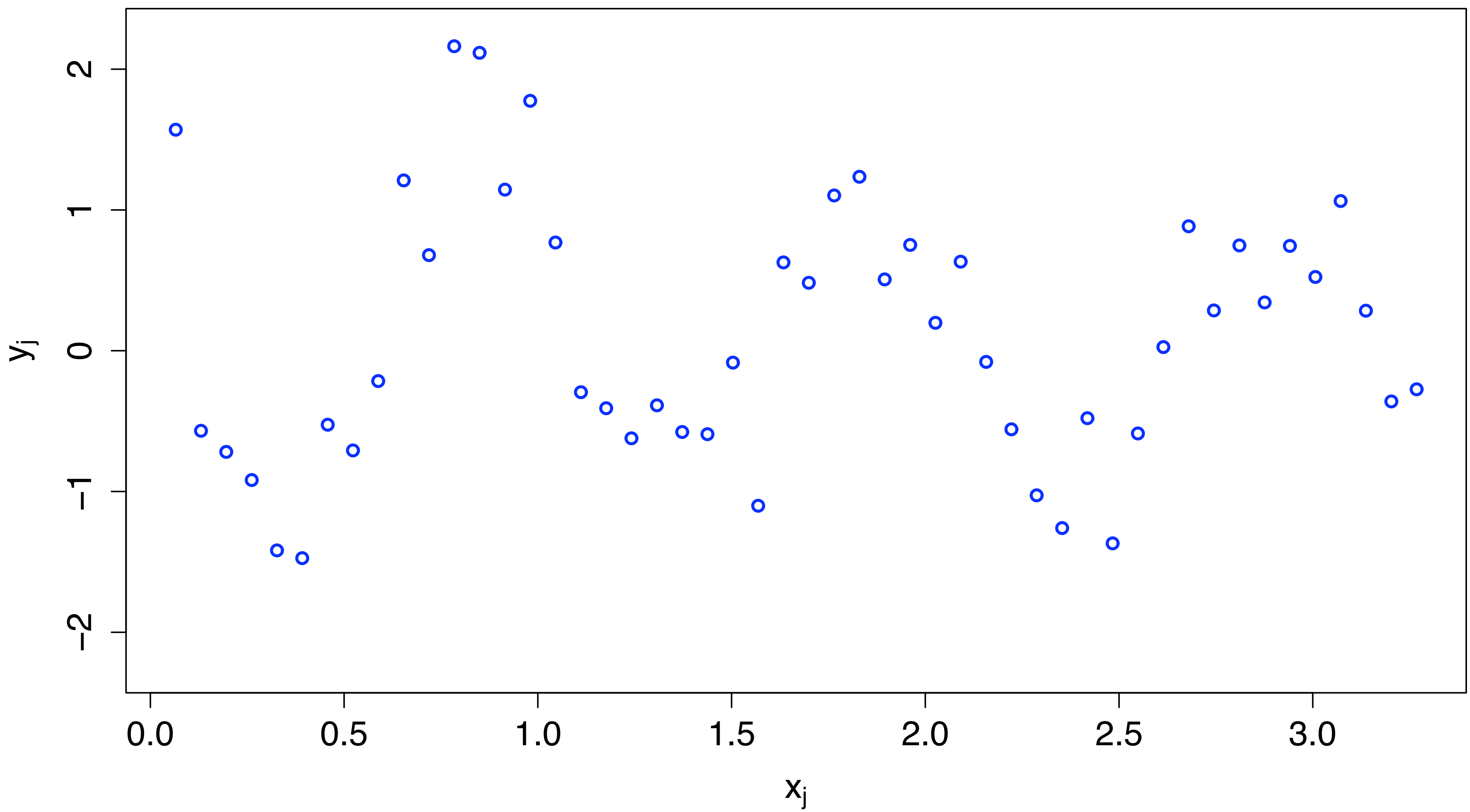
Toy Parametric Model with Regular Sampling



Toy Parametric Model with Additive Noise



Toy Parametric Model with Additive Noise



Background: III

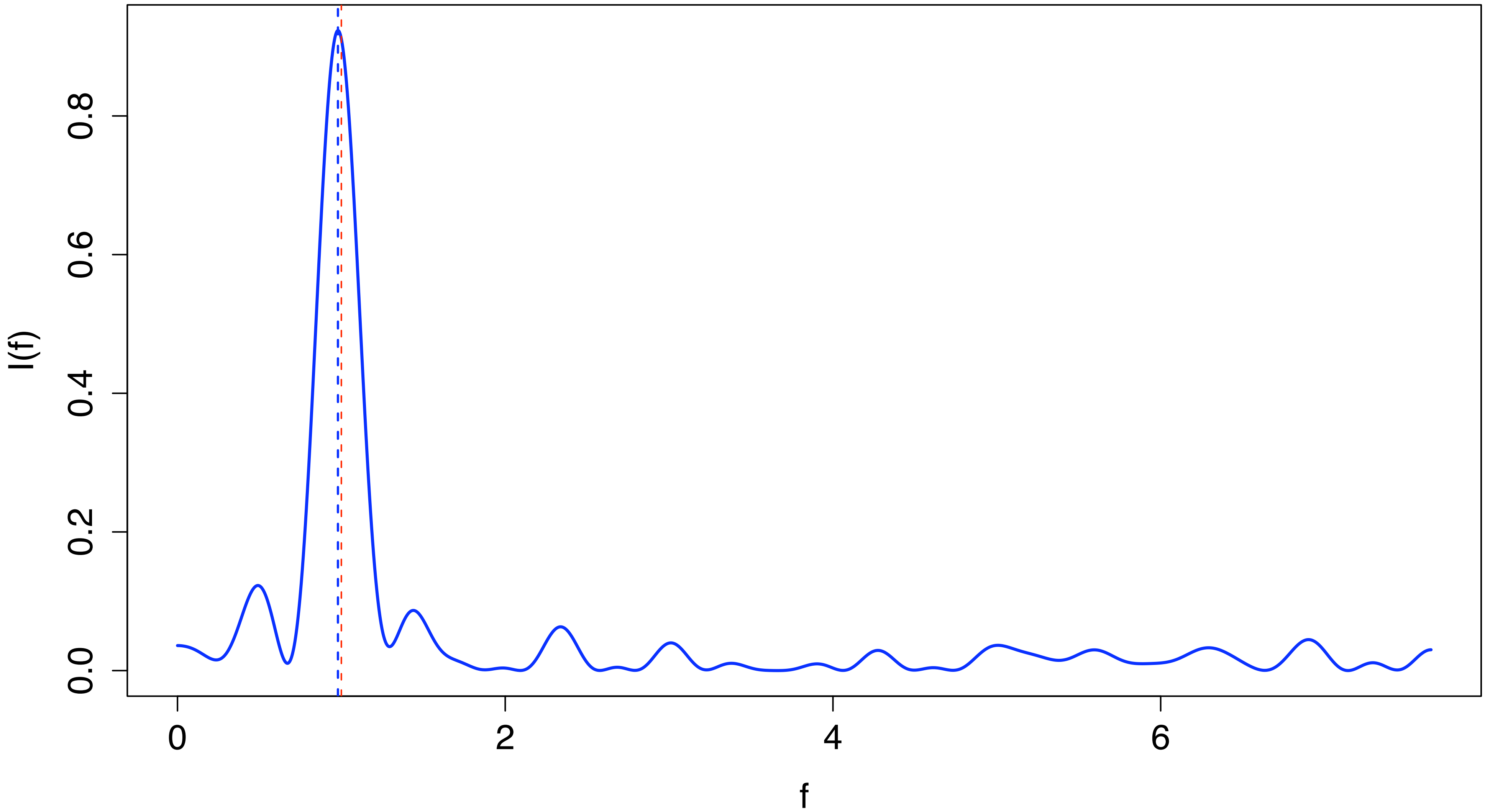
- one solution to toy problem: form periodogram

$$I(f) = \frac{\Delta}{n} \left| \sum_{j=1}^{n} x_j e^{-i2\pi f j \Delta} \right|^2$$

(direct function of frequency f; indirect function of period $1/f$)

- let $\hat{f}$ be frequency at which $I(f)$ is maximized over $(0, 1/[2\,\Delta])$ so that $1/\hat{f}$ is estimate of period p_0

Periodogram for Example of Toy Parametric Model



Background: IV

- here $\hat{f} = 0.98$, which is very close to truth $f_0 = 1/p_0 = 1$
- Q: with just $n = 50$ observations, did we just get lucky?
- Whittle (1952) and Walker (1971) show

$$E\{\hat{f}\} \approx f_0 \text{ and } \operatorname{var}\{\hat{f}\} \approx \frac{C}{n^3}, \text{ where } C = \frac{6\sigma_\epsilon^2}{(A\pi\,\Delta)^2},$$

  with approximations improving with increasing n
- contrast with problem of estimating mean of IID RVs ζ_j with variance σ_ζ^2 using sample mean $\hat{\zeta} = \sum_j \zeta_j/n$, for which

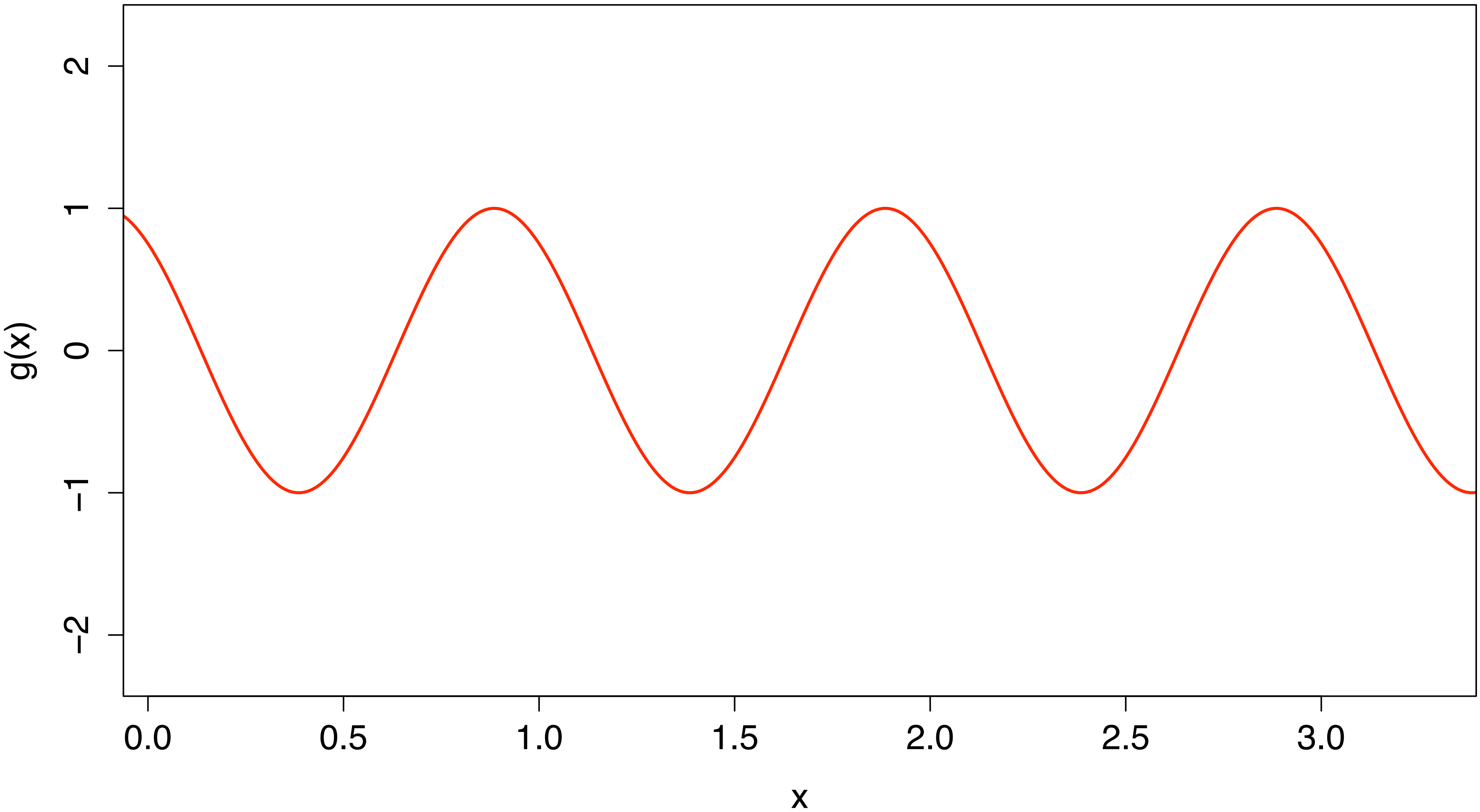
$$\operatorname{var}\{\hat{\zeta}\} = \frac{\sigma_\zeta^2}{n}$$

- note: can also estimate $g(x)$ in obvious ways

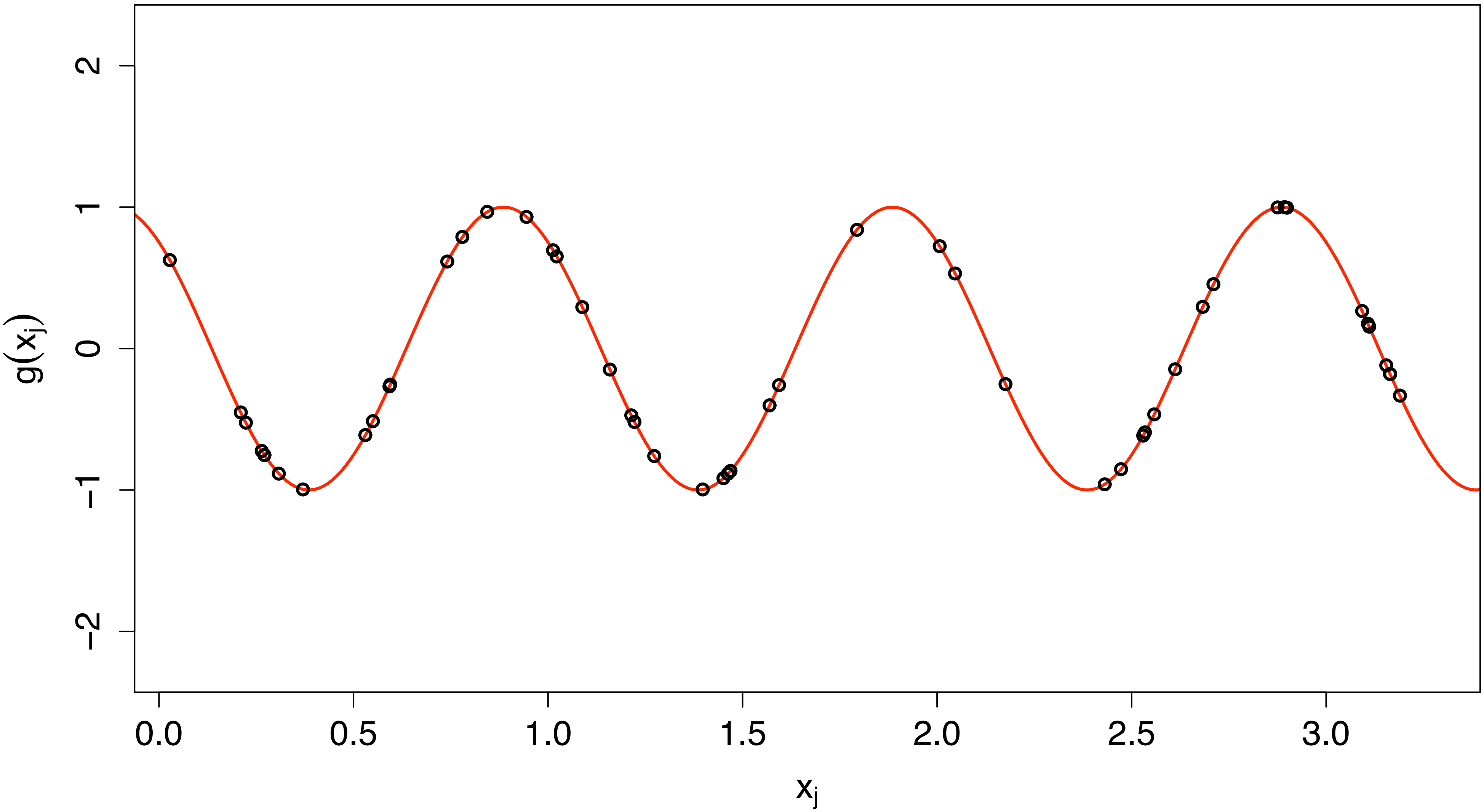
Background: V

- more realistic nonparametric version of problem: assume
  - $g(x)$ has r bounded derivatives
  - $x_1 \leq x_2 \leq \cdots \leq x_n$ are irregularly spaced, but with spacing dictated by IID nonnegative RVs (some conditions needed)
  - ϵ_j's same as in toy problem
- consider same example as before, with sample points dictated by uniform distribution

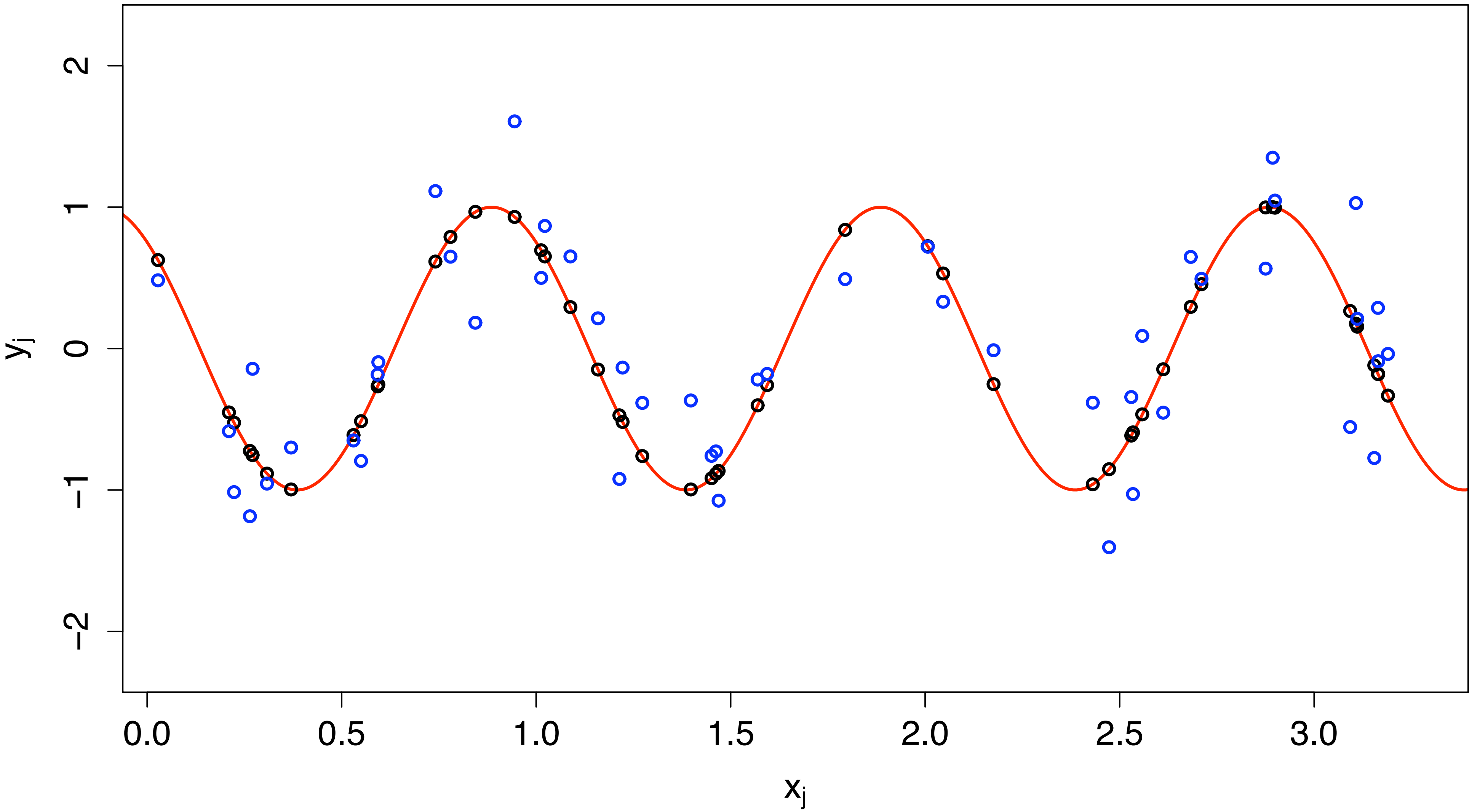
Toy Parametric Model



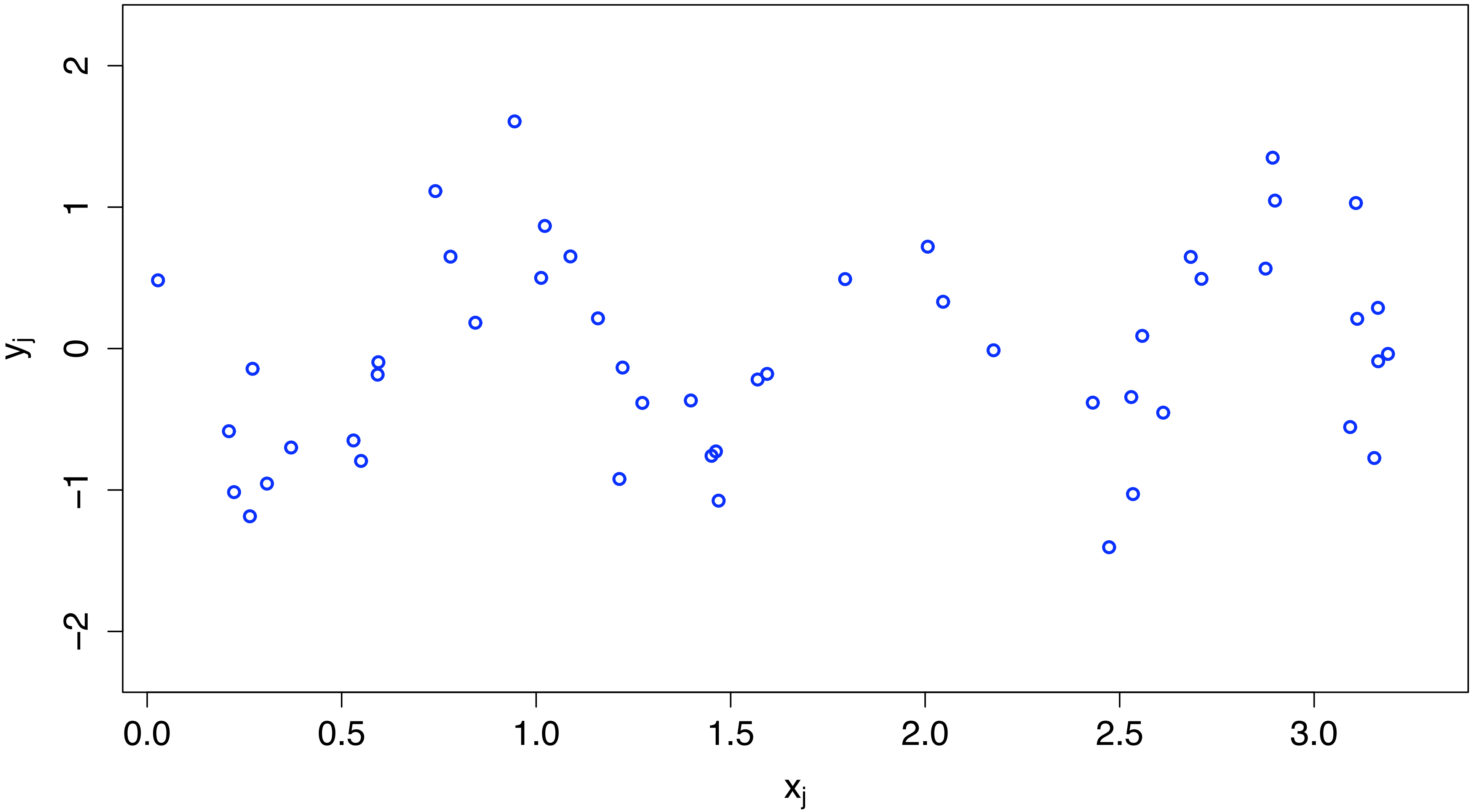
Toy Parametric Model with Irregular Sampling



Toy Parametric Model with Additive Noise



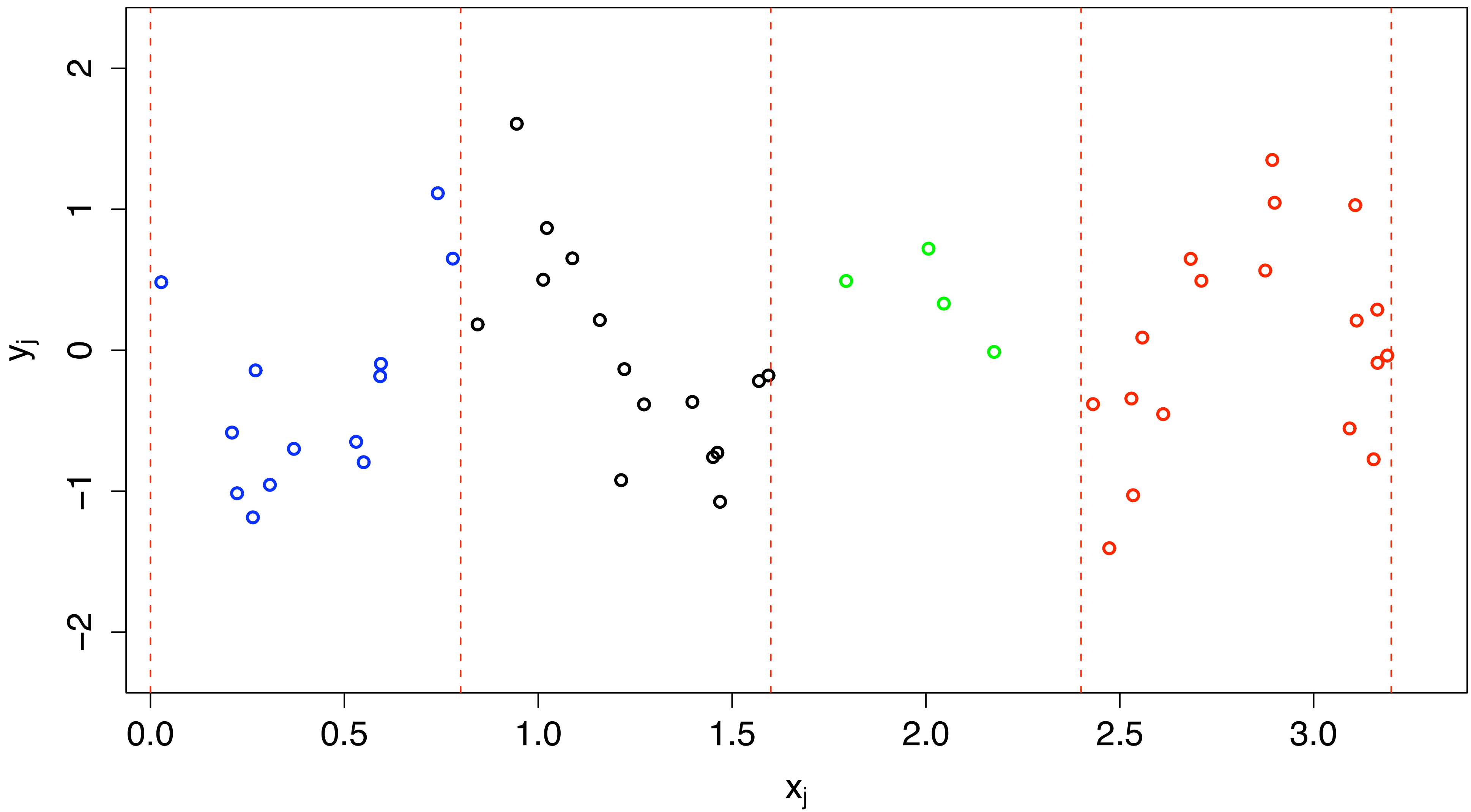
Toy Parametric Model with Additive Noise



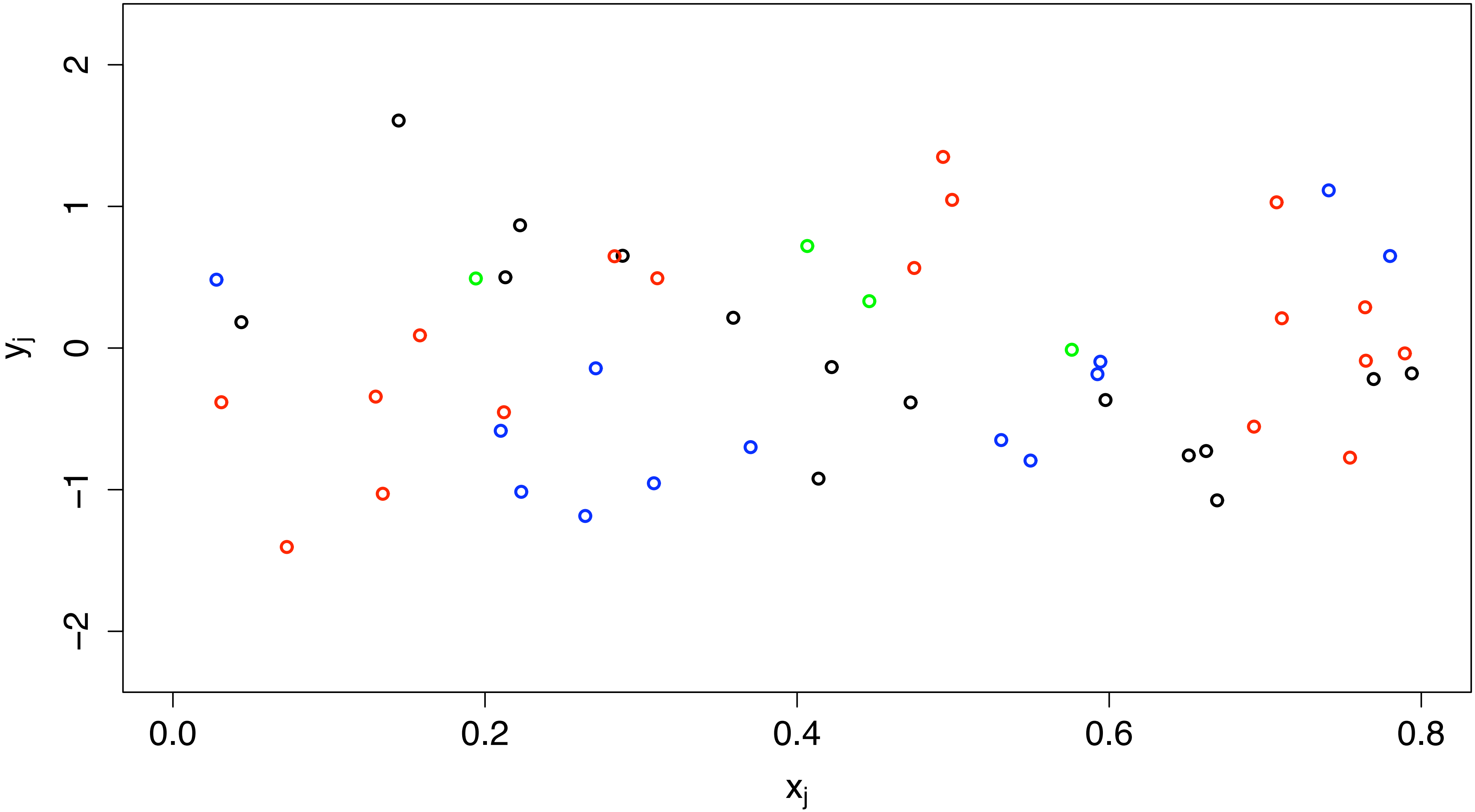
Background: VI

- for given p, let $\tilde{x}_j = x_j - p\lfloor x_j/p \rfloor$, where $\lfloor x \rfloor$ is largest integer strictly less than x; i.e., wrap x_j's circularly to get $\tilde{x}_j$'s

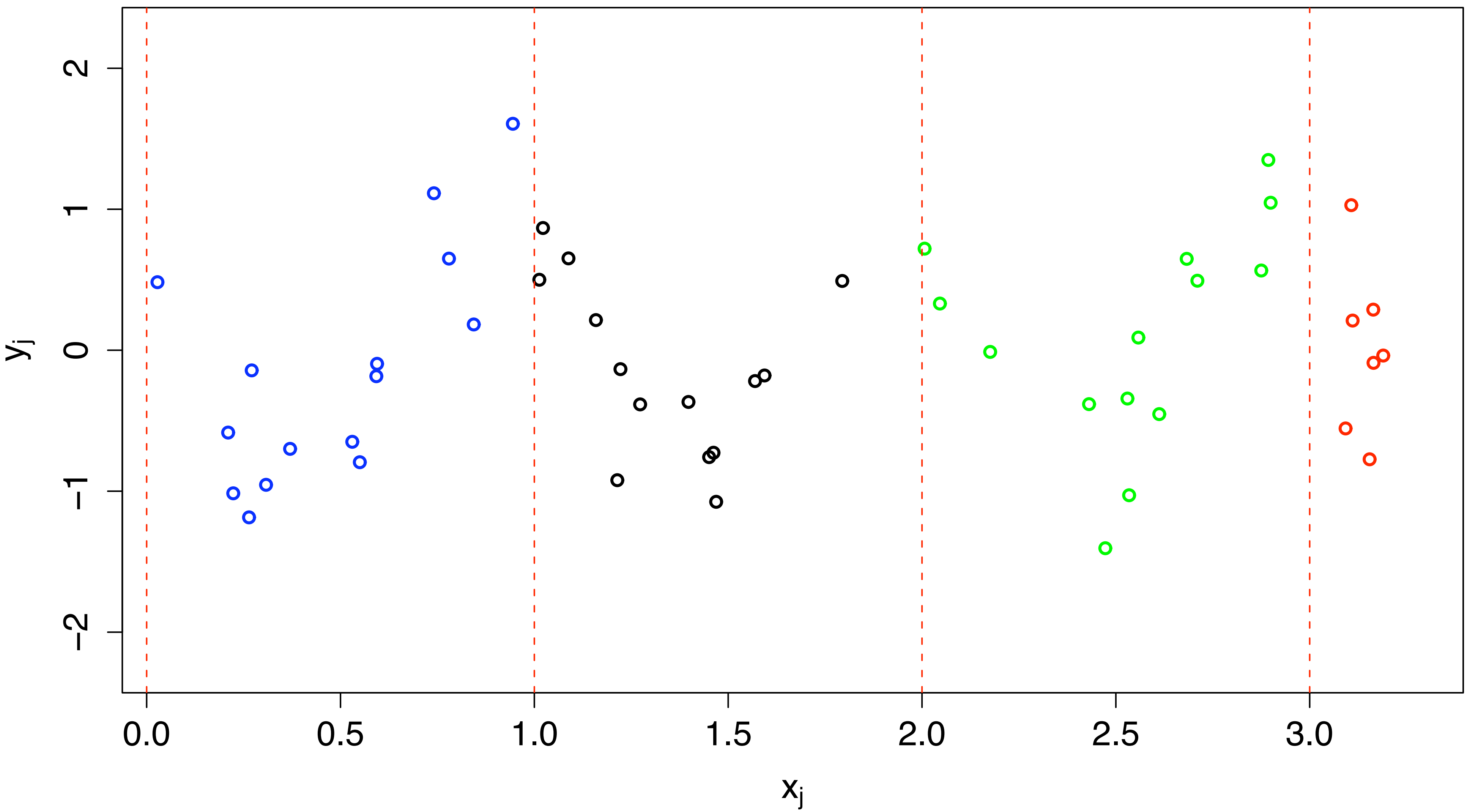
Example of Nonparametric Model ($p = 0.8$)



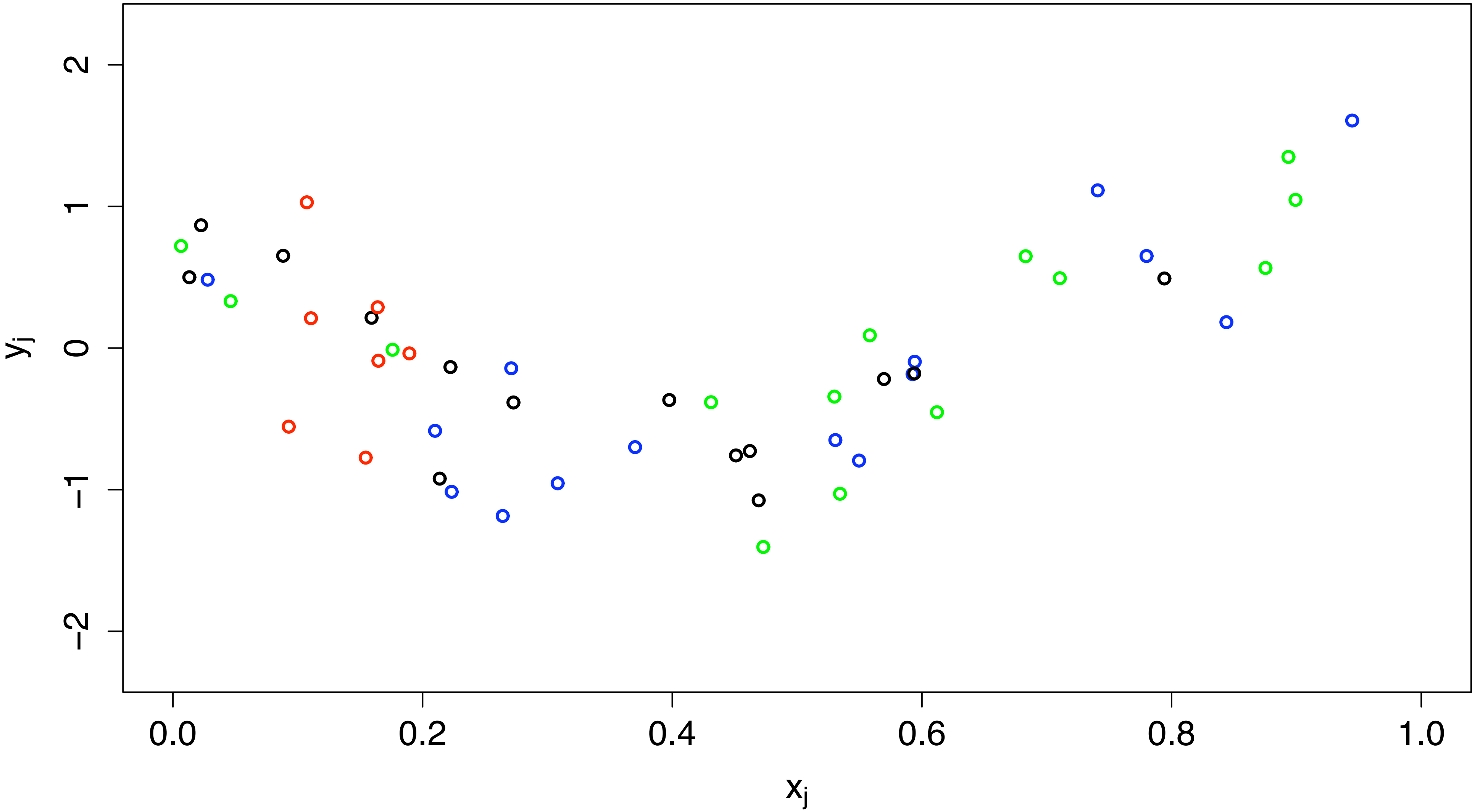
Example of Nonparametric Model ($p = 0.8$)



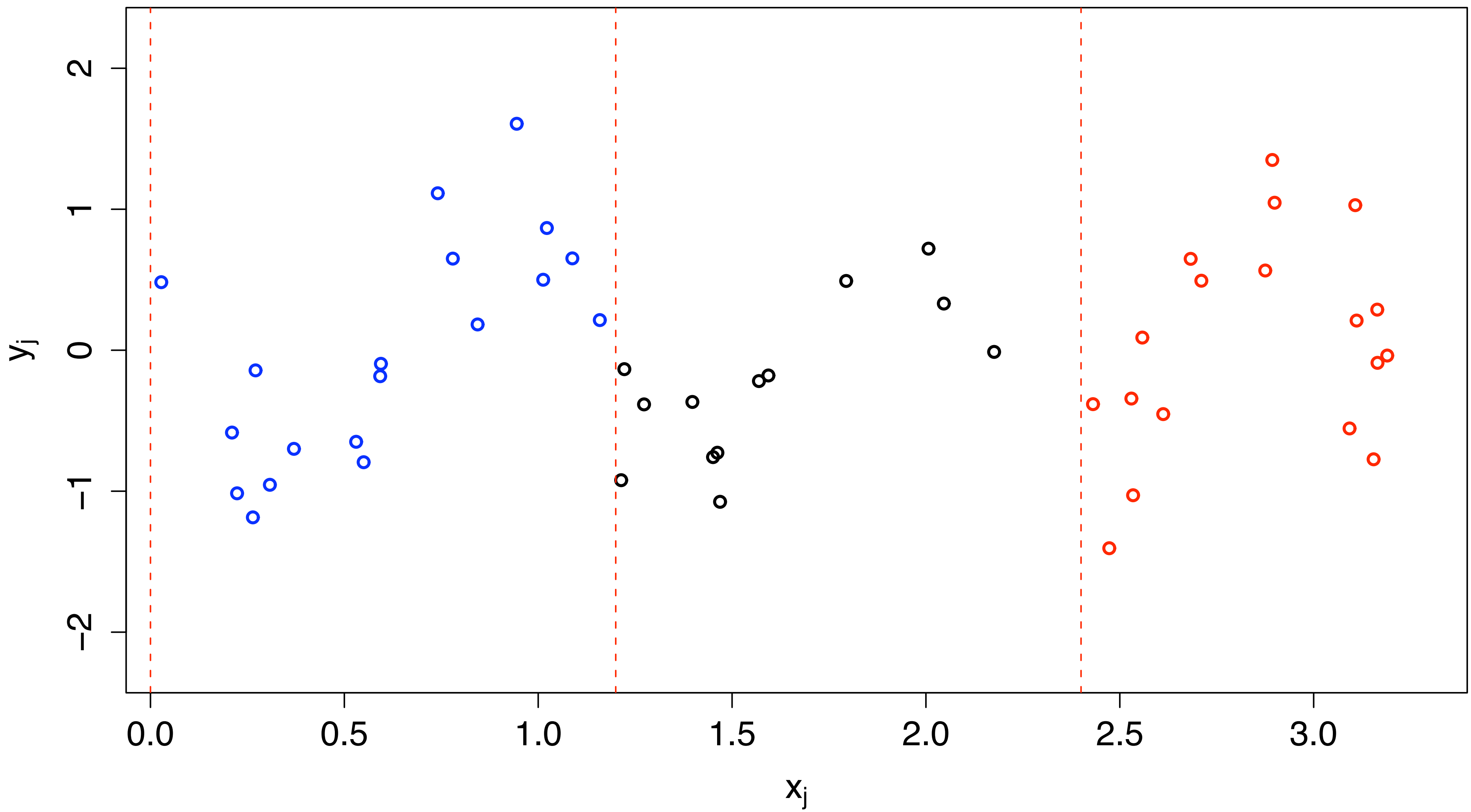
Example of Nonparametric Model ($p = 1.0$)



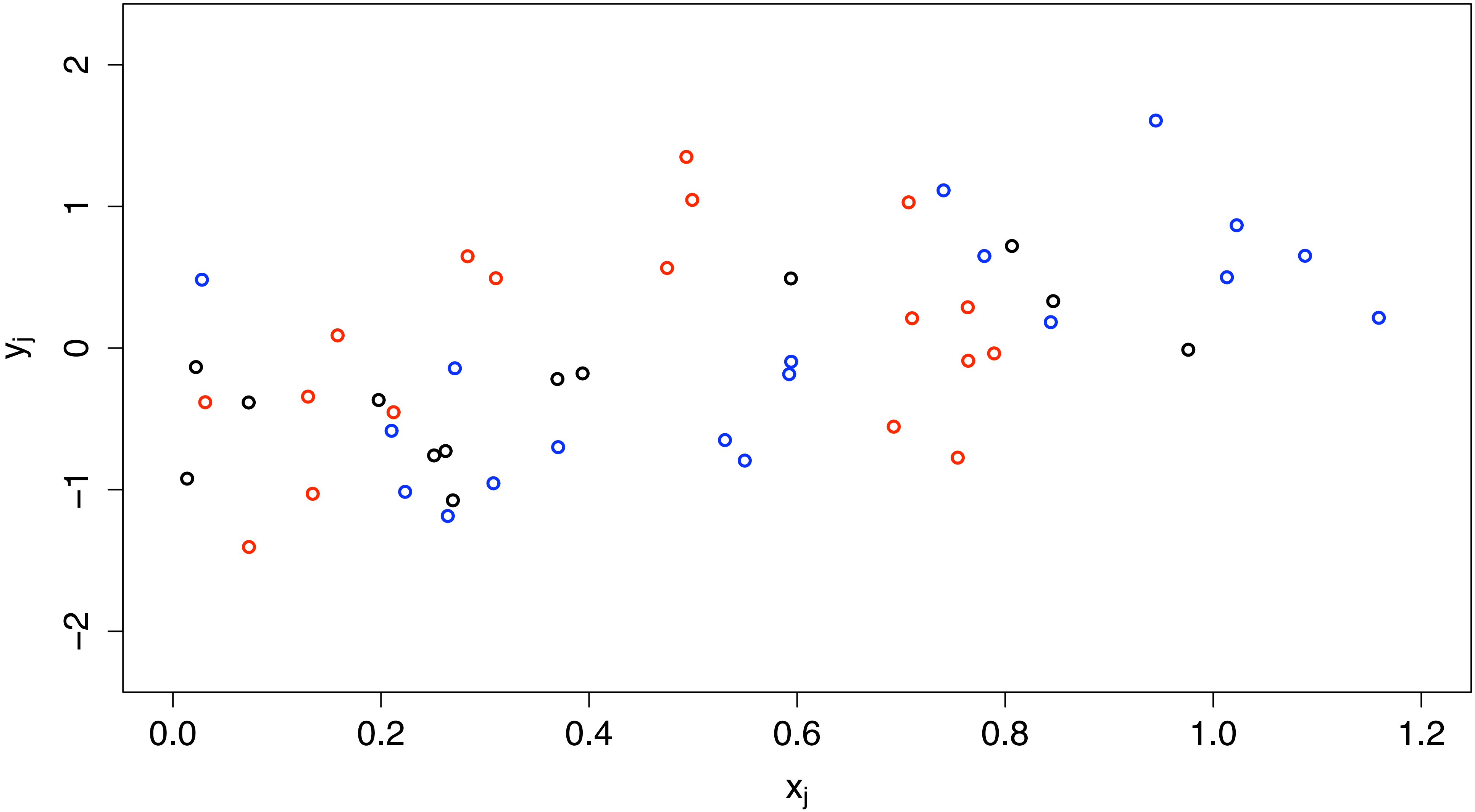
Example of Nonparametric Model ($p = 1.0$)



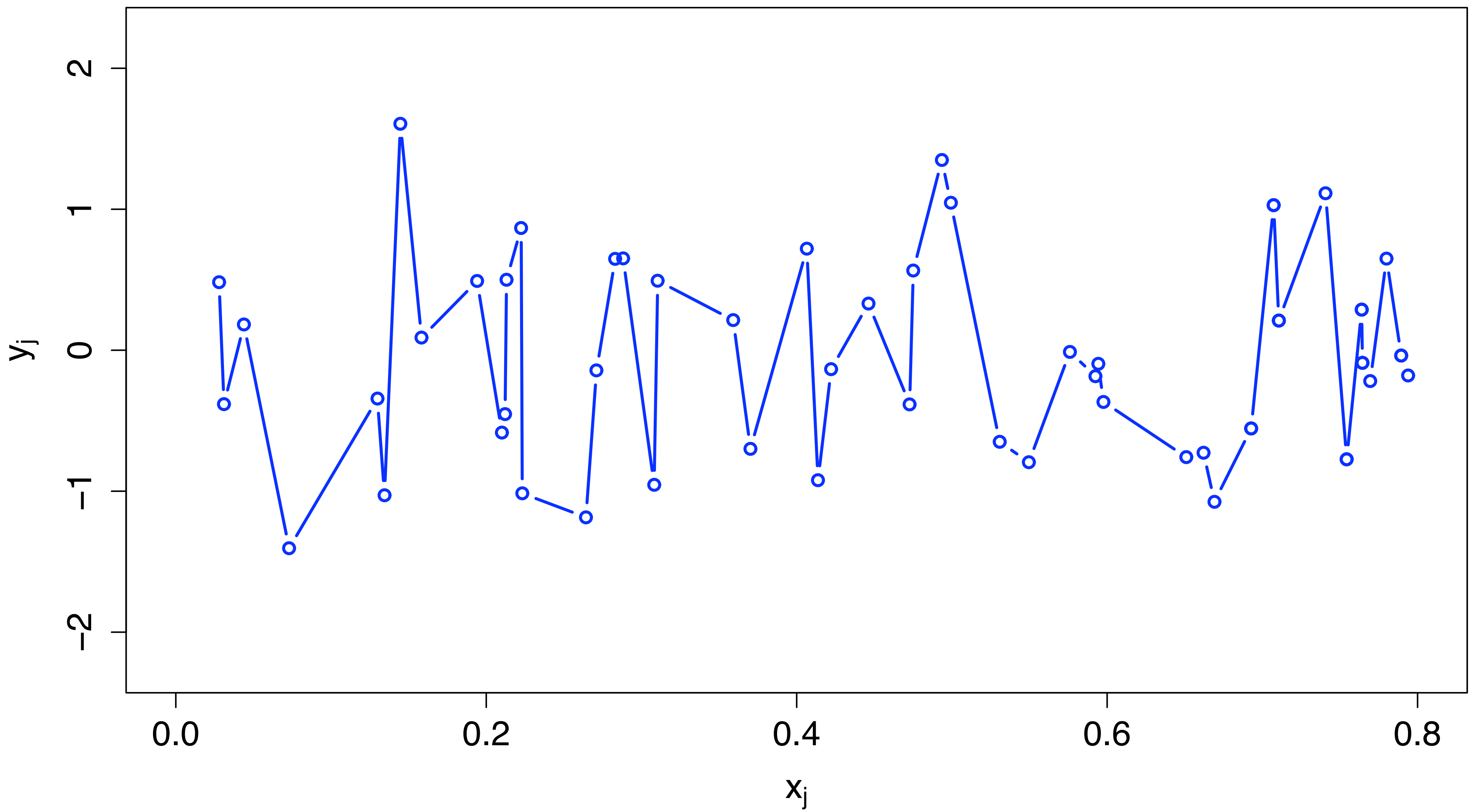
Example of Nonparametric Model ($p = 1.2$)



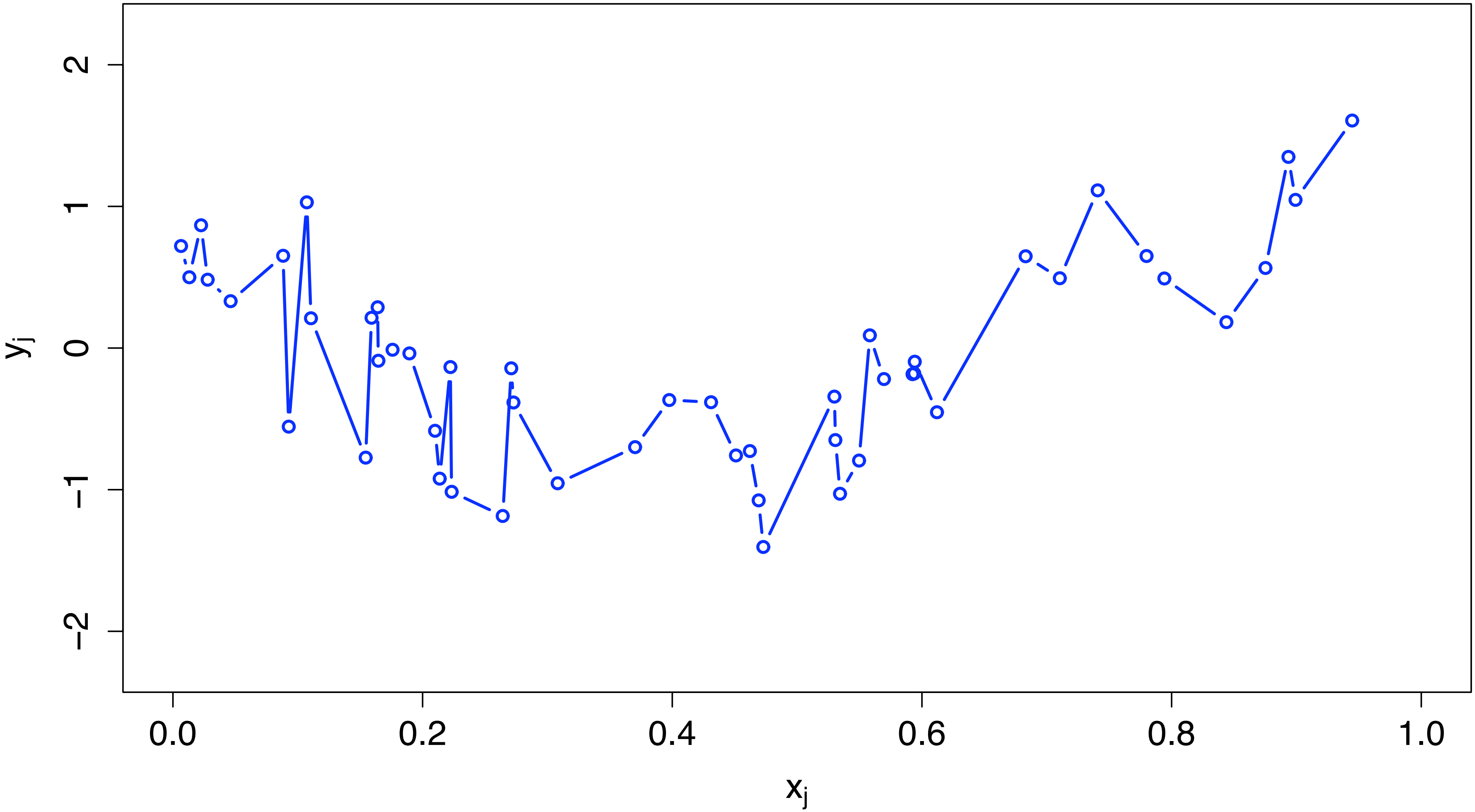
Example of Nonparametric Model ($p = 1.2$)



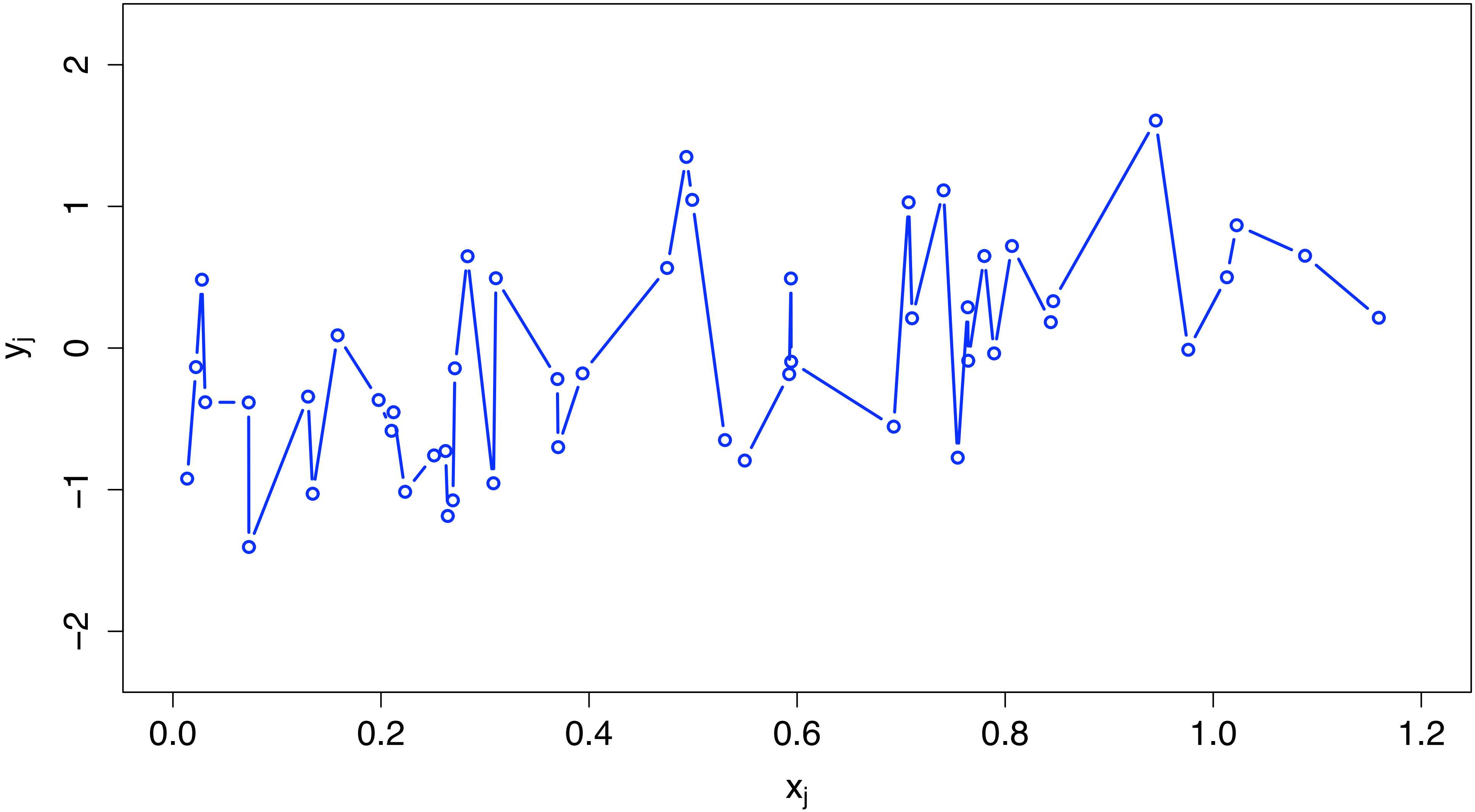
Example of Nonparametric Model ($p = 0.8$)



Example of Nonparametric Model ($p = 1.0$)



Example of Nonparametric Model ($p = 1.2$)



Background: VII

- define Nadaraya–Watson estimator:
$$\hat{g}(x|p) = \frac{\sum_j y_j K_j(x|p)}{\sum_j K_j(x|p)}, \ \text{ where } \ K_j(x|p) = K([x - \tilde{x}_j]/h),$$
K is a kernel (smoother), and h is its bandwidth
- define squared-error criterion:
$$S(p) = \sum_{j=1}^{n} \left(y_j - \hat{g}(x_j|p)\right)^2$$
- $\hat{p} = \text{argmin } S(p)$ is a consistent estimator of p_0
- under regularity conditions,
$$E\{\hat{p}\} \approx p_0 \ \text{ and } \ \text{var}\,\{\hat{p}\} \approx \frac{C}{n^3}$$

Background: VIII

- same n^3 convergence rate as in toy parametric problem (!), and yields estimate of $g(x)$ also
- mission accomplished . . . or is it?
- many questions/issues:
  - computationally feasible with data explosion?
  - nonlinear optimization?
  - choice of bandwidth?
  - asymptotics versus finite sample?
  - outliers? correlation? multivariate? highly irregular? . . .
  - apologizes needed for advocating a non-Bayesian approach!?
- bottom line: room for other procedures, but Hall et al. (2000) is serious competition (ain't sexy, but snickers will be ignored!)

Two New Approaches

- correntropy
  - approach is similar in spirit to Hall et al. in use of circular wrapping, but has different procedure for selecting p_0
  - heuristic approach whose theoretical underlinings need further exploration
- Gaussian processes
  - fundamentally different from circular wrapping
  - adapts technique originally designed for 2D spatial processes to irregularly sampled time series
  - incorporates Bayesian ideas, so has firmer theoretical backing
- bottom line: both approaches are of interest!

Correntropy: I

- for regularly sampled time series (i.e., $x_j = j\,\Delta$), periodogram

$$I(f) = \frac{\Delta}{n}\left|\sum_{j=1}^{n} x_j e^{-i2\pi f j\,\Delta}\right|^2$$

is Fourier transform of biased estimator of autocovariance sequence:

$$\hat{s}_\tau = \frac{1}{n}\sum_{j=1}^{n-\tau} x_j x_{j+\tau}, \quad 0 \le \tau \le n-1,$$

with $\hat{s}_\tau = 0$ for $\tau \ge n$, and $\hat{s}_{-\tau} = \hat{s}_\tau$.

- regarding x_j as realization of RV X_j, $\hat{s}_\tau$ is estimate of $s_\tau = E\{X_j X_{j+\tau}\}$ for second-order stationary process

Correntropy: II

- correntropy idea is to consider $E\{\kappa(X_j, X_{j+\tau})\}$, where

$$\kappa(x_j, x_k) = \frac{1}{\sqrt{2\pi\sigma_\kappa^2}} \exp\left(-\frac{(x_j - x_k)^2}{2\sigma_\kappa^2}\right)$$

- σ_κ is user-settable parameter controlling influence of higher-order moments (25 settings entertained in IEEE SPL paper)
- assume $E\{\kappa(X_j, X_{j+\tau})\}$ is independent of j (sufficient condition is strict stationarity)
- unbiased estimator of $V_\tau = E\{\kappa(X_j, X_{j+\tau})\}$ is

$$\widehat{V}_\tau = \frac{1}{n-\tau} \sum_{j=1}^{n-\tau} \kappa(x_j, x_{j+\tau}), \quad 0 \le \tau \le n-1,$$

with $\widehat{V}_{-\tau} = \widehat{V}_\tau$ (note: biased estimator performs poorly)

Correntropy: III

- correntropy spectral density (CSD) is taken to be

$$P(f) = \sum_{\tau=-(n-1)}^{n-1} (\widehat{V}_\tau - \overline{V}) e^{-i2\pi f \tau \Delta}$$

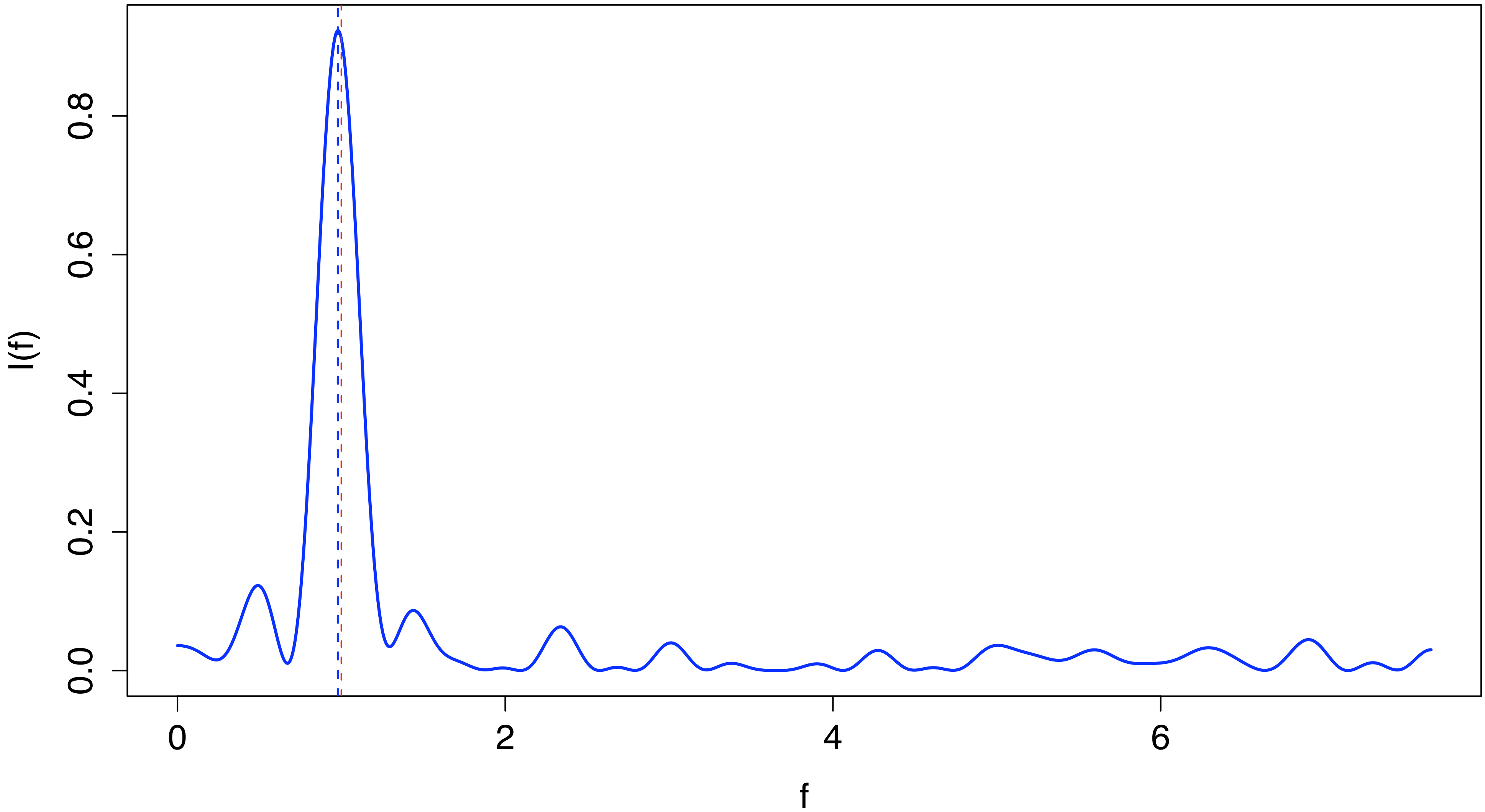
where

$$\overline{V} = \frac{1}{2n-1} \sum_{\tau=-(n-1)}^{n-1} \widehat{V}_\tau$$

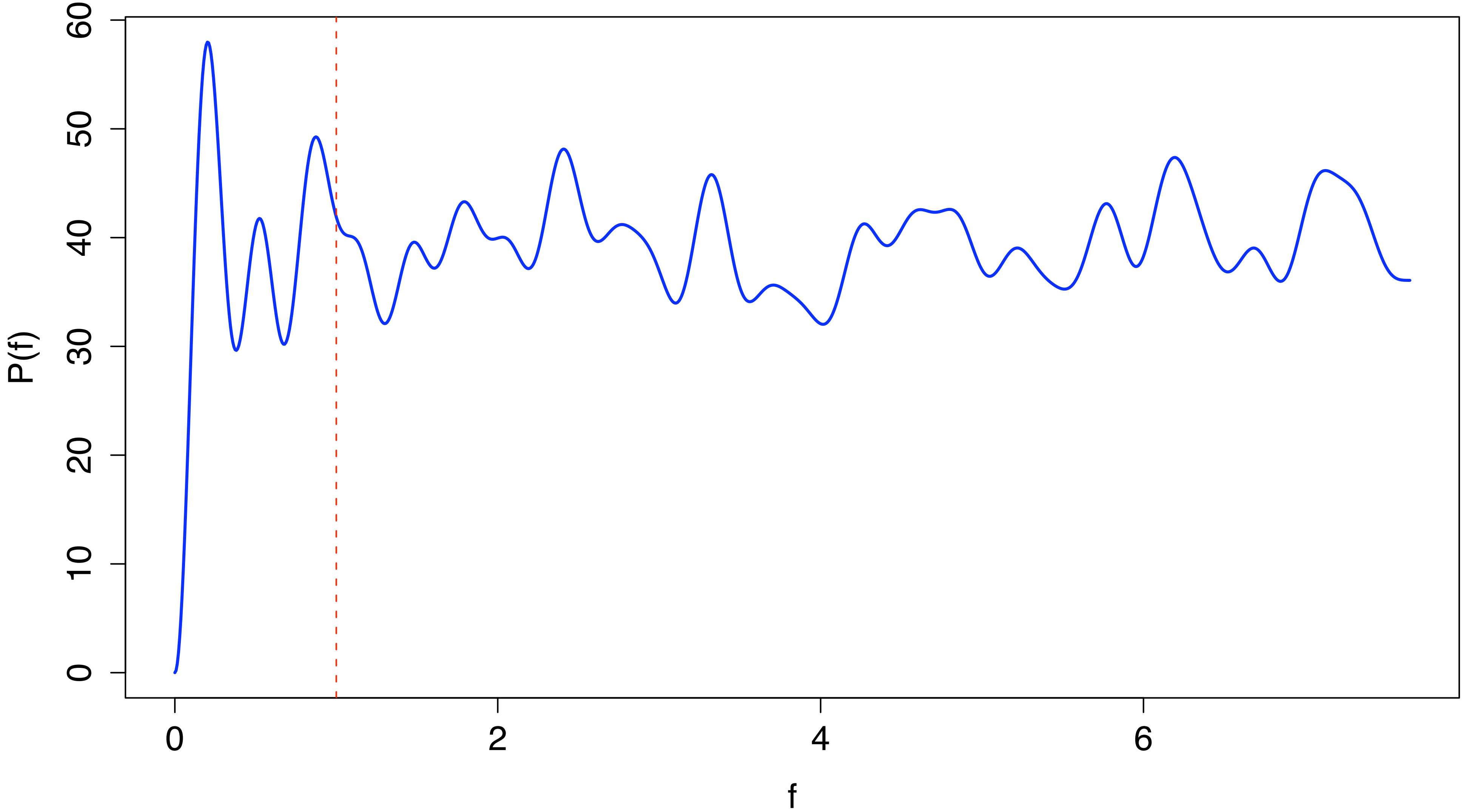
- as with periodogram, search for peaks in $P(f)$
- as a first test, reconsider regularly sampled toy time series

$$y_j = \cos\left(2\pi j \tfrac{1}{15.3} + 0.72\right) + \epsilon_j$$

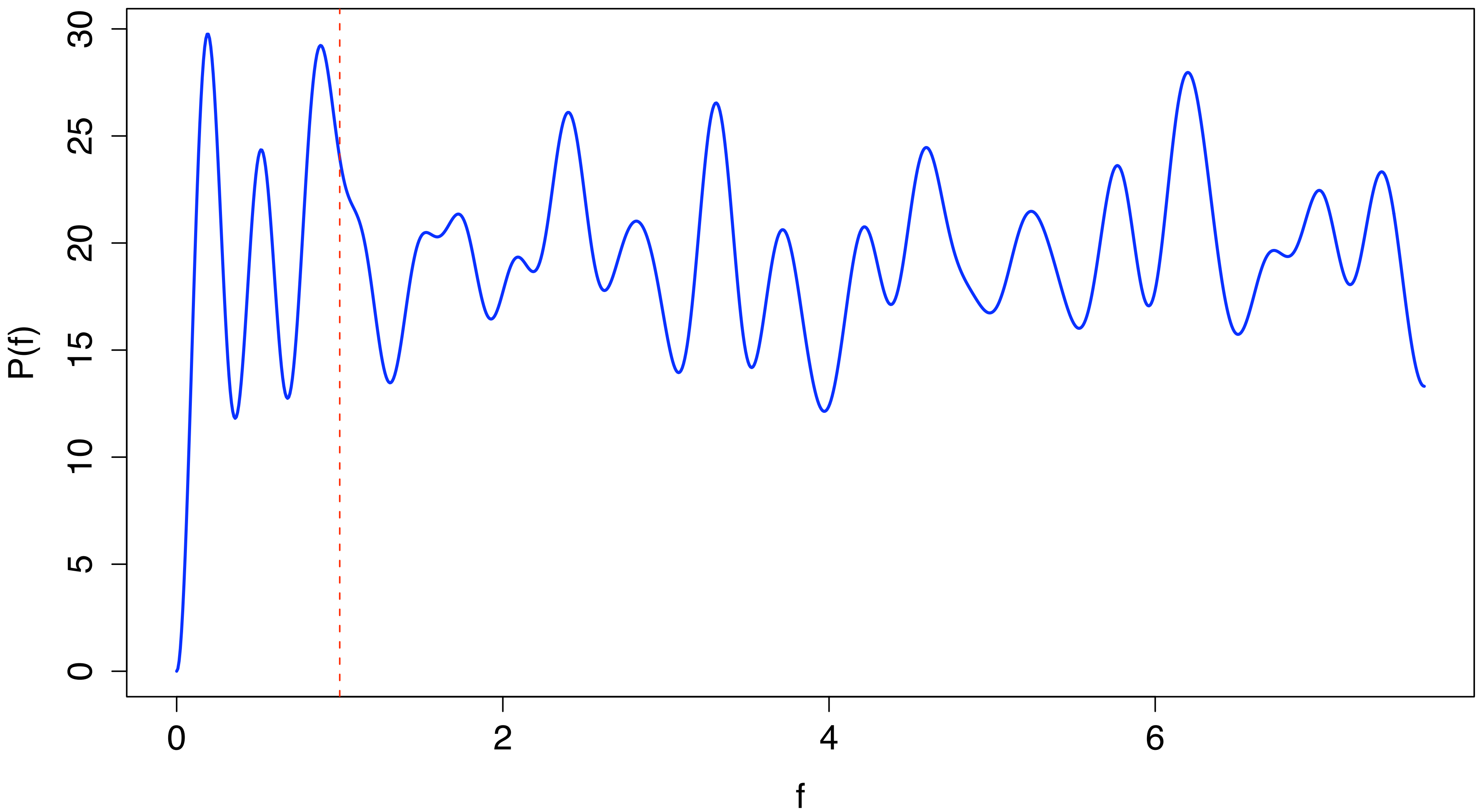
Periodogram for Example of Toy Parametric Model



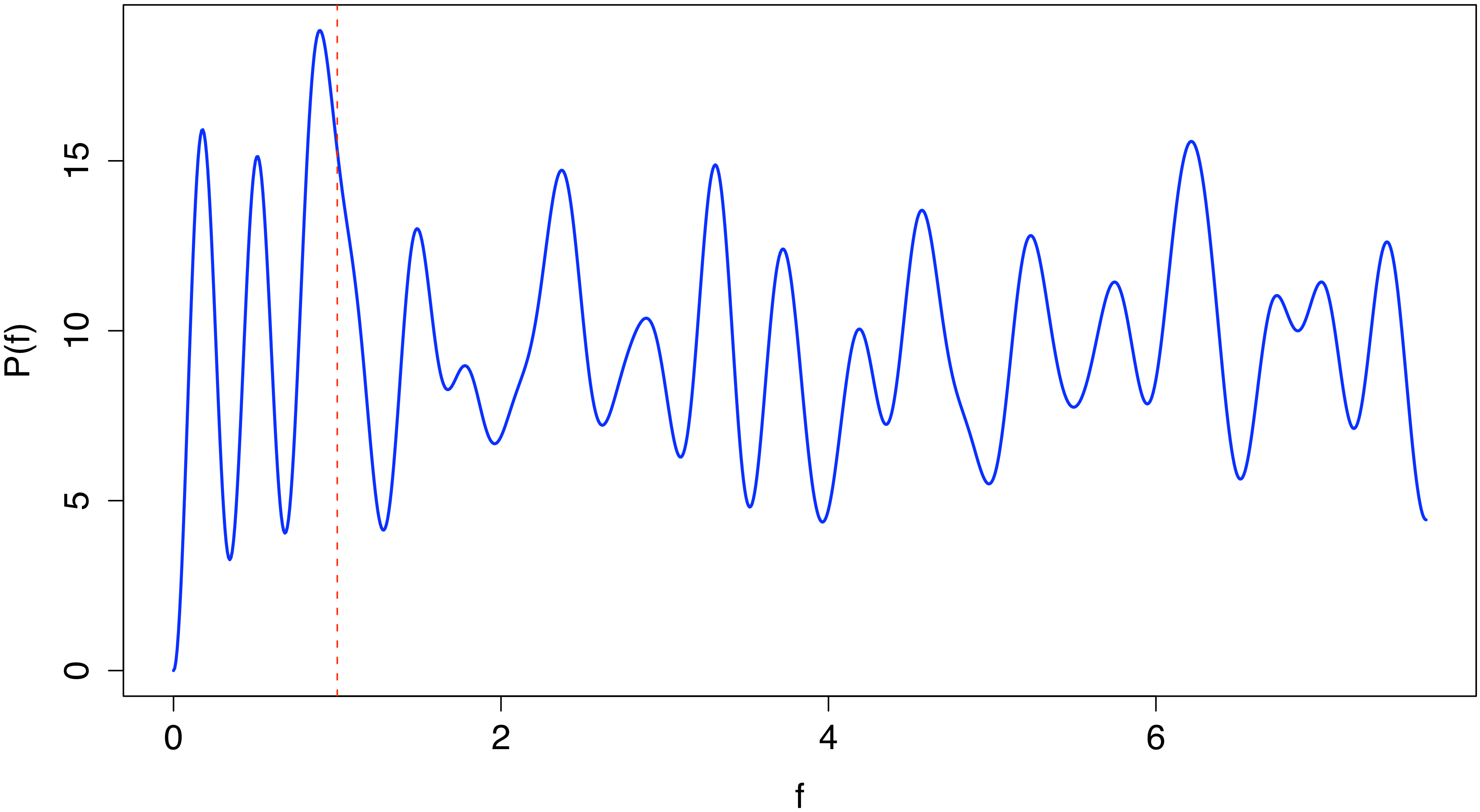
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.01$



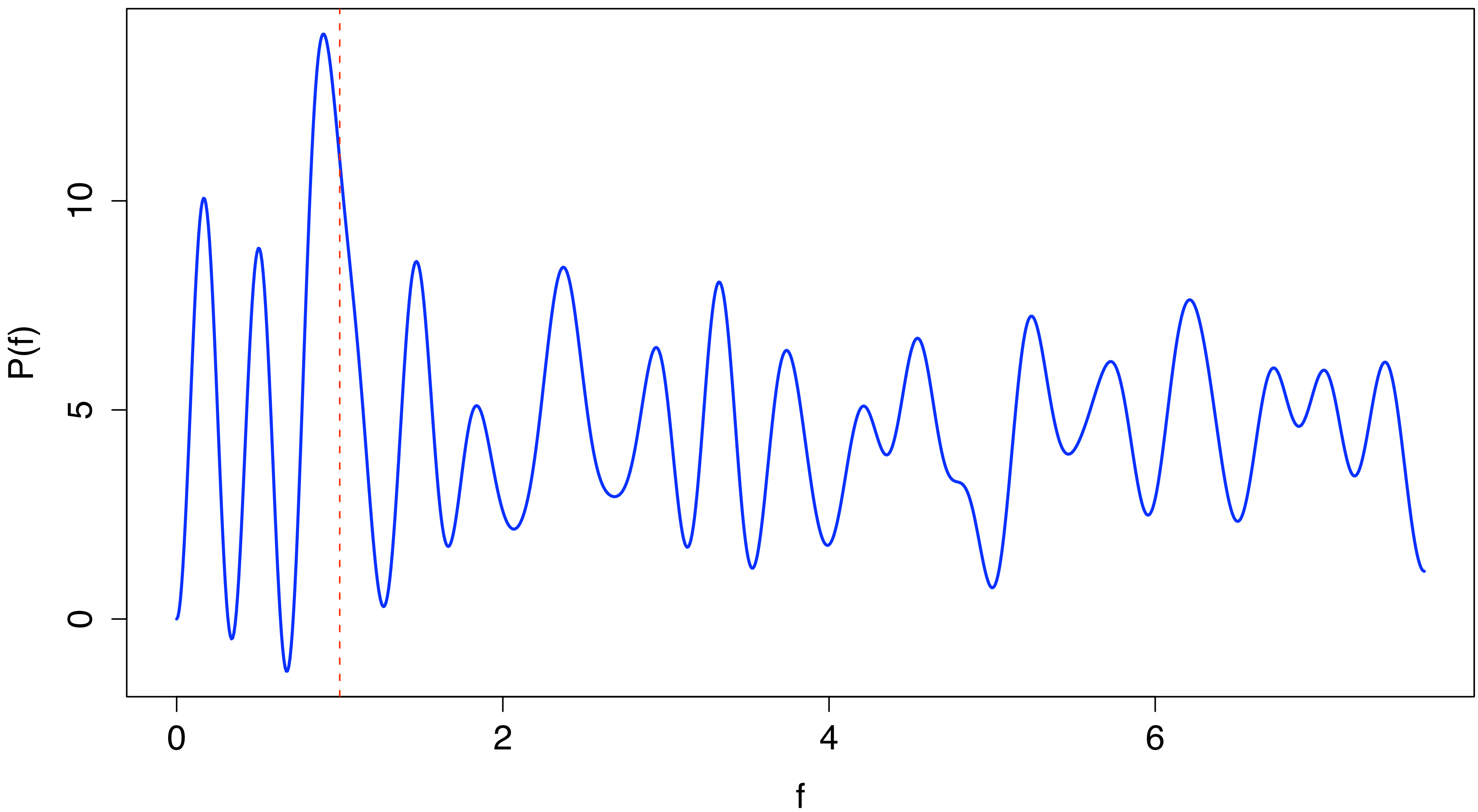
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.02$



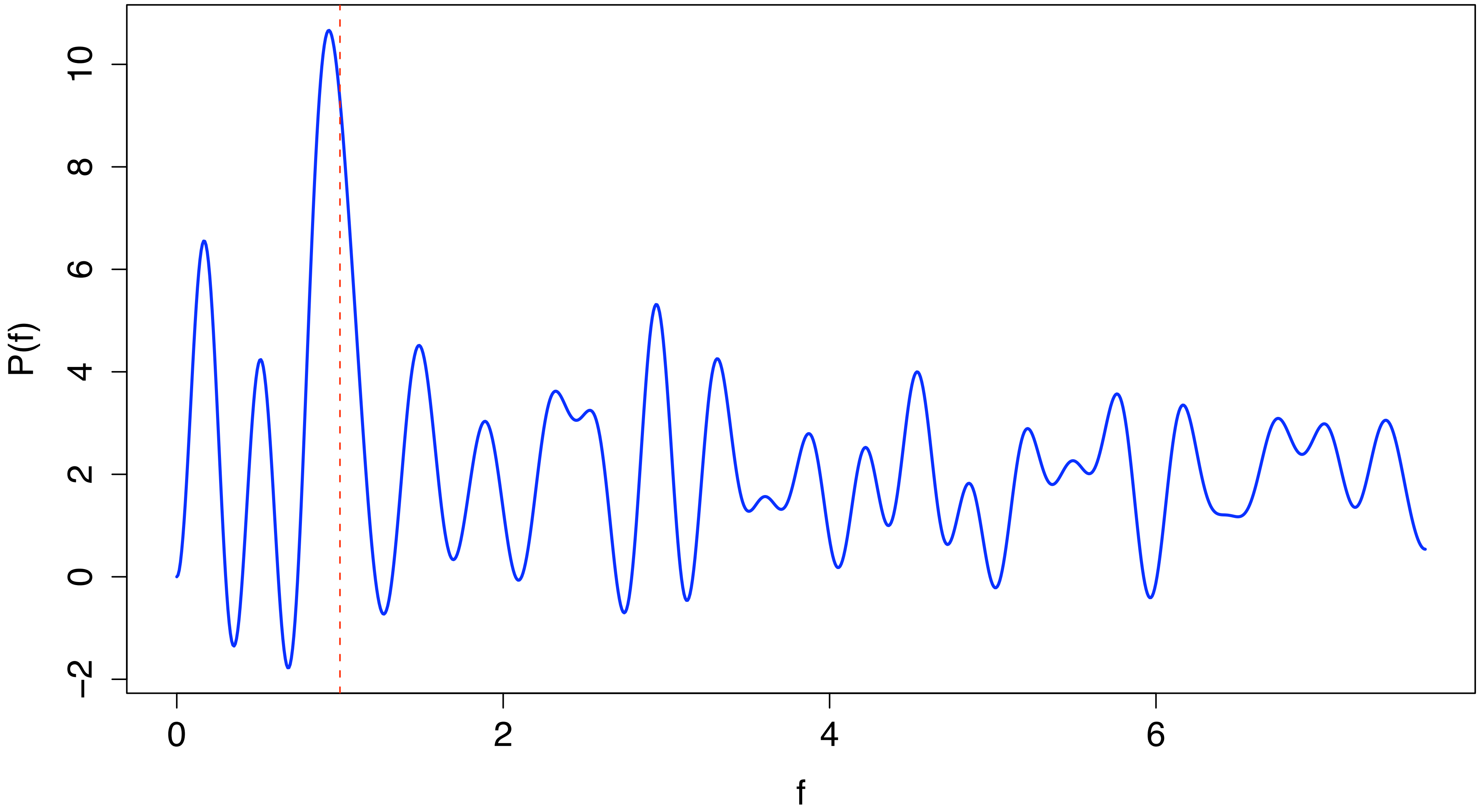
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.04$



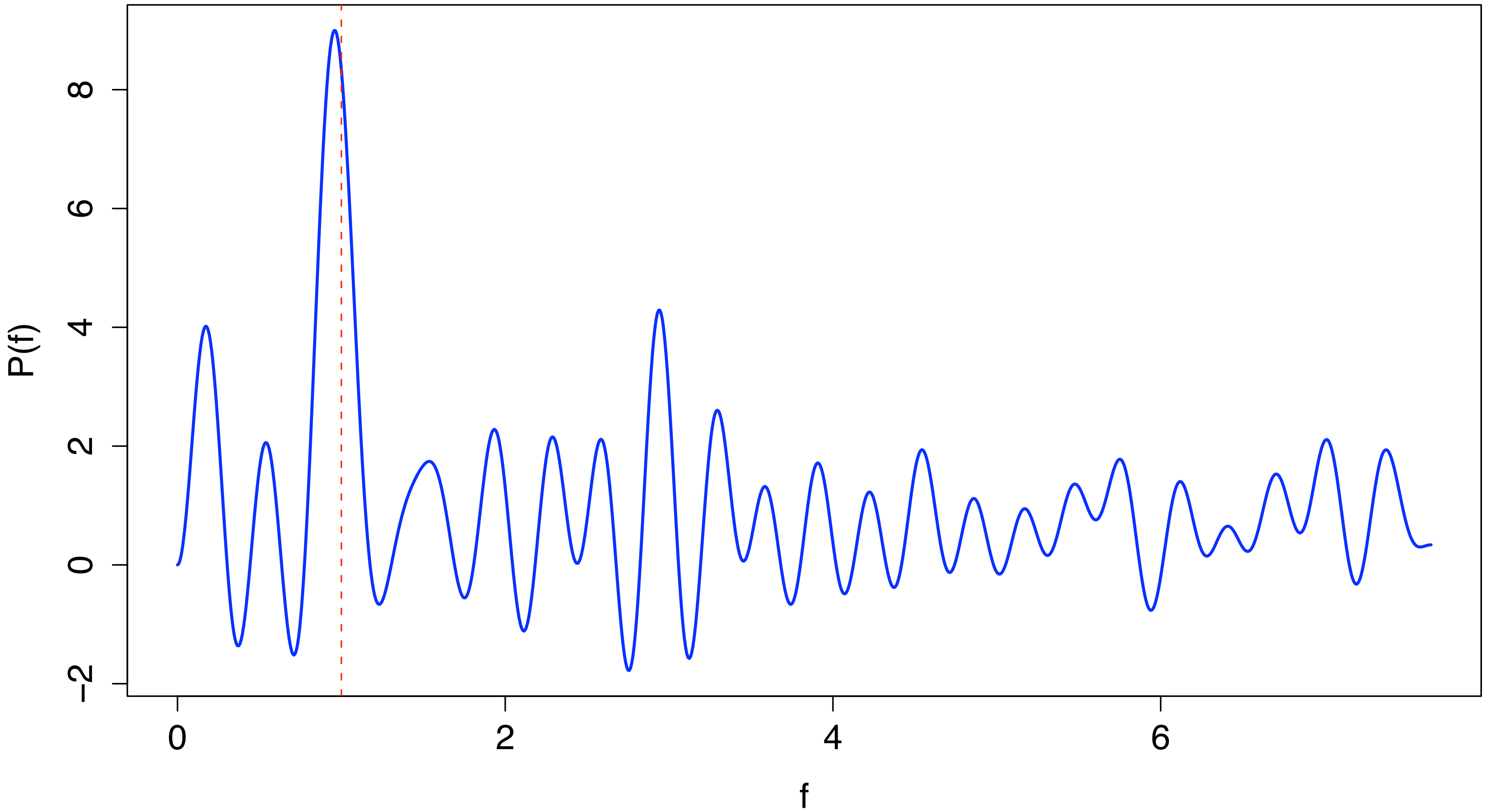
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.08$



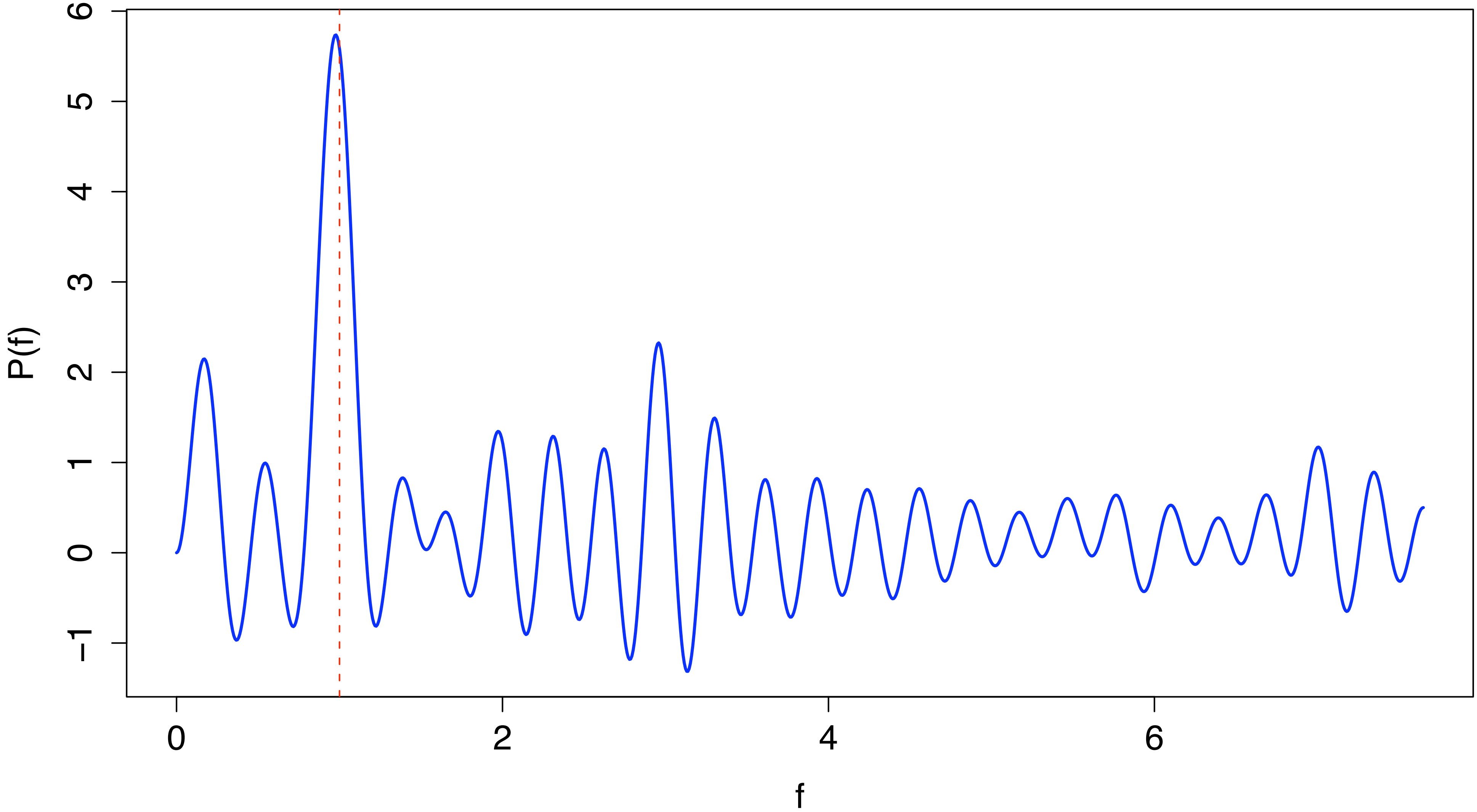
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.16$



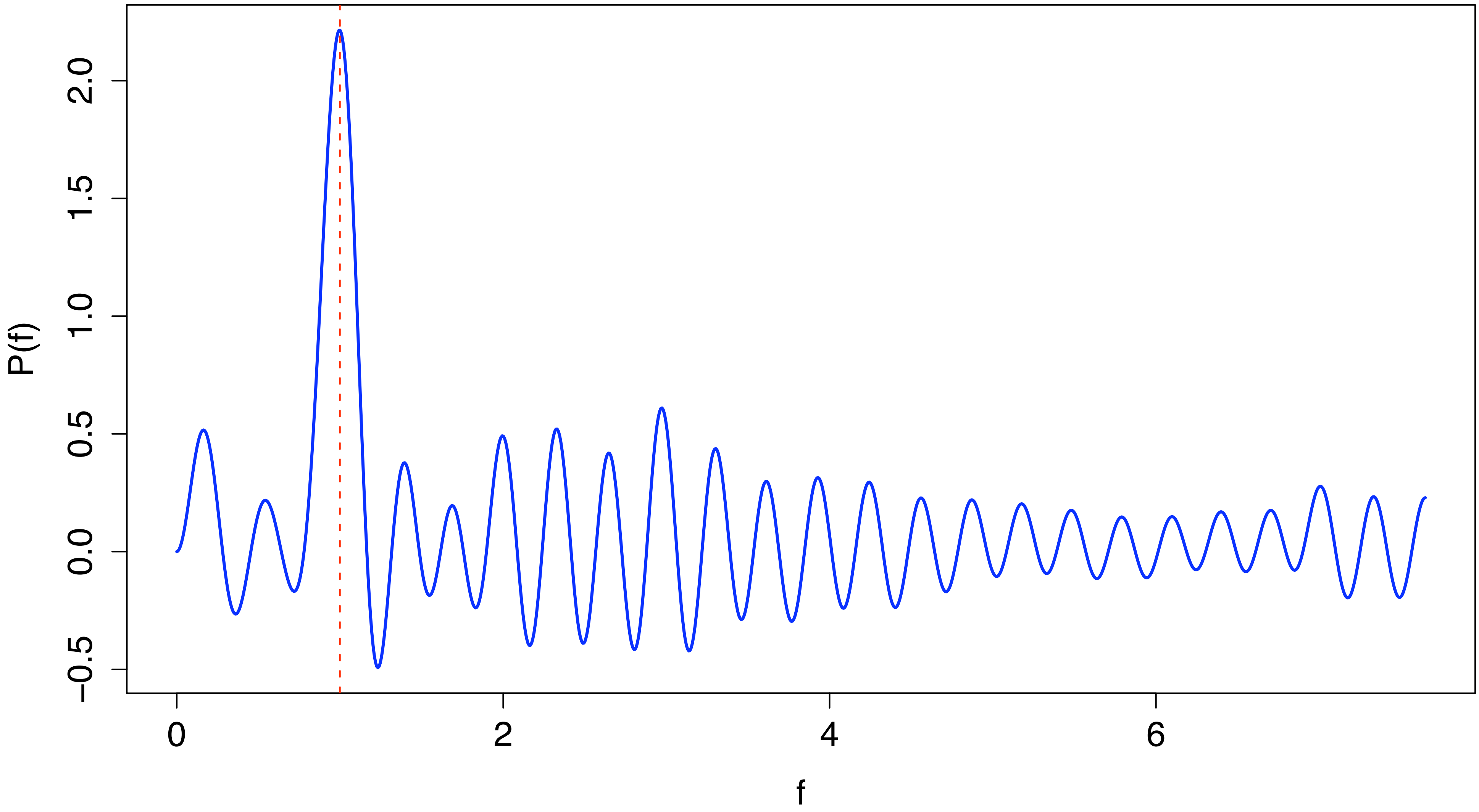
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.32$



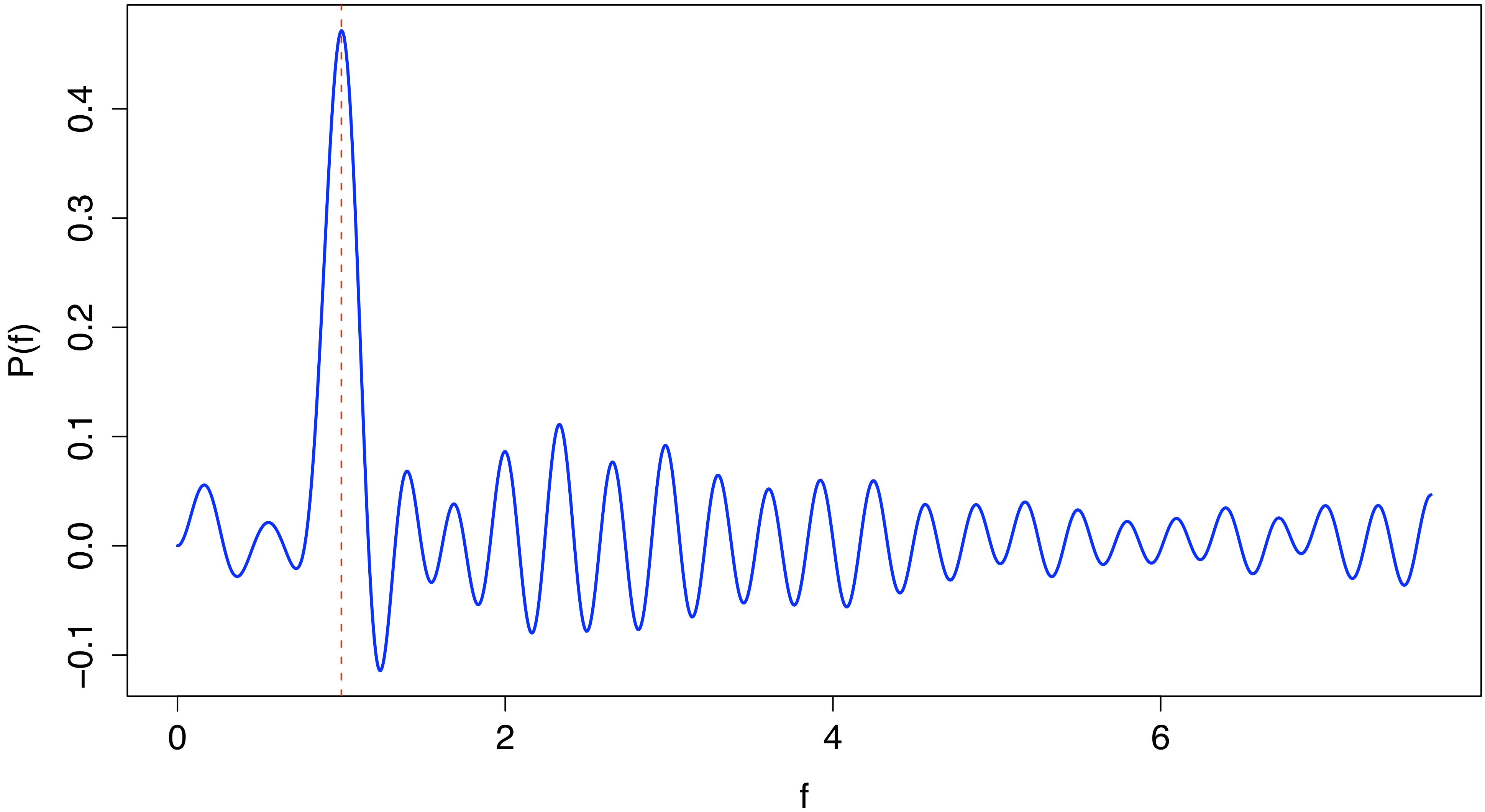
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 0.64$



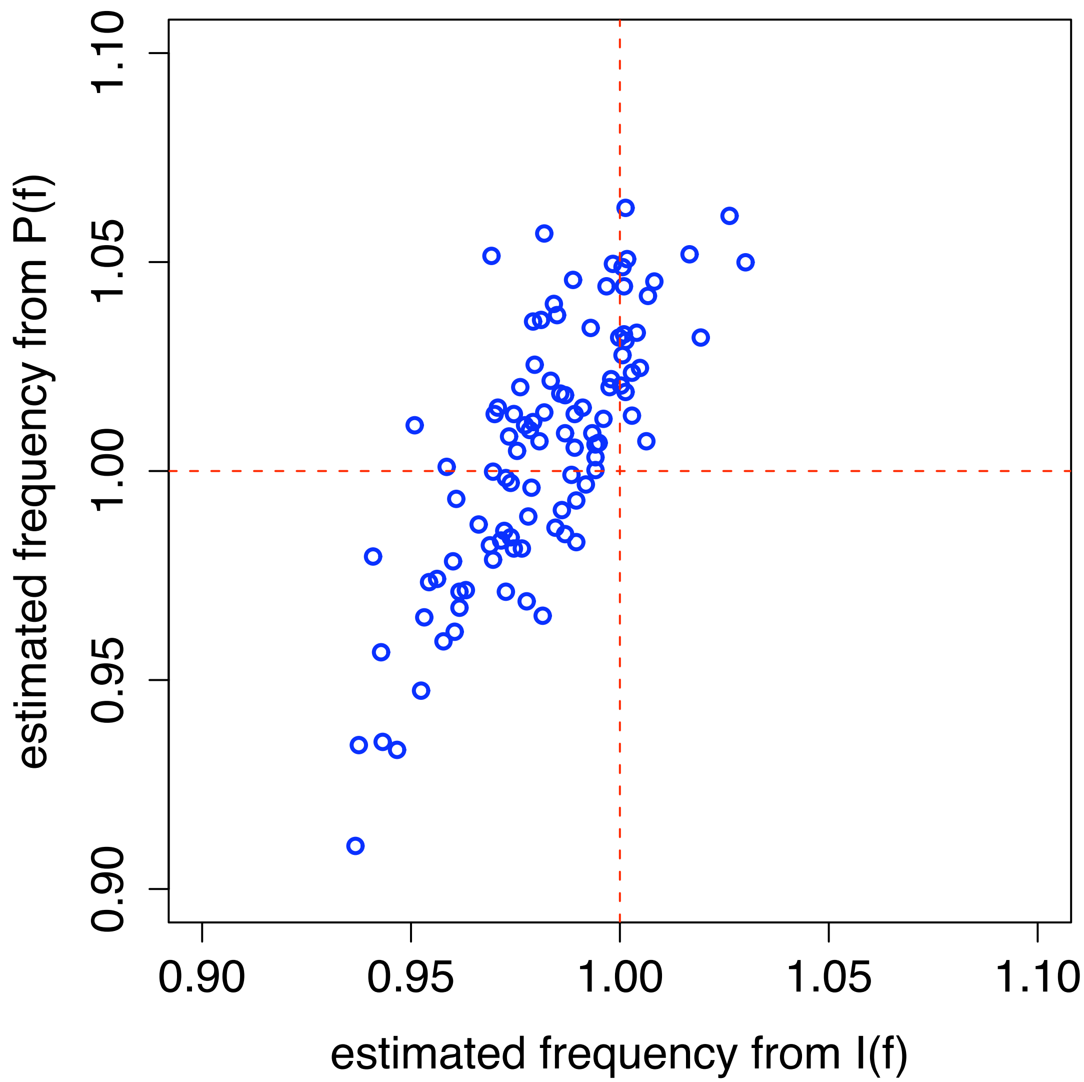
CSD for Example of Toy Parametric Model, $\sigma_\kappa = 1.28$



CSD for Example of Toy Parametric Model, $\sigma_\kappa = 2.56$



Estimation of f_0: CSD $P(f)$ vs. Periodogram $I(f)$



	$P(f)$	$I(f)$
bias	-0.003	-0.020
SD	0.065	0.019
RMSE	0.065	0.028

(10,000 replications)

Correntropy: IV

- Q: what is best way to pick σ_K?
- Q: how are peaks in CSD related to harmonic variations?
- Q: evidently $P(f)$ can be negative, so in what sense can it be regarded as a density function?
- for irregularly sampled series, must resort to slotting procedure (unappealing, but necessary)
- note: locations of peaks in CSD refined through maximization of information potential (IP) metric, so inaccuracies in CSD might be irrelevant
- Q: computational issues?
- Q: robustness?

Gaussian Processes Approach: I

- similar to before, let $y_j = g(x_j) + \epsilon_j$ be model for observations y_j of brightness of a star at times x_j
- letting $\mathbf{y} = [y_1, \ldots, y_n]^T$ & $\mathbf{g} = [g(x_1), \ldots, g(x_n)]^T$, consider

$$\mathbf{y} \sim \mathcal{N}(\mathbf{g}, \sigma_\epsilon^2 I_n), \text{ while } \mathbf{g} \,|\, \boldsymbol{\theta} \sim \mathcal{N}(\mathbf{0}, K_{\boldsymbol{\theta}}),$$

  where $K_{\boldsymbol{\theta}}$ is an $n \times n$ covariance matrix whose (j, k)th element is given by

$$K_{\boldsymbol{\theta},j,k} = \beta \exp\left(-\sin^2(2\pi f_0[x_j - x_k])/l^2\right) \text{ with } \boldsymbol{\theta} = [\beta, f_0, l]^T$$

- $K_{\boldsymbol{\theta}}$ is covariance matrix for samples from a continuous parameter stationary process with autocovariance function

$$\gamma(\tau) = \beta \exp\left(-\sin^2(2\pi f_0 \tau)/l^2\right)$$

Gaussian Processes Approach: II

- to estimate f_0, need to compute potentially messy likelihood:
$$-\frac{1}{2}\mathbf{y}^T\left(K_{\boldsymbol{\theta}}+\sigma^2 I_n\right)^{-1}\mathbf{y}-\frac{1}{2}\log\left|K_{\boldsymbol{\theta}}+\sigma^2 I_n\right|+C,$$
- alternatively, could specify Gaussian process capturing periodic or quasi-periodic variations indirectly within a state-space framework, thus allowing use of Kalman filter (KF)
- can adjust basic structural model (BSM) used in econometrics, which assumes regular sampling, to work with irregular sampling (can be handled readily by state space models and KF)
- BSM has terms for periodic variations (seasonal component γ_j) and quasi-periodic variations (business cycle ψ_j):
$$y_j = \mu_j + \gamma_j + \psi_j + \epsilon_j,$$
where μ_j is trend, and ϵ_j is an error term

Gaussian Processes Approach: III

- dynamics of seasonal component dictated by pair(s) of state variables patterned as follows:

$$\begin{pmatrix} \gamma_{j+1} \\ \gamma_{j+1}^* \end{pmatrix} = \begin{pmatrix} \cos(\lambda_s) & \sin(\lambda_s) \\ -\sin(\lambda_s) & \cos(\lambda_s) \end{pmatrix} \begin{pmatrix} \gamma_j \\ \gamma_j^* \end{pmatrix} + \begin{pmatrix} \omega_j \\ \omega_j^* \end{pmatrix},$$

  where

$$\begin{pmatrix} \omega_j \\ \omega_j^* \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma_\omega^2 I_2),$$

  with ω_j and ω_j^* being uncorrelated with previous disturbances

- $\sigma_\omega^2 = 0$ yields harmonic process with frequency dictated by λ_s, which can be treated as known or estimated

- $\sigma_\omega^2 > 0$ yields nonstationary process; i.e., amplitudes & phases can evolve

Gaussian Processes Approach: IV

- business cycle dictated by single pair of state variables:

$$\begin{pmatrix} \psi_{j+1} \\ \psi^*_{j+1} \end{pmatrix} = \rho \begin{pmatrix} \cos(\lambda_c) & \sin(\lambda_c) \\ -\sin(\lambda_c) & \cos(\lambda_c) \end{pmatrix} \begin{pmatrix} \psi_j \\ \psi^*_j \end{pmatrix} + \begin{pmatrix} \kappa_j \\ \kappa^*_j \end{pmatrix},$$

  where

$$\begin{pmatrix} \kappa_j \\ \kappa^*_j \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, \sigma^2_\kappa I_2),$$

  with κ_j and κ^*_j being uncorrelated with previous disturbances
- for $|\rho| < 1$, component is a stationary ARMA(2,1) process with associated period $2\pi/\lambda_c$
- above can (presumably!!!) be adjusted to work with irregular sampling, thus providing a Gaussian process with covariance matrix implicitly defined by ρ, λ_c and σ^2_κ (have interpretations analogous to those for l, f_0 and β in $\gamma(\tau)$ dictating $K_{\boldsymbol{\theta}}$)

Gaussian Processes Approach: V

- for problem of interest here, entertain either

$$y_j = \gamma_j + \epsilon_j \text{ or } y_j = \psi_j + \epsilon_j$$

  as appropriate models for capturing periodicity

- advantage is ability to evaluate likelihood function in a computationally easy manner (maintains Bayesian framework)
- Gaussian process approach does not appear to have any computational advantages over Hall et al. (2000)

Conclusions

- period estimation is in theory an easy statistical problem (n^3 rate of convergence), so could entertain strategy of trading efficiency for computation speed-up (direction for future research?)
- correntropy and Gaussian process approaches are of interest, but need to compare carefully with existing methods
- looking forward to new developments from Protopapas et al.!

Thanks to . . .

- Eric and Jogesh for the invitation to participate and for their considerable efforts in putting SCMA V together

References

- J. Durbin and S. J. Koopman (2001), *Time Series Analysis by State Space Models*, Oxford, England: Oxford University Press
- P. Hall (2008), 'Nonparametric Methods for Estimating Periodic Functions, with Applications in Astronomy,' in *COMPSTAT 2008: Proceedings in Computational Statistics*, 18th Symposium, Porto, Portugal, Paula Brito, editor, Physica-Verlag, Heidelberg, 2008 pp. 3–18
- P. Hall, J. Reimann and J. Rice (2000), 'Nonparametric Estimation of a Periodic Function,' *Biometrika*, **87**, pp. 545–557
- A. C. Harvey (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, England: Cambridge University Press
- P. Huijse, P. A. Estévez, P. Zegers, J. C. Principe and P. Protopapas (2011), 'Period Estimation in Astronomical Time Series Using Slotted Correntropy,' *IEEE Signal Processing Letters*, **18**, pp. 371–374
- S. J. Koopman and K. M. Lee (2009), 'Seasonality with Trend and Cycle Interactions in Unobserved Components Models,' *Journal of the Royal Statistical Society – Series C (Applied Statistics)*, **58**, pp. 427–448