Estimation of ACVF and ACF: I

• given a time series presumed to be a realization of a portion $X_1, X_2, \ldots, X_n$ of a stationary process, overheads II–62 and II–63 stated definitions for a sample ACVF and ACS that are realizations of the RVs

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

• here we look at these estimators of the ACVF $\gamma(h)$ and the ACF $\rho(h) = \gamma(h)/\gamma(0)$ in more detail

• prior to doing so, let’s review three basic properties of the ACVF we’ve mentioned already, and then introduce a fourth
Four Basic Properties of ACVF $\{\gamma(h)\}$: I

1. $\gamma(0) \geq 0$ (since $\gamma(0) = \text{var} \{X_t\}$, restatement of $\text{var} \{X_t\} \geq 0$)
2. $|\gamma(h)| \leq \gamma(0)$ for all $h$ (since $\rho(h) = \gamma(h)/\gamma(0)$ and $|\rho(h)| \leq 1$ because it is a correlation coefficient – see overhead II–15)
3. $\gamma(-h) = \gamma(h)$, i.e., $\gamma(h)$ is an even function (overhead II–49)
4. sequence $\{\gamma(h)\}$ is nonnegative definite – by definition this means that, if $t_1, t_2, \ldots, t_n$ are any $n$ integers and if $a_1, a_2, \ldots, a_n$ are any $n$ real-valued numbers, then we must have

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(t_i - t_j) \geq 0$$

• note: same concept sometimes called positive semidefinite
Four Basic Properties of ACVF \( \{\gamma(h)\} \): II

- to see that property 4 is true, consider

\[ Y \equiv \sum_{i=1}^{n} a_i X_{t_i}, \]

i.e., a linear combination of \( n \) RVs arbitrarily picked from \( \{X_t\} \)

- recalling that \( \text{var} \{Y\} \) must be nonnegative, note that

\[
\text{var} \{Y\} = \text{cov} \{Y, Y\} = \text{cov} \left\{ \sum_{i=1}^{n} a_i X_{t_i}, \sum_{j=1}^{n} a_j X_{t_j} \right\}
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \text{cov} \{X_{t_i}, X_{t_j}\}
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \gamma(t_i - t_j),
\]

thus showing that \( \{\gamma(h)\} \) is nonnegative definite
Four Basic Properties of ACVF \( \{\gamma(h)\} \): III

- theorem: a real-valued function defined on the integers is the ACVF for some stationary process if and only if it is even and nonnegative definite

- previous overhead establishes one part of theorem (if \( \{\gamma(h)\} \) is an ACVF, then it is nonnegative definite)

- second part (if \( \{\gamma(h)\} \) is nonnegative definite, then there is a stationary process that has \( \{\gamma(h)\} \) as its ACVF) is more difficult to establish

  – in fact, can show that, if \( \{\gamma(h)\} \) is nonnegative definite, there exists a Gaussian stationary process having \( \{\gamma(h)\} \) as its ACVF (i.e., any finite collection of RVs from \( \{X_t\} \) obeys a multivariate normal distribution)
Four Basic Properties of ACVF \( \{\gamma(h)\} \): IV

- given an arbitrary even function \( \{\kappa(h)\} \) defined on the integers, it is usually quite difficult to show that it is nonnegative definite based directly on the definition of this concept (Example 2.1.1 of B&D gives a rare instance where this approach works)

- two approaches used in practice to show that \( \{\kappa(h)\} \) is non-negative definite:
  1. find a stationary process that has \( \{\kappa(h)\} \) as its ACVF, which means \( \{\kappa(h)\} \) must be nonnegative definite (thus \( \{\cos (ch)\} \) is such because it is the ACVF for stationary process \( X_t = Z_1 \cos (ct) + Z_2 \sin (ct) \) considered in Problem 2(b))
  2. show that \( \{\kappa(h)\} \) arises from an integrated spectrum and appeal to a powerful theorem relating such spectra to non-negative definite functions (discussed in Stat/EE 520)
Estimation of ACVF and ACF: II

• recall that, for $|h| \leq n - 1$,

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \bar{X}_n)(X_t - \bar{X}_n)$$

• expression for $E\{\hat{\gamma}(h)\}$ is messy, so let’s consider instead

$$\bar{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \mu)(X_t - \mu),$$

for which

$$E\{\bar{\gamma}(h)\} = \frac{1}{n} \sum_{t=1}^{n-|h|} E\{(X_{t+|h|} - \mu)(X_t - \mu)\} = \frac{n - |h|}{n} \gamma(h) \neq \gamma(h)$$

in general when $h \neq 0$; i.e., $\bar{\gamma}(h)$ is a biased estimator.
Estimation of ACVF and ACF: III

• rather than using $\bar{\gamma}(h)$, might seem more natural to consider

$$\tilde{\gamma}(h) = \frac{1}{n - |h|} \sum_{t=1}^{n-|h|} (X_{t+|h|} - \mu)(X_t - \mu) = \frac{n}{n - |h|} \bar{\gamma}(h)$$

which differs from $\bar{\gamma}(h)$ only in its divisor, and for which

$$E\{\tilde{\gamma}(h)\} = \frac{1}{n - |h|} \sum_{t=1}^{n-|h|} E\{(X_{t+|h|} - \mu)(X_t - \mu)\} = \gamma(h);$$

i.e., $\tilde{\gamma}(h)$ is a unbiased estimator

• returning now to $\hat{\gamma}(h)$ (i.e., we use $\bar{X}$ rather than $\mu$), could consider using $\frac{n}{n-|h|}\hat{\gamma}(h)$ in view of above result

• $\hat{\gamma}(h) & \frac{n}{n-|h|}\hat{\gamma}(h)$ called, respectively, biased & unbiased ACVF estimators (even though latter is actually biased in general!)
Wind Speed Time Series $\{x_t\}$
Biased & Unbiased Sample ACVF for Wind Speed
Estimation of ACVF and ACF: IV

• as sample ACVFs for wind speed time series demonstrate, unbiased ACVF estimate \( \{ \frac{n}{n-|h|} \hat{\gamma}(h) \} \) need not satisfy requirements needed to be a theoretical ACVF
• by contrast, biased ACVF estimate \( \{ \hat{\gamma}(h) \} \) always satisfies the requirements (in particular, it is always nonnegative definite)
  – McLeod and Jiménez (1984, 1985) present a clever proof based upon the ACVF for moving average processes
• for time series models in common use, biased estimator typically has smaller mean square error than unbiased estimator, which provides additional rationale for preferring \( \hat{\gamma}(h) \)
Estimation of ACVF and ACF: V

- for time series models we will be considering later on,
  \[ \hat{\rho}_k = [\hat{\rho}(1), \ldots, \hat{\rho}(k)]' \]
is approximately \( \mathcal{N}(\rho_k, W/n) \) for large \( n \) (\( k \) fixed & \( k/n \) small),
where
\[ \rho_k = [\rho(1), \ldots, \rho(k)]', \]
and \((i, j)\)th element of \( k \times k \) covariance matrix \( W \) is given by
Bartlett’s formula:
\[
\begin{align*}
  w_{i,j} &= \sum_{h=1}^{\infty} \left[ \rho(h + i) + \rho(h - i) - 2\rho(i)\rho(h) \right] \\
  &\quad \times \left[ \rho(h + j) + \rho(h - j) - 2\rho(j)\rho(h) \right]
\end{align*}
\]
- for IID noise, \( w_{i,j} = 1 \) if \( i = j \) and \( = 0 \) otherwise, leading to
previously stated result that \( \hat{\rho}(1), \ldots, \hat{\rho}(k) \) are approximately
IID \( \mathcal{N}(0, 1/n) \) RVs (see overhead II–64)
Example – Bartlett’s Formula for MA(1) Process

- for MA(1) model $X_t = Z_t + \theta Z_{t-1}$ with $Z_t \sim \text{WN}(0, \sigma^2)$, have

$$w_{h,h} = \begin{cases} 
1 - 3\rho^2(1) + 4\rho^4(1), & h = 1, \\
1 + 2\rho^2(1), & h > 1,
\end{cases}$$

so $\hat{\rho}(h)$ is approximately $\mathcal{N}(\rho(h), w_{h,h}/n)$ for large $n$ (recall that $\rho(1) = \theta/(1 + \theta^2)$ and $\rho(h) = 0$ for $h \geq 2$)

- consider $n = 200$ observations from simulated Gaussian MA(1) process with $\theta = 0.8$ and $\sigma^2 = 1$

- true ACF is

$$\rho(h) = \begin{cases} 
1, & h = 0, \\
0.8/(1 + 0.8^2) \approx 0.4878, & h = \pm 1, \\
0, & \text{otherwise}
\end{cases}$$
Simulated MA(1) Time Series

\[ x_t \]

\[ t \]

0 50 100 150 200

\(-4\) \(-2\) \(0\) \(2\)
True & Sample ACFs with 95% Confidence Bounds
**Example – Bartlett’s Formula for AR(1) Process**

- for AR(1) model $X_t = \phi X_{t-1} + Z_t$ with $Z_t \sim \text{WN}(0, \sigma^2)$ and $|\phi| < 1$, have
  $$w_{h,h} = \frac{(1 - \phi^{2h})(1 + \phi^2)}{1 - \phi^2} - 2h\phi^{2h}, \quad h = 1, 2, \ldots$$

  so $\hat{\rho}(h)$ is approximately $\mathcal{N}(\rho(h), w_{h,h}/n)$ for large $n$ (recall that $\rho(h) = \phi^h$ for $h \geq 1$)

- consider two examples
  - residuals $\{r_t\}$ from least squares line fit to Lake Huron levels
  - wind speed series $\{x_t\}$

- for both examples, will compare sample ACF with AR(1) model based on setting $\phi$ to $\hat{\rho}(1)$

- yields $\hat{\phi} = 0.762$ for $\{r_t\}$ and $\hat{\phi} = 0.891$ for $\{x_t\}$
Residuals $r_t = x_t - \hat{c}_0 - \hat{c}_1 t$ from Least Squares Fit
Model & Sample ACFs with 95% Confidence Bounds
Model & Sample ACFs with 95% Confidence Bounds
References
