STA 542 – Homework 3

Due: Wednesday, September 23, at the start of class

Credit. This homework is graded for effort rather than correctness. A substantive attempt addresses every assigned part, shows the reasoning behind each answer, and interprets any computational output. Full effort credit does not require a correct solution: show what you tried, including unsuccessful approaches, and identify precisely where you remain stuck.

Submission. Submit one ordered hard copy, secured with a staple or paper clip, with your name clearly written on the first page. Mathematical and statistical reasoning must be written legibly by hand. R code, numerical output, tables, and plots may be printed and appended, or sketched when exact output is unnecessary. Label every computational exhibit, refer to it in the handwritten response, and explain what it shows. Use material through Workbook 6.

Problems

Problem 1. Likelihood, moments, and long-run variance.

We observe a time series and want to estimate its mean and the uncertainty of the sample mean. Write $x_1, \ldots, x_T$ for the observed values of $X_1, \ldots, X_T$. Initially, assume the stationary causal AR(1) model

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t, \qquad |\phi| < 1,$$

where the unobserved noise variables Z_t are independent and identically distributed with Laplace density

$$f_Z(z; \sigma) = \frac{1}{\sqrt{2}\sigma} \exp\left(-\frac{\sqrt{2}|z|}{\sigma} \right), \qquad z \in \mathbb{R}, \quad \sigma > 0.$$

This noise distribution has mean zero and variance σ^2. The parameters μ, ϕ, and σ^2 are unknown.

(a) **Write the likelihood.** For candidate parameter values, define the residual

$$r_t(\mu, \phi) = x_t - \mu - \phi(x_{t-1} - \mu), \qquad t = 2, \ldots, T.$$

Explain why Z_t is independent of $X_1, \ldots, X_{t-1}$, and write the conditional density of X_t at x_t using f_Z and $r_t(\mu, \phi)$. Use the chain rule for conditional densities to write the likelihood of $X_2, \ldots, X_T$ given $X_1 = x_1$.

(b) **Find the maximum-likelihood estimates.** Use the log likelihood to show that, for fixed σ, maximizing over μ and ϕ minimizes

$$\sum_{t=2}^{T} |r_t(\mu, \phi)|.$$

Explain how you would numerically obtain the estimates of μ and ϕ. Next hold μ and ϕ fixed, assume the residuals are not zero, and differentiate with respect to σ. Find the maximizing value, and give an expression for the resulting estimate of σ^2.

(c) **Estimate using white-noise moments.** Keep the stationary causal AR(1) equation $X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$, but now assume only $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$. Also assume (X_t) is ACVF-learnable through lag one (Workbook 5, Definition 5.8).

Consider the sample ACVF from Workbook 5, Definition 5.5,

$$\widehat{\gamma}_T(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X}_T)(X_{t+h} - \bar{X}_T).$$

Use the AR(1) covariance equations from Workbook 3, Example 3.10, to express ϕ and σ^2 in terms of $\gamma(0)$ and $\gamma(1)$. Replace these by their sample estimates to obtain moment estimators, and justify their consistency using Proposition 5.9.

(d) **Compare estimation accuracy.** Implement the log likelihood and both estimators from parts (a)–(c).

Set the seed to 542. Simulate 500 independent records from the Laplace AR(1) model in part (a), with $\mu = 0$, $\phi = 0.8$, and $\sigma^2 = 1$. For each record, start at $X_0 = 0$, discard the first 500 observations, and retain the next $T = 2000$. Estimate all three parameters using both methods; for the moment method, estimate μ by the sample mean. Keep the first record for part (e).

For each parameter and method, report the mean estimate and root mean squared error across the 500 records, using the generating parameter as the true value. Which method estimates each parameter more accurately in this experiment?

Now let E_t be independent $\mathrm{Exp}(1)$ variables and define

$$Z_t = \frac{E_{t-1}(E_t - 1)}{\sqrt{2}}, \qquad X_t = 0.8 X_{t-1} + Z_t.$$

Verify that (Z_t) is white noise with variance one. Repeat the simulation and accuracy comparison using this noise, retaining the same Laplace fitting criterion and moment estimators. The true parameters remain $\mu = 0$, $\phi = 0.8$, and $\sigma^2 = 1$.

How does the accuracy comparison change? Explain why the moment equations from part (c) remain valid, although the conditional likelihood from part (a) is no longer correctly specified.

(e) **Estimate long-run variance.** Recall that

$$\tau^2 = \gamma(0) + 2\sum_{h=1}^{\infty} \gamma(h).$$

Under the AR(1) assumption, express τ^2 using only $\gamma(0)$ and $\gamma(1)$. Replace these covariances by their sample estimates to obtain an estimator of τ^2.

Compare this with the estimator using sample covariances through lag m, where $1 \le m < T$:

$$\widehat{\tau}^2_{m,T} := \widehat{\gamma}_T(0) + 2\sum_{h=1}^{m} \widehat{\gamma}_T(h).$$

For the Laplace AR(1) process in part (d), calculate τ^2 and express the covariance tail omitted beyond lag m as a function of m. Using fixed-lag consistency, find the probability limit of $\widehat{\tau}^2_{m,T}$ when m stays fixed as T grows. Does increasing T remove the truncation error?

For the first Laplace record saved in part (d), report a table containing the true long-run variance, the AR(1) estimate, and $\widehat{\tau}^2_{m,T}$ for $m \in \{2, 3, 5, 10\}$. Relate the comparison to your population calculation.

(f) **An incorrect AR(1) model.** Now consider

$$X_t = Z_t + \tfrac{1}{4}Z_{t-1} + \tfrac{1}{2}Z_{t-2} + \tfrac{3}{4}Z_{t-3},$$

where the noise variables Z_t are i.i.d. Laplace with mean zero and variance one. Derive the ACVF and long-run variance of this process.

Using fixed-lag consistency, find the probability limit of the AR(1) long-run variance estimate from part (e). For which fixed values of m is $\widehat{\tau}^2_{m,T}$ consistent for the actual long-run variance? Explain why the conclusion differs from part (e).

Set the seed to 543, generate independent Laplace noise values $Z_1, \ldots, Z_{2003}$ of variance one, and form $X_4, \ldots, X_{2003}$ using the equation above. Repeat the numerical comparison from part (e) for this record. Plot the sample ACVF and the AR(1) ACVF obtained from the moment estimates in part (c), at lags zero through ten. What dependence does the fitted AR(1) ACVF miss?

Problem 2. Estimating the mean deviation from a target.

The `TSA::deere2` data are 102 consecutive deviations from a target in an industrial process at Deere & Co. We want to estimate the mean deviation from the target and quantify the uncertainty of that estimate.

Model the observations as a weakly stationary process with unknown mean μ and ACVF γ. Estimate μ by $\bar{X}_T$. The question is how the assumed covariance structure affects our estimate of $\operatorname{Var}(\bar{X}_T)$.

(a) **Start with the observed moments.** Make a time-series plot of the observations in `deere2.csv`, using the column `deviation_micrometres`, with deviation from the target (in micrometers) on the vertical axis. Report the sample mean and compute the sample ACVF at lags zero through 20 using Workbook 5, Definition 5.5. Distinguish $\widehat{\gamma}_T(0)$ from an estimate of $\operatorname{Var}(\bar{X}_T)$: what does each quantify?

(b) **Fit the AR(1) candidate.** Use your moment estimator from Problem 1. For the conditional Gaussian AR(1) fit, use `lm()` to regress each observation on its predecessor with an intercept. Estimate ϕ by the slope and σ^2 by the mean squared residual, using divisor $T-1$. Report a two-row table (moments and conditional Gaussian likelihood) giving the estimated ϕ and innovation variance σ^2. Using the Gaussian fit, estimate the AR(1) ACVF using the covariance formula in Workbook 3, Example 3.10. Plot the sample ACVF at lags zero through 15 and overlay the fitted AR(1) ACVF. Label the covariance axis in squared micrometers and identify both estimates in a legend. Use the Gaussian fit for all AR(1) comparisons below.

(c) **Estimate the uncertainty of the average.** Implement the Bartlett estimator from Workbook 6, Definition 6.7, at $H \in \{0, 2, 5, 10, 20\}$:

$$\widehat{\tau}^2_{H,T} := \widehat{\gamma}_T(0) + 2\sum_{h=1}^{H}\left(1 - \frac{h}{H+1}\right)\widehat{\gamma}_T(h).$$

At $H = 0$, the sum is empty; this estimate ignores serial covariances. Report a six-row table: the five bandwidths and the fitted AR(1) method. For each row, give the estimated long-run variance and the resulting estimated standard error of the mean, $\sqrt{\widehat{\tau}^2/T}$. Label their units. The sample mean being assessed is the same in every row.

(d) **Interpret the comparison.** Write a short paragraph explaining how the covariance assumption affects the estimated uncertainty of the sample mean. Refer to specific lags in the covariance plot and values in the uncertainty table. Explain why similar AR(1) parameter estimates from two fitting methods need not validate the AR(1) covariance pattern. Does a smaller reported standard error establish that a method is more accurate?

Then use Proposition 6.8 to state the additional conditions under which a Bartlett estimate with a bandwidth sequence H_T consistently estimates the actual long-run variance. Explain why these calculations on one record do not establish those conditions or identify the true long-run variance.

Problem 3. What would help you understand this homework better?

Look back over your work on Problems 1 and 2. Identify three points that you would like to understand better, using the categories below. Each response should refer to a specific part of this homework.

(a) **A mathematical step.** Choose a calculation or proof step that you found difficult or could not fully justify. This might concern differentiating the log likelihood, summing an ACVF, or justifying a probability limit.

(b) **A statistical idea.** Choose a statistical idea used in your solution that remains unclear. For example, why can likelihood and moment estimates differ, or does agreement between two estimates support the model assumptions? The idea should arise in this homework, even if it also applies to independent data.

(c) **A time-series idea.** Choose something about dependence or learning from one observed path that you would like to understand better. This could arise when comparing the AR(1) covariance formula with the sample ACVF in Problem 1, or estimating the uncertainty of the sample mean in Problem 2.

For each response, write a short paragraph addressing the following:

- Identify the subpart and formulate your own question. Explain what you understood or could do, and where your reasoning became uncertain.
- Suggest how you could explore the question using a specific result, proof, example, or exercise from Workbooks 1–6. Describe a possible activity, such as working through a simpler case or discussing your attempted reasoning with a TA.
- Describe a short calculation, explanation, or example you could attempt without notes to check your understanding.

These are possible approaches to further study; you are not being asked to complete additional work for this assignment.