STA 542 – Homework 1

Due: Wednesday, September 9, 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.

Problems

Problem 1. One AR equation, two predictive laws. Fix $\phi = 0.8$ and let $(X_t)_{t\in\mathbb{Z}}$ be the stationary causal solution of

$$X_t = \phi X_{t-1} + Z_t.$$

Compare the following two specifications for the innovations:

$$\begin{aligned} &\text{Model G:} \quad Z_t \overset{\text{iid}}{\sim} N(0,1), \\ &\text{Model B:} \quad \mathbb{P}(Z_t = -1) = \mathbb{P}(Z_t = 1) = \tfrac{1}{2}, \quad (Z_t) \text{ i.i.d.} \end{aligned}$$

(a) For each model, compute $\mathbb{E}[X_t]$ and the autocovariance function

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}).$$

Under Model G, use the causal representation

$$X_t = \sum_{j=0}^{\infty} \phi^j Z_{t-j}$$

to explain why (X_t) is a Gaussian process and identify its mean function and covariance kernel. Under Model B, explain why the process is not Gaussian even though it has the same mean and autocovariance function. *Hint: consider* $Z_t = X_t - \phi X_{t-1}$.

(b) Suppose $(x_1, \ldots, x_T)$ has been observed. Express your answers in terms of the most recent observation x_T. Under each model, find the best linear predictor of X_{T+1} from the entire observed history and compute its mean squared forecast error. Then compute the conditional mean of X_{T+1} given the observed history. Explain why the i.i.d. assumption is enough to make the conditional mean agree with the best linear predictor in both models, and why the earlier observations do not enter.

(c) Derive the conditional distribution of X_{T+1} given the observed history under each model. Under Model G, construct the equal-tail 95% conditional prediction interval; use $z_{0.975} \approx 1.96$. Under Model B, give the conditional support as a prediction set, then find the shortest interval having conditional coverage at least 95%. State its actual conditional coverage.

(d) In two or three complete sentences, explain how the two models can have the same AR equation, autocovariance function, and best linear predictor while producing different prediction sets.

Problem 2. A plug-in forecast under two noise specifications. The R series `datasets::co2` contains monthly atmospheric carbon dioxide measurements from 1959 through 1997 at the Mauna Loa Observatory in Hawaii.[1] Concentration is reported in parts per million (ppm): a level of 360 ppm means approximately 360 carbon dioxide molecules per million molecules of dry air. Thus the series records the composition of sampled air, not the amount of carbon dioxide emitted during a month. The monthly readings have a strong seasonal cycle. We average the twelve readings within each year because the target here is annual mean concentration, not the month-to-month seasonal pattern.

The following code averages the monthly readings within each year, selects the observations available through 1989, and forms their eleven annual changes in time order.

```
co2_annual <- aggregate(co2, nfrequency = 1, FUN = mean)
x_observed <- window(co2_annual, start = 1978, end = 1989)
d_observed <- diff(as.numeric(x_observed))
x_1989 <- as.numeric(tail(x_observed, 1))
d_bar <- mean(d_observed)
```

The annual changes from 1959–1977 have standard deviation about 0.5 ppm per year, so throughout Problems 2 and 3 we use $\sigma = 0.5$ ppm as a fixed working value. The 1978–1989 observations form the window available when the forecast is made. Do not inspect the 1990–1997 observations until the forecast has been constructed in part (c).

Let X_t denote annual mean concentration in year t, and treat $X_{1978} = x_{1978}$ as the fixed starting level. Let $D_t := X_t - X_{t-1}$ denote the annual change. At a fixed candidate drift μ, use the model

$$D_t = \mu + Z_t, \qquad Z_t \overset{\text{iid}}{\sim} N(0, \sigma^2), \qquad \sigma = 0.5 \text{ ppm}.$$

The same μ governs the observed and future changes. A *plug-in forecast* estimates it by the mean $\widehat{\mu} = \overline{d}$ of the observed 1979–1989 changes and then computes the forecast as if $\mu = \widehat{\mu}$ were known. The procedure does not make the drift known; treating the estimate as exact is the approximation under study.

(a) Run the displayed data-preparation code. Report x_{1989} and the sample mean $\overline{d}$ of the eleven values in `d_observed`. Interpret $\overline{d}$ in the units of the data.

(b) At a fixed μ, derive the conditional distribution of X_{1989+h} for $h = 1, \ldots, 8$. Substitute $\widehat{\mu} = \overline{d}$ to obtain the plug-in distribution. Report its center and equal-tail 95% interval for the 1997 annual mean; use $z_{0.975} \approx 1.96$.

[1] Keeling and Whorf, Scripps Institution of Oceanography; see Scripps CO_2 data. The R series reports the preliminary 1997 SIO manometric mole fraction scale; three missing months in 1964 were linearly interpolated.

(c) In R, create the eight plug-in forecast centers and standard deviations from your formulas in part (b). Only then reveal the held-out observations by running

```
x_held_out <- window(co2_annual, start = 1990, end = 1997)
```

Report the recorded 1997 value and the forecast error $x_{1997} - \widehat{x}_{1997|1989}$. Did the plug-in interval contain the recorded value? In one sentence, state which source of uncertainty the interval includes and which source it omits.

(d) Now weaken the innovation assumption to $(Z_t) \sim \mathrm{WN}(0, \sigma^2)$, while keeping μ fixed, and write $W := (X_{1978}, \ldots, X_{1989})$. Show that the future forecast error is uncorrelated with every component of W. Hence find the best linear predictor of X_{1989+h} from W and its mean squared error. Express $\mathbb{E}[X_{1989+h} \mid W]$ in terms of conditional expectations of the future innovations. Explain why white noise does not determine this conditional mean, whereas i.i.d. innovations would.

(e) Continue with the white-noise specification and treat the observed $\overline{d}$ as the fixed drift, as in the plug-in calculation. Use Chebyshev's inequality and the mean squared error from part (d) to construct a 95% marginal prediction interval for X_{1997} under this fitted second-order law. Report its endpoints and compare its half-width with the Gaussian interval from part (b). For each interval, state whether its guarantee is exact or conservative and conditional or marginal. Finish by stating which uncertainty both intervals omit.

Problem 3. A Bayesian forecast for an uncertain drift. Problem 2 replaced the unknown drift by the estimate $\widehat{\mu} = \overline{d}$ and then forecast as if that estimate were the true value. A Bayesian analysis instead uses a distribution to describe our uncertainty about the one unknown drift. Before the recent observations are used, this distribution is the *prior*. The increment model supplies the *likelihood* of the recent observations at each possible drift. Bayes' theorem combines the prior and likelihood to obtain the *posterior*. The posterior predictive distribution then averages the fixed-drift forecasts from Problem 2 over the posterior.

The parameter μ is one common drift shared by every observed and future increment. It is not redrawn each year. Its prior describes uncertainty about this one value before the 1978–1989 observations are used.

The annual changes from 1959–1977 had mean about 1.0 ppm per year. We summarize that earlier information by the prior

$$\mu \sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right), \qquad \mu_0 = 1.0, \qquad n_0 = 4.$$

The choice $n_0 = 4$ treats the older record as four effective observations. It is a modeling choice that discounts the older information.

Relabel the $n = 11$ recent increments as $D_1, \ldots, D_n$. At each possible value of μ, the conditional model is

$$D_1, \ldots, D_n \mid \mu \overset{\text{iid}}{\sim} N(\mu, \sigma^2).$$

This model supplies the likelihood. Conditional on the same μ, the future increments follow this normal law independently of the observed increments. The older years enter through the prior, the recent increments update that prior, and the resulting posterior weights the fixed-drift forecasts.

(a) Write the likelihood of the observed recent increments $D_1 = d_1, \ldots, D_n = d_n$. Combine it with the prior to derive

$$\mu \mid D_1 = d_1, \ldots, D_n = d_n \sim N(m, s^2).$$

Express m and s^2 in terms of $n_0, \mu_0, n, \overline{d}$, and σ^2, then evaluate them. State how the earlier information and recent data enter the posterior mean.

(b) Use the fixed-drift forecast from Problem 2 and average it over the posterior for μ to show that, for $h = 1, \ldots, 8$,

$$X_{1989+h} \mid D_1 = d_1, \ldots, D_n = d_n \sim N(x_{1989} + hm, h\sigma^2 + h^2 s^2).$$

Identify the source of each variance term. Report the posterior predictive center and equal-tail 95% interval for the 1997 annual mean.

(c) Use R to compute the posterior mean and variance of μ, followed by the eight Bayesian forecast centers and standard deviations from part (b). Write your own R code to produce a two-panel figure. In both panels, show the observations through 1989 in black and the held-out observations in red, using the same axes. Add the plug-in mean and pointwise 95% bounds to the left panel, and the Bayesian mean and pointwise 95% bounds to the right panel. Label the axes, panels, and series, and submit the figure. Compare the 1997 plug-in and Bayesian centers and interval half-widths, and state which center is closer to the recorded value.

(d) Let $I_{\mathrm{B}}(d)$ denote the 1997 posterior predictive interval from part (b). Write the conditional probability statement that makes it a 95% posterior predictive interval. In two sentences, state what remains random after the increments have been observed and explain why one held-out path cannot verify a 95% probability claim.