PS1 Solutions

Jingle Fu

Problem 1: Bayesian inference in a stationary AR(1)

Model. We observe a stationary AR(1) initialized in the infinite past,

$$y_t = \phi\, y_{t-1} + u_t, \qquad u_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1), \quad |\phi| < 1,$$

with prior $\phi \sim \text{Unif}(-1,1)$.

(a) Conditional and unconditional likelihoods

From the model $y_t \mid y_{t-1}, \phi \sim \mathcal{N}(\phi y_{t-1}, 1)$, we have:

$$p(y_t \mid y_{t-1}, \phi) = (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(y_t - \phi y_{t-1})^2\right\}.$$

Given the Markov property of the AR(1) process, the joint density of the sample conditional on y_0 is the product of the conditional densities:

$$\begin{aligned}
p(Y_{1:T} \mid \phi, y_0) &= \prod_{t=1}^{T} p(y_t \mid y_{t-1}, \phi) \\
&= \prod_{t=1}^{T} (2\pi)^{-1/2} \exp\left\{-\frac{1}{2}(y_t - \phi y_{t-1})^2\right\} \\
&= (2\pi)^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2\right\}.
\end{aligned}$$

The unconditional likelihood treats y_0 as a random variable. Since the process is stationary ($|\phi| < 1$) and initialized in the infinite past, the unconditional distribution is:

$$y_0 \mid \phi \sim \mathcal{N}\left(0, \frac{1}{1-\phi^2}\right).$$

Thus,

$$p(y_0 \mid \phi) = \left(\frac{1-\phi^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}(1-\phi^2)y_0^2\right\}.$$

The unconditional likelihood is then:

$$
\begin{aligned}
p(Y_{1:T}, y_0 \mid \phi) &= p(Y_{1:T} \mid \phi, y_0)\, p(y_0 \mid \phi) \\
&= (2\pi)^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2\right\} \left(\frac{1-\phi^2}{2\pi}\right)^{1/2} \exp\left\{-\frac{1}{2}(1-\phi^2)y_0^2\right\} \\
&= (2\pi)^{-(T+1)/2}(1-\phi^2)^{1/2} \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2 + (1-\phi^2)y_0^2\right]\right\}.
\end{aligned}
$$

(b) Posterior (conditional likelihood + uniform prior)

The posterior is proportional to the likelihood times the prior:

$$
p(\phi \mid Y_{1:T}, y_0) \propto p(Y_{1:T} \mid \phi, y_0)\, p(\phi).
$$

Substituting the expressions:

$$
\begin{aligned}
p(\phi \mid Y_{1:T}, y_0) &\propto \left[(2\pi)^{-T/2} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2\right\}\right] \left[\frac{1}{2}\mathbf{1}\{-1<\phi<1\}\right] \\
&\propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2\right\} \mathbf{1}\{-1<\phi<1\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\left(\sum_t y_{t-1}^2\right)\phi^2 - 2\left(\sum_t y_t y_{t-1}\right)\phi\right]\right\} \mathbf{1}\{-1<\phi<1\}.
\end{aligned}
$$

Let:

$$
\hat{\phi}_{ML} = \frac{\sum_{t=1}^{T} y_{t-1}y_t}{\sum_{t=1}^{T} y_{t-1}^2}, \quad V = \frac{1}{\sum_{t=1}^{T} y_{t-1}^2}.
$$

Then the posterior is proportional to:

$$
p(\phi \mid Y_{1:T}, y_0) \propto \exp\left\{-\frac{1}{2V}(\phi - \hat{\phi}_{ML})^2\right\} \mathbf{1}\{-1<\phi<1\}.
$$

The full normalized posterior density is:

$$
p(\phi \mid Y_{1:T}, y_0) = \frac{\frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{1}{2V}(\phi - \hat{\phi}_{ML})^2\right\} \mathbf{1}\{-1<\phi<1\}}{\int_{-1}^{1} \frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{1}{2V}(x - \hat{\phi}_{ML})^2\right\} dx}.
$$

The denominator is the probability that a $\mathcal{N}(\hat{\phi}_{ML}, V)$ random variable falls in $(-1, 1)$, which we can write as:

$$
P(-1<\phi<1) = \Phi\left(\frac{1-\hat{\phi}_{ML}}{\sqrt{V}}\right) - \Phi\left(\frac{-1-\hat{\phi}_{ML}}{\sqrt{V}}\right),
$$

where $\Phi(\cdot)$ is the standard Normal CDF.
Then the posterior density is:

$$p(\phi \mid Y_{1:T}, y_0) = \frac{\frac{1}{\sqrt{2\pi V}} \exp\left\{-\frac{1}{2V}(\phi - \hat{\phi}_{ML})^2\right\}}{\Phi\left(\frac{1-\hat{\phi}_{ML}}{\sqrt{V}}\right) - \Phi\left(\frac{-1-\hat{\phi}_{ML}}{\sqrt{V}}\right)} \mathbf{1}\{-1 < \phi < 1\}.$$

If we used an improper prior $p(\phi) \propto c$, the posterior would be:

$$p(\phi \mid Y_{1:T}, y_0) \propto \exp\left\{-\frac{1}{2V}(\phi - \hat{\phi}_{ML})^2\right\},$$

which is the kernel of an *unrestricted* $\mathcal{N}(\hat{\phi}_{ML}, V)$ distribution.

(c) 95% credible set under $|\phi| < 1$

We are given:

$$\hat{\phi}_{ML} = 0.95, \qquad \sum y_t y_{t-1} = 20.$$

From the definition:

$$\hat{\phi}_{ML} = \frac{\sum_{t=1}^T y_t y_{t-1}}{\sum_{t=1}^T y_{t-1}^2} \quad \Rightarrow \quad \sum_{t=1}^T y_{t-1}^2 = \frac{\sum_{t=1}^T y_t y_{t-1}}{\hat{\phi}_{ML}} = \frac{20}{0.95} \approx 21.0526.$$

The posterior variance is:

$$V = \frac{1}{\sum_{t=1}^T y_{t-1}^2} \approx \frac{1}{21.0526} \approx 0.0475.$$

So the posterior is $\phi \mid Y \sim \mathcal{N}(0.95, 0.0475)$ truncated to $(-1, 1)$.
A 95% credible set is an interval $[a, b]$ such that:

$$\mathbb{P}(a \leq \phi \leq b \mid Y) = 0.95.$$

Since our posterior is a truncated Normal and the mean (0.95) is close to the upper truncation point (1), the HPD interval will not be symmetric.
Let F_{TN}^{-1} be the inverse CDF of the truncated Normal. Then:

$$a = F_{\text{TN}}^{-1}(0.05), \qquad b = 1,$$

since the density is decreasing beyond the mean towards 1.
Let $\mu = 0.95$, $\sigma = \sqrt{V} \approx \sqrt{0.0475} \approx 0.2179$. The CDF of the truncated Normal is:

$$F_{\text{TN}}(x) = \frac{\Phi\left(\frac{x-\mu}{\sigma}\right) - \Phi\left(\frac{-1-\mu}{\sigma}\right)}{\Phi\left(\frac{1-\mu}{\sigma}\right) - \Phi\left(\frac{-1-\mu}{\sigma}\right)}.$$

We want $F_{\text{TN}}(a) = 0.05$.
First, compute the denominator (normalization constant):

$$\Phi\left(\frac{1-0.95}{0.2179}\right) - \Phi\left(\frac{-1-0.95}{0.2179}\right) = \Phi(0.2294) - \Phi(-8.947) \approx 0.5907 - 0 = 0.5907.$$

Now,

$$F_{\text{TN}}(a) = \frac{\Phi\left(\frac{a-0.95}{0.2179}\right) - 0}{0.5907} = 0.05 \quad \Rightarrow \quad \Phi\left(\frac{a-0.95}{0.2179}\right) = 0.05 \times 0.5907 \approx 0.02954.$$

Then

$$\frac{a-0.95}{0.2179} = \Phi^{-1}(0.02954) \approx -1.89 \quad \Rightarrow \quad a \approx 0.95 - 1.89 \times 0.2179 \approx 0.538.$$

Thus, the 95% credible set is approximately

$$[0.538, 1].$$

(d) Impact of using the unconditional likelihood

If we use the unconditional likelihood $p(Y_{1:T}, y_0 \mid \phi)$, the posterior becomes:

$$p(\phi \mid Y_{1:T}, y_0) \propto p(Y_{1:T}, y_0 \mid \phi)\, p(\phi).$$

Substituting the expression from part (a):

$$p(\phi \mid Y_{1:T}, y_0) \propto (1-\phi^2)^{1/2} \exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{T}(y_t - \phi y_{t-1})^2 + (1-\phi^2)y_0^2\right]\right\} \mathbf{1}\{-1 < \phi < 1\}.$$

The term $(1-\phi^2)^{1/2}$ and the extra term $(1-\phi^2)y_0^2$ in the exponent mean the posterior is no longer a truncated Normal distribution. The kernel is not a simple exponential of a quadratic in ϕ. This makes analytical derivation of the posterior impossible. We would have to resort to numerical methods (e.g., importance sampling, MCMC) to approximate the posterior distribution. In large T, the additional term is $O_p(1)$ and differences vanish asymptotically.

Problem 2: Bayesian model selection in a Gaussian LM

Solution (a).
Data: $y = X_k\beta_k + u$, with $u \sim \mathcal{N}(0, I_n)$ and prior $\beta_k \sim \mathcal{N}(0, I_k)$. We consider nested

models M_k using the first k columns of X for $k = 1, \ldots, 10$. The marginal likelihood is

$$p(y \mid M_k) = \mathcal{N}\big(y \mid 0,\ I_n + X_k X_k'\big).$$

Using determinant and Woodbury identities leads to the numerically stable form

$$\log p(y \mid M_k) = -\tfrac{n}{2}\log(2\pi) - \tfrac{1}{2}\log\big|I_k + X_k' X_k\big| - \tfrac{1}{2}\Big(y'y - y'X_k\,(I_k + X_k'X_k)^{-1}X_k'y\Big).$$

With equal model priors, posterior model probabilities are proportional to $p(y \mid M_k)$. Following the instruction to do a *thin QR* on a Gaussian design matrix Z and use the Q factor as X, the k columns of X_k are orthonormal:

$$X_k'X_k = I_k.$$

Plugging this into the marginal likelihood yields a transparent model-comparison score:

$$\log p(y \mid M_k) = \text{const} \ + \ \underbrace{\tfrac{1}{4}\sum_{j=1}^{k}(x_j'y)^2}_{\text{fit gain}} \ - \ \underbrace{\tfrac{k}{2}\log 2}_{\text{Occam penalty}} \quad .$$

Thus, moving from model $k-1$ to model k changes the score by

$$\Delta_k = \tfrac{1}{4}(x_k'y)^2 - \tfrac{1}{2}\log 2.$$

Table 1: Average posterior probability by k and sample size

k	N=50	N=100	N=500
1	0.141	0.139	0.139
2	0.128	0.127	0.128
3	0.118	0.120	0.119
4	0.111	0.113	0.112
5	0.100	0.101	0.100
6	0.090	0.092	0.092
7	0.084	0.084	0.085
8	0.080	0.079	0.079
9	0.075	0.074	0.075
10	0.073	0.071	0.072

Asymptotic behavior of the posterior model probabilities.
With this particular design (thin-QR X so that $X'X = I$ for every n, and a prior $\beta \sim \mathcal{N}(0, I)$ that does *not* scale with n):

- The fit term $\frac{1}{4}\sum_{j=1}^{k}(x_j'y)^2$ is $O_p(1)$, because each projection $x_j'y$ has variance 1 regardless of n (the columns are unit-norm).

Table 2: Selection frequency (share) by k and sample size

k	N=50	N=100	N=500
1	0.428	0.402	0.412
2	0.118	0.132	0.142
3	0.085	0.107	0.100
4	0.097	0.098	0.089
5	0.055	0.057	0.053
6	0.042	0.043	0.053
7	0.046	0.035	0.038
8	0.049	0.053	0.034
9	0.039	0.033	0.037
10	0.041	0.040	0.042

- The penalty term $\frac{k}{2}\log 2$ is constant in n.

Therefore, as $n \to \infty$, the posterior model probabilities do *not* concentrate on the true k_0; instead, they converge to a stable distribution dominated by small k, which explains the near-invariance you observe going from $n = 50$ to $n = 500$. Equivalently, increasing n does not add information per column when the columns remain orthonormal with unit length; the projections determining the Bayes factors remain $O_p(1)$.

```
set.seed(2025)

log_marglik <- function(y, X) {
    n <- nrow(X); K <- ncol(X)
    yty <- sum(y*y)
    out <- numeric(K)
    for (k in 1:K) {
        Xk <- X[, 1:k, drop=FALSE]
        Sk <- diag(k) + crossprod(Xk) # I_k + X'X
        L  <- chol(Sk)
        logdet <- 2*sum(log(diag(L)))
        v <- forwardsolve(t(L), crossprod(Xk, y))  # L z = X' y
        w <- backsolve(L, v)  # L' w = z => w = S^{-1} X' y
        quad <- sum((Xk %*% w) * y)  # y' X S^{-1} X' y
        out[k] <- -0.5*n*log(2*pi) - 0.5*(yty - quad) - 0.5*logdet
      }
    out
  }

run_once <- function(n, kmax=10, k0=4, M=1000) {
    Z <- matrix(rnorm(n*kmax), n, kmax)
    qrZ <- qr(Z)
    X <- qr.Q(qrZ)  # thin Q with orthonormal columns
    beta <- c(rep(0.5, k0), rep(0, kmax-k0))
    mu <- as.vector(X %*% beta)
```

```r
        counts <- integer(kmax)
        post_sum <- numeric(kmax)
        for (m in 1:M) {
            y <- mu + rnorm(n)
            lml <- log_marglik(y, X)
            w <- exp(lml - max(lml))
            post <- w / sum(w)
            kstar <- which.max(post)
            counts[kstar] <- counts[kstar] + 1
            post_sum <- post_sum + post
          }
        data.frame(k=1:kmax,
        freq_selected=counts,
        avg_posterior=post_sum/M,
        freq_selected_share=counts/M)
      }

    res50  <- run_once(50,  kmax=10, k0=4, M=1000)
    res100 <- run_once(100, kmax=10, k0=4, M=1000)
    res500 <- run_once(500, kmax=10, k0=4, M=1000)

    library(tidyverse)
    library(knitr)

    res50$N  <- 50
    res100$N <- 100
    res500$N <- 500
    res_all <- bind_rows(res50, res100, res500)

    # 1) Average posterior probability
    avg_table <- res_all %>%
    select(k, N, avg_posterior) %>%
    pivot_wider(names_from = N, names_prefix = "N=", values_from = avg_
   posterior) %>%
    arrange(k)

    avg_latex <- knitr::kable(
    avg_table,
    format = "latex",
    booktabs = TRUE,
    caption = "Average posterior probability by $k$ and sample size",
    digits = 3,
    label = "tab:avg_posterior"
    )

    # 2) Selection frequency (share)
    freq_table <- res_all %>%
    select(k, N, freq_selected_share) %>%
```

```
  pivot_wider(names_from = N, names_prefix = "N=", values_from = freq_selected_share) %>%
  arrange(k)

freq_latex <- knitr::kable(
  freq_table,
  format = "latex",
  booktabs = TRUE,
  caption = "Selection frequency (share) by $k$ and sample size",
  digits = 3,
  label = "tab:freq_selected"
)

writeLines(as.character(avg_latex), con = "PS1_avg_posterior.tex")
writeLines(as.character(freq_latex), con = "PS1_freq_selected.tex")

```

Listing 1: R code for Problem 2