Bayesian Inference with Stan: End-to-End Workflow

This report presents a portfolio-ready Bayesian workflow implemented in R with Stan: synthetic data generation, posterior sampling via MCMC, convergence diagnostics (trace plots), and posterior uncertainty visualization (density plots).

R Implementation

The following script shows the core implementation used to simulate data, run Stan sampling, extract posterior draws, and generate diagnostics and posterior density plots.

```r
# ============================================================
# Bayesian Inference with Stan (R + rstan) — Portfolio Script
# ============================================================

# ---- Packages ----
library(rstan)
library(tibble)
library(ggplot2)

# ---- Reproducibility ----
set.seed(123)

# ---- Synthetic data generator ----
# Model used for simulation:
#   theta_i  ~ Normal(mu, 1/tau)
#   sigma2_i ~ Inverse-Gamma(a0, b0) implemented as 1 / Gamma(a0, b0)
#   y_i      ~ Normal(theta_i, sigma2_i)
generate_sample <- function(N, mu, tau, a0, b0) {
  theta  <- rnorm(N, mean = mu, sd = sqrt(1 / tau))
  sigma2 <- 1 / rgamma(N, shape = a0, rate = b0)
  y <- vapply(seq_len(N), function(i) {
    rnorm(1, mean = theta[i], sd = sqrt(sigma2[i]))
  }, numeric(1))
  data.frame(y = y, theta = theta, sigma2 = sigma2)
}

# ---- Simulate data ----
true_mu  <- 0.5
true_tau <- 5
a0 <- 1
b0 <- 1

simulated_data <- generate_sample(N = 100, mu = true_mu, tau = true_tau, a0 = a0, b0 = b0)

# ---- Stan data ----
stan_data <- list(
  N  = nrow(simulated_data),
  x  = simulated_data$y,
  v0 = 0.5,  # keep only if your Stan program uses v0
  a0 = a0,
  b0 = b0
)

# ---- Compile and sample ----
rstan_options(auto_write = TRUE)
options(mc.cores = parallel::detectCores())

# Option A: separate file
# my_mod <- stan_model("modelo.stan")

# Option B: inline Stan code (example placeholder)
# stan_code <- "data { ... } parameters { ... } model { ... }"
# my_mod <- stan_model(model_code = stan_code)

my_mod <- stan_model("modelo.stan")
```

```
my_fit <- sampling(my_mod, data = stan_data, iter = 2000, chains = 4, seed = 123)

# ---- Extract draws ----
my_chains <- as.data.frame(my_fit)

# ---- Quick comparison (true vs estimated) ----
true_values <- c(true_mu, true_tau)
estimated_values <- c(mean(my_chains$mu), mean(my_chains$tau))

comparison <- tibble(
  Parameter       = c("mu", "tau"),
  True_Value      = true_values,
  Estimated_Value = estimated_values,
  Absolute_Error  = abs(true_values - estimated_values)
)

print(comparison)

# ---- Trace plots (diagnostics) ----
trace_plot_mu <- ggplot(my_chains, aes(x = seq_along(mu), y = mu)) +
  geom_line() +
  labs(title = "Trace Plot — mu (μ)", x = "Iteration", y = "mu (μ)") +
  theme_minimal()

trace_plot_tau <- ggplot(my_chains, aes(x = seq_along(tau), y = tau)) +
  geom_line() +
  labs(title = "Trace Plot — tau (τ)", x = "Iteration", y = "tau (τ)") +
  theme_minimal()

print(trace_plot_mu)
print(trace_plot_tau)

# ---- Posterior density plots ----
density_mu <- ggplot(my_chains, aes(x = mu)) +
  geom_density(alpha = 0.4) +
  geom_vline(aes(xintercept = mean(mu)), linetype = "dashed", linewidth = 1) +
  geom_vline(aes(xintercept = quantile(mu, 0.025)), linetype = "dotted", linewidth = 1) +
  geom_vline(aes(xintercept = quantile(mu, 0.975)), linetype = "dotted", linewidth = 1) +
  labs(title = "Posterior Density — mu (μ)", x = "mu (μ)", y = "Density") +
  theme_minimal()

density_tau <- ggplot(my_chains, aes(x = tau)) +
  geom_density(alpha = 0.4) +
  geom_vline(aes(xintercept = mean(tau)), linetype = "dashed", linewidth = 1) +
  geom_vline(aes(xintercept = quantile(tau, 0.025)), linetype = "dotted", linewidth = 1) +
  geom_vline(aes(xintercept = quantile(tau, 0.975)), linetype = "dotted", linewidth = 1) +
  labs(title = "Posterior Density — tau (τ)", x = "tau (τ)", y = "Density") +
  theme_minimal()

print(density_mu)
print(density_tau)
```
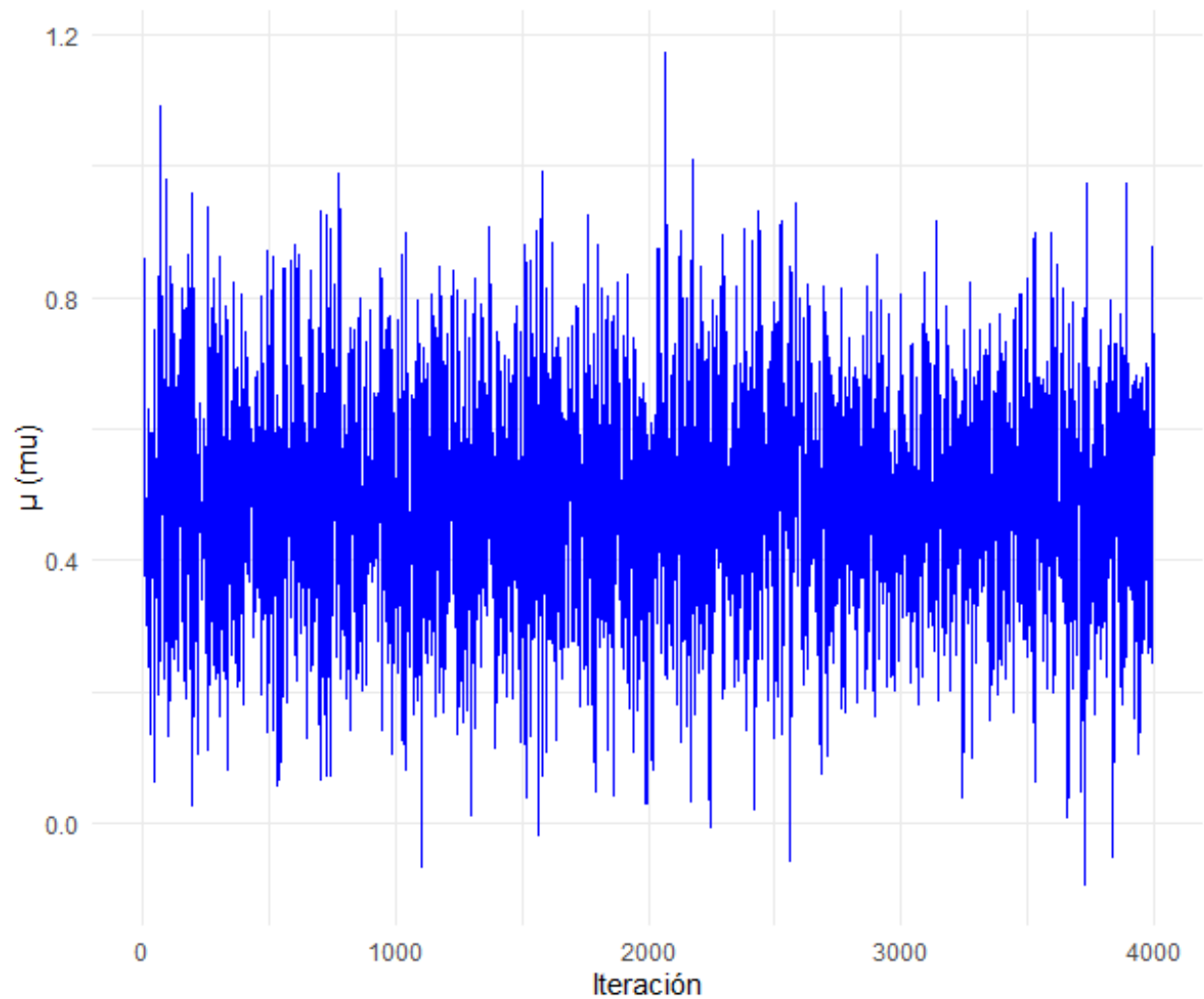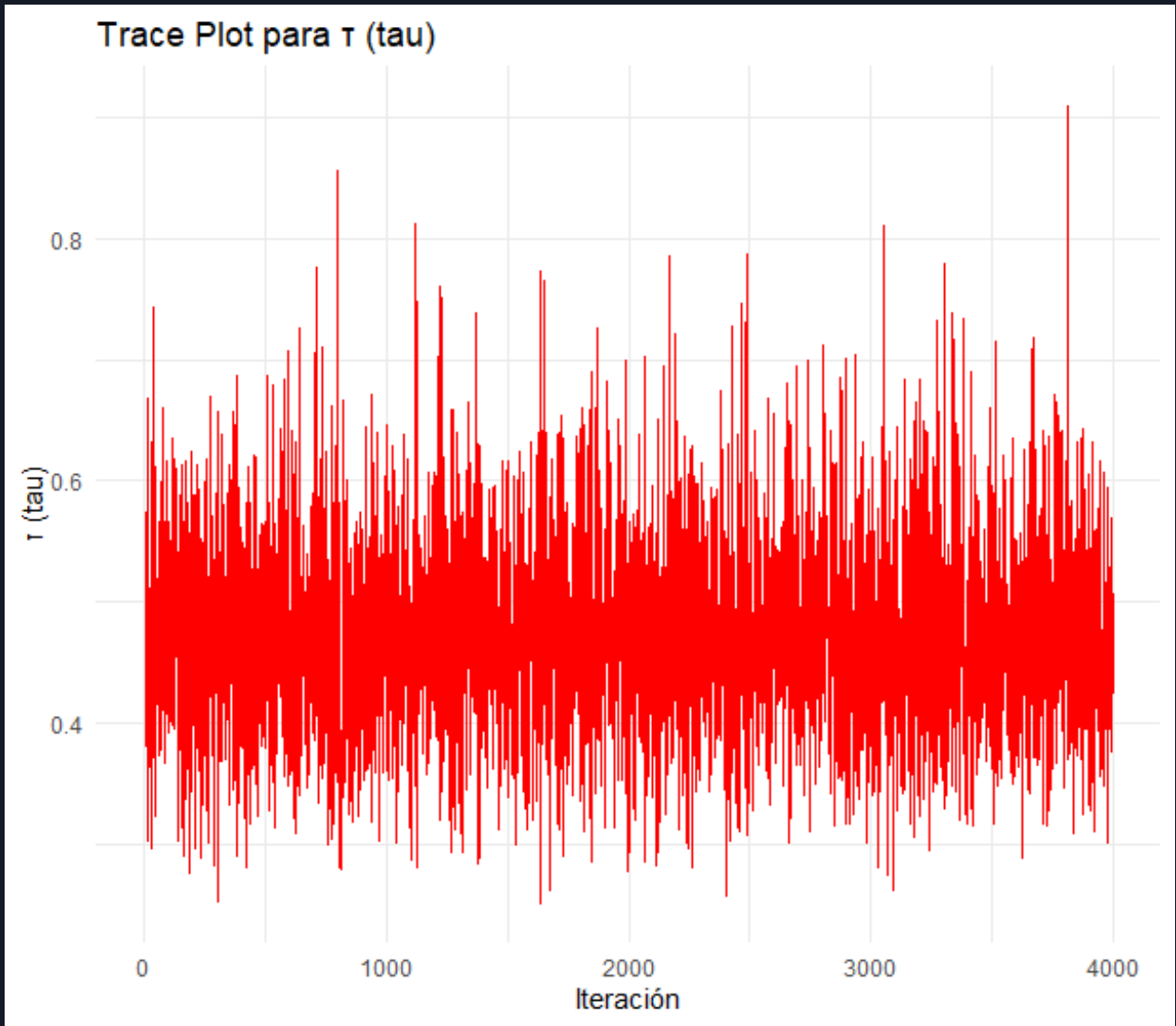
Diagnostics — Trace Plots

Trace plots help validate MCMC convergence and chain mixing. Stable fluctuations around a consistent region, without long-term trends, support reliable posterior inference.
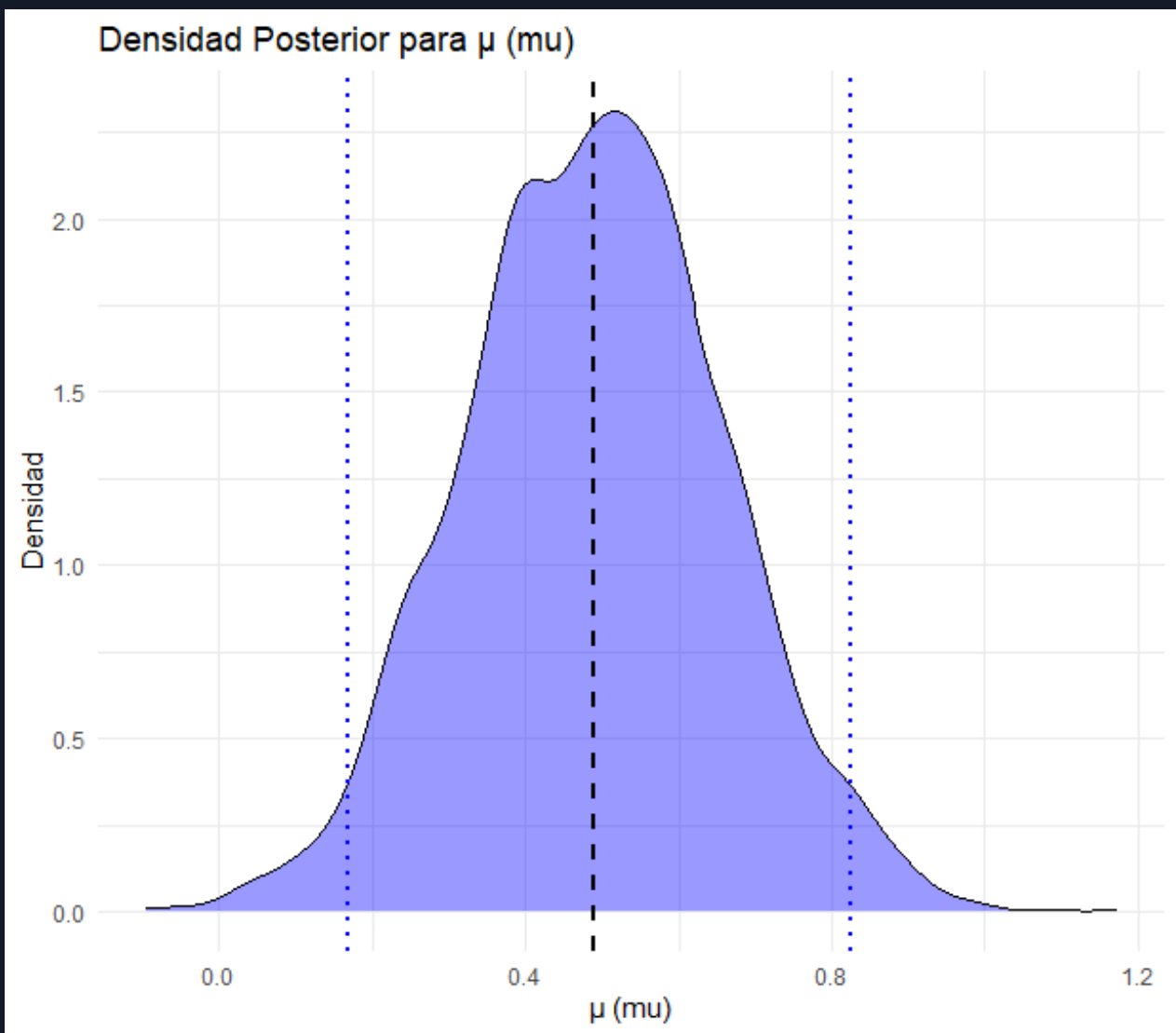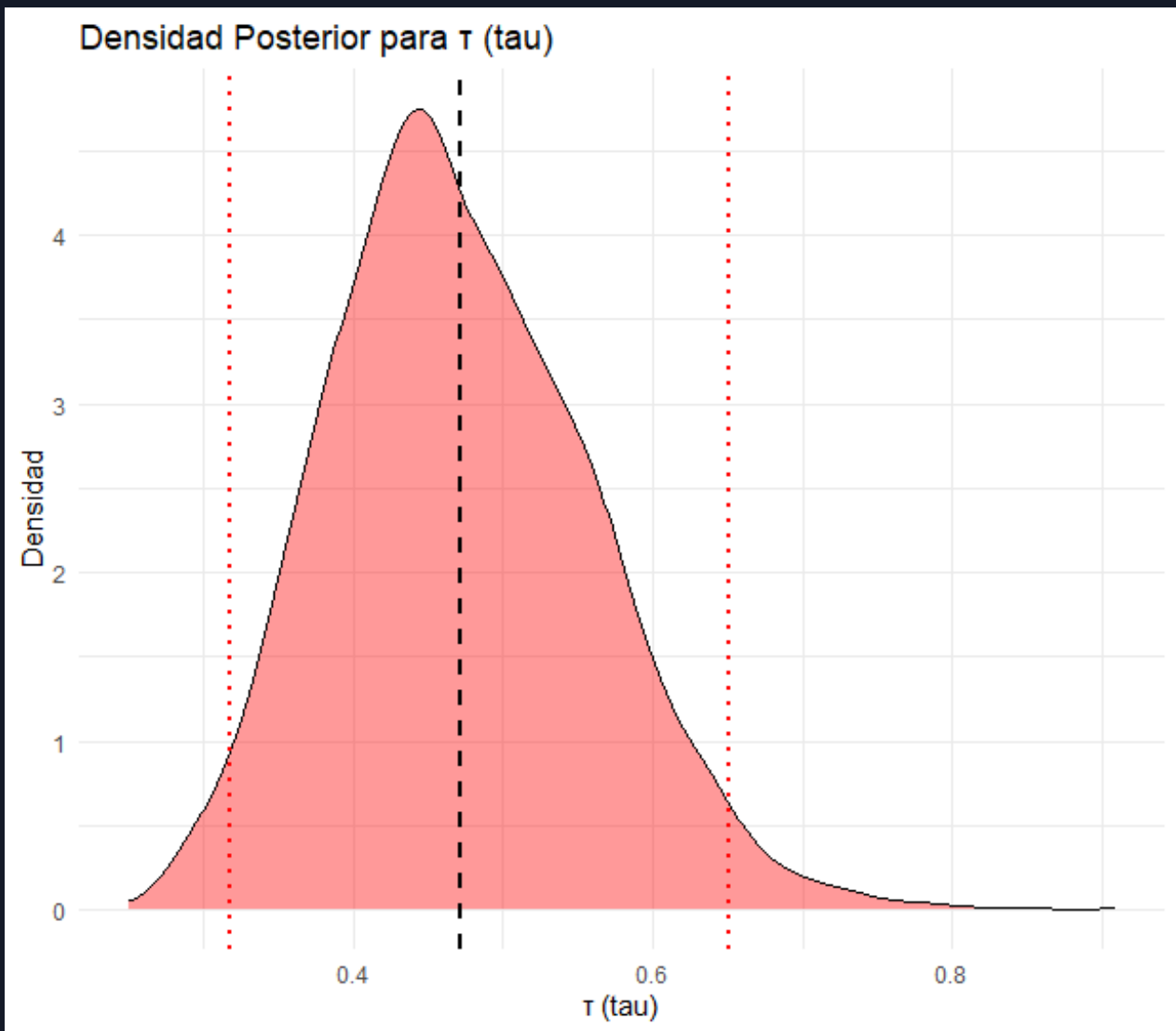
Trace Plot para μ (mu)

Trace Plot para τ (tau)

Posterior Uncertainty — Density Plots

Posterior densities summarize uncertainty in the estimated parameters. The dashed line indicates the posterior mean, while dotted lines indicate the 95% credible interval bounds.

Densidad Posterior para μ (mu)

Densidad Posterior para τ (tau)

Technologies & Tools