Bayesian Analysis of Hurricane Data (1975–2021)

Saurabh Gupta

February 17, 2025

Abstract

We analyze hurricane data from the `storms` dataset (1975–2021) to determine if the distribution of maximum wind speed and annual hurricane counts differ between Period 1 (1975–1999) and Period 2 (2000–2021). For maximum wind speed, we model log(mxspd) using a Normal likelihood with a conjugate Normal–Inverse-Gamma prior; for counts, we assume a Poisson likelihood with a Gamma prior. Posterior summaries reveal statistically significant increases in both metrics in Period 2. Sensitivity analyses show the findings are robust to prior specification.

1 Introduction

The `storms` dataset records hourly wind speeds for Atlantic storms. We focus on hurricanes (wind speed $\geq$ 65 knots) and compute for each storm its year and maximum wind speed (`mxspd`), as well as annual hurricane counts (`cnt`). Our goal is to test whether the distributions of `mxspd` and `cnt` differ over time.

2 Methods

2.1 Likelihood and Graphical Assessment

For maximum wind speed, we treat `mxspd` as continuous. Exploratory plots of `mxspd` and its logarithm are generated. The log transformation is used to better approximate normality.

$$\log(\text{mxspd}_i) \sim \mathcal{N}(\mu, \sigma^2).$$

For annual hurricane counts, we assume a Poisson model: $Y \sim \text{Poisson}(\lambda)$.

2.2 Prior Distributions

For log(mxspd), we choose the following conjugate priors:

$$\mu \mid \sigma^2 \sim \mathcal{N}(\mu_0, \sigma^2/\tau_0) \quad \text{and} \quad \sigma^2 \sim \text{Inverse-Gamma}(\alpha_0, \beta_0),$$

where we set $\mu_0 =$ sample mean(log(mxspd)), $\tau_0 = 0.001$, $\alpha_0 = 0.001$, and $\beta_0 = 0.001$. The normal-inverse-gamma conjugate prior is a standard choice for Bayesian inference involving normal likelihoods with unknown mean and variance.

For hurricane counts, we use:

$$\lambda \sim \text{Gamma}(a_0, b_0)$$

with $a_0 = 0.01$ and $b_0 = 0.01$. The gamma prior is conjugate to the Poisson likelihood.

2.3 Posterior Analysis

For log(mxspd), the conjugate analysis gives the following posterior parameters:

$$\tau_n = \tau_0 + n, \quad \mu_n = \frac{\tau_0\mu_0 + n\bar{x}}{\tau_0 + n}, \quad \alpha_n = \alpha_0 + \frac{n}{2}, \quad \beta_n = \beta_0 + \frac{1}{2}(n-1)s^2 + \frac{\tau_0\, n(\bar{x} - \mu_0)^2}{2(\tau_0 + n)}.$$

Posterior samples are drawn for each period and the difference (Period 2 minus Period 1) is computed to obtain a point estimate and standard error.

For hurricane counts, the posterior is:

$$\lambda \mid Y \sim \text{Gamma}\Big(a_0 + \sum Y,\ b_0 + n\Big).$$

2.4 Sensitivity Analysis

Alternative hyperparameters are used to check robustness. For the log(mxspd) model, one might use $\tau_0 = 0.01$, $\alpha_0 = 0.01$, $\beta_0 = 0.01$, and for the Poisson model, $a_0 = 0.1$, $b_0 = 0.1$.

3 Results

3.1 Maximum Wind Speed

Compute the difference for each sample as:

$$\Delta_i = \mu^{(i)}_{\text{period2}} - \mu^{(i)}_{\text{period1}}, \quad i = 1, \dots, n.$$

The point estimate of the difference is computed by taking the mean of these differences:

$$\widehat{\Delta} = \frac{1}{n}\sum_{i=1}^{n} \Delta_i,$$

and the standard error is given by the standard deviation of the differences:

$$\text{SE} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\Big(\Delta_i - \widehat{\Delta}\Big)^2}.$$

Estimated difference (Period 2 − Period 1): ≈ 0.049 (SE ≈ 0.029).
On the log scale, Period 2 storms have higher maximum wind speeds by about 0.049. Converting to a percentage increase, $\exp(0.049) - 1 \approx 5\%$.

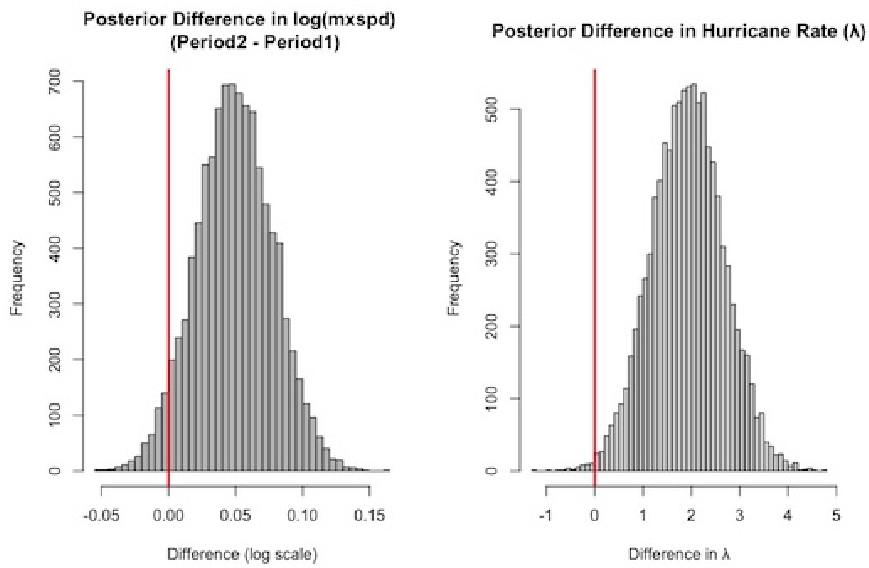

Figure 1: Posterior Differences in log(mxspd) and Hurricane Rate

3.2 Hurricane Counts

The posterior difference in λ (the annual rate) was estimated at approximately 1.92 (SE $\approx$ 0.75), suggesting an increase in the number of hurricanes per year in Period 2. The average number of hurricanes per year is roughly 2 higher in Period 2 compared to Period 1.

3.3 Sensitivity Analysis

log(mxspd) alternate priors yield a difference of ≈ 0.049 (SE ≈ 0.028), nearly identical to the main analysis. Hurricane Counts difference ≈ 1.93 (SE ≈ 0.75), again very close to the main analysis.

4 Discussion

In layman's terms, our analysis shows that both the average maximum wind speed (after log transformation) and the annual hurricane count are significantly higher in the recent period (2000–2021) compared to the earlier period (1975–1999). Sensitivity analyses confirm that these results remain robust under alternative prior specifications.

5 Conclusion

Both maximum wind speed (on log scale) and annual hurricane counts show statistically significant increases in more recent period, with consistent findings under alternative priors.

The Bayesian analysis provides strong evidence that hurricane intensity and frequency have increased over time. These findings are consistent across various prior choices, indicating robust model inferences.

```r
# =========================================================
# ST 540 Applied Bayesian Analysis { Exam 1
# Due Date: February 17, 2025
# Author: Saurabh Gupta
# =========================================================

# ----- 1. Likelihood and Graphical Assessment -----
library(dplyr)
data("storms")
# Only analyze hurricanes
storms <- storms[!is.na(storms$category),]

# Extract basic variables from storms
year <- storms$year
name <- paste0(storms$name, "_", storms$year)
wind <- storms$wind

# Create unique storm identifiers and extract max wind speed per storm
uni <- unique(name)
n <- length(uni)
year_storm <- rep(0, n)
mxspd <- rep(0, n)
for(i in 1:length(uni)) {
  # For each unique storm, use all observations with that identifier
  year_storm[i] <- min(year[name == uni[i]])
  mxspd[i] <- max(wind[name == uni[i]])
}

# Compute annual hurricane counts
year_counts <- as.numeric(names(table(year_storm)))
cnt <- as.vector(table(year_storm))

# Plot maximum wind speed by storm year and annual counts (base R)
plot(year_storm, mxspd, xlab="Year", ylab="Max wind speed (knots)",
     main="Max Wind Speed per Storm", pch=16, col="blue")
plot(year_counts, cnt, xlab="Year", ylab="Number of hurricanes",
     main="Annual Hurricane Counts", type="o", col="green", pch=16)
```

```r
# Split data into Period1: 1975 - 1999 and Period2: 2000 - 2021
period <- ifelse(year_storm < 2000, "Period1", "Period2")
# For annual counts, find period based on year
period_year <- ifelse(year_counts < 2000, "Period1", "Period2")

# For maximum wind speed analysis, find log transformed mxspd
log_mxspd <- log(mxspd)

# Plot histograms and QQ plots for each period (raw and log transformed)
par(mfrow=c(2,2))
for(p in c("Period1", "Period2")) {
  idx <- (period == p)
  raw_data <- mxspd[idx]
  log_data <- log_mxspd[idx]

  hist(raw_data, main=paste("Histogram of mxspd", p),
       xlab="Max Wind Speed (knots)", col="blue", breaks=20)
  qqnorm(raw_data, main=paste("QQ Plot of mxspd", p))
  qqline(raw_data, col="red", lwd=2)

  hist(log_data, main=paste("Histogram of log(mxspd)", p),
       xlab="log(mxspd)", col="green", breaks=20)
  qqnorm(log_data, main=paste("QQ Plot of log(mxspd)", p))
  qqline(log_data, col="red", lwd=2)
}
par(mfrow=c(1,1))

# --------------------------------------------------------------
# 2. Conjugate Uninformative Prior Distributions
# --------------------------------------------------------------
#
# For Maximum Wind Speed (log-transformed):
#   Likelihood: log(mxspd_i) ~ N( , ²)
#
#   Priors (vague, conjugate):
#       | ² ~ N( , ²/ )   where   = mean(all_log_mxspd),   = 0.001
#     ²     ~ Inverse-Gamma( , )   with   = 0.001,   = 0.001
#
# For Hurricane Counts:
#   Likelihood: Y ~ Poisson( )
#
#   Prior:
```

```r
#       ~ Gamma(a, b) with a = 0.01 and b = 0.01

# Set priors for log (mxspd) analysis
all_log <- log_mxspd
mu0 <- mean(all_log)
tau0 <- 0.001
alpha0 <- 0.001
beta0 <- 0.001

# Set priors for hurricane counts
a0_pois <- 0.01
b0_pois <- 0.01

# ------------------------------------------------------------
# 3. Posterior Summaries and Testing for Difference Between Periods
# ------------------------------------------------------------
# For log(mxspd):
#
#   Posterior (conjugate for Normal likelihood):
#       =  + n
#       = ( + n·x)/(+n)
#       =  + n/2
#       =  + 0.5*(n-1)*s² + (·n*(x-)²)/(2*(+n))
#
# Find posterior samples for Period1 and Period2 and then find the difference.
#
# For Hurricane Counts
#   Posterior:
#       | Y ~ Gamma(a + Y, b + n)
#
# Find posterior samples for each period and find the difference.

# Find posterior parameters for log(mxspd)
calculate_posterior <- function(data, mu0, tau0, alpha0, beta0) {
  n <- length(data)
  xbar <- mean(data)
  s2 <- var(data)
  tau_n <- tau0 + n
  mu_n <- (tau0 * mu0 + n * xbar) / tau_n
  alpha_n <- alpha0 + n / 2
  beta_n <- beta0 + 0.5 * (n-1) * s2 + (tau0 * n * (xbar - mu0)^2) / (2 * tau_n)
  return(list(mu_n=mu_n, tau_n=tau_n, alpha_n=alpha_n, beta_n=beta_n, n=n))
```

```r
}

# Separate log(mxspd) for each period
log_period1 <- log_mxspd[period == "Period1"]
log_period2 <- log_mxspd[period == "Period2"]

# Compute posterior parameters for each period
posterior_log_period1 <- calculate_posterior(log_period1, mu0,tau0,alpha0,beta0)
posterior_log_period2 <- calculate_posterior(log_period2, mu0,tau0,alpha0,beta0)

set.seed(123)
n_samples <- 10000

# For Period 1
sigma2_period1 <- 1 / rgamma(n_samples, shape = posterior_log_period1$alpha_n,
                             rate = posterior_log_period1$beta_n)
mu_samples_period1 <- rnorm(n_samples, mean = posterior_log_period1$mu_n,
                            sd = sqrt(sigma2_period1 / posterior_log_period1$tau_n))

# For Period 2
sigma2_period2 <- 1 / rgamma(n_samples, shape = posterior_log_period2$alpha_n,
                             rate = posterior_log_period2$beta_n)
mu_samples_period2 <- rnorm(n_samples, mean = posterior_log_period2$mu_n,
                            sd = sqrt(sigma2_period2 / posterior_log_period2$tau_n))

# Posterior difference (Period2 - Period1)
diff_mu <- mu_samples_period2 - mu_samples_period1
point_estimate_diff <- mean(diff_mu)
standard_error_diff <- sd(diff_mu)
print(paste("Posterior difference in mean log(mxspd):", point_estimate_diff))
print(paste("Standard error:", standard_error_diff))

# For hurricane counts
# Create indices for each period
index1 <- year_counts < 2000
index2 <- year_counts >= 2000

# Separate annual counts by period
count_period1 <- cnt[index1]
count_period2 <- cnt[index2]

# Number of years in each period
```

```r
n_period1 <- sum(index1)
n_period2 <- sum(index2)

# Compute posterior parameters for  (Poisson likelihood)
posterior_count_period1 <- list(a_post = a0_pois + sum(count_period1),
                    b_post = b0_pois + n_period1)
posterior_count_period2 <- list(a_post = a0_pois + sum(count_period2),
                    b_post = b0_pois + n_period2)

# Draw posterior samples for  in each period
lambda_samples_p1 <- rgamma(n_samples, shape = posterior_count_period1$a_post,
                            rate = posterior_count_period1$b_post)
lambda_samples_p2 <- rgamma(n_samples, shape = posterior_count_period2$a_post,
                            rate = posterior_count_period2$b_post)

# Posterior difference for hurricane counts:
diff_lambda <- lambda_samples_p2 - lambda_samples_p1
point_estimate_lambda <- mean(diff_lambda)
standard_error_lambda <- sd(diff_lambda)
print(paste("Posterior difference in  :", point_estimate_lambda))
print(paste("Standard error:", standard_error_lambda))

par(mfrow=c(1,2))
hist(diff_mu, breaks=50, main="Posterior Difference in log(mxspd)
     (Period2 - Period1)",
     xlab="Difference (log scale)", col="gray")
abline(v=0, col="red", lwd=2)

hist(diff_lambda, breaks=50, main="Posterior Difference in Hurricane Rate ()",
     xlab="Difference in ", col="lightgray")
abline(v=0, col="red", lwd=2)
par(mfrow=c(1,1))

# -------------------------------------------------------------
# 4. Sensitivity Analysis: Check if results are sensitive to priors
# -------------------------------------------------------------
# For log(mxspd): Use alternative hyperparameters
tau0_alternate <- 0.01;
alpha0_alternate <- 0.01;
beta0_alternate <- 0.01

posterior_log_period1_alt <- calculate_posterior(log_period1, mu0,
```

```
                             tau0_alternate, alpha0_alternate, beta0_alternate)
posterior_log_period2_alt <- calculate_posterior(log_period2, mu0,
                             tau0_alternate, alpha0_alternate, beta0_alternate)

# Find posterior samples for  in each period
sigma2_period1_alt <- 1 / rgamma(n_samples,
                            shape = posterior_log_period1_alt$alpha_n,
                            rate = posterior_log_period1_alt$beta_n)
mu_samples_period1_alt <- rnorm(n_samples, mean =posterior_log_period1_alt$mu_n,
                sd = sqrt(sigma2_period1_alt / posterior_log_period1_alt$tau_n))

sigma2_period2_alt <- 1 / rgamma(n_samples,
                            shape = posterior_log_period2_alt$alpha_n,
                            rate = posterior_log_period2_alt$beta_n)
mu_samples_period2_alt <- rnorm(n_samples, mean =posterior_log_period2_alt$mu_n,
                sd = sqrt(sigma2_period2_alt / posterior_log_period2_alt$tau_n))

diff_mu_alt <- mu_samples_period2_alt - mu_samples_period1_alt
print(paste("Sensitivity Analysis for log(mxspd) - Difference Alternate Prior:",
            mean(diff_mu_alt)))
print(paste("Sensitivity Analysis for log(mxspd)- Std Error:", sd(diff_mu_alt)))

# For hurricane counts, alternative prior hyperparameters
a0_pois_alt <- 0.1
b0_pois_alt <- 0.1

index1 <- year_counts < 2000
index2 <- year_counts >= 2000
n_period1 <- sum(index1)
n_period2 <- sum(index2)
count_period1 <- cnt[index1]
count_period2 <- cnt[index2]
total_period1 <- sum(count_period1)
total_period2 <- sum(count_period2)

# Find posterior parameters for each period
posterior_count_period1_alt <- list(a_post = a0_pois_alt + total_period1,
                        b_post = b0_pois_alt + n_period1)
posterior_count_period2_alt <- list(a_post = a0_pois_alt + total_period2,
                        b_post = b0_pois_alt + n_period2)

# Posterior samples for lambda
```

```r
lambda_samples_p1_alt <- rgamma(n_samples,
                                shape = posterior_count_period1_alt$a_post,
                                rate = posterior_count_period1_alt$b_post)
lambda_samples_p2_alt <- rgamma(n_samples,
                                shape = posterior_count_period2_alt$a_post,
                                rate = posterior_count_period2_alt$b_post)

diff_lambda_alt <- lambda_samples_p2_alt - lambda_samples_p1_alt
print(paste("Sensitivity Analysis (Hurricane Counts) - Difference (Alt Prior):",
            mean(diff_lambda_alt)))
print(paste("Sensitivity Analysis (Hurricane Counts) - Std Error:",
            sd(diff_lambda_alt)))

# ------------------------------------------------------------
# End of code
# Execute the code to genrate plots and graphs
# print metrices for reference -
#Posterior difference in mean log(mxspd): 0.0491857616376029
#Standard error: 0.028559582614 73
#Posterior difference in  : 1.9217695599 92
#Standard error: 0.7447848374554
#Sensitivity Analysis for log(mxspd) - Difference Alternate Prior: 0.048791232424 79
#Sensitivity Analysis for log(mxspd)- Std Error: 0.02845322583773
#Sensitivity Analysis (Hurricane Counts) - Difference (Alt Prior): 1.928157992365
#Sensitivity Analysis (Hurricane Counts) - Std Error: 0.745620105880

# ------------------------------------------------------------
```