

677 final project for CASI Chapter 6

Yang Xiao

2024-05-04

1 Introduction

In the ever-evolving landscape of statistical methods, the adoption of computer-aided techniques has transformed traditional approaches into more dynamic and empirically driven practices. Chapter 6 of *Computer Age Statistical Inference*, titled “Empirical Bayes,” exemplifies this transformation. This chapter delves into the Empirical Bayes method, a statistical technique that has gradually matured alongside advances in computational power and data availability. The purpose of these study notes is to elucidate the main themes and methodologies discussed by Efron and Hastie, providing a detailed examination of how Empirical Bayes can be leveraged in modern statistical practices. By dissecting the computational methods, mathematical foundations, and practical implications, these notes aim to equip readers with a comprehensive understanding of the chapter and inspire further exploration into the subject.

2 Chapter Summary

2.1 Main Concepts of Empirical Bayes

Empirical Bayes methods represent a unique intersection between traditional Bayesian inference and frequentist statistics. By utilizing observed data to estimate “prior” distributions—typically unknown in classic Bayesian frameworks—Empirical Bayes facilitates more robust statistical inference. This chapter introduces the concept by revisiting historical statistical challenges and demonstrating how Empirical Bayes addresses them with modern computational resources.

2.2 Methodologies and Applications

The chapter begins with a discussion on the historical constraints of classical statistics, constrained by the computational technologies of the era. With the advent of electronic computation post-World War II, statistical methodology witnessed a significant shift, allowing for more expansive data analysis techniques such as Empirical Bayes. The authors provide an insightful exploration of Robbins’ formula, a pivotal development in Empirical Bayes methods, applied to a dataset from a European automobile insurance company. This example illustrates how Empirical Bayes can be used to predict future claims based on past data without a known prior distribution.

2.3 Further Developments and Modern Applications

As the chapter progresses, it delves deeper into various applications of Empirical Bayes, particularly in contexts where traditional methods fall short. One notable application discussed is the “Missing-Species Problem,” where Empirical Bayes methods are used to estimate the number of unseen species in ecological studies. This example not only highlights the method’s utility in dealing with incomplete data sets but also showcases its broader applicability in fields ranging from insurance to ecology and beyond.

3 Mathematical Foundations

3.1 Theoretical Background of Empirical Bayes

Empirical Bayes methods blend frequentist techniques with Bayesian principles to create robust statistical inferences where traditional priors are unknown. This approach effectively uses the data itself to estimate the prior distributions, enabling a form of Bayesian inference without complete prior information. The central concept relies on the assumption that, although the parameters of the prior distribution are unknown, they can be estimated from the data.

3.2 Key Formulas and Models

3.2.1 Robbins' Formula:

A cornerstone of Empirical Bayes in Chapter 6 is Robbins' formula, which addresses the problem of estimating future probabilities or parameters when the prior distribution is not explicitly known. For instance, in the case of the automobile insurance company data, the formula helps predict future claims based on past claims data. The formula is expressed as follows:

$$\hat{E}(\theta|x) = \frac{(x+1)f(x+1)}{f(x)}$$

where:

- θ is the parameter of interest,
- x is the observed data,
- $f(x)$ is the probability mass function of the observed data.

This formula illustrates the empirical update of the expected value of the parameter given new data when the prior is estimated from the data itself.

3.2.2 Bayesian Updating:

The chapter also discusses Bayesian updating in the context of unknown priors, which is foundational for Empirical Bayes methods. The updating process can be mathematically described by:

$$E[\theta|x] = \frac{\int \theta p(x|\theta)g(\theta)d\theta}{\int p(x|\theta)g(\theta)d\theta}$$

Here, $p(x|\theta)$ denotes the likelihood of observing x given the parameter θ , and $g(\theta)$ represents the estimated prior distribution of θ . This integral demonstrates how empirical data informs the updating of beliefs about the parameter.

3.2.3 Poisson-Gamma Model and Derivations

The Poisson-Gamma model is a standard example of a conjugate prior setup in Bayesian statistics. This model is particularly useful for data that are counts of events, such as the number of claims filed, the number of diseases detected, or the number of species observed.

Model Setup Consider a scenario where we observe a series of independent events that occur at a constant average rate. Let X denote the number of events (e.g., claims, diseases) observed, which follows a Poisson distribution:

$$X \sim \text{Poisson}(\lambda)$$

where λ is the rate parameter of the Poisson distribution, representing the average number of events per interval.

The Gamma distribution is used as the prior for λ due to its conjugacy with the Poisson distribution, which greatly simplifies the Bayesian updating calculations. The prior distribution for λ is given by:

$$\lambda \sim \Gamma(\alpha, \beta)$$

where α is the shape parameter and β is the rate parameter of the Gamma distribution.

Posterior Distribution Given the observed data x from the Poisson distribution, the posterior distribution of λ can be derived using Bayes' theorem. The likelihood of observing x given λ is:

$$p(x|\lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$$

The prior distribution is:

$$p(\lambda) = \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Applying Bayes' theorem, the posterior distribution $p(\lambda|x)$ is proportional to the product of the likelihood and the prior:

$$p(\lambda|x) \propto p(x|\lambda)p(\lambda) = \frac{e^{-\lambda} \lambda^x}{x!} \cdot \frac{\beta^\alpha}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

Simplifying this expression, we obtain:

$$p(\lambda|x) \propto \lambda^{x+\alpha-1} e^{-\lambda(\beta+1)}$$

Recognizing this as the kernel of a Gamma distribution, we can write the posterior distribution of λ explicitly as:

$$\lambda|x \sim \Gamma(x + \alpha, \beta + 1)$$

3.3 Implications and Applications:

This posterior distribution of λ provides a Bayesian update based on the observed data. The updated parameters, $x + \alpha$ and $\beta + 1$, reflect how the prior beliefs are modified in light of new evidence. This model can be applied to predict future event occurrences and can be extended to multiple data points and complex scenarios involving hierarchical Bayesian models.

4 Computational Methods

4.1 Summary

The theoretical foundations of the Empirical Bayes methods laid out in the previous sections lend themselves well to practical applications using statistical software. This section demonstrates how to implement these methods using both Python and R, focusing on the Poisson-Gamma model as a case study.

4.2 example for Poisson-Gamma

```

# Set prior parameters for the Gamma distribution
alpha <- 2.0 # Shape parameter
beta <- 1.0  # Rate parameter

# Observed data (e.g., number of events)
x_observed <- 10

# Update posterior parameters based on the observed data
alpha_post <- alpha + x_observed # Update shape
beta_post <- beta + 1            # Update rate

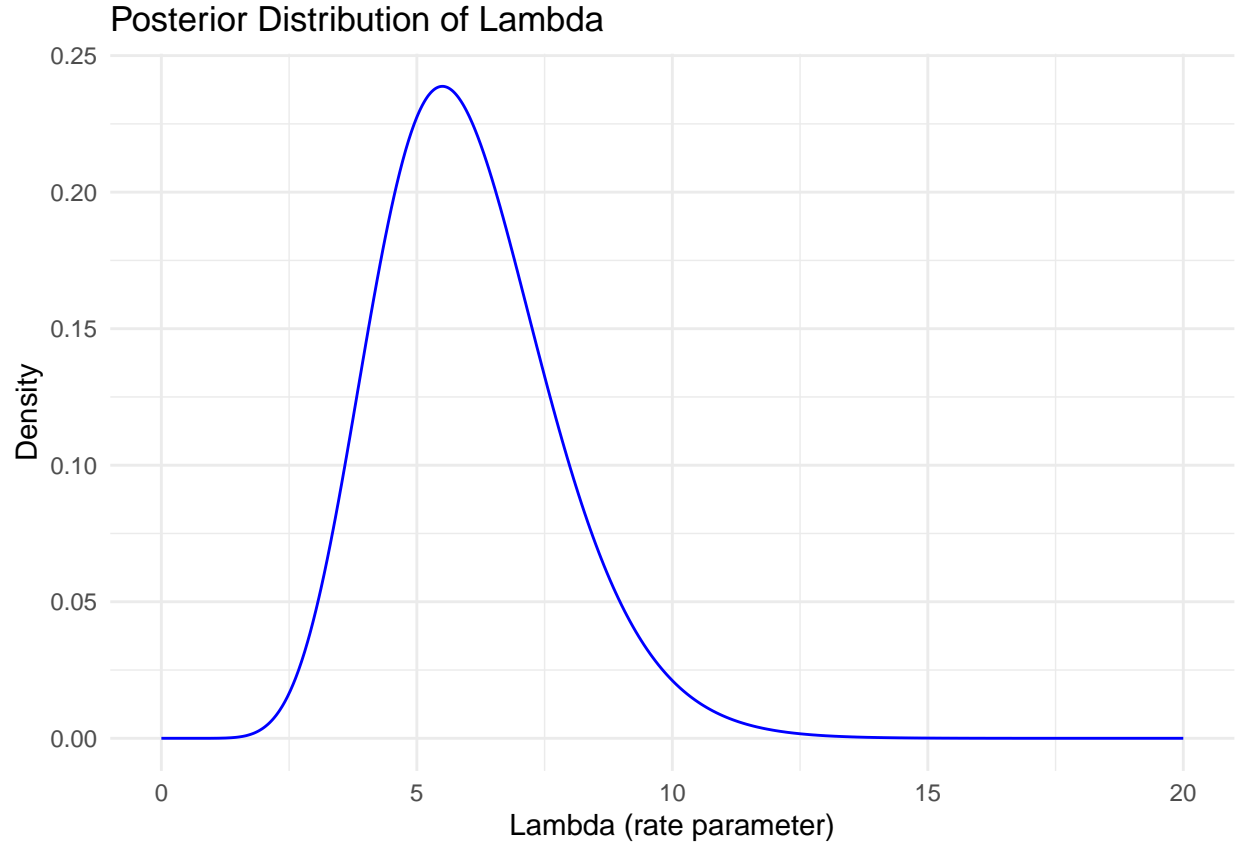
# Define the range for lambda values
lambda <- seq(0, 20, length.out = 1000)

# Calculate the posterior distribution
posterior <- dgamma(lambda, shape = alpha_post, rate = beta_post)

# Create a data frame for plotting
posterior_data <- data.frame(lambda, Density = posterior)

# Plot the posterior distribution using ggplot2
ggplot(posterior_data, aes(x = lambda, y = Density)) +
  geom_line(color = "blue") +
  labs(title = "Posterior Distribution of Lambda",
       x = "Lambda (rate parameter)",
       y = "Density") +
  theme_minimal()

```



- **Prior Distribution Parameters:** The shape parameter α is set to 2.0 and the rate parameter β is set to 1.0. This defines the prior beliefs about the distribution of λ before observing any data.
- **Observed Data:** You've specified that the number of observed events (`x_observed`) is 10. This data is used to update the prior and obtain the posterior distribution.
- **Posterior Parameters Update:**
 - The shape parameter of the posterior gamma distribution, α_{post} , is calculated as $\alpha + x_{observed}$, resulting in 12.0.
 - The rate parameter, β_{post} , is updated by adding 1 to the prior rate, resulting in 2.0.
- **Posterior Distribution Calculation:** Using these updated parameters, the gamma distribution function (`dgamma`) calculates the probability density values across a sequence of λ values ranging from 0 to 20.
- **Plotting:** The `ggplot2` package is used to plot these values, showing how the density of the posterior distribution changes with λ . The plot you have displayed indicates that the distribution of λ is concentrated around a particular value, suggesting a strong posterior belief about the rate parameter after considering the observed data.

4.3 Hierarchical Poisson-Gamma coding

This Stan model represents a hierarchical Poisson-Gamma model, typically used for analyzing count data with a hierarchical structure. Let me explain each part:

1. **data{}**: This section defines the necessary data for the model. In this model, there are two data items:
 - **N**: Indicates the number of groups (or populations) in the data.
 - **y**: Is an integer vector of length **N**, containing the observed count data for each group.
2. **parameters{}**: This section defines the model parameters. In this model, there are three parameters:
 - **alpha**: The shape parameter of the gamma distribution, controlling the shape of the distribution.
 - **beta**: The rate parameter of the gamma distribution, controlling the scale of the distribution.
 - **lambda**: A real vector of length **N**, representing the rate parameters of the Poisson distribution for each group.
3. **model{}**: This section specifies the probabilistic model.
 - Both **alpha** and **beta** have gamma prior distributions. Here, a weakly informative prior with shape and rate parameters set to 0.1 is chosen, allowing the data to have more influence on the prior distribution.
 - **lambda[i]** is the rate parameter of the Poisson distribution for each group, drawn from a shared gamma distribution.
 - **y[i]** represents the observed data for each group, modeled using a Poisson distribution.

The objective of this model is to estimate the rate parameters **lambda** for each group's Poisson distribution based on the observed data **y**, while considering the uncertainty in the shared gamma distribution parameters **alpha** and **beta**.

```
options(mc.cores = parallel::detectCores())
rstan_options(auto_write = TRUE)
# Define the Stan model for a hierarchical Poisson-Gamma model
stan_model_code <- "
data {
  int<lower=0> N;           // Number of groups
  int<lower=0> y[N];        // Observed data (counts)
}

parameters {
  real<lower=0> alpha;      // Hyper-parameter for Gamma distribution (shape)
  real<lower=0> beta;       // Hyper-parameter for Gamma distribution (rate)
  real<lower=0> lambda[N];  // Rate parameters for each group
}

model {
  alpha ~ gamma(0.1, 0.1); // Weakly informative prior for alpha
  beta ~ gamma(0.1, 0.1);  // Weakly informative prior for beta

  for (i in 1:N) {
    lambda[i] ~ gamma(alpha, beta); // Each group rate drawn from a common Gamma
    y[i] ~ poisson(lambda[i]);      // Data likelihood
  }
}
"

stan_model <- stan_model(model_code = stan_model_code)
```

```

N <- 5
y <- c(5, 12, 3, 16, 8)

# Fit the model
fit <- sampling(stan_model, data = list(N = N, y = y), iter = 2000, chains = 4)

```

```

## Warning: There were 106 divergent transitions after warmup. See
## https://mc-stan.org/misc/warnings.html#divergent-transitions-after-warmup
## to find out why this is a problem and how to eliminate them.

```

```

## Warning: Examine the pairs() plot to diagnose sampling problems

```

```

## Warning: Tail Effective Samples Size (ESS) is too low, indicating posterior variances and tail quantiles may be unreliable.
## Running the chains for more iterations may help. See
## https://mc-stan.org/misc/warnings.html#tail-ess

```

```

print(fit)

```

```

## Inference for Stan model: anon_model.
## 4 chains, each with iter=2000; warmup=1000; thin=1;
## post-warmup draws per chain=1000, total post-warmup draws=4000.
##
##               mean se_mean   sd  2.5%  25%   50%   75%  97.5% n_eff Rhat
## alpha         4.35    0.20 3.85  0.55  1.75  3.16  5.53 15.36   370 1.01
## beta          0.50    0.02 0.47  0.05  0.19  0.36  0.65  1.86   388 1.01
## lambda[1]     6.14    0.06 2.27  2.50  4.47  5.90  7.55 11.24  1588 1.00
## lambda[2]    11.16    0.06 3.02  6.18  9.00 10.84 12.91 17.95  2932 1.00
## lambda[3]     4.71    0.05 2.06  1.43  3.20  4.46  6.02  9.44  1803 1.00
## lambda[4]    13.89    0.09 3.52  7.94 11.38 13.58 16.08 21.46  1567 1.00
## lambda[5]     8.26    0.05 2.52  4.06  6.51  7.94  9.71 14.25  2563 1.00
## lp__          50.65    0.07 2.10 45.78 49.44 50.94 52.19 53.82  1040 1.00
##
## Samples were drawn using NUTS(diag_e) at Sat May  4 14:50:10 2024.
## For each parameter, n_eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor on split chains (at
## convergence, Rhat=1).

```

1. Hyperparameters (Alpha and Beta)

The estimates for alpha and beta suggest that there's a relatively broad range for the shape parameter (alpha), as indicated by its wider credible interval. This variability in alpha can affect the dispersion of the rate parameters (lambda) across groups. The beta parameter, being closer to 1, indicates a moderate rate of occurrence per group on average but varies less than alpha.

2. Group-Specific Parameters (Lambda)

The lambda parameters represent the rate of events for each group and are directly influenced by the estimated alpha and beta.

3. Diagnostics

- **n_eff (Effective Sample Size)**: Indicates the effective number of independent samples for estimating the parameters. Higher values are better as they suggest less autocorrelation.
- **Rhat (Potential Scale Reduction Factor)**: Should be close to 1.0, indicating convergence. In this output, all parameters have an Rhat of 1.00, which is excellent and suggests that the model has converged well.

4. Interpretation

This hierarchical model effectively accounts for variations both within and across groups by modeling the rates of events (e.g., disease occurrences, insurance claims) in a Bayesian framework. The variability among the lambda values demonstrates the necessity of considering group-specific dynamics rather than assuming a single rate across all groups.

The posterior distributions for the group-specific lambdas, informed by the data through the Empirical Bayes setup, provide a robust way to estimate these rates while accounting for uncertainty. Such models are particularly useful in fields like epidemiology, where event rates can vary significantly by demographic or geographic groupings.

5. Conclusion

The results indicate that the model is well-calibrated and the estimates are reliable, given the effective sample sizes and convergence diagnostics. This model can serve as a powerful tool for making informed decisions based on grouped count data, with applications spanning healthcare, insurance, and beyond.

5. Historical Context

Empirical Bayes methods trace their roots back to the work of Herbert Robbins in the 1950s, who introduced the concept in a series of papers. The development of these methods was motivated by the need for effective statistical techniques that could incorporate prior information without requiring a fully specified Bayesian prior. This need became increasingly significant with the advent of large-scale data collection and computational advances post-World War II.

The evolution of Empirical Bayes reflects broader trends in the statistical community, particularly the shift from purely theoretical approaches to more data-driven, computational methods. Over the decades, as computational power increased and more complex data became available, Empirical Bayes methods grew in popularity and sophistication, providing a practical framework for addressing real-world statistical problems where traditional Bayesian methods were impractical due to the lack of complete prior knowledge.

6. Statistical Practice Implications

The integration of Empirical Bayes methods into modern statistical practice has had profound implications across various fields. In medicine, these methods are used to adjust for varying risk factors across different patient groups, improving the accuracy of clinical trials and epidemiological studies. In ecology, Empirical Bayes allows researchers to estimate population parameters where only incomplete data are available, crucial for conservation efforts.

Furthermore, the flexibility of Empirical Bayes methods in handling large datasets has made them particularly valuable in the era of big data. They allow statisticians to efficiently incorporate historical data into

current analyses, enhancing decision-making processes in business and finance. The ability to update prior distributions dynamically based on new data aligns well with the needs of industries that require adaptive, real-time analytical methodologies.

7. Conclusion

This exploration of Empirical Bayes methods through Chapter 6 of “Computer Age Statistical Inference” has highlighted the rich theoretical underpinnings and practical applications of this statistical approach. From its early conceptualization by Herbert Robbins to its sophisticated computational implementations in modern statistics, Empirical Bayes has proven to be a versatile and powerful tool. It bridges the gap between traditional Bayesian and frequentist methods, providing a pragmatic solution to real-world problems involving uncertain priors and large datasets.

The hierarchical Poisson-Gamma model example underscores the method’s adaptability and relevance in handling complex, structured data scenarios. As data continues to grow in volume and complexity, the role of Empirical Bayes in statistical practice is expected to expand, continuing to provide robust, data-driven insights across diverse scientific, industrial, and technological domains.

8. References

1. Goldstein, Michael, and David Wooff. *Bayes Linear Statistics: Theory and Methods*. John Wiley & Sons, 2007.
2. Johnson, Valen E. “Bayes Factors Based on Test Statistics.” *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 5, 2005, pp. 689-701.
3. Robbins, Herbert. “The Empirical Bayes Approach to Statistical Decision Problems.” *The Annals of Mathematical Statistics*, vol. 35, no. 1, 1964, pp. 1-20.