# Chapter 6 Empirical Bayes

Chenxuan Xiong

2024-05-06

## What is Empirical Bayes

Empirical Bayes methods are procedures for statistical inference in which the *prior probability distribution is estimated from the data*. This approach stands in contrast to standard Bayesian methods, for which the prior distribution is fixed before any data are observed.

It's a theory that originated in the 1940s, but even now the main problems it faces remain largely unresolved —— *data collection*. At pre-computer times, amassing large data sets needs enormous efforts which might be a full life's labor (Shakespeare word counts case). Even now, coming by a dependable data is still hard, and it's also statistician's job to pry it out of enormous database.

## Examples

### Claims data and Robbins' Formula

Overview: If $x$ is a random variable follows poisson distribution with parameter $\theta$, we can estimate the $\theta$ for next time period based on the observations in current time period without knowing the prior distribution for $\theta$, $g(\theta)$.

Table below shows one year of claims data for a European automobile insurance company; 7840 of the 9461 policy holders made no claims during the year, 1317 made a single claim, 239 made two claims each, etc. The insurance company is concerned about the claims each policy holder will make in the next year.

| **Claims $x$** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|---|
| Counts $y_x$ | 7840 | 1317 | 239 | 42 | 14 | 4 | 4 | 1 |

```
claims_x <- 0:7
counts_yx <- c(7840, 1317, 239, 42, 14, 4, 4, 1)
```

We suppose that $x_k$, the number of claims to be made in a single year by policy holder $k$, follows a Poisson distribution with parameter $\theta_k$,

$$Pr\{x_k = x\} = p_{\theta_k} = e^{-\theta_k}\theta_k^x/x!, \qquad (1)$$

where $\theta_k$ is the expected value of $x_k$, the smaller $\theta_k$ the better.

Suppose the prior density for customer's $\theta$ values is $g(\theta)$. By Bayes's rule:

$$g(\mu|x) = g(\mu)f_\mu(x)/f(x), \qquad \mu \in \Omega$$

, where $f(x)$ is the *marginal density* of $x$,

$$f(x) = \int_\Omega f_\mu(x)g(\mu)d(\mu).$$

We have

$$E\{\theta|x\} = \frac{E(\theta, x)}{E(x)} = \frac{\int_0^\infty \theta p_\theta(x)g(\theta)d\theta}{\int_0^\infty p_\theta(x)g(\theta)d\theta} \qquad (2)$$

Since $\theta = E(x)$, where $x$ means make $x$ claims in a single year. The number of claims $X$ to expect the next year from the same customer$E(X|x)$ is also $E(\theta|x)$ ($\theta$ being the expectation of $X$)

If we plug formula (1) into (3), formula (3) becomes

$$\begin{aligned} E\{\theta|x\} &= \frac{\int_0^\infty \theta[e^{-\theta}\theta^x/x!]g(\theta)d\theta}{\int_0^\infty [e^{-\theta}\theta^x/x!]g(\theta)d\theta} \\ &= \frac{\int_0^\infty [e^{-\theta}\theta^{x+1}/x!]g(\theta)d\theta}{\int_0^\infty [e^{-\theta}\theta^x/x!]g(\theta)d\theta} \\ &= \frac{\int_0^\infty (x+1)[e^{-\theta}\theta^{x+1}/(x+1)!]g(\theta)d\theta}{\int_0^\infty [e^{-\theta}\theta^x/x!]g(\theta)d\theta}, \end{aligned} \qquad (3)$$

where the denominator is the *marginal density* of $x$, we denote it by $f(x)$. It gives *Robbins' formula*,

$$E\{\theta|x\} = (x+1)f(x+1)/f(x) \qquad (4)$$

With this formula, we can estimate $E\{\theta|x\}$ without knowing the prior density $g(\theta)$. The obvious estimate of the marginal density $f(x)$ is the proportion of total counts in category $x$,

$$\hat{f}(x) = y_x/N, \qquad with \quad N = \sum_x y_x,$$

and the *Robbins' formula* has a empirical version of

$$\begin{aligned} E\{\theta|x\} &= (x+1)\hat{f}(x+1)/\hat{f}(x) \\ &= (x+1)y_{x+1}/y_x \end{aligned}$$

Calculate $E\{\theta|0\}, E\{\theta|1\}...E\{\theta|6\}$:

```
claims_x <- 0:7
counts_yx <- c(7840, 1317, 239, 42, 14, 4, 4, 1)
exp_theta <- (claims_x + 1) * c(counts_yx[-1], NA) / counts_yx
exp_theta
```

[1] 0.1679847 0.3629461 0.5271967 1.3333333 1.4285714 6.0000000 1.7500000 [8] NA

**Note:** When $x = 6$, the estimated value calculated by R is different from the result showed in the book.

| **Claims x** | **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
|---|---|---|---|---|---|---|---|---|
| Counts yx | 7840 | 1317 | 239 | 42 | 14 | 4 | 4 | 1 |
| Result from book | .168 | .363 | .527 | 1.33 | 1.43 | 6.00 | 1.25 | |
| Result calculated here | 0.1679847 | 0.3629461 | 0.5271967 | 1.3333333 | 1.4285714 | 6.0000000 | 1.7500000 | |

The empirical bayes sometimes might not stable in some cases (like the right end of the table), we can use a parametric approach: give $g(\theta)$ a distribution. Assume $g(\theta)$ for $\theta_k$ has a gamma form

$$g(\theta) = \frac{\theta^{\nu-1} e^{-\theta/\sigma}}{\sigma^{\nu}\Gamma(\nu)}, \quad \text{for } \theta \geq 0 \qquad (5)$$

with $\nu$ and $\sigma$ unknown.

Estimate $\nu$ and $\sigma$ with MLE:

The likelihood function:

$$L(\nu, \sigma) = \prod_{i=1}^{n} \frac{\theta_i^{\nu-1} e^{-\theta_i/\sigma}}{\sigma^{\nu}\Gamma(\nu)}$$

$$\log L(\nu, \sigma) = \sum_{i=1}^{n} \left[ (\nu - 1) \log \theta_i - \frac{\theta_i}{\sigma} - \nu \log \sigma - \log \Gamma(\nu) \right]$$

Derivative with respect to $\nu$:

$$\frac{\partial \log L}{\partial \nu} = \sum_{i=1}^{n} \left[ \log \theta_i - \log \sigma - \psi(\nu) \right] = 0$$

Here, $\psi$ is the digamma function, which is the derivative of the logarithm of the gamma function, $log\Gamma(\nu)$

Derivative with respect to $\sigma$:

$$\frac{\partial \log L}{\partial \sigma} = \sum_{i=1}^{n} \left[ -\frac{\theta_i}{\sigma^2} + \frac{\nu}{\sigma} \right] = 0$$

R implementation

```
library(fitdistrplus)
library(MASS)

observed_claims <- rep(claims_x, times=counts_yx)

adjusted_claims <- observed_claims + 0.04

gamma_fit <- fitdistr(adjusted_claims, "gamma")

print(gamma_fit$estimate)
```

```
shape     rate
```

0.5014174 1.9713406

Calculate parametrixally estumated marginal density with estimated $\nu$ and $\sigma$. (Based on formula given in the appendix)

$$\hat{f}(x) = f_{\hat{\nu},\hat{\sigma}}(x) = \frac{\gamma^{x+\hat{\nu}}\Gamma(\hat{\nu}+x)}{\hat{\sigma}^{\hat{\nu}}\Gamma(\hat{\nu})x!}$$

```
nu <- gamma_fit$estimate["shape"]
sigma <- gamma_fit$estimate["rate"]
gamma <- sigma / (1 + sigma)

f_hat_x <- dnbinom(claims_x, size = nu, prob = gamma)

print(f_hat_x)
```

[1] 8.140518e-01 1.373722e-01 3.470708e-02 9.739361e-03 2.869207e-03 [6] 8.693382e-04 2.682623e-04 8.385274e-05

Calculate $\hat{y} = Nf_{\hat{\nu},\hat{\sigma}}(x)$

```
y_hat = sum(counts_yx)*f_hat_x

print(y_hat)
```

[1] 7701.7442028 1299.6788479 328.3636377 92.1440991 27.1455709 [6] 8.2248090 2.5380298 0.7933308

Calculate $E_{\hat{\nu},\hat{\sigma}}\{\theta|x\} = (x+1)\hat{y}_{x+1}/\hat{y}_x$

```
exp_theta_para <- (claims_x + 1) * c(y_hat[-1], NA) / y_hat
exp_theta_para
```

[1] 0.1687512 0.5052997 0.8418481 1.1783965 1.5149449 1.8514933 2.1880418 [8] NA

**Note:**The estimated value calculated by R is different from the result showed in the book. Since the gamma MLE only accept positive values, some transformations are needed for claim 0.

| Claims x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| Counts yx | 7840 | 1317 | 239 | 42 | 14 | 4 | 4 | 1 |
| Result from book | .164 | .398 | .633 | .87 | 1.10 | 1.34 | 1.57 | |
| Result calculated | 0.169 | 0.505 | 0.842 | 1.178 | 1.515 | 1.851 | 2.188 | |

# The Missing-Species Problem

Overview: Empirical Bayes can be use in finding estimation of exists but unseen observations.

Butterfly data; number y of species seen x times each in *two* years of trapping; 118 species trapped just once, 74 trapped twice each, etc.

| x | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 118 | 74 | 44 | 24 | 29 | 22 | 20 | 19 | 20 | 15 | 12 | 14 |

| x | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| y | 6 | 12 | 6 | 9 | 9 | 6 | 10 | 10 | 11 | 5 | 3 | 3 |

Suppose there are $S$ species in all, seen or unseen, $x_k$ is the number of times species $k$ is trapped in one time unit, follows,

$$x_k \sim Poi(\theta_k) = \frac{\theta_k^x e^{-\theta_k}}{x!}, \qquad \text{for } k = 1, 2, 3,..., \text{S}$$

Question Interested: if Corbet trapped for one additional year, how many new species would he expect to capture?

Key Assumption:

$$x_k(t) \sim Pos(\theta_k, t) = \frac{(\theta_k t)^x e^{-\theta_k t}}{x!}$$

is independent from $x_k$. That is, any one species is trapped independently over time at a rate proportional to its parameter $\theta_k$.

$$P(\text{species k is not seen in the initial trapping period but is seen in the new period})$$
$$= P(x_k = 0, x_k(t) > 0) = e^{-\theta_k}(1 - e^{-\theta_k t})$$

The expectation number of new species seen in new trapping period:

$$E(t) = \sum_{k=1}^{S} e^{-\theta_k}(1 - e^{-\theta_k t})$$
$$= S \int_0^\infty e^{-\theta}(1 - e^{-\theta t}) g(\theta)\, d\theta$$

where $g(\theta)$ is the "empirical density" putting probability 1/S on each of the $\theta_k$ values.

Expanding $1 - e^{-\theta t}$ gives

$$E(t) = \int_0^\infty \left[e^{-\theta}(\theta t - (\theta t)^2/2! + (\theta t)^3/3! - \dots)\right] g(\theta)\, d\theta \qquad (6)$$

where

$$e_x = E\{y_x\} = \sum_{k=1}^{S} e^{-\theta_k}\theta_k^x/x!$$
$$= S \int_0^\infty e^{-\theta}\theta^x/x!\, g(\theta)\, d\theta \qquad (7)$$

Comparing (6) and (7) we have

$$E(t) = e_1 t - e_2 t^2 + e_3 t^3 - \dots$$

estimate $e_x$ by $y_x$:

$$\hat{E}(t) = y_1 t - y_2 t^2 + y_3 t^3 - \dots$$

and t = 1/2, we have

$$\hat{E}(1/2) = 118(1/2) - 74(1/2)^2 + 44(1/2)^3 - \dots = 45.2$$

Implmentation with R:

```
library(knitr)
library(kableExtra)
y_values <- c(118, 74, 44, 24, 29, 22, 20, 19, 20, 15, 12, 14, 6, 12, 6, 9, 9, 6, 10, 1
0, 11, 5, 3, 3)

t_values <- seq(0, 1, by = 0.1)

calculate_E_hat <- function(t, y_values) {
  terms <- length(y_values)
  series_terms <- sapply(1:terms, function(k) y_values[k] * t^k * (-1)^(k+1))
  sum(series_terms)
}

E_hat_values <- sapply(t_values, calculate_E_hat, y_values = y_values)

et = data.frame(t = t_values, E_hat = E_hat_values)
et %>%
  kable("html") %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F)
```

| t | E_hat |
|---:|---:|
| 0.0 | 0.00000 |
| 0.1 | 11.10187 |
| 0.2 | 20.96169 |
| 0.3 | 29.79148 |
| 0.4 | 37.79271 |
| 0.5 | 45.17149 |
| 0.6 | 52.14693 |
| 0.7 | 58.92833 |
| 0.8 | 65.57362 |
| 0.9 | 71.55992 |
| 1.0 | 75.00000 |

The same as the claim example, the estimations become unstable when t become large (t>1). We can use a parametric impirical Bayes on this case too. The prior of $g(\theta)$ for $\theta_k$. Then $E(t)$ will be

$$E(t) = e_1 \left\{1 - (1 + \gamma t)^{-\nu}\right\} / (\gamma \nu),$$

where $\gamma = \frac{\sigma}{1+\sigma}$. Taking $\hat{e}_1 = y_1$, the MLE gives

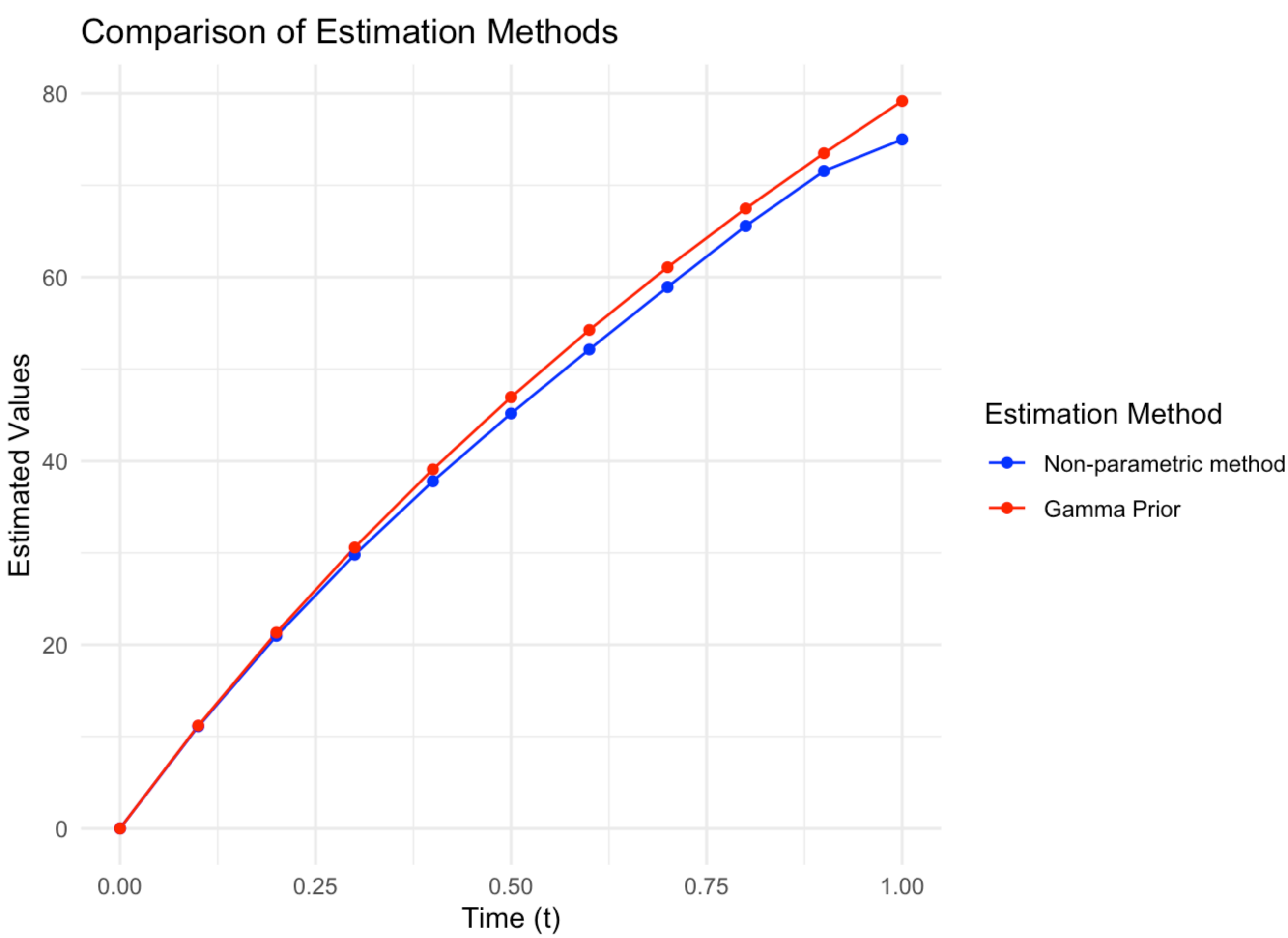
$$\hat{\nu} = 0.104 \qquad \text{and} \qquad \hat{\sigma} = 89.79$$

Implmentation in R

```
nu_hat <- 0.104
sigma_hat <- 89.79

gamma <- sigma_hat / (1 + sigma_hat)

E_t <- function(t, e1, gamma, nu) {
  numerator <- e1 * (1 - (1 + gamma * t)^(-nu))
  denominator <- gamma * nu
  return(numerator / denominator)
}

e1 <- 118  # Assuming the first nonparametric estimate y1 = 118

t_values <- seq(0, 1, by = 0.1)
E_t_values <- sapply(t_values, E_t, e1 = e1, gamma = gamma, nu = nu_hat)

et_gamma <- data.frame(t = t_values, E_t = E_t_values)
et_gamma %>%
  kable("html") %>%
  kable_styling(bootstrap_options = c("striped", "hover"), full_width = F)
```

| t | E_t |
|---|---|
| 0.0 | 0.00000 |
| 0.1 | 11.19732 |
| 0.2 | 21.33347 |
| 0.3 | 30.58504 |
| 0.4 | 39.08842 |
| 0.5 | 46.95109 |
| 0.6 | 54.25922 |
| 0.7 | 61.08287 |
| 0.8 | 67.47981 |
| 0.9 | 73.49817 |
| 1.0 | 79.17850 |

```
library(tidyr)
library(ggplot2)
data <- merge(et,et_gamma)
data_long <- pivot_longer(data, cols = c("E_hat", "E_t"), names_to = "Series", values_to
= "Values")

ggplot(data_long, aes(x = t, y = Values, color = Series, group = Series)) +
  geom_line() +  # Adds line elements
  geom_point() +  # Adds point elements
  scale_color_manual(
    name = "Estimation Method",  # Custom legend title
    values = c("blue", "red"),  # Colors for each line
    labels = c("Non-parametric method", "Gamma Prior")  # Custom labels for legend
  ) +
  labs(x = "Time (t)", y = "Estimated Values", title = "Comparison of Estimation Method
s") +
  theme_minimal()
```




# Indirect Evidence:

## Concept of Indirect Evidence

Definition: A good statistical argument often combines many small, sometimes contradictory, pieces of evidence to produce a coherent conclusion. This is particularly important in contexts like clinical trials, where outcomes don't always consistently support the efficacy of a new drug.

Direct vs. Indirect Evidence: Direct evidence in statistics directly answers the question of interest (e.g., success or failure of a clinical trial subject directly related to the treatment). Indirect evidence, meanwhile, leverages prior experiences or related cases to infer about a current case, not relying directly on it but using related data to inform conclusions.

### Role of Bayesian Inference

Incorporating Experience: Bayesian inference is crucial for integrating indirect evidence into statistical reasoning. It uses prior experience (through prior probability distributions) to inform current inferences.

Empirical Bayes Approach: This approach simplifies traditional Bayesian methods by removing the need for a specified prior distribution. Instead, it uses the data itself to estimate the prior, which is then used to influence the estimate of the parameters of interest. An example given discusses how an Empirical Bayes estimate of a parameter$\theta$ is derived and compares it with a direct estimate.

# Summary

This chapter shows how to use empirical Bayes in problems that seem impossible to be solved and indicates a growing acceptance of using indirect evidence through Empirical Bayes methods and other Bayesian approaches to address complex, real-world problems in statistical inference.