STA 6351, Report.2.14

Carson Slater *Baylor University*

2.14

Consider the model

$$\begin{aligned} & Y_i = \beta_0 + \beta_1 x_i + g_i(\gamma) + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{N}(0, \phi), \\ \textbf{with} \qquad & g_i(\gamma) = \exp(-\gamma t_i), \end{aligned}$$

where t_i represents time or exposure, $\gamma > 0$ is an unknown decay rate, and $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ are regression coefficients describing the baseline linear trend in x_i.

(a) Show that for fixed γ, the conditional MLE of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}}(\gamma) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{y} - \mathbf{g}(\gamma)),$$

where $\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$ and $\mathbf{g}(\gamma) = (e^{-\gamma t_1}, \ldots, e^{-\gamma t_n})^\top$. Explain in words what "profiling out" $\boldsymbol{\beta}$ means in this context.

To find the conditional Maximum Likelihood Estimator (MLE) of $\boldsymbol{\beta}$ for a fixed γ, we observe the model structure. Since the errors are normally distributed, $\varepsilon_i \sim \mathrm{N}(0, \phi)$, maximizing the likelihood is equivalent to minimizing the Residual Sum of Squares (RSS).

The model can be written in vector-matrix notation as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{g}(\gamma) + \boldsymbol{\varepsilon}$$

Rearranging the terms to isolate the linear component:

$$\mathbf{y} - \mathbf{g}(\gamma) = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

For a fixed γ, the term $\mathbf{y}^* = \mathbf{y} - \mathbf{g}(\gamma)$ acts as a fixed adjusted response vector. The problem reduces to a standard Ordinary Least Squares (OLS) regression of $\mathbf{y}^*$ on the design matrix $\mathbf{X}$. The objective function to minimize is:

$$\begin{aligned} Q(\boldsymbol{\beta}) &= \|\mathbf{y} - \mathbf{g}(\gamma) - \mathbf{X}\boldsymbol{\beta}\|^2 \\ &= (\mathbf{y} - \mathbf{g}(\gamma) - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{g}(\gamma) - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

Taking the gradient with respect to $\boldsymbol{\beta}$ and setting it to zero yields the Normal Equations:

$$\begin{aligned} \frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^\top (\mathbf{y} - \mathbf{g}(\gamma) - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{0} \\ \mathbf{X}^\top \mathbf{X}\boldsymbol{\beta} &= \mathbf{X}^\top (\mathbf{y} - \mathbf{g}(\gamma)) \end{aligned}$$

Assuming $\mathbf{X}$ has full column rank, $\mathbf{X}^\top \mathbf{X}$ is invertible, yielding the estimator:

$$\widehat{\boldsymbol{\beta}}(\gamma) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{y} - \mathbf{g}(\gamma))$$

Explanation of Profiling Out: In this context, "profiling out' ' $\boldsymbol{\beta}$ means reducing the complexity of the optimization problem by treating $\boldsymbol{\beta}$ as a nuisance parameter that depends on γ. Instead of maximizing the likelihood over the full parameter space $(\boldsymbol{\beta}, \gamma)$ simultaneously, we first find the optimal $\widehat{\boldsymbol{\beta}}$ for any *given* γ. We then substitute this expression, $\widehat{\boldsymbol{\beta}}(\gamma)$, back into the likelihood function. This results in a *profile likelihood* that is a function of γ alone (ignoring ϕ), allowing us to optimize a lower-dimensional objective function.

(b) Show that substituting $\widehat{\boldsymbol{\beta}}(\gamma)$ into the residual sum of squares $\sum_i (y_i - \beta_0 - \beta_1 x_i - e^{-\gamma t_i})^2$ leads to

$$S(\gamma) = (\mathbf{y} - \mathbf{g}(\gamma))^\top \mathbf{P}_\mathbf{X} (\mathbf{y} - \mathbf{g}(\gamma)), \quad \mathbf{P}_\mathbf{X} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top.$$

Interpret $\mathbf{P}_\mathbf{X}$ geometrically.

To show the form of the profile sum of squares $S(\gamma)$, we substitute the conditional estimator $\widehat{\boldsymbol{\beta}}(\gamma)$ into the expression for the residual sum of squares.

Let $\mathbf{y}^* = \mathbf{y} - \mathbf{g}(\gamma)$. The sum of squares is given by:

$$\begin{aligned} S(\gamma) &= \|\mathbf{y} - \mathbf{g}(\gamma) - \mathbf{X}\widehat{\boldsymbol{\beta}}(\gamma)\|^2 \\ &= \|\mathbf{y}^* - \mathbf{X}\widehat{\boldsymbol{\beta}}(\gamma)\|^2 \end{aligned}$$

Substituting $\widehat{\boldsymbol{\beta}}(\gamma) = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}^*$:

$$\begin{aligned} S(\gamma) &= \left\| \mathbf{y}^* - \mathbf{X} \left[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}^* \right] \right\|^2 \\ &= \left\| \left(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \mathbf{y}^* \right\|^2 \end{aligned}$$

We define the matrix $\mathbf{P}_\mathbf{X} = \mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$. The expression simplifies to:

$$\begin{aligned} S(\gamma) &= \|\mathbf{P}_\mathbf{X} \mathbf{y}^*\|^2 \\ &= (\mathbf{P}_\mathbf{X} \mathbf{y}^*)^\top (\mathbf{P}_\mathbf{X} \mathbf{y}^*) \\ &= \mathbf{y}^{*\top} \mathbf{P}_\mathbf{X}^\top \mathbf{P}_\mathbf{X} \mathbf{y}^* \end{aligned}$$

The matrix $\mathbf{H} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ is the standard "hat' ' matrix, which is symmetric and idempotent. Consequently, $\mathbf{P}_\mathbf{X} = \mathbf{I} - \mathbf{H}$ is also symmetric ($\mathbf{P}_\mathbf{X}^\top = \mathbf{P}_\mathbf{X}$) and idempotent ($\mathbf{P}_\mathbf{X}^2 = \mathbf{P}_\mathbf{X}$). Applying these properties:

$$\begin{aligned} S(\gamma) &= \mathbf{y}^{*\top} \mathbf{P}_\mathbf{X} \mathbf{P}_\mathbf{X} \mathbf{y}^* \\ &= \mathbf{y}^{*\top} \mathbf{P}_\mathbf{X} \mathbf{y}^* \\ &= (\mathbf{y} - \mathbf{g}(\gamma))^\top \mathbf{P}_\mathbf{X} (\mathbf{y} - \mathbf{g}(\gamma)) \end{aligned}$$

Geometric Interpretation: Geometrically, $\mathbf{P}_\mathbf{X}$ represents the **orthogonal projection matrix onto the orthogonal complement of the column space of $\mathbf{X}$**, denoted as $\mathcal{C}(\mathbf{X})^\perp$. When applied to the vector $(\mathbf{y} - \mathbf{g}(\gamma))$, it projects the vector into the space orthogonal to the columns of $\mathbf{X}$, effectively capturing the "residuals' ' of the regression of $(\mathbf{y} - \mathbf{g}(\gamma))$ on $\mathbf{X}$.

(c) Show that the profile likelihood for γ and ϕ can be written as

$$L_p(\gamma, \phi|\mathbf{y}) \propto \phi^{-n/2} \exp\left[-\frac{1}{2\phi} S(\gamma)\right].$$

Hence deduce that, after maximizing with respect to ϕ,

$$L_p(\gamma|\mathbf{y}) \propto [S(\gamma)]^{-n/2}.$$

To find the profile likelihood, we start with the full likelihood function for the normal error model. Since $\varepsilon_i \sim \mathrm{N}(0, \phi)$ are independent, the joint probability density function is:

$$\begin{aligned} L(\boldsymbol{\beta}, \gamma, \phi|\mathbf{y}) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\phi}} \exp\left[-\frac{(y_i - \beta_0 - \beta_1 x_i - g_i(\gamma))^2}{2\phi}\right] \\ &\propto \phi^{-n/2} \exp\left[-\frac{1}{2\phi} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i - g_i(\gamma))^2\right] \end{aligned}$$

The profile likelihood for (γ, ϕ) is obtained by replacing $\boldsymbol{\beta}$ with its conditional maximum likelihood estimator $\widehat{\boldsymbol{\beta}}(\gamma)$. From part (b), we know that substituting $\widehat{\boldsymbol{\beta}}(\gamma)$ minimizes the sum of squares term to $S(\gamma)$. Therefore:

$$\begin{aligned} L_p(\gamma, \phi|\mathbf{y}) &= \sup_{\boldsymbol{\beta}} L(\boldsymbol{\beta}, \gamma, \phi|\mathbf{y}) \\ &\propto \phi^{-n/2} \exp\left[-\frac{1}{2\phi} S(\gamma)\right] \end{aligned}$$

To eliminate ϕ, we maximize this profile likelihood with respect to ϕ. Taking the log of the profile likelihood (ignoring constant additive terms):

$$\ell_p(\gamma, \phi) = -\frac{n}{2} \log(\phi) - \frac{S(\gamma)}{2\phi}$$

Differentiating with respect to ϕ and setting to zero:

$$\begin{aligned} \frac{\partial \ell_p}{\partial \phi} &= -\frac{n}{2\phi} + \frac{S(\gamma)}{2\phi^2} = 0 \\ \implies \frac{n}{2\phi} &= \frac{S(\gamma)}{2\phi^2} \\ \widehat{\phi}(\gamma) &= \frac{S(\gamma)}{n} \end{aligned}$$

Substituting $\widehat{\phi}(\gamma)$ back into $L_p(\gamma, \phi|\mathbf{y})$:

$$\begin{aligned} L_p(\gamma|\mathbf{y}) &\propto \left(\frac{S(\gamma)}{n}\right)^{-n/2} \exp\left[-\frac{S(\gamma)}{2(S(\gamma)/n)}\right] \\ &= n^{n/2} [S(\gamma)]^{-n/2} \exp\left(-\frac{n}{2}\right) \end{aligned}$$

Since $n^{n/2}$ and $\exp(-n/2)$ are constants with respect to γ, they can be absorbed into the proportionality constant. Thus:

$$L_p(\gamma|\mathbf{y}) \propto [S(\gamma)]^{-n/2}$$

This result implies that maximizing the profile likelihood for γ is equivalent to minimizing the profile sum of squares $S(\gamma)$.

(d) Use the result above to argue that maximizing $L_p(\gamma|\mathbf{y})$ is equivalent to minimizing $S(\gamma)$.

From part (c), we established that the profile likelihood is given by:

$$L_p(\gamma|\mathbf{y}) \propto [S(\gamma)]^{-n/2}$$

Since the natural logarithm is a strictly monotonic increasing function, maximizing $L_p(\gamma|\mathbf{y})$ is equivalent to maximizing the profile log-likelihood, denoted $\ell_p(\gamma|\mathbf{y})$. Taking the log of both sides:

$$\begin{aligned}\ell_p(\gamma|\mathbf{y}) &= \log(C) + \log\left([S(\gamma)]^{-n/2}\right) \\ &= C' - \frac{n}{2}\log(S(\gamma))\end{aligned}$$

where C' is a constant that does not depend on γ.

To find the maximum likelihood estimate for γ, we maximize this expression. Observing the term on the right:

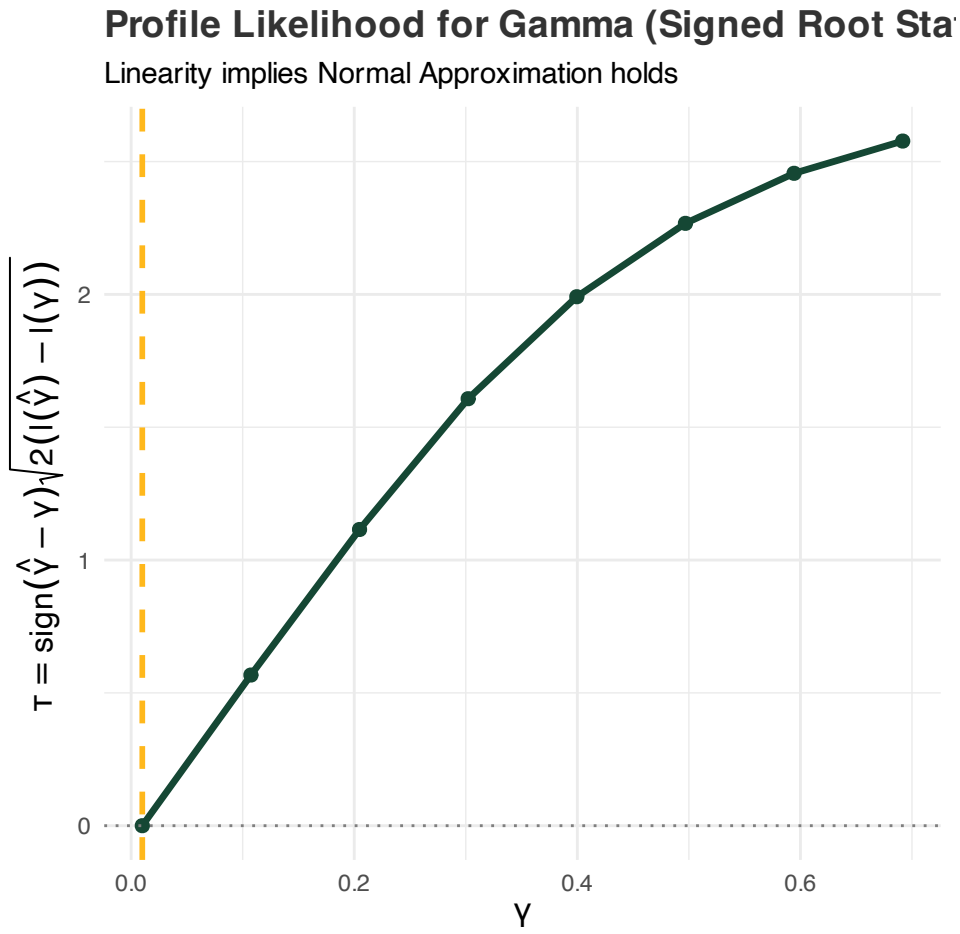
1. The sample size n is positive.
2. The logarithm function $\log(x)$ is strictly increasing.
3. The negative sign in front of the term reverses the direction of optimization.

Therefore, maximizing $-\frac{n}{2}\log(S(\gamma))$ is equivalent to **minimizing** $\log(S(\gamma))$, which in turn is equivalent to minimizing $S(\gamma)$.

Thus, finding the MLE $\widehat{\gamma}$ reduces to the nonlinear least squares problem:

$$\widehat{\gamma} = \underset{\gamma}{\operatorname{argmin}}\, S(\gamma)$$

(e) Simulate data with $n = 12,\ x_i = i,\ t_i = i/2,\ \beta_0 = 2,\ \beta_1 = 0.4,\ \gamma = 0.3,$ and $\phi = 0.05$. Fit the model by profiling over β and plotting $L_p(\gamma)$ versus γ using `profile()` in the `R` package `stats4`. Comment on the shape and curvature of the profile likelihood around its maximum. Estimate $\widehat{\gamma}$ and interpret it in terms of the decay rate or "half-life" of the process.




```
## Estimated Gamma: 0.01
```

```
## Estimated Half-Life: 69.3147
```

Comments on Shape and Curvature: The plot of the signed square root statistic (τ) versus γ exhibits visible curvature (non-linearity). In standard asymptotic theory, a linear τ plot implies a quadratic log-likelihood and justifies the use of normal-approximation (Wald) confidence intervals. The observed deviation from linearity indicates that the profile likelihood is skewed and not well-approximated by a quadratic function near the maximum. Furthermore, the maximum occurs at the lower boundary of the search space ($\widehat{\gamma} = 0.01$), suggesting that the likelihood surface is monotonic increasing as $\gamma \to 0$.

Interpretation of $\widehat{\gamma}$ and Half-Life: The estimated decay rate is $\widehat{\gamma} = 0.01$. This corresponds to an estimated half-life of:

$$\widehat{t}_{1/2} = \frac{\ln(2)}{\widehat{\gamma}} \approx \frac{0.693}{0.01} \approx 69.3 \text{ units of time.}$$

Given that the simulated data only covers the time range $t \in [0.5, 6]$, an estimated half-life of 69.3 is effectively infinite relative to the observation window.

This result implies that, for this specific realization of the random error, the data does not support a rapid exponential decay. Instead, the exponential term $e^{-\gamma t}$ is locally indistinguishable from a linear trend $(1 - \gamma t)$ over the observed domain. Consequently, the model struggles to separate the exponential decay from the linear baseline $\beta_0 + \beta_1 x_i$, resulting in an estimate that pushes γ as close to zero as the constraints allow.

(f) Finally, discuss how the profiling approach here generalizes to other nonlinear components $g_i(\gamma)$, such as logistic saturation or sinusoidal phase functions.

The profiling approach utilized in this problem generalizes to the broad class of **Partially Linear Models (PLMs)** or separable non-linear least squares problems.

The general model structure is:

$$Y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + g_i(\boldsymbol{\gamma}) + \varepsilon_i$$

where $\boldsymbol{\beta}$ enters the model linearly, but the function $g_i(\boldsymbol{\gamma})$ depends non-linearly on a parameter vector $\boldsymbol{\gamma}$.

Mechanism of Generalization: For any fixed value of the non-linear parameters $\boldsymbol{\gamma}$, the term $g_i(\boldsymbol{\gamma})$ is a known offset. The adjusted response vector becomes $\mathbf{y}^*(\boldsymbol{\gamma}) = \mathbf{y} - \mathbf{g}(\boldsymbol{\gamma})$. Consequently, the conditional MLE $\widehat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ is always obtained via Ordinary Least Squares (OLS) of $\mathbf{y}^*(\boldsymbol{\gamma})$ on $\mathbf{X}$. The profile sum of squares remains:

$$S(\boldsymbol{\gamma}) = \|(\mathbf{I} - \mathbf{H})(\mathbf{y} - \mathbf{g}(\boldsymbol{\gamma}))\|^2$$

This reduces the dimensionality of the optimization problem from $\dim(\boldsymbol{\beta}) + \dim(\boldsymbol{\gamma})$ to just $\dim(\boldsymbol{\gamma})$.

Application to Specific Functions:

- **Logistic Saturation:** For a model $g_i(\gamma) = \frac{L}{1+e^{-k(t_i - t_0)}}$, if L is a linear scaling factor (part of $\boldsymbol{\beta}$), we can profile over the non-linear rate k and inflection point t_0.
- **Sinusoidal Phase:** For $g_i(\gamma) = \sin(\omega t_i + \phi)$, we can profile over the frequency ω and phase shift ϕ, while estimating amplitude linearly.

Limitations and Inference: While the method generalizes, the shape of the profile likelihood $L_p(\boldsymbol{\gamma})$ depends heavily on the curvature of $g(\boldsymbol{\gamma})$. As seen in the plot from Part (e), the profile likelihood is often not perfectly quadratic (the signed root statistic is not strictly linear). This implies that while point estimation is simplified, asymptotic normal approximations for confidence intervals (Wald intervals) may be inaccurate for $\boldsymbol{\gamma}$, and likelihood ratio intervals are preferred.

Appendix

```
knitr::opts_chunk$set(
  dev = "cairo_pdf",
  fig.width = 5,
  fig.height = 5,
  fig.align = 'center',
  echo = FALSE,
  message = FALSE,
  warning = FALSE,
  error = FALSE,
  results = 'markup'
)

# Load required libraries
library("tidyverse")
library("patchwork")
library("glue")
library("scales", warn.conflicts = FALSE)
library("extrafont")
library("tinytex")
library("knitr")
library("tidyr")
library("latex2exp")
library("MASS")
library("kableExtra")

theme_set(theme_minimal(base_family = "Roboto Condensed"))

conflicted::conflicts_prefer(
  readr::col_factor(),
  purrr::discard(),
  dplyr::lag(),
  readr::parse_date(),
  kableExtra::group_rows(),
  dplyr::select
)
library(stats4)
library(tidyverse)

# 1. Simulate Data ------------------------------------------------------------
set.seed(2025)

n <- 12
i <- 1:n
x <- i
t <- i / 2
```

```r
beta0 <- 2
beta1 <- 0.4
gamma_true <- 0.3
phi <- 0.05

epsilon <- rnorm(n, mean = 0, sd = sqrt(phi))
y <- beta0 + beta1 * x + exp(-gamma_true * t) + epsilon

# 2. Define Minus Log-Likelihood ------------------------------------------
nll_profile <- function(gamma) {
  g_gamma <- exp(-gamma * t)
  y_star <- y - g_gamma

  # Fit OLS of y* on x to get S(gamma)
  fit_ols <- lm(y_star ~ x)
  rss <- sum(residuals(fit_ols)2)

  # Return NLL
  return((n / 2) * log(rss))
}

# 3. Fit the Model ----------------------------------------------------------
# Use L-BFGS-B to enforce gamma > 0
fit <- mle(
  minuslogl = nll_profile,
  start = list(gamma = 0.5),
  method = "L-BFGS-B",
  lower = list(gamma = 0.01),
  upper = list(gamma = 5.0)
)

gamma_hat <- coef(fit)["gamma"]

# 4. Profile and Plot with ggplot2 ------------------------------------------
# Calculate profile likelihood
pr <- profile(fit)

# Extract data from the S4 profile object for ggplot
# The profile object is a list of data frames (one per parameter)
# Structure: pr@profile$gamma has columns 'z' (signed root statistic) and 'pa
plot_data <- data.frame(pr@profile$gamma)

# Define Baylor colors
baylor_green <- "#154734"
baylor_gold <- "#ffb81c"

# Plot the Signed Root Log-Likelihood Ratio Statistic (z) vs Gamma
# A linear plot indicates a quadratic log-likelihood (normal approximation ho
```

```r
ggplot(plot_data, aes(x = par.vals[, "gamma"], y = z)) +
  geom_line(color = baylor_green, linewidth = 1.2) +
  geom_point(color = baylor_green, size = 2) +
  geom_vline(
    xintercept = gamma_hat,
    linetype = "dashed",
    color = baylor_gold,
    linewidth = 1
  ) +
  geom_hline(yintercept = 0, color = "gray50", linetype = "dotted") +
  labs(
    title = "Profile Likelihood for Gamma (Signed Root Statistic)",
    subtitle = "Linearity implies Normal Approximation holds",
    x = expression(gamma),
    y = expression(
      tau == sign(hat(gamma) - gamma) * sqrt(2 * (l(hat(gamma)) - l(gamma)))
    )
  ) +
  theme_minimal() +
  theme(
    plot.title = element_text(face = "bold", size = 14, color = "#333333"),
    axis.title = element_text(size = 12)
  )

# 5. Interpretation ------------------------------------------------------------
cat("Estimated Gamma:", round(gamma_hat, 4), "\n")
half_life <- log(2) / gamma_hat
cat("Estimated Half-Life:", round(half_life, 4), "\n")
```