STA 6351, Report.2.8

Carson Slater *Baylor University*

2.8

Suppose $Y_i \sim$ Poisson(λ_i) with $\log(\lambda_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, $i = 1, \dots, n$, where $\boldsymbol{x}_i \in \mathbb{R}^p$ are fixed covariates.

(a) Write down the log-likelihood $\ell(\beta; \boldsymbol{y})$ in exponential-family form.

The probability mass function for $Y_i \sim \text{Poisson}(\lambda_i)$ is given by:

$$f(y_i; \lambda_i) = \frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}$$

Given the link function $\log(\lambda_i) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$, we have $\lambda_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$.

The log-likelihood for the full sample $\boldsymbol{y} = (y_1, \dots, y_n)^\top$ is:

$$\begin{aligned}
\ell(\boldsymbol{\beta}; \boldsymbol{y}) &= \sum_{i=1}^{n} \log\left(\frac{\lambda_i^{y_i} e^{-\lambda_i}}{y_i!}\right) \\
&= \sum_{i=1}^{n} [y_i \log(\lambda_i) - \lambda_i - \log(y_i!)] \\
&= \sum_{i=1}^{n} \left[y_i(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \log(y_i!)\right]
\end{aligned}$$

To express this in exponential-family form (canonical form), we rearrange terms to isolate the inner product of the natural parameter and the sufficient statistic:

$$\begin{aligned}
\ell(\boldsymbol{\beta}; \boldsymbol{y}) &= \sum_{i=1}^{n} y_i \boldsymbol{x}_i^\top \boldsymbol{\beta} - \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \sum_{i=1}^{n} \log(y_i!) \\
&= \boldsymbol{\beta}^\top \left(\sum_{i=1}^{n} y_i \boldsymbol{x}_i\right) - \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \sum_{i=1}^{n} \log(y_i!)
\end{aligned}$$

Here, the sufficient statistic is $\sum_{i=1}^{n} y_i \boldsymbol{x}_i = \mathbf{X}^\top \boldsymbol{y}$.

(b) Derive the score vector $U(\beta) = \nabla_{\boldsymbol{\beta}} \ell(\beta; \boldsymbol{y})$.

We start with the log-likelihood function derived in part (a):

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) = \sum_{i=1}^{n} \left[y_i(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \log(y_i!)\right]$$

The score vector $U(\boldsymbol{\beta})$ is the gradient (vector of partial derivatives) of $\ell(\boldsymbol{\beta}; \boldsymbol{y})$ with respect to the vector

$\boldsymbol{\beta}$. We can differentiate term by term:

$$\begin{aligned}
U(\boldsymbol{\beta}) &= \nabla_{\boldsymbol{\beta}} \ell(\boldsymbol{\beta}; \boldsymbol{y}) \\
&= \frac{\partial}{\partial \boldsymbol{\beta}} \sum_{i=1}^{n} \left[y_i(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \right] \\
&= \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}} \left[y_i(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \log(y_i!) \right]
\end{aligned}$$

We use two vector calculus rules: $\frac{\partial}{\partial \boldsymbol{v}}(\boldsymbol{a}^\top \boldsymbol{v}) = \boldsymbol{a}$ and the chain rule $\frac{\partial}{\partial \boldsymbol{v}} f(\boldsymbol{a}^\top \boldsymbol{v}) = f'(\boldsymbol{a}^\top \boldsymbol{v}) \cdot \boldsymbol{a}$.

$$\begin{aligned}
U(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \left[y_i \left(\frac{\partial}{\partial \boldsymbol{\beta}} \boldsymbol{x}_i^\top \boldsymbol{\beta} \right) - \left(\frac{\partial}{\partial \boldsymbol{\beta}} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \right) - \left(\frac{\partial}{\partial \boldsymbol{\beta}} \log(y_i!) \right) \right] \\
&= \sum_{i=1}^{n} \left[y_i \boldsymbol{x}_i - \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) \cdot \boldsymbol{x}_i - \mathbf{0} \right] \\
&= \sum_{i=1}^{n} [y_i \boldsymbol{x}_i - \lambda_i \boldsymbol{x}_i] \quad (\text{since } \lambda_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})) \\
&= \sum_{i=1}^{n} (y_i - \lambda_i) \boldsymbol{x}_i
\end{aligned}$$

This can also be written compactly in matrix form, where $\boldsymbol{y}$ and $\boldsymbol{\lambda}$ are $n \times 1$ vectors and $\mathbf{X}$ is the $n \times p$ design matrix:

$$U(\boldsymbol{\beta}) = \mathbf{X}^\top (\boldsymbol{y} - \boldsymbol{\lambda})$$

(c) Show that $U(\boldsymbol{\beta})$ depends on the data only through $T(\boldsymbol{y}) = \mathbf{X}^\top \boldsymbol{y}$, where $\mathbf{X}$ is the $n \times p$ design matrix.

From part (b), the score vector is:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} (y_i - \lambda_i) \boldsymbol{x}_i$$

We can split this summation into two parts:

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} y_i \boldsymbol{x}_i - \sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i$$

Let's analyze each term:

1. **The first term**, $\sum_{i=1}^{n} y_i \boldsymbol{x}_i$, is the product of the data vector $\boldsymbol{y}$ and the covariate matrix $\mathbf{X}$. This can be written in matrix notation as:

   $$\sum_{i=1}^{n} y_i \boldsymbol{x}_i = \mathbf{X}^\top \boldsymbol{y}$$

   This is precisely the statistic $T(\boldsymbol{y})$ defined in the problem.

2. **The second term**, $\sum_{i=1}^{n} \lambda_i \boldsymbol{x}_i$, involves $\lambda_i = \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$. The λ_i values depend only on the fixed covariates $\boldsymbol{x}_i$ and the parameter vector $\boldsymbol{\beta}$. They do not depend on the observed data $\boldsymbol{y}$. Therefore, this entire second term is a function of $\mathbf{X}$ and $\boldsymbol{\beta}$, but it is not a function of the data $\boldsymbol{y}$.

Substituting these back into the expression for $U(\boldsymbol{\beta})$:

$$U(\boldsymbol{\beta}) = \underbrace{\mathbf{X}^\top \boldsymbol{y}}_{T(\boldsymbol{y})} - \underbrace{\sum_{i=1}^{n} \lambda_i(\boldsymbol{\beta}) \boldsymbol{x}_i}_{\text{Function of } \boldsymbol{\beta} \text{ and } \mathbf{X} \text{ only}}$$

Thus, the score vector $U(\boldsymbol{\beta})$ depends on the data $\boldsymbol{y}$ only through the statistic $T(\boldsymbol{y}) = \mathbf{X}^\top \boldsymbol{y}$.

(d) Conclude that $T(\boldsymbol{y})$ is the minimal sufficient statistic for β. *Hint.* Compare each step with the logistic regression example. The algebra differs, but the structure of sufficiency is the same.

From part (a), we established that the log-likelihood for the Poisson regression model is:

$$\ell(\boldsymbol{\beta}; \boldsymbol{y}) = \boldsymbol{\beta}^\top (\mathbf{X}^\top \boldsymbol{y}) - \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta}) - \sum_{i=1}^{n} \log(y_i!)$$

This is in the canonical form of an exponential family, $\ell(\boldsymbol{\theta}; \boldsymbol{y}) = \boldsymbol{\theta}^\top T(\boldsymbol{y}) - A(\boldsymbol{\theta}) + h(\boldsymbol{y})$, with:

- Natural parameter: $\boldsymbol{\theta} = \boldsymbol{\beta}$
- Sufficient statistic: $T(\boldsymbol{y}) = \mathbf{X}^\top \boldsymbol{y}$
- Log-partition function: $A(\boldsymbol{\theta}) = \sum_{i=1}^{n} \exp(\boldsymbol{x}_i^\top \boldsymbol{\beta})$
- Base measure term: $h(\boldsymbol{y}) = -\sum_{i=1}^{n} \log(y_i!)$

The parameter space for the natural parameter $\boldsymbol{\theta} = \boldsymbol{\beta}$ is $\mathbb{R}^p$. Since this parameter space contains an open set in $\mathbb{R}^p$ (as $\mathbb{R}^p$ is itself an open set), a standard theorem of exponential families states that the corresponding sufficient statistic $T(\boldsymbol{y})$ is **minimal sufficient**.

Therefore, we conclude that $T(\boldsymbol{y}) = \mathbf{X}^\top \boldsymbol{y}$ is the minimal sufficient statistic for $\boldsymbol{\beta}$.

Appendix

```r
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
)
```