STA 6351, Report.2.10

Carson Slater *Baylor University*

2.10

(a) Derive $\partial\ell/\partial\alpha$ and $\partial\ell/\partial\beta$ from $\ell(\alpha, \beta; \mathrm{X})$, which is expressed as,

$$\ell(\alpha, \beta; \mathbf{x}) = n \log \alpha - n\alpha \log \beta + (\alpha - 1) \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha}$$

The partial derivatives are:

$$\frac{\partial \ell}{\partial \alpha} = \frac{n}{\alpha} - n \log \beta + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha} \log x_i + \log \beta \sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha}$$

$$\frac{\partial \ell}{\partial \beta} = \frac{\alpha}{\beta} \left[\sum_{i=1}^{n} \left(\frac{x_i}{\beta}\right)^{\alpha} - n\right]$$

(b) (Vector first-order LR) Fix $\boldsymbol{\theta_0} = (\boldsymbol{\alpha_0}, \boldsymbol{\beta_0})$. Show that

$$\frac{\mathbf{L}(\boldsymbol{\theta_1})}{\mathbf{L}(\boldsymbol{\theta_0})} \approx \mathbf{exp} \left\{ (\boldsymbol{\alpha_1} - \boldsymbol{\alpha_0}) \frac{\boldsymbol{\partial \ell}}{\boldsymbol{\partial \alpha}}\bigg|_{\boldsymbol{\theta_0}} + (\boldsymbol{\beta_1} - \boldsymbol{\beta_0}) \frac{\boldsymbol{\partial \ell}}{\boldsymbol{\partial \beta}}\bigg|_{\boldsymbol{\theta_0}} \right\}.$$

Identify explicitly which sample sums appear in the exponent at $\boldsymbol{\theta_0}$.

The approximation is based on a first-order Taylor expansion of the log-likelihood function, $\ell(\theta) = \log L(\theta)$, around θ_0.

$$\ell(\theta_1) \approx \ell(\theta_0) + (\theta_1 - \theta_0)^T \nabla \ell(\theta_0)$$
$$\ell(\alpha_1, \beta_1) \approx \ell(\alpha_0, \beta_0) + (\alpha_1 - \alpha_0) \frac{\partial \ell}{\partial \alpha}\bigg|_{\theta_0} + (\beta_1 - \beta_0) \frac{\partial \ell}{\partial \beta}\bigg|_{\theta_0}$$

Rearranging and recognizing that $\ell(\theta_1) - \ell(\theta_0) = \log L(\theta_1) - \log L(\theta_0) = \log(L(\theta_1)/L(\theta_0))$:

$$\log \left(\frac{L(\theta_1)}{L(\theta_0)}\right) \approx (\alpha_1 - \alpha_0) \frac{\partial \ell}{\partial \alpha}\bigg|_{\theta_0} + (\beta_1 - \beta_0) \frac{\partial \ell}{\partial \beta}\bigg|_{\theta_0}$$

Exponentiating both sides gives the desired result:

$$\frac{L(\theta_1)}{L(\theta_0)} \approx \exp \left\{ (\alpha_1 - \alpha_0) \frac{\partial \ell}{\partial \alpha}\bigg|_{\theta_0} + (\beta_1 - \beta_0) \frac{\partial \ell}{\partial \beta}\bigg|_{\theta_0} \right\}.$$

To identify the sample sums, we evaluate the score functions (from part a) at $\theta_0 = (\alpha_0, \beta_0)$:

$$\left.\frac{\partial \ell}{\partial \alpha}\right|_{\theta_0} = \frac{n}{\alpha_0} - n\log\beta_0 + \sum_{i=1}^{n} \log x_i - \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} \log x_i + \log\beta_0 \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0}$$

$$\left.\frac{\partial \ell}{\partial \beta}\right|_{\theta_0} = \frac{\alpha_0}{\beta_0} \left[\sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} - n\right]$$

The sample sums that appear in the exponent at θ_0 are:

$$\sum_{i=1}^{n} \log x_i, \quad \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0}, \quad \text{and} \quad \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} \log x_i$$

(c) (Second-order refinement) Compute the observed Hessian $H(\theta_0; \mathbf{X})$. Write the second-order approximation to logLR and interpret the quadratic form in terms of local curvature (information). Which additional sample combinations appear through H?

(c) The observed Hessian $H(\theta; \mathbf{X})$ is the matrix of second partial derivatives of the log-likelihood. Let $\theta = (\alpha, \beta)$. The components, evaluated at $\theta_0 = (\alpha_0, \beta_0)$, are:

$$H_{11} = \left.\frac{\partial^2 \ell}{\partial \alpha^2}\right|_{\theta_0} = -\frac{n}{\alpha_0^2} - \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} (\log x_i - \log\beta_0)^2$$

$$H_{22} = \left.\frac{\partial^2 \ell}{\partial \beta^2}\right|_{\theta_0} = \frac{\alpha_0}{\beta_0^2} \left[n - (\alpha_0 + 1) \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0}\right]$$

$$H_{12} = \left.\frac{\partial^2 \ell}{\partial \alpha \partial \beta}\right|_{\theta_0} = \frac{1}{\beta_0} \left[\sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} - n\right] + \frac{\alpha_0}{\beta_0} \sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} (\log x_i - \log\beta_0)$$

$$H(\theta_0; \mathbf{X}) = \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix}$$

The second-order approximation to the log-likelihood ratio is found by extending the Taylor expansion from part (b). Let $\Delta\theta = \theta_1 - \theta_0$:

$$\begin{aligned} \log\left(\frac{L(\theta_1)}{L(\theta_0)}\right) &= \ell(\theta_1) - \ell(\theta_0) \\ &\approx (\theta_1 - \theta_0)^T \nabla \ell(\theta_0) + \frac{1}{2} (\theta_1 - \theta_0)^T H(\theta_0) (\theta_1 - \theta_0) \end{aligned}$$

The quadratic form $\frac{1}{2}\Delta\theta^T H(\theta_0)\Delta\theta$ is the second-order refinement. The Hessian $H(\theta_0)$ describes the local curvature of the log-likelihood surface at θ_0. A "sharper" peak (more negative curvature) corresponds to a larger observed information $J(\theta_0) = -H(\theta_0)$, implying that the data provides more information to distinguish θ_0 from nearby parameter values θ_1.

The additional sample combination that appears through H (specifically, in the H_{11} component) and was not present in the first-order approximation (the score vector) is:

$$\sum_{i=1}^{n} \left(\frac{x_i}{\beta_0}\right)^{\alpha_0} (\log x_i - \log\beta_0)^2$$

(This can also be expressed as a linear combination of $\sum(x_i/\beta_0)^{\alpha_0}(\log x_i)^2$, $\sum(x_i/\beta_0)^{\alpha_0}\log x_i$, and $\sum(x_i/\beta_0)^{\alpha_0}$).

(d) (Reparameterization tip) Repeat part (b) using $\tilde{\boldsymbol{\theta}} = (\alpha, \eta)$ with $\boldsymbol{\eta} = \log \boldsymbol{\beta}$. Comment on numerical stability and the appearance of $\sum \boldsymbol{x_i^{\alpha_0}}$ and $\sum \log \boldsymbol{x_i}$ in the exponent.

We reparameterize the log-likelihood using $\tilde{\theta} = (\alpha, \eta)$, where $\eta = \log \beta$ (so $\beta = e^\eta$).

$$\begin{aligned}\tilde{\ell}(\alpha, \eta; \mathbf{x}) &= n \log \alpha - n\alpha \log(e^\eta) + (\alpha - 1)\sum_{i=1}^n \log x_i - \sum_{i=1}^n \left(\frac{x_i}{e^\eta}\right)^\alpha \\ &= n \log \alpha - n\alpha\eta + (\alpha - 1)\sum_{i=1}^n \log x_i - e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha\end{aligned}$$

The new score vector $\nabla\tilde{\ell} = (\partial\tilde{\ell}/\partial\alpha, \partial\tilde{\ell}/\partial\eta)^T$ has components:

$$\begin{aligned}\frac{\partial\tilde{\ell}}{\partial\alpha} &= \frac{n}{\alpha} - n\eta + \sum_{i=1}^n \log x_i - \frac{\partial}{\partial\alpha}\left(e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha\right) \\ &= \frac{n}{\alpha} - n\eta + \sum_{i=1}^n \log x_i - \left[(-\eta e^{-\alpha\eta})\sum_{i=1}^n x_i^\alpha + e^{-\alpha\eta}\sum_{i=1}^n (x_i^\alpha \log x_i)\right] \\ &= \frac{n}{\alpha} - n\eta + \sum_{i=1}^n \log x_i + \eta e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha - e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha \log x_i\end{aligned}$$

$$\begin{aligned}\frac{\partial\tilde{\ell}}{\partial\eta} &= -n\alpha - \frac{\partial}{\partial\eta}\left(e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha\right) \\ &= -n\alpha - \left((-\alpha e^{-\alpha\eta})\sum_{i=1}^n x_i^\alpha\right) \\ &= \alpha\left(e^{-\alpha\eta}\sum_{i=1}^n x_i^\alpha - n\right)\end{aligned}$$

Fixing $\tilde{\theta}_0 = (\alpha_0, \eta_0)$ with $\eta_0 = \log\beta_0$, the first-order approximation from part (b) is:

$$\log\left(\frac{L(\tilde{\theta}_1)}{L(\tilde{\theta}_0)}\right) \approx (\alpha_1 - \alpha_0)\left.\frac{\partial\tilde{\ell}}{\partial\alpha}\right|_{\tilde{\theta}_0} + (\eta_1 - \eta_0)\left.\frac{\partial\tilde{\ell}}{\partial\eta}\right|_{\tilde{\theta}_0}$$

The exponent now explicitly contains the sums $\sum \log x_i$ and $\sum x_i^{\alpha_0}$ (as well as $\sum x_i^{\alpha_0}\log x_i$). The score components evaluated at $\tilde{\theta}_0$ are:

$$\begin{aligned}\left.\frac{\partial\tilde{\ell}}{\partial\alpha}\right|_{\tilde{\theta}_0} &= \frac{n}{\alpha_0} - n\eta_0 + \sum_{i=1}^n \log x_i + \eta_0 e^{-\alpha_0\eta_0}\sum_{i=1}^n x_i^{\alpha_0} - e^{-\alpha_0\eta_0}\sum_{i=1}^n x_i^{\alpha_0}\log x_i \\ \left.\frac{\partial\tilde{\ell}}{\partial\eta}\right|_{\tilde{\theta}_0} &= \alpha_0\left(e^{-\alpha_0\eta_0}\sum_{i=1}^n x_i^{\alpha_0} - n\right)\end{aligned}$$

This parameterization is common and preferred for numerical stability. The scale parameter β is constrained to be positive ($\beta > 0$), while $\eta = \log \beta$ is unconstrained, i.e., $\eta \in (-\infty, \infty)$. Numerical optimization routines (like Newton-Raphson or BFGS) are more stable and efficient on unconstrained domains as they do not need to handle boundary conditions. The sums $\sum x_i^{\alpha_0}$ and $\sum \log x_i$ now appear directly, rather than the scaled $\sum (x_i/\beta_0)^{\alpha_0}$ from part (b).

(e) (Conceptual) Explain why, despite these informative approximations, a fixed low-dimensional *exact* sufficient statistic does not exist for the two-parameter Weibull.[1]

Despite the informative approximations in parts (b)-(d), a fixed low-dimensional exact sufficient statistic does not exist for the two-parameter Weibull distribution because the Weibull is not a member of the exponential family when both parameters are unknown.

The **Pitman-Koopman-Darmois theorem** states that if a distribution admits a fixed-dimensional sufficient statistic (whose dimension does not grow with sample size n) for all sample sizes, then the distribution must belong to the exponential family. Conversely, distributions outside the exponential family cannot have fixed-dimensional sufficient statistics.

The two-parameter Weibull density is:

$$f(x; \alpha, \beta) = \frac{\alpha}{\beta} \left(\frac{x}{\beta} \right)^{\alpha - 1} \exp \left\{ - \left(\frac{x}{\beta} \right)^{\alpha} \right\}$$

For a distribution to be in the exponential family, it must have the form:

$$f(x; \theta) = h(x) \exp\{\eta(\theta)^T T(x) - A(\theta)\}$$

where $T(x)$ is a fixed-dimensional sufficient statistic that does not depend on the parameter values.

For the Weibull, the likelihood involves terms like $\sum_{i=1}^{n} x_i^{\alpha}$, where the *power* α is itself an unknown parameter. This means the sufficient statistic would need to include $\sum x_i^{\alpha}$ for *every possible value* of α, requiring infinitely many components. As parts (b) and (d) show, different values of α_0 require different sample sums ($\sum x_i^{\alpha_0}$, $\sum x_i^{\alpha_0} \log x_i$), so no fixed finite-dimensional statistic can summarize the data for all (α, β).

Therefore, while the score and Hessian provide useful *local* approximations at a specific θ_0, we cannot reduce the full sample $\mathbf{X} = (x_1, \ldots, x_n)$ to a fixed low-dimensional sufficient statistic that captures all information about (α, β) for arbitrary parameter values.

[1] Relate your argument to the Pitman–Koopman–Darmois theorem (look it up!) and the non-exponential-family structure of the two-parameter Weibull.