STAT 855. Lifetime Data Analysis

X. Joan Hu

Department of Statistics and Actuarial Science
Simon Fraser University

Spring 2022

What to do today?

I.2. Review of Likelihood Methods

Likelihood methods are inference procedures based on the likelihood function with a population model.

- Formulation.
  - Consider a population under consideration, quantified by a random variable Y.
  - Assume $Y \sim f(y|\theta)$, where θ is the population characteristics of interest.
    - If Y is discrete, $f(y|\theta) = P(Y = y|\theta)$, the pmf.
    - If Y is continuous, $f(y|\theta)$ is the pdf with $P(Y \leq y) = \int_{-\infty}^{y} f(u|\theta)du$.
- Sample data.
  - Collect n iid (independent and identically distributed) observations on Y: $y_1, \ldots, y_n$.
- Likelihood function of θ based on the sample data:

$$L(\theta|\mathbf{y}) = \prod_{i=1}^{n} f(y_i|\theta)$$

A measure of the possibility to have the data with the parameter θ.

I.2. Review of Likelihood Methods

Maximum Likelihood Estimation. (MLE)
The estimator $\hat{\theta}$ is the MLE of θ if

$$L(\hat{\theta}|\mathbf{y}) = \max_{\theta} L(\theta|\mathbf{y})$$

- rationale
- implementation
  - Often to obtain $\hat{\theta}$ by solving

$$\frac{\partial}{\partial\theta} \log L(\theta|\mathbf{y}) = 0.$$

  - Computational algorithms: optimization or solving equation e.g. Newton-Raphson algorithm (eg *R: nlm()*)

Part I.2 Review of Likelihood Based Approaches

likelihood related quantities:

- score function $U(\theta)$
- Fisher information $I(\theta)$

I.2. Review of Likelihood Methods

Properties of MLE

- Invariance property
  If $\phi = h(\theta)$, $\hat{\phi} = h(\hat{\theta})$ is the MLE of ϕ.

- Large sample properties (when $n >> 1$): provided with some regularity conditions,
  - MLE is consistent,
  - MLE has the asymptotic normality: $\hat{\theta} \sim N(\theta, AV_n(\theta))$ approximately
  - MLE is asymptotically most efficient,
  - MLE's variance can be estimated by

$$\hat{V}(\hat{\theta}) = -\Big(\frac{\partial^2 \log L(\theta|\mathbf{y})}{\partial \theta^2}\Big)^{-1}\Big|_{\theta=\hat{\theta}}$$

I.2. Review of Likelihood Methods

Inference with MLE when $n >> 1$

- Testing with MLE: $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$
  - Wald test:
    $Z = \frac{\hat{\theta}-\theta_0}{\sqrt{\hat{V}(\hat{\theta})}} \sim N(0,1)$ approximately under H_0
  - Score test:
    $U(\theta) = \frac{\partial \log L(\theta|\mathbf{y})}{\partial \theta}$ and $\frac{U(\theta_0)}{\sqrt{Var\big(U(\theta_0)\big)}} \sim N(0,1)$ approximately under H_0
  - Likelihood ratio test:
    $\Lambda = \frac{L(\theta_0|\mathbf{y})}{L(\hat{\theta}|\mathbf{y})}$ and $-2\log\Lambda \sim \chi^2(1)$ approxmately under H_0

I.2. Review of Likelihood Methods

Inference with MLE when $n >> 1$

- Confidence interval with MLE:
  - Wald type: CI with level of $1-\alpha$

$$\hat{\theta} \pm Z_{1-\alpha/2}\sqrt{\hat{V}(\hat{\theta})}$$

  - Score based: CI with level of $1-\alpha$

$$\left\{\theta : \left|U(\theta)/\sqrt{Var\big(U(\theta)\big)}\right| \leq Z_{1-\alpha/2}\right\}$$

  - Likelihood ratio based: CI with level of $1-\alpha$

$$\left\{\theta : -2\log\big(L(\theta|\mathbf{y})/L(\hat{\theta}|\mathbf{y})\big) \leq \chi^2_{1-\alpha}(1)\right\}$$

Part I.2 Review of Likelihood Based Approaches

Consider r.v. $Y \sim f(\cdot;\theta)$, $\theta \in \Theta$

with a set of observations: $Y_1, \ldots, Y_n$

If the observations are not indpt, $L(\theta|data) =$**?**

- assume the joint distn of $(Y_1, \ldots, Y_n)$ ...

e.g. with a frailty model?

e.g. with a copula model?

Part I.2 Review of Likelihood Based Approaches

Consider r.v. $Y \sim f(\cdot;\theta)$, $\theta \in \Theta$
with a set of observations: $Y_1, \ldots, Y_n$

If some of $Y_1, \ldots, Y_n$ are not available, $L(\theta|data) =$?

- e.g. assume the available observations form a random sample?
- e.g. use other information?

Part I.2 Review of Likelihood Based Approaches

What if consider r.v. $Y \sim F(\cdot)$, unspecified $F(\cdot)$
with a set of iid observations: $Y_1, \ldots, Y_n$

How to estm $F(\cdot)$ based on $L(F)$?

- the nonparametric MLE: empirical distn?

- local likelihood estimation: local polynomial approximation?

Part I.2 Review of Likelihood Based Approaches

What if

- consider $Y|X = x \sim f(y|x;\theta)$, $\theta \in \Theta$?

  with a set of observations: $(Y_1, X_1), \ldots, (Y_n, X_n)$
  $L(\theta|data) =?$

- $f(y|x;\theta)$ is unspecified but assume $E(Y|X = x) = \mu(x;\theta)$ and $V(Y|X = x) = \sigma^2(x;\theta)$?

conditional likelihood, quasi-likelihood, pseudo-likelihood, partial likelihood, composite likelihood, ...

Part II. Parametric Inference in LIDA

Part II.1 Some Basic Concepts

Consider continuous r.v. $T \geq 0$, time to an event

- probability density function (pdf) $f(t), t \geq 0$
- cumulative distribution function (cdf) $F(t) = P(T \leq t), t \geq 0$
- survivor function $S(t) = P(T \geq t), t \geq 0$

Part II.1 Some Basic Concepts

Consider continuous r.v. $T \geq 0$, time to an event

- hazard function

$$h(t) = \lim_{\Delta t \to 0+} \frac{P\big(T \in [t, t + \Delta t) \big| T \geq t\big)}{\Delta t}$$

  interpretation, analytic properties, ...

- cumulative hazard function $H(t) = \int_0^t h(u)du, t \geq 0$

Part II.2 Commonly Used Parametric Distributions

Exponential distribution $T \sim NE(\lambda), \lambda > 0$

$$f(t;\lambda) = \lambda \exp(-\lambda t), \quad t > 0$$

(or $f(t;\theta) = \frac{1}{\theta}\exp(-t/\theta),\ t > 0$ with $\theta > 0$)

- $E(T)$
- $S(t)$
- $h(t)$
- memoryless property: $P(T > a + b | T > a) = P(T > b)$
- the central role

Part II.2 Commonly Used Parametric Distributions

Weibull distribution $T \sim Weibull(k, \theta)$:

$$f(t;\theta,k) = \frac{k}{\theta}\Big(\frac{t}{\theta}\Big)^{k-1} \exp\big(-(t/\theta)^k\big), \quad t > 0$$

with $\theta > 0$ and $k > 0$.

- $E(T) = \theta\Gamma(1 + 1/k)$, $Var(T) = \theta^2\Big[\Gamma(1+\frac{2}{k}) - \Gamma(1+\frac{1}{k})^2\Big]$
- $S(t)$
- $h(t) = \frac{k}{\theta}\Big(\frac{t}{\theta}\Big)^{k-1}$
- $T^k \sim NE(1/\theta)$

A translated Weibull distribution: $T \sim Weibull(k, \theta, \alpha)$:

$$f(t;\theta,k) = \frac{k}{\theta}\Big(\frac{t-\alpha}{\theta}\Big)^{k-1} \exp\Big\{-\Big(\frac{t-\alpha}{\theta}\Big)^k\Big\}, \quad t > \alpha$$

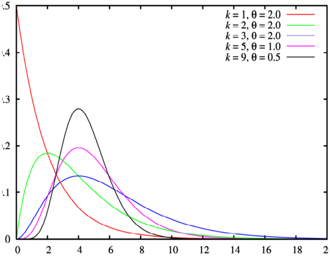
Part II.2 Commonly Used Parametric Distributions

Gamma distribution $T \sim \Gamma(k, \theta)$:

$$f(t; \theta, k) = \frac{1}{\theta\Gamma(k)}\left(\frac{t}{\theta}\right)^{k-1} \exp\big(-t/\theta\big), \quad t > 0$$

with $\theta > 0$ and $k > 0$.

- $E(T) = k\theta$, $Var(T) = k\theta^2$
- $S(t)$
- $h(t)$
- T_1, T_2 indpt and $T_j \sim \Gamma(\alpha_j, \theta)$: $T_1 + T_2 \sim \Gamma(\alpha_1 + \alpha_2, \theta)$
- $\frac{2}{\theta} T \sim \chi^2(2k)$



– *Wikimedia Commons*

Part II.2 Commonly Used Parametric Distributions

Log-normal distribution $T \sim logN(\mu, \sigma^2)$, ie $logT \sim N(\mu, \sigma^2)$:

$$T \sim \frac{1}{t\sigma\sqrt{2\pi}} \exp\big(-\frac{(logt-\mu)^2}{2\sigma^2}\big),\ \ t > 0$$

with $\sigma > 0$.

- $E(T) = \exp(\mu + \sigma^2/2)$
- $S(t)$
- $h(t)$

Part II.2 Commonly Used Parametric Distributions

- **Extreme value distribution**
- **Gumbel distribution**
- **... ...**

See books on reliability, such as Lawless (2003), for more examples of parametric models for event time

What to study next?

Part I. Preliminaries

- *I.1 Introduction*
- *I.2 Review of Likelihood Based Approaches*

Part II. Parametric Inference

- *II.1 Some basic concepts*
- *II.2 Commonly used parametric distributions*
- **II.3 Various incomplete data structures**
- **II.4 Parametric analysis with right-censored data**