Estimating Equations

Now we will consider "robustifying" inference so that misspecification does not invalidate the resulting inference.

"M-estimation" "Misspecified models".

Motivating example.

Example: Consider the $\boldsymbol{Z} = (Z_1, \ldots, Z_5)^\top$ with cdf

$$F(\boldsymbol{z}; \alpha) = \exp\left\{-\left(z_1^{-\frac{1}{\alpha}} + z_2^{-\frac{1}{\alpha}} + z_3^{-\frac{1}{\alpha}} + z_4^{-\frac{1}{\alpha}} + z_5^{-\frac{1}{\alpha}}\right)^\alpha\right\}, \quad \boldsymbol{z} \geq \boldsymbol{0}, \alpha \in (0, 1].$$

If $\alpha = 1 \longrightarrow$ independence

$\alpha \to 0$ complete dependence ($Z_i = Z_j$ w.p. 1).

Marginal:

$$P(Z_i \le z) = \exp\left(-\left(z^{-1/\alpha}\right)^\alpha\right] = \exp(-z^{-1})$$

"Unit Frechet"

Comments:

1. F is max-stable. $\rightarrow$ suitable for multivariate extreme value data w/ high dimension.

   def'n: $[F(n\underset{\sim}{z})]^n = F(\underset{\sim}{z})$

$$[F(n\underset{\sim}{z})]^n = \left(\exp\left[-\left\{(nz_1)^{-1/\alpha} + \ldots + (nz_5)^{-1/\alpha}\right\}^\alpha\right]\right)^n$$

$$= \left(\exp\left[-\left\{n^{-1/\alpha}\left(z_1^{-1/\alpha} + \ldots + z_5^{-1/\alpha}\right)\right\}^\alpha\right]\right)^n$$

$$= \left(\exp\left[-n^{-1}\left(z_1^{-1/\alpha} + \ldots + z_5^{-1/5}\right)^\alpha\right]\right)^n$$

$$= \exp\left[\left(z_1^{-1/\alpha} + \ldots + z_5^{-1/5}\right)^\alpha\right] = F(\underset{\sim}{z}) \checkmark$$

2. $Z_1, \ldots, Z_5$ are exchangeable. order doesn't matter

$$P(z_1, \ldots, z_5) = P(z_3, z_2, z_4, z_5, z_1) \text{ etc.}$$

Realistic? Maybe not.

But this gives us equal pairwise dependence, which can help us reduce the # of parameters.

And illustrate the concept of an estimating equation.

Let's consider the likelihood.

we want MLE
$\Rightarrow$ we want likelihood

Suppose we observe $\underline{Z}_i = (Z_{i1}, \dots, Z_{i5})^T$ $i = 1, \dots, n$, $\overset{iid}{\sim} F$. We want to estimate α.

$\Rightarrow$ Need to find the density, i.e. $\dfrac{\partial^5 F}{\partial z_1 \partial z_2 \cdots \partial z_5}$

$$\frac{\partial F}{\partial z_1} \overset{\text{chain rule}}{=} \exp\left[-\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha}\right] \times \left\{-\alpha\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha-1}\right\} \times \left\{-\frac{1}{\alpha} z_1^{-1/\alpha - 1}\right\}$$

$$\frac{\partial^2 F}{\partial z_1 \partial z_2} \overset{\text{product rule}}{=} \exp\left[-\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha}\right] \times \left\{-\alpha\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha-1}\right\}^2 \times \left\{-\frac{1}{\alpha} z_2^{-1/\alpha - 1}\right\} \times \left\{-\frac{1}{\alpha} z_1^{-1/\alpha - 1}\right\} +$$
$$\exp\left[-\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha}\right] \times \left\{\alpha(\alpha-1)\left(z_1^{-1/\alpha} + \dots + z_5^{-1/\alpha}\right)^{\alpha-2}\right\} \times \left\{-\frac{1}{\alpha} z_2^{-1/\alpha - 1}\right\} \times \left\{-\frac{1}{\alpha} z_1^{-1/\alpha - 1}\right\}$$

$$\frac{\partial^2 F}{\partial z_1 \partial z_2 \partial z_3} = \dots \text{ product rule on each of 2 term} \longrightarrow 4 \text{ terms}$$

by the time we get the $\dfrac{\partial^5 F}{\partial z_1 \cdots \partial z_5}$ this is gross just to write the likelihood!

" composite " or " pairwise " likelihood.

How about if we were to just use pairs of points to estimate α?

$$F_{Z_1 Z_2}(z_1, z_2) = \exp\left[-\left(z_1^{-1/\alpha} + z_2^{-1/\alpha}\right)^{\alpha}\right]$$

$$\frac{\partial^2 F}{\partial z_1 \partial z_2} = \exp\left[-\left(z_1^{-1/\alpha} + z_2^{-1/\alpha}\right)^{\alpha}\right] (z_1 z_2)^{-1/\alpha - 1} \left\{\left(\frac{1}{\alpha} - 1\right)\left(z_1^{-1/\alpha} + z_2^{-1/\alpha}\right)^{\alpha-2} + \left(z_1^{-1/\alpha} + z_2^{-1/\alpha}\right)^{2\alpha-2}\right\}$$

If we just used $(z_{1i}, z_{2i}), i = 1, \ldots, n$ would the likelihood based on the bivariate density be a good estimator for α?

Yes: unbiased

No: inefficient (not using all data).

What if we took all $\binom{5}{2} = 10$ pairs? $(z_{i1}, z_{i2}), (z_{i1}, z_{i3}), (z_{i1}, z_{i4}) \ldots$

Yes: unbiased, efficient (using all data)

No: it's not the right likelihood.