Theory and Methods of Inference

Daniele Zago

February 7, 2022

Contents

Part I
Preliminaries

0.1 Location and scale families

Def. (Location family)
A ***location family*** is a parametric family of distributions indexed by $\mu \in \mathbb{R}^p$ such that

$$p_Y(y|\mu) = p_0(y - \mu),$$

where $p_0(\cdot)$ is a given pdf. The parameter μ is called the ***location parameter***

Location families A location family can be obtained by $Y = \mu + Y_0$, where Y_0 has density $p_0(\cdot)$.

M.g.f. In a location family, if Y_0 has mgf $M_0(t) = \mathbb{E}[e^{tY_0}]$, then

$$M_Y(t|\mu) = e^{\mu t} M_0(t).$$

Sample A random sample $y_1, \ldots, y_n$ from Y has joint distribution

$$p_Y(y|\mu) = \prod_{i=1}^{n} p_0(y_i - \mu),$$

and $\overline{Y}_n$ also belongs to a location family. More generally, if $t : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$t(y_1 + a, \ldots, y_n + a) = a + t(y_1, \ldots, y_n),$$

then $t(Y_1, \ldots, Y_n)$ belongs to a location family.

Example (Location families)
Notable examples of location families are:

› $Y \sim \mathrm{U}(\vartheta, \vartheta + 1)$ with density

$$p_Y(t|\vartheta) = \mathbb{1}_{[\vartheta, \vartheta+1]}(y), \quad \vartheta \in \mathbb{R}$$

› The location family generated by $Y_0 \sim \mathrm{Exp}(1)$,

$$p_Y(y|\mu) = e^{-(y-\mu)} \cdot \mathbb{1}_{[\mu, +\infty)}.$$

› Laplace, Cauchy, and Normal distribution with fixed σ.

Def. (Scale family)
A ***scale family*** is a parametric family of distributions indexed by $\sigma \in \mathbb{R}^+$ such that

$$p_Y(y|\sigma) = \sigma^{-1} p_0(y/\sigma),$$

where $p_0(\cdot)$ is a given pdf. The parameter σ is called the ***scale parameter***

Scale family A scale family can be obtained by $Y = \sigma Y_0$, where Y_0 has density $p_0(\cdot)$.

M.g.f. In a scale family, if Y_0 has mgf $M_0(t) = \mathbb{E}[e^{tY_0}]$, then

$$M_Y(t|\sigma) = M_0(\sigma t).$$

Sample A random sample $y_1, \dots, y_n$ from Y has joint distribution

$$p_Y(y|\mu) = \sigma^{-n} \prod_{i=1}^{n} p_0(y_i/\sigma), \quad \sigma \in \mathbb{R}^+,$$

and $\overline{Y}_n$ also belongs to a scale family. More generally, if $t : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$t(by_1, \dots, by_n) = bt(y_1, \dots, y_n),$$

then $t(Y_1, \dots, Y_n)$ belongs to a scale family.
Combining the above two definition, we obtain the scale and location families, which play a major role in mathematical statistics.

Def. (Scale and location family)
A scale and location family is a parametric family of distributions such that

$$p_Y(y|\mu, \sigma) = \frac{1}{\sigma} p_0\left(\frac{y-\mu}{\sigma}\right),$$

where p_0 is a given pdf. μ is called the ***location parameter***, while σ is called the ***scale parameter***.

Location-scale family A location and scale family can be obtained by $Y = \mu_0 + \sigma Y_0$, where Y_0 has density $p_0(\cdot)$.

M.g.f. In a location and scale family, if Y_0 has mgf $M_0(t) = \mathbb{E}[e^{tY_0}]$, then

$$M_Y(t|\mu, \sigma) = e^{-\mu t} M_0(\sigma t).$$

Sample A random sample $y_1, \dots, y_n$ from Y has joint distribution

$$p_Y(y|\mu) = \sigma^{-n} \prod_{i=1}^{n} p_0\big((y_i - \mu)/\sigma\big), \quad \sigma \in \mathbb{R}^+,$$

and $\overline{Y}_n$ also belongs to a location and scale family. More generally, if $t : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$t(by_1 + a, \ldots, by_n + a) = a + bt(y_1, \ldots, y_n),$$

then $t(Y_1, \ldots, Y_n)$ belongs to a location and scale family.

Example (Notable location-scale families)

Notable examples of location and scale families include

- $Y \sim \mathrm{Unif}(\vartheta_1, \vartheta_2)$ with density

$$p_Y(y|\vartheta) = \frac{1}{\vartheta_2 - \vartheta_1} \mathbb{1}_{[\vartheta_1, \vartheta_2]}, \quad \vartheta_1 < \vartheta_2.$$

- $Y \sim \mathrm{Exp}(\lambda) + \mu$ with density

$$p_Y(y|\mu, \lambda) = \lambda e^{-\lambda(y-\mu)} \mathbb{1}_{[\mu, +\infty)}.$$

- The translated Gamma distribution with a fixed shape parameter.
- The Laplace and Normal distributions.
- $Y \sim \mathrm{Logistic}(\mu, \sigma)$ with density

$$p_Y(y|\mu, \sigma) = \frac{1}{\sigma} \frac{e^{-(y-\mu)/\sigma}}{\left[1 + e^{-(y-\mu)/\sigma}\right]^2}$$

- $Y \sim \mathrm{Cauchy}(\mu, \sigma)$ with density

$$p_Y(y|\mu, \sigma) = \frac{1}{\sigma\pi} \frac{1}{1 + \left(\frac{y-\mu}{\sigma}\right)^2}.$$

- $Y \sim \mathrm{EV}(\mu, \sigma)$ with density

$$p_Y(y|\mu, \sigma) = \frac{1}{\sigma} \exp\left\{\frac{y-\mu}{\sigma} - e^{(y-\mu)/\sigma}\right\}.$$

0.2 Exponential families

Def. (Exponential family)

An ***exponential family*** is a family of distributions for y multivariate or univariate with parameter $\vartheta \in \Theta \subseteq \mathbb{R}^p$, and density

$$p_Y(y|\vartheta) = c(\vartheta) h(y) \exp\left\{\psi(\vartheta)^\top t(y)\right\}, \tag{1}$$

where $h(\cdot) \geq 0$, $\psi(\vartheta) = \big(\psi_1(\vartheta), \ldots, \psi_k(\vartheta)\big)$ is a function of Θ with image $\operatorname{Im}\psi = \Psi \subseteq \mathbb{R}^k$. The distributions are either all discrete or all continuous.

Normalization The function $c(\vartheta)$ is the normalizing constant which depends on the value of the parameter ϑ.

Support The support of Y is the closure of $\{y \in \mathbb{R}^d : h(y) > 0\}$, hence it is the same for all elements of the family.

Identifiability For ϑ to be an identifiable parameter, the function $\psi(\vartheta)$ must be injective.

Representation Under some requirements, the family (1) is a ***minimal representation***, i.e. it involves the minimum possible number of function $\psi_j(\vartheta)$ and associated statistics $t_j(y)$. This is satisfied for instance if

1. Θ contains at least $k+1$ elements.
2. $1, \psi_1, \psi_2, \ldots, \psi_k$ are linearly independent functions.
3. $1, t_1, t_2, \ldots, t_k$ are linearly independent functions.

We call k the ***order*** of the family and $t(y) = \big(t_1(y), t_2(y), \ldots, t_n(y)\big)$ the ***canonical statistic*** of $\mathcal{F}$. The parameter $\psi = \psi(\vartheta)$ is called ***canonical parameter***.

Prop. 1 (Density of the canonical statistic)
In the canonical parametrization ψ, the canonical statistic t has density

$$P_T(t;\psi) = c\big(\vartheta(\psi)\big)\tilde{h}(t)\exp\left\{\psi^\top t\right\}, \tag{2}$$

where

$$\tilde{h}(t) = \int_{y \in S_Y : t(y) = t} \mathrm{d}H(y).$$

Def. (Regular exponential family)
If Ψ is an open set such that $\psi \in \Psi$ are all the values for which the function

$$\exp\left\{\psi^\top t\right\}\tilde{h}(t)$$

is integrable, then the exponential family with density (1) is called ***regular***.

Prop. 2 (Closure under random sampling)
If $y = (y_1, \ldots, y_n)$ is a random sample of Y_i i.i.d from an exponential family, then y is again an exponential family with density

$$p_Y(y,\vartheta) = c(\vartheta)^n \prod_{i=1}^n h(y_i)\exp\left\{\psi(\vartheta)^\top \sum_{i=1}^n t(y_i)\right\},$$

and the canonical statistic of y is $t^{(n)}(y) = \sum_{i=1}^n t(y_i)$.

Lecture 1

2022-01-27