Chapter 1 : statistical vs. real models

- Statistical models
- Quantities of interest
- Exponential families

Statistical models

For most of the course, we assume that the data is a random sample $x_1, \ldots, x_n$ and that

$$X_1, \ldots, X_n \sim F(x)$$

as i.i.d. variables or as transforms of i.i.d. variables

[observations versus Random Variables]

Motivation:

Repetition of observations increases information about F, by virtue of probabilistic limit theorems (LLN, CLT)

Statistical models

For most of the course, we assume that the data is a random sample $x_1, \ldots, x_n$ and that

$$X_1, \ldots, X_n \sim F(x)$$

as i.i.d. variables or as transforms of i.i.d. variables

Motivation:

Repetition of observations increases information about F, by virtue of probabilistic limit theorems (LLN, CLT)

Warning 1: Some aspects of F may ultimately remain unavailable

Statistical models

For most of the course, we assume that the data is a random sample $x_1, \ldots, x_n$ and that

$$X_1, \ldots, X_n \sim F(x)$$

as i.i.d. variables or as transforms of i.i.d. variables

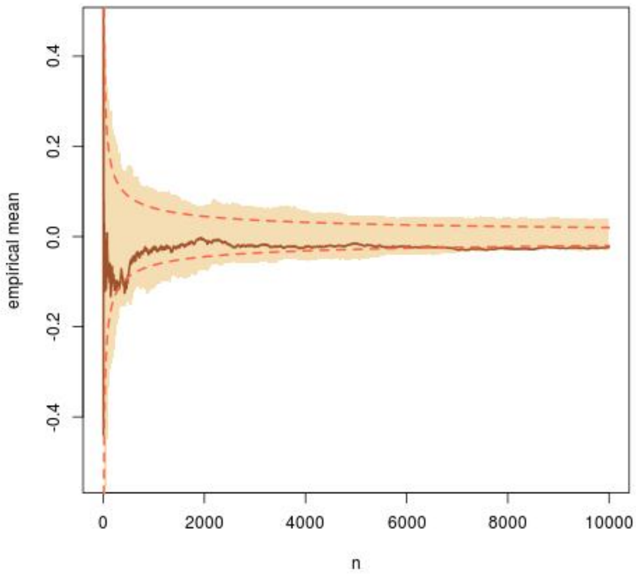
Motivation:

Repetition of observations increases information about F, by virtue of probabilistic limit theorems (LLN, CLT)

Warning 2: The model is always wrong, even though we behave as if...

Limit of averages

Case of an iid sequence $X_1, \ldots, X_n \sim \mathcal{N}(0,1)$



Evolution of the range of $\bar{X}_n$ across 1000 repetitions, along with one random sequence and the theoretical 95% range

Limit theorems

Law of Large Numbers (LLN)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{prob}} \mathbb{E}[X]$$

[proof: see Terry Tao's "What's new", 18 June 2008]

Limit theorems

Law of Large Numbers (LLN)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X]$$

[proof: see Terry Tao's "What's new", 18 June 2008]

Limit theorems

Law of Large Numbers (LLN)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{\text{a.s.}} \mathbb{E}[X]$$

Central Limit Theorem (CLT)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$ and a finite variance $\sigma^2 = \mathrm{var}(X)$,

$$\sqrt{n}\left\{\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X]\right\} \xrightarrow{\text{dist.}} N(0, \sigma^2)$$

[proof: see Terry Tao's "What's new", 5 January 2010]

Limit theorems

Central Limit Theorem (CLT)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$ and a finite variance $\sigma^2 = \mathrm{var}(X)$,

$$\sqrt{n}\left\{\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X]\right\} \xrightarrow{\text{dist.}} N(0, \sigma^2)$$

[proof: see Terry Tao's "What's new", 5 January 2010]

Continuity Theorem

If

$$X_n \xrightarrow{\text{dist.}} a$$

and g is continuous at a, then

$$g(X_n) \xrightarrow{\text{dist.}} g(a)$$

Limit theorems

Central Limit Theorem (CLT)

If $X_1, \ldots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$ and a finite variance $\sigma^2 = \mathrm{var}(X)$,

$$\sqrt{n}\left\{\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}[X]\right\} \xrightarrow{\text{dist.}} N(0, \sigma^2)$$

[proof: see Terry Tao's "What's new", 5 January 2010]

Slutsky's Theorem

If X_n, Y_n, Z_n converge in distribution to X, a, and b, respectively, then

$$X_n Y_n + Z_n \xrightarrow{\text{dist.}} aX + b$$

Limit theorems

Central Limit Theorem (CLT)

If $X_1, \dots, X_n$ are i.i.d. random variables, with a well-defined expectation $\mathbb{E}[X]$ and a finite variance $\sigma^2 = \mathrm{var}(X)$,

$$\sqrt{n}\left\{\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X]\right\} \xrightarrow{\text{dist.}} N(0, \sigma^2)$$

[proof: see Terry Tao's "What's new", 5 January 2010]

Delta method's Theorem

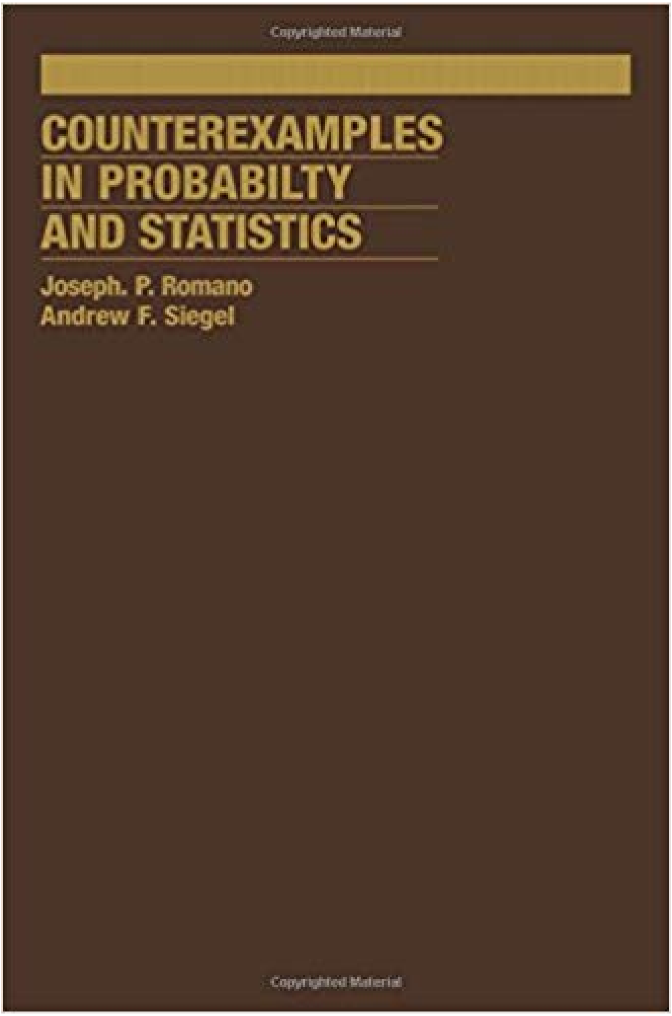
If

$$\sqrt{n}\{X_n - \mu\} \xrightarrow{\text{dist.}} N_p(0, \Omega)$$

and $g : \mathbb{R}^p \to \mathbb{R}^q$ is a continuously differentiable function on a neighbourhood of $\mu \in \mathbb{R}^p$, with a non-zero gradient $\nabla g(\mu)$, then

$$\sqrt{n}\{g(X_n) - g(\mu)\} \xrightarrow{\text{dist.}} N_q(0, \nabla g(\mu)^T \Omega \nabla g(\mu))$$

Entertaining read



Exemple 1: Binomial sample

Case # 1: Observation of i.i.d. Bernoulli variables

$$X_i \sim \mathcal{B}(p)$$

with unknown parameter p (e.g., opinion poll)
Case # 2: Observation of independent Bernoulli variables

$$X_i \sim \mathcal{B}(p_i)$$

with unknown and different parameters p_i (e.g., opinion poll, flu epidemics)
Transform of i.i.d. $U_1, \ldots, U_n$:

$$X_i = \mathbb{I}(U_i \leqslant p_i)$$

Exemple 1: Binomial sample

Case # 1: Observation of i.i.d. Bernoulli variables

$$X_i \sim \mathcal{B}(p)$$

with unknown parameter p (e.g., opinion poll)
Case # 2: Observation of conditionally independent Bernoulli variables

$$X_i | z_i \sim \mathcal{B}(p(z_i))$$

with covariate-driven parameters $p(z_i)$ (e.g., opinion poll, flu epidemics)
Transform of i.i.d. $U_1, \ldots, U_n$:

$$X_i = \mathbb{I}(U_i \leqslant p_i)$$

Parametric versus non-parametric

Two classes of statistical models:

- Parametric when F varies within a family of distributions indexed by a parameter θ that belongs to a finite dimension space Θ:

  $$F \in \{F_\theta,\ \theta \in \Theta\}$$

  and to "know" F is to know which θ it corresponds to (identifiability);
- Non-parametric all other cases, i.e. when F is not constrained in a parametric way or when only some aspects of F are of interest for inference

Trivia: Machine-learning does not draw such a strict distinction between classes

Parametric versus non-parametric

Two classes of statistical models:

- Parametric when F varies within a family of distributions indexed by a parameter θ that belongs to a finite dimension space Θ:

$$F \in \{F_\theta,\ \theta \in \Theta\}$$

  and to "know" F is to know which θ it corresponds to (identifiability);
- Non-parametric all other cases, i.e. when F is not constrained in a parametric way or when only some aspects of F are of interest for inference

Trivia: Machine-learning does not draw such a strict distinction between classes

Non-parametric models

In non-parametric models, there may still be constraints on the range of F's as for instance

$$\mathbb{E}_F[Y|X = x] = \Psi(\beta^T x), \ \text{var}_F(Y|X = x) = \sigma^2$$

in which case the statistical inference only deals with estimating or testing the constrained aspects or providing prediction.
Note: Estimating a density or a regression function like $\Psi(\beta^T x)$ is only of interest in a restricted number of cases

Parametric models

When $F = F_\theta$, inference usually covers the whole of the parameter θ and provides

- point estimates of θ, i.e. values substituting for the unknown "true" θ
- confidence intervals (or regions) on θ as regions likely to contain the "true" θ
- testing specific features of θ (true or not?) or of the whole family (goodness-of-fit)
- predicting some other variable whose distribution depends on θ

$$z_1, \ldots, z_m \sim G_\theta(z)$$

Inference: all those procedures depend on the sample $(x_1, \ldots, x_n)$

Parametric models

When $F = F_\theta$, inference usually covers the whole of the parameter θ and provides

- point estimates of θ, i.e. values substituting for the unknown "true" θ
- confidence intervals (or regions) on θ as regions likely to contain the "true" θ
- testing specific features of θ (true or not?) or of the whole family (goodness-of-fit)
- predicting some other variable whose distribution depends on θ

$$z_1, \ldots, z_m \sim G_\theta(z)$$

Inference: all those procedures depend on the sample $(x_1, \ldots, x_n)$

Example 1: Binomial experiment again

Model: Observation of i.i.d. Bernoulli variables

$$X_i \sim \mathcal{B}(p)$$

with unknown parameter p (e.g., opinion poll)
Questions of interest:

1. likely value of p or range thereof
2. whether or not p exceeds a level p_0
3. how many more observations are needed to get an estimation of p precise within two decimals
4. what is the average length of a "lucky streak" (1's in a row)

Exemple 2: Normal sample

Model: Observation of i.i.d. Normal variates

$$X_i \sim \mathrm{N}(\mu, \sigma^2)$$

with unknown parameters μ and $\sigma > 0$ (e.g., blood pressure)

Questions of interest:

1. likely value of μ or range thereof
2. whether or not μ is above the mean η of another sample $y_1, \ldots, y_m$
3. percentage of extreme values in the next batch of m x_i's
4. how many more observations to exclude $\mu = 0$ from likely values
5. which of the x_i's are outliers

Quantities of interest

Statistical distributions (incompletely) characterised by (1-D) moments:

- central moments

$$\mu_1 = \mathbb{E}\left[X\right] = \int x dF(x) \quad \mu_k = \mathbb{E}\left[(X - \mu_1)^k\right] \;\; k > 1$$

- non-central moments

$$\xi_k = \mathbb{E}\left[X^k\right] \;\; k \geqslant 1$$

- α quantile

$$\mathbb{P}(X < \zeta_\alpha) = \alpha$$

and (2-D) moments

$$\mathrm{cov}(X^i, X^j) = \int (x^i - \mathbb{E}[X^i])(x^j - \mathbb{E}[X^j]) dF(x^i, x^j)$$

Note: For parametric models, those quantities are transforms of the parameter θ

Example 1: Binomial experiment again

Model: Observation of i.i.d. Bernoulli variables

$$X_i \sim \mathcal{B}(p)$$

Single parameter p with

$$\mathbb{E}[X] = p \quad \mathrm{var}(X) = p(1-p)$$

[somewhat boring...]

Median and mode

Example 1: Binomial experiment again

Model: Observation of i.i.d. Binomial variables

$$X_i \sim \mathcal{B}(n, p) \ \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Single parameter p with

$$\mathbb{E}[X] = np \ \mathrm{var}(X) = np(1-p)$$

[somewhat less boring!]

Median and mode

Example 2: Normal experiment again

Model: Observation of i.i.d. Normal variates

$$X_i \sim N(\mu, \sigma^2) \quad i = 1, \ldots, n\,,$$

with unknown parameters μ and $\sigma > 0$ (e.g., blood pressure)

$$\mu_1 = \mathbb{E}[X] = \mu \;\; \mathrm{var}(X) = \sigma^2 \;\; \mu_3 = 0 \;\; \mu_4 = 3\sigma^4$$

Median and mode equal to μ

Exponential families

Class of parametric densities with nice analytic properties

Start from the normal density:

$$\begin{aligned}\varphi(x;\theta) &= \frac{1}{\sqrt{2\pi}} \exp\left\{x\theta - {}^{x^2}\!/_2 - {}^{\theta^2}\!/_2\right\} \\ &= \frac{\exp\{-\theta^2/2\}}{\sqrt{2\pi}} \underbrace{\exp\{x\theta\}}_{x \text{ meets } \theta} \exp\left\{-x^2/2\right\}\end{aligned}$$

where θ and x only interact through single exponential product

Exponential families

Class of parametric densities with nice analytic properties

Definition

A parametric family of distributions on $\mathcal{X}$ is an exponential family if its density with respect to a measure ν satisfies

$$f(x|\theta) = c(\theta)h(x)\underbrace{\exp\{T(x)^{\mathrm{T}}\tau(\theta)\}}_{\text{scalar product}}, \theta \in \Theta,$$

where $T(\cdot)$ and $\tau(\cdot)$ are k-dimensional functions and $c(\cdot)$ and $h(\cdot)$ are positive unidimensional functions.

Function $c(\cdot)$ is redundant, being defined by normalising constraint:

$$c(\theta)^{-1} = \int_{\mathcal{X}} h(x)\exp\{T(x)^{\mathrm{T}}\tau(\theta)\}d\nu(x)$$

Exponential families (examples)

Example 1: Binomial experiment again

Binomial variable

$$X \sim \mathcal{B}(n, p) \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

can be expressed as

$$\mathbb{P}(X = k) = (1-p)^n \binom{n}{k} \exp\{k \log(p/(1-p))\}$$

hence

$$c(p) = (1-p)^n\,,\ h(x) = \binom{n}{x}\,,\ T(x) = x\,,\ \tau(p) = \log(p/(1-p))$$

Exponential families (examples)

Example 1: Binomial experiment again

Binomial variable

$$X \sim \mathcal{B}(n, p) \quad \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

can be expressed as

$$\mathbb{P}(X = k) = (1-p)^n \binom{n}{k} \exp\{k \log(p/(1-p))\}$$

hence

$$c(p) = (1-p)^n\,,\ h(x) = \binom{n}{x}\,,\ T(x) = x\,,\ \tau(p) = \log(p/(1-p))$$

Exponential families (examples)

Example 2: Normal experiment again

Normal variate

$$X \sim \mathcal{N}(\mu, \sigma^2)$$

with parameter $\theta = (\mu, \sigma^2)$ and density

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-(x-\mu)^2/2\sigma^2\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-x^2/2\sigma^2 + x\mu/\sigma^2 - \mu^2/2\sigma^2\} \\ &= \frac{\exp\{-\mu^2/2\sigma^2\}}{\sqrt{2\pi\sigma^2}} \exp\{-x^2/2\sigma^2 + x\mu/\sigma^2\} \end{aligned}$$

hence

$$c(\theta) = \frac{\exp\{-\mu^2/2\sigma^2\}}{\sqrt{2\pi\sigma^2}}, \quad T(x) = \begin{pmatrix} x^2 \\ x \end{pmatrix}, \quad \tau(\theta) = \begin{pmatrix} -1/2\sigma^2 \\ \mu/\sigma^2 \end{pmatrix}$$

natural exponential families

reparameterisation induced by the shape of the density:

Definition

In an exponential family, the natural parameter is $\tau(\theta)$ and the natural parameter space is

$$\Theta = \left\{ \tau \in \mathbb{R}^k ; \int_{\mathcal{X}} h(x) \exp\{T(x)^{\mathrm{T}} \tau\} d\nu(x) < \infty \right\}$$

Example For the $\mathcal{B}(m, p)$ distribution, the natural parameter is

$$\theta = \log\{p/(1-p)\}$$

and the natural parameter space is $\mathbb{R}$

natural exponential families

reparameterisation induced by the shape of the density:

Definition

In an exponential family, the natural parameter is $\tau(\theta)$ and the natural parameter space is

$$\Theta = \left\{ \tau \in \mathbb{R}^k ; \int_{\mathcal{X}} h(x) \exp\{T(x)^{\mathrm{T}} \tau\} d\nu(x) < \infty \right\}$$

Example For the $\mathcal{B}(m, p)$ distribution, the natural parameter is

$$\theta = \log\{p/(1-p)\}$$

and the natural parameter space is $\mathbb{R}$

regular and minimal exponential families

Possible to add and (better!) delete useless components of T:

Definition

A regular exponential family corresponds to the case where Θ is an open set.
A minimal exponential family corresponds to the case when the $T_i(X)$'s are linearly independent, i.e.

$$\mathbb{P}_\theta(\alpha^T T(X) = \text{const.}) = 0 \quad \text{for } \alpha \neq 0 \quad \theta \in \Theta$$

Also called non-degenerate exponential family
Usual assumption when working with exponential families

regular and minimal exponential families

Possible to add and (better!) delete useless components of T:

Definition

A regular exponential family corresponds to the case where Θ is an open set.
A minimal exponential family corresponds to the case when the $T_i(X)$'s are linearly independent, i.e.

$$\mathbb{P}_\theta(\alpha^T T(X) = \text{const.}) = 0 \quad \text{for } \alpha \neq 0 \quad \theta \in \Theta$$

Also called non-degenerate exponential family

Usual assumption when working with exponential families

Illustrations

- For a Normal $\mathcal{N}(\mu, \sigma^2)$ distribution,

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} \exp\{-{}^{x^2}/_{2\sigma^2} + {}^{\mu}/_{\sigma^2}\, x - {}^{\mu^2}/_{2\sigma^2}\}$$

  means this is a two-dimensional minimal exponential family
- For a fourth-power distribution

$$f(x|\mu) = C(\theta) \exp\{-(x-\theta)^4\}\} \propto e^{-x^4}\, e^{4\theta^3 x - 6\theta^2 x^2 + 4\theta x^3 - \theta^4}$$

  implies this is a three-dimensional minimal exponential family

[Exercise: find C]

convexity properties

Highly regular densities

Theorem

The natural parameter space Θ of an exponential family is convex and the inverse normalising constant $c^{-1}(\theta)$ is a convex function.

Example For $\mathcal{B}(n, p)$, the natural parameter space is $\mathbb{R}$ and the inverse normalising constant $(1 + \exp(\theta))^n$ is convex

convexity properties

Highly regular densities

Theorem

The natural parameter space Θ of an exponential family is convex and the inverse normalising constant $c^{-1}(\theta)$ is a convex function.

Example For $\mathcal{B}(n, p)$, the natural parameter space is $\mathbb{R}$ and the inverse normalising constant $(1 + \exp(\theta))^n$ is convex

analytic properties

Lemma

If the density of X has the minimal representation

$$f(x|\theta) = c(\theta)h(x)\exp\{T(x)^{\mathrm{T}}\theta\}$$

then the natural statistic $Z = T(X)$ is also distributed from an exponential family and there exists a measure ν_T such that the density of Z $[= T(X)]$ against ν_T is

$$f(z;\theta) = c(\theta)\exp\{z^{\mathrm{T}}\theta\}$$

analytic properties

Theorem

If the density of $Z = T(X)$ against ν_T is $c(\theta)\exp\{z^T\theta\}$, if the real value function φ is measurable, with

$$\int |\varphi(z)| \exp\{z^T\theta\}\, d\nu_T(z) < \infty$$

on the interior of Θ, then

$$f: \ \theta \to \int \varphi(z) \exp\{z^T\theta\}\, d\nu_T(z)$$

is an analytic function on the interior of Θ and

$$\nabla f(\theta) = \int z\varphi(z) \exp\{z^T\theta\}\, d\nu_T(z)$$

moments of exponential families

Normalising constant $c(\cdot)$ generating all moments

Proposition

If $T(\cdot): \mathcal{X} \rightarrow \mathbb{R}^d$ and the density of $Z = T(X)$ is $\exp\{z^T\theta - \psi(\theta)\}$, then

$$\mathbb{E}_\theta\left[\exp\{T(x)^T u\}\right] = \exp\{\psi(\theta + u) - \psi(\theta)\}$$

and $\psi(\cdot)$ is the cumulant generating function.

[Laplace transform]

moments of exponential families

Normalising constant $c(\cdot)$ generating all moments

Proposition

If $T(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^d$ and the density of $Z = T(X)$ is $\exp\{z^T\theta - \psi(\theta)\}$, then

$$\mathbb{E}_\theta[T_i(X)] = \frac{\partial\psi(\theta)}{\partial\theta_i} \quad i = 1, \ldots, d,$$

and

$$\mathbb{E}_\theta\left[T_i(X)\,T_j(X)\right] = \frac{\partial^2\psi(\theta)}{\partial\theta_i\partial\theta_j} \quad i, j = 1, \ldots, d$$

Sort of integration by part in parameter space:

$$\int \left\{T_i(x) + \frac{\partial}{\partial\theta_i}\log c(\theta)\right\} c(\theta)h(x)\exp\{T(x)^T\theta\}d\nu(x) = \frac{\partial}{\partial\theta_i}1 = 0$$

Sample from exponential families

Take an exponential family

$$f(x|\theta) = h(x)\, \exp\left\{\tau(\theta)^{\mathrm{T}} T(x) - \psi(\theta)\right\}$$

and id sample $x_1, \ldots, x_n$ from $f(x|\theta)$.

Then

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} h(x_i)\, \exp\left\{\tau(\theta)^{\mathrm{T}} \sum_{i=1}^{n} T(x_i) - n\psi(\theta)\right\}$$

Remark

For an exponential family with summary statistic $T(\cdot)$, the statistic

$$S(X_1, \ldots, X_n) = \sum_{i=1}^{n} T(X_i)$$

is sufficient for describing the joint density

Sample from exponential families

Take an exponential family

$$f(x|\theta) = h(x)\ \exp\left\{\tau(\theta)^{\mathrm{T}} T(x) - \psi(\theta)\right\}$$

and id sample $x_1, \ldots, x_n$ from $f(x|\theta)$.
Then

$$f(x_1, \ldots, x_n|\theta) = \prod_{i=1}^{n} h(x_i)\ \exp\left\{\tau(\theta)^{\mathrm{T}} \sum_{i=1}^{n} T(x_i) - n\psi(\theta)\right\}$$

Remark

For an exponential family with summary statistic $T(\cdot)$, the statistic

$$S(X_1, \ldots, X_n) = \sum_{i=1}^{n} T(X_i)$$

is sufficient for describing the joint density

connected examples of exponential families

Example

Chi-square χ^2_k distribution corresponding to distribution of $X_1^2 + \ldots + X_k^2$ when $X_i \sim \mathcal{N}(0, 1)$, with density

$$f_k(z) = \frac{z^{k/2-1} \exp\{-z/2\}}{2^{k/2}\Gamma(k/2)} \; z \in \mathbb{R}_+$$

connected examples of exponential families

Counter-Example

Non-central chi-square $\chi^2_k(\lambda)$ distribution corresponding to distribution of $X_1^2 + \ldots + X_k^2$ when $X_i \sim \mathcal{N}(\mu, 1)$, with density

$$f_{k,\lambda}(z) = {}^1\!/_2 \left({}^z\!/_\lambda\right)^{k/4-1/2} \exp\{-(z+\lambda)/2\} I_{k/2-1}(\sqrt{z\lambda}) \;\; z \in \mathbb{R}_+$$

where $\lambda = k\mu^2$ and I_ν Bessel function of second order

connected examples of exponential families

Counter-Example

Fisher $\mathcal{F}_{n,m}$ distribution
corresponding to the ratio

$$Z = \frac{Y_n/n}{Y_m/m} \quad Y_n \sim \chi^2_n,\ Y_m \sim \chi^2_m\,,$$

with density

$$f_{m,n}(z) = \frac{(n/m)^{n/2}}{B(n/2, m/2)} z^{n/2-1} \left(1 + n/m z\right)^{-n+m/2} \quad z \in \mathbb{R}_+$$

connected examples of exponential families

Example

Ising $\mathcal{B}e(n/2, m/2)$ distribution corresponding to the distribution of

$$Z = \frac{nY}{nY + m} \text{ when } Y \sim \mathcal{F}_{n,m}$$

has density

$$f_{m,n}(z) = \frac{1}{B(n/2, m/2)} z^{n/2-1} (1-z)^{m/2-1} \quad z \in (0,1)$$

connected examples of exponential families

Counter-Example

Laplace double-exponential $\mathcal{L}(\mu, \sigma)$ distribution corresponding to the rescaled difference of two exponential $\mathcal{E}(\sigma^{-1})$ random variables,

$$Z = \mu + X_1 - X_2 \text{ when } X_1, X_2 \overset{\sim}{\text{iid}} \mathcal{E}(\sigma^{-1})$$

has density

$$f(z; \mu, \sigma) = \frac{1}{\sigma} \exp\{-\sigma^{-1}|x - \mu|\}$$