1.5 Exponential family of distributions

Exponential family of distributions

Definition:
Let Y be a random variable with values in $\mathbb{R}$ (or $\mathbb{N}_0$ or $\mathbb{Z}$),
and suppose its probability distribution $f(y;\theta)$ depends on a single parameter θ. Then the distribution belongs to the *exponential family* if it admits the form

$$f(y;\theta) = s(y)t(\theta)e^{a(y)b(\theta)}$$

for some functions a, b, s, and t.

Another way of writing this is

$$f(y;\theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)] \tag{1.5.1}$$

where $s(y) = e^{d(y)}$ and $t(\theta) = e^{c(\theta)}$.

- If $a(y) = y$, the distribution of Y is called **canonical**
- If the distribution is in its canonical form, $b(\theta)$ is called the **natural parameter** of the distribution.
- If there are other parameters in addition to θ, they are usually called **nuisance parameters**, and they can be functions of a, b, c and d.

Notice that many well-known distributions belong to the exponential family.

Binomial distribution

Suppose $\mathrm{Y} \sim Bin(n, p)$. The *probability of getting exactly y successes in n trials* is given by the **probability mass function**:

$$f(y|p) = \binom{n}{y} p^y (1-p)^{n-y},$$

for $y \in \{0, 1, \ldots, n\}$.
Assume that *p is the parameter of interest* which indicates the probability of success in a single experiment, and that we know n.

> **Think:** Does the Binomial belong to the Exponential family of distributions?

The probability function can be rewritten as

$$f(y|p) = \exp\left[y \log p - y \log(1-p) + n \log(1-p) + \log\binom{n}{y}\right]$$

this is of the form in $(1.5.1)$ with

$$a(y) = y, \quad b(p) = \log\frac{p}{1-p}, \quad c(p) = n\log(1-p), \quad d(y) = \log\binom{n}{y}. \qquad (1.5.2)$$

The binomial distribution is used as a model for observations of a process with binary outcomes:

1. The number of candidates who pass a test - the possible outcomes for each candidate being to pass or to fail;
2. The number of patients with some disease who are alive at a specified time since diagnosis - the possible outcomes being survival or death.

> **Practice:** Draw the probability mass function of a Binomial random variable with $n = 40$ and $p = 0.5$ in R.

Practice using R

Draw the probability mass function of a Binomial random variable with $n = 40$ and $p = 0.5$ in `R`.

Normal distribution

A random variable Y with values in $\mathbb{R}$ is **normal**, i.e. $\mathrm{Y} \sim N(\mu, \sigma^2)$, if it has **probability density function**

$$f(y|\mu) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right).$$

Think: Does the Normal belong to the Exponential family of distributions?

The pdf can be rewritten in exponential family form:

$$f(y|\mu) = \exp\left[-\frac{y^2}{2\sigma^2} + \frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right]$$

where

$$a(y) = y, \quad b(\mu) = \mu/\sigma^2, \quad c(\mu) = -\frac{\mu^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2), \quad d(y) = -\frac{y^2}{2\sigma^2} \tag{1.5.3}$$

(alternatively, the term $-\frac{1}{2}\log(2\pi\sigma^2)$ can be included in $d(y)$.

The normal distribution is used to model *continuous data with symmetric distribution*. The normal distribution is common in applications, since:

1. Many phenomena are well described by the normal distribution, ex. height or blood pressure of people;
2. Even if data are not normal, the average or total of a random sample of values will be approximately normally distributed (Central Limit Theorem);
3. Statistical theory is developed in a large extent for the Normal distribution.

Practice: Draw the probability density function of a Normal random variable with $\mu = 4$ and $\sigma = 1.5$ in R.

Practice using R

Draw the probability density function of a Normal random variable with $\mu = 4$ and $\sigma = 1.5$ in $\mathbf{R}$.

Poisson distribution

A random variable Y with positive values is **Poisson** with mean $\lambda > 0$, i.e. $\mathrm{Y} \sim \mathrm{Pois}(\lambda)$, if it has **probability density function**,

$$f(y|\lambda) = \frac{\lambda^y}{e^\lambda y!}.$$

Think: Does the Normal belong to the Exponential family of distributions?

The pdf can be rewritten in exponential family form:

$$f(y|\lambda) = \exp(y \log \lambda - \lambda - \log y!)$$

where

$$a(y) = y, \quad b(\lambda) = \log \lambda, \quad c(\lambda) = -\lambda, \quad d(y) = -\log y!. \tag{1.5.4}$$

The Poisson distribution provides a suitable model for **count data** and expresses *the probability of a given number of events occurring in a fixed interval of time and/or space* if these events occur with a known average rate and independently of the time since the last event.

1. The number of medical conditions reported by a person;
2. The number of tropical cyclones during a season;
3. The number of spelling mistakes on the page of a newspaper;
4. The number of faulty components in a computer or in a batch of manufactured items.

Practice: Draw the probability density function of a Poisson random variable with $\lambda = 1$ in R.

Practice using R

Draw the probability density function of a Poisson random variable with $\lambda = 1$ in `R`.

Properties of distributions in the exponential family

The *expected value* and *variance* of $a(\mathrm{Y})$ are given by

$$\mathbb{E}[a(\mathrm{Y})] = -\frac{c'(\theta)}{b'(\theta)} \tag{1.5.5}$$

and

$$\mathbb{V}ar[a(\mathrm{Y})] = \frac{b''(\theta)c'(\theta) - c''(\theta)b'(\theta)}{b'(\theta)^3} \tag{1.5.6}$$

You may want to see how these results were derived.

Score and information

We now formally introduce the statistics **score** and **information**.

The **log-likelihood function** for the exponential family is

$$\ell(\theta; y) = a(y)b(\theta) + c(\theta) + d(y).$$

The **score statistic** is defined as

$$\mathrm{U}(\theta; y) = \frac{d\ell(\theta; y)}{d\theta} = a(y)b'(\theta) + c'(\theta).$$

It depends on y and may hence be interpreted as a random variable:

$$\mathrm{U} := \mathrm{U}(\theta; Y) = a(\mathrm{Y})b'(\theta) + c'(\theta).$$

The *expected value of the score statistic* U is

$$\mathbb{E}(\mathrm{U}) = b'(\theta)\mathbb{E}[a(\mathrm{Y})] + c'(\theta) = 0 \tag{1.5.7}$$

because, if we use $(1.5.5)$,

$$\mathbb{E}(\mathrm{U}) = b'(\theta)\left[-\frac{c'(\theta)}{b'(\theta)}\right] + c'(\theta) = 0$$

The *variance of the score statistic* U is called the **information**, denoted by $\mathcal{I}$:

$$\mathcal{I} = \mathbb{V}ar(\mathrm{U}) = b'(\theta)^2\mathbb{V}ar[a(\mathrm{Y})] = \frac{b''(\theta)c'(\theta)}{b'(\theta)} - c''(\theta). \tag{1.5.8}$$

Another property of the score function U is

$$\mathbb{E}(\mathrm{U}^2) = \mathbb{V}\mathrm{ar}(\mathrm{U}) = -\mathbb{E}\left(\frac{d\mathrm{U}}{d\theta}\right).$$

The first equality follows from the general result $\mathbb{V}\mathrm{ar}(\mathrm{X}) = \mathbb{E}(\mathrm{X}^2) - [\mathbb{E}(\mathrm{X})]^2$, valid for any r.v., and $\mathbb{E}(\mathrm{U}) = 0$.

To see the second equality, note that

$$
\begin{aligned}
\mathbb{E}\left(\frac{d\mathrm{U}}{d\theta}\right) &= \mathbb{E}\left(a(\mathrm{Y})b''(\theta) + c''(\theta)\right) = b''(\theta)\mathbb{E}[a(\mathrm{Y})] + c''(\theta) \\
&= b''(\theta)\left[-\frac{c'(\theta)}{b'(\theta)}\right] + c''(\theta) = -\mathbb{V}\mathrm{ar}(\mathrm{U}) = -\mathcal{I}
\end{aligned}
$$

Pressure example

Let's consider back the example with the times to failure of Kevlar epoxy strand pressure.

The **Weibull distribution** belongs to the exponential family since its pdf

$$f(y;\lambda,\theta)=\frac{\lambda y^{\lambda-1}}{\theta^{\lambda}}\exp\left[-\left(\frac{y}{\theta}\right)^{\lambda}\right]$$

can be rewritten as

$$f(y;\theta)=\exp\left[\log\lambda+(\lambda-1)\log y-\lambda\log\theta-\left(\frac{y}{\theta}\right)^{\lambda}\right]$$

with

$$a(y)=y^{\lambda},\quad b(\theta)=-\theta^{-\lambda},\quad c(\theta)=\log\lambda-\lambda\log\theta,\quad d(y)=(\lambda-1)\log y$$

where λ is a nuisance parameter.

Check your understanding

This is a non-assessed self-practice. Work through each of the questions and press submit to be able to see the solution.

Question 1 *Submitted Jan 22nd 2024 at 8:57:25 am*

[See slide: Exponential family of distributions]

Use $\mathbf{R}$ to complete the following tasks:

1. Draw the probability mass function of a Binomial random variable with $n = 40$ and $p = 0.5$.
2. Draw the probability density function of a Normal random variable with $\mu = 4$ and $\sigma = 1.5$.
3. Draw the probability density function of a Poisson random variable with $\lambda = 1$.

```
# Question 1
# Probaility mass function of a Binomial random variable with n=40 and p=0.5
xseq <- 0:40
plot(xseq, dbinom(xseq, size=40, prob=0.5), ylab=expression(f(x)), pch=16)

# Probaility density function of a Normal random variable with mu=4 and sigma=1.5
xseq <- seq(from=-10, to=10, length=100)
plot(xseq, dnorm(xseq, mean=4, sd=1.5), type="l", ylab=expression(f(x)))

# Probaility density function of a Poisson random variable with mu=4 and sigma=1.5
xseq <- 0:10
plot(xseq, dpois(xseq, lambda=1), ylab=expression(f(x)), pch=16)
```

Question 2 *Submitted Jan 22nd 2024 at 9:00:05 am*

[Dobson and Barnett (2018, Exercise 3.2)]

If the random variable Y has the Gamma distribution with scale parameter β, which is the parameter of interest, and a known shape parameter α, then its probability density function is:

$$f(y;\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta}.$$

Show that this distribution belongs to the exponential family and find the natural parameter. Find also $\mathbb{E}(\mathrm{Y})$ and $\mathbb{V}\mathrm{ar}(\mathrm{Y})$.

Solution

We can re-write the pdf as:

$$\begin{aligned} f(y;\beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-y\beta} \\ &= \exp\left\{\log\left(\frac{\beta^\alpha}{\Gamma(\alpha)} y^{\alpha-1}\right) - y\beta\right\} \\ &= \exp\left\{\alpha\log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log(y) - y\beta\right\}, \end{aligned}$$

so the Gamma distribution does belong to the exponential family with: $a(y) = y$, natural parameter $b(\beta) = -\beta$, $c(\beta) = \alpha\log(\beta)$ and $d(y) = (\alpha-1)\log(y) - \log(\Gamma(\alpha))$.

As a consequence we have:

$$\begin{aligned} \mathbb{E}(a(\mathrm{Y})) &= \mathbb{E}(\mathrm{Y}) \\ &= -\frac{c'(\beta)}{b'(\beta)} \\ &= \frac{\alpha}{\beta} \end{aligned}$$

and

$$\begin{aligned} \mathbb{V}\mathrm{ar}(a(\mathrm{Y})) &= \mathbb{V}\mathrm{ar}(\mathrm{Y}) \\ &= \frac{b''(\beta)c'(\beta) - c''(\beta)b'(\theta)}{b'(\beta)^3} \\ &= \frac{\alpha}{\beta^2}. \end{aligned}$$