

## Chapter 2

# The Statistical Approach

### 2.1 The Setup

Assume that we have observed data  $D = x$  which was the result of a random experiment  $X$  (or can be approximated as such). The data are then modelled using

1. A sample space,  $\mathcal{X}$  for the observed value of  $x$
2. A probability density function for  $X$  at  $x$ ,  $f(x; \theta)$
3. A parameter space for  $\theta$ ,  $\Theta$

The **inference problem** is to use  $x$  to infer properties of  $\theta$ .

### 2.2 Approaches to Statistical Inference

The major approaches to statistical inference are:

1. **Frequentist** or classical
2. **Bayesian**
3. **Likelihood**

### 2.3 Types of Statistical Inference

There are four major statistical inferences:

1. **Estimation:** Select one value of  $\theta$ , the estimate, to be reported. Some measure of reliability is assumed to be reported as well.

2. **Testing:** Compare two values (or sets of values) of  $\theta$  and choose one of them as better.
3. **Interval Estimation:** Select a region of  $\theta$  values as being consistent, in some sense, with the observed data.
4. **Prediction:** Use the observed data to predict a new result of the experiment.

Note that the first three inferences can be defined as functions from the sample space to subsets of the parameter space. Thus estimation of  $\theta$  is achieved by defining

$$\hat{\theta} : \mathcal{X} \mapsto \Theta$$

Then the observation of  $x$  results in  $\hat{\theta}(x)$  as the estimated value of  $\theta$  for the observed data. Similarly hypothesis testing maps  $\mathcal{X}$  into  $\{\Theta_0, \Theta_1\}$  and interval estimation maps  $\mathcal{X}$  into subsets (intervals) of  $\Theta$ .

## 2.4 Statistics and Combinants

### 2.4.1 *Statistics and Sampling Distributions*

Since inferences are defined by functions on the sample space it is convenient to have some nomenclature.

**Definition 2.4.1.** A **statistic** is a real or vector-valued function defined on the sample space of a statistical model.

The sample mean, sample variance, sample median, and sample correlation are all statistics.

**Definition 2.4.2.** The probability distribution of a statistic is called its **sampling distribution**.

A major problem in standard or frequentist statistical theory is the determination of sampling distributions:

1. Either exactly (using probability concepts)
2. Approximately (using large sample results)
3. By simulation (using R or similar statistical software)

### 2.4.2 *Combinants*

**Definition 2.4.3.** A **combinant** is a real or vector-valued function defined on the sample space and the parameter space such that for each fixed  $\theta$  it is a statistic.

Thus a combinant is defined for pairs  $(x, \theta)$  where  $x$  is in the sample space and  $\theta$  is in the parameter space. For each  $\theta$  it is required to be a statistic.

The density function  $f(x; \theta)$  is a combinant, as are the likelihood and functions of the likelihood.

**Definition 2.4.4.** If  $f(x; \theta)$  is the density of  $x$  the **score function** is the combinant defined by

$$s(\theta; x) = \frac{\partial f(x; \theta)}{\partial \theta}$$

(This assumes differentiation with respect to  $\theta$  is defined.)

**Definition 2.4.5.** The **score equation** is the equation (in  $\theta$ ) defined by

$$s(\theta; x) = \frac{\partial f(x; \theta)}{\partial \theta} = 0$$

The solution to this equation gives the maximum likelihood estimate, MLE, of  $\theta$ .

Combinants are used to determine estimates, interval estimates, and tests as well as to investigate the frequency properties of likelihood-based quantities.

### 2.4.3 Frequentist Inference

In the **frequentist paradigm** inference is the process of connecting the observed data and the inference (statements about the parameters) using the **sampling distribution** of a statistic. Note that the sampling distribution is determined by the density function  $f(x; \theta)$ .

### 2.4.4 Bayesian Inference

In the **Bayesian paradigm** inference is the process of connecting the observed data and the inference (statements about the parameters) using the **posterior distribution** of the parameter values. The **posterior distribution** is determined by the model density and the **prior distribution** of  $\theta$  using Bayes theorem (this implicitly treats  $f(x; \theta)$  as the conditional  $f(x|\theta)$  of  $X$  given  $\theta$ ):

$$p(\theta|x) = \frac{f(x; \theta)\text{prior}(\theta)}{f(x)}$$

where  $f(x)$  is the marginal distribution of  $X$  at  $x$ .

$$f(x) = \int_{\Theta} f(x; \theta) \text{prior}(\theta) d\theta$$

### 2.4.5 Likelihood Inference

In the **likelihood paradigm** inference is the process of evaluating the statistical evidence for parameter values provided by the likelihood function.

The statistical evidence for  $\theta_2$  vis-a-vis  $\theta_1$  is defined by

$$\text{Ev}(\theta_2 : \theta_1; x) = \frac{f(x; \theta_2)}{f(x; \theta_1)}$$

Values for this ratio of 8, 16, and 32 are taken as moderate, strong, and very strong evidence, respectively.

Note that if we define the **likelihood** of  $\theta$  as

$$\mathcal{L}(\theta; x) = \frac{f(x; \theta)}{f(x; \hat{\theta})}$$

where  $\hat{\theta}$  is the maximum likelihood estimate of  $\theta$ , then the statistical evidence for  $\theta_2$  vs  $\theta_1$  can be expressed as

$$\text{Ev}(\theta_2 : \theta_1; x) = \frac{\mathcal{L}(\theta_2; x)}{\mathcal{L}(\theta_1; x)}$$

and the posterior of  $\theta$  can then be expressed as

$$p(\theta|x) = \frac{\mathcal{L}(\theta; x) \text{prior}(\theta)}{f(x)}$$

i.e., the posterior is proportional to the product of the likelihood and the prior.

## 2.5 Exercises

As pointed out in the text if  $f(\mathbf{x}; \theta)$  is the density function of the observed data  $(x_1, x_2, \dots, x_n)$  and  $\theta$  is the parameter, then

(a) The **likelihood**,  $\mathcal{L}(\theta; \mathbf{x})$ , is

$$\mathcal{L}(\theta) = \frac{f(\mathbf{x}; \theta)}{f(\mathbf{x}; \hat{\theta})}$$

where  $\hat{\theta}$  maximizes  $f(\mathbf{x}; \theta)$  and is called the maximum likelihood estimate of  $\theta$ .

(b) The **score function** is

$$\frac{\partial \ln[f(\mathbf{x}; \theta)]}{\partial \theta}$$

(c) The **observed Fisher information** is

$$J(\theta) = - \frac{\partial^2 \ln[f(\mathbf{x}; \theta)]}{\partial \theta^2}$$

evaluated at  $\theta = \hat{\theta}$ .

(d) The **expected Fisher information**,  $I(\theta)$ , is the expected value of  $J(\theta)$ , i.e.,

$$I(\theta) = -\mathbb{E} \left\{ \frac{\partial^2 \ln[f(\mathbf{x}; \theta)]}{\partial \theta^2} \right\}$$

1. Find the likelihood, the maximum likelihood estimate, the score function, and the observed and expected Fisher information when  $x_1, x_2, \dots, x_n$  represent the results of a random sample from

- (i) A normal distribution with expected value  $\theta$  and known variance  $\sigma^2$
- (ii) A Poisson distribution with parameter  $\theta$
- (iii) A Gamma distribution with known parameter  $\alpha$  and  $\theta$

2. For each of the problems in (1) generate a random sample of size 25, i.e.:

- (i) Take  $\sigma^2 = 1$  and  $\theta = 3$ .
- (ii) Take  $\theta = 5$ .
- (iii) Take  $\alpha = 3$  and  $\theta = 2$ .

For (i)–(iii) plot the likelihood functions.

3. Suppose that  $Y_i$  for  $i = 1, 2, \dots, n$  are independent, each normal with expected value  $\beta x_i$  and variance  $\sigma^2$  where  $\sigma^2$  is known and the  $x_i$  are known constants.

(i) Show that the joint density is

$$f(\mathbf{y}; \beta) = (2\pi\sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2 \right\}$$

- (ii) Find the score function.
- (iii) Show that the maximum likelihood estimate for  $\beta$  is

$$\hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

(iv) Find the observed Fisher information.

- (v) Using (iii) find the likelihood for  $\beta$ .
- (vi) Find the sampling distribution of  $\hat{\beta}$ . Remember that the sum of independent normal random variables is also normal.
- (vii) Show that the sampling distribution of  $-2 \ln[\mathcal{L}(\beta; \mathbf{y})]$  is chi-square with 1 degree of freedom.

Introductory Statistical Inference with the Likelihood  
Function

Rohde, C.A.

2014, XVI, 332 p. 12 illus., Hardcover

ISBN: 978-3-319-10460-7