statistical model

Canonical name	StatisticalModel
Date of creation	2013-03-22 14:33:18
Last modified on	2013-03-22 14:33:18
Owner	CWoo (3771)
Last modified by	CWoo (3771)
Numerical id	10
Author	CWoo (3771)
Entry type	Definition
Classification	msc 62A01
Defines	identifiable parameterization
Defines	parameter space

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a random vector with a given realization $\mathbf{X}(\omega) = (x_1, \ldots, x_n)$, where ω is the outcome (of an observation or an experiment) in the sample space Ω. A *statistical model* $\mathcal{P}$ based on $\mathbf{X}$ is a set of probability distribution functions of $\mathbf{X}$:

$$\mathcal{P} = \{F_{\mathbf{X}}\}.$$

If it is known in advance that this family of distributions comes from a set of continuous distributions, the statistical model $\mathcal{P}$ can be equivalently defined as a set of probability density functions:

$$\mathcal{P} = \{f_{\mathbf{X}}\}.$$

As an example, a coin is tossed n times and the results are observed. The probability of landing a head during one toss is p. Assume that each toss is independent of one another. If $\mathbf{X} = (X_1, \ldots, X_n)$ is defined to be the vector of the n ordered outcomes, then a statistical model based on $\mathbf{X}$ can be a family of Bernoulli distributions

$$\mathcal{P} = \{\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}\},$$

where $X_i(\omega) = x_i$ and $x_i = 1$ if ω is the outcome that the ith toss lands a head and $x_i = 0$ if ω is the outcome that the ith toss lands a tail.

Next, suppose X is the number of tosses where a head is observed, then a statistical model based on X can be a family binomial distributions:

$$\mathcal{P} = \{\binom{n}{x} p^x(1-p)^{n-x}\},$$

where $X(\omega) = x$, where ω is the outcome that x heads (out of n tosses) are observed.

A statistical model is usually *parameterized* by a function, called a *parameterization*

$$\Theta \to \mathcal{P} \text{ given by } \theta \mapsto F_\theta \text{ so that } \mathcal{P} = \{F_\theta \mid \theta \in \Theta\},$$

where Θ is called a *parameter space*. Θ is usually a subset of $\mathbb{R}^n$. However, it can also be a function space.

In the first part of the above example, the statistical model is parameterized by

$$p \mapsto \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}.$$

If the parameterization is a one-to-one function, it is called an *identifiable parameterization* and θ is called a *parameter*. The p in the above example is a parameter.