From Peter Hall, "The Bootstrap and Edgeworth Expansion", Springer-Verlag, 1992

1.2 The Main Principle

We begin by describing a convenient physical analogy, and later (following equation (1.5)) give an explicit statement of the principle.

A Russian "matryoshka" doll is a nest of wooden figures, usually with slightly different features painted on each. Call the outer figure "doll 0," the next figure "doll 1," and so on; see Figure 1.1. Suppose we are not allowed to observe doll 0 — it represents the population in a sampling scheme. We wish to estimate the number n_0 of freckles on its face. Let n_i denote the number of freckles on the face of doll i. Since doll 1 is smaller than doll 0, n_1 is likely to be an underestimate of n_0, but it seems reasonable to suppose that the ratio of n_1 to n_2 should be close to the ratio of n_0 to n_1. That is, $n_1/n_2 \simeq n_0/n_1$, so that $\hat{n}_0 = n_1^2/n_2$ might be a reasonable estimate of n_0.

The key feature of this argument is our hypothesis that the relationship between n_2 and n_1 should closely resemble that between n_1 and the unknown n_0. Under the (fictitious) assumption that the relationships are *identical*, we equated the two ratios and obtained our estimate $\hat{n}_0$. We could refine the argument by delving more deeply into the nest of dolls, adding correction terms to $\hat{n}_0$ so as to take account of the relationship between doll i and doll $i+1$ for $i \geq 2$. We shall have more to say about that in Section 1.4. But for the time being we study the more obvious implications of the "Russian doll principle."

Much of statistical inference amounts to describing the relationship between a sample and the population from which the sample was drawn. Formally, given a functional f_t from a class $\{f_t : t \in \mathcal{T}\}$, we wish to determine that value t_0 of t that solves an equation

$$E\{f_t(F_0, F_1) \mid F_0\} = 0, \tag{1.1}$$

where F_0 denotes the population distribution function and F_1 the distribution function "of the sample." An explicit definition of F_1 will be given shortly. Conditioning on F_0 in (1.1) serves to stress that the expectation is taken with respect to the distribution F_0. We call (1.1) the *population equation* because we need properties of the population if we are to solve this equation exactly.

For example, let $\theta_0 = \theta(F_0)$ denote a true parameter value, such as the rth power of a mean,

$$\theta_0 = \left\{ \int x \, dF_0(x) \right\}^r .$$

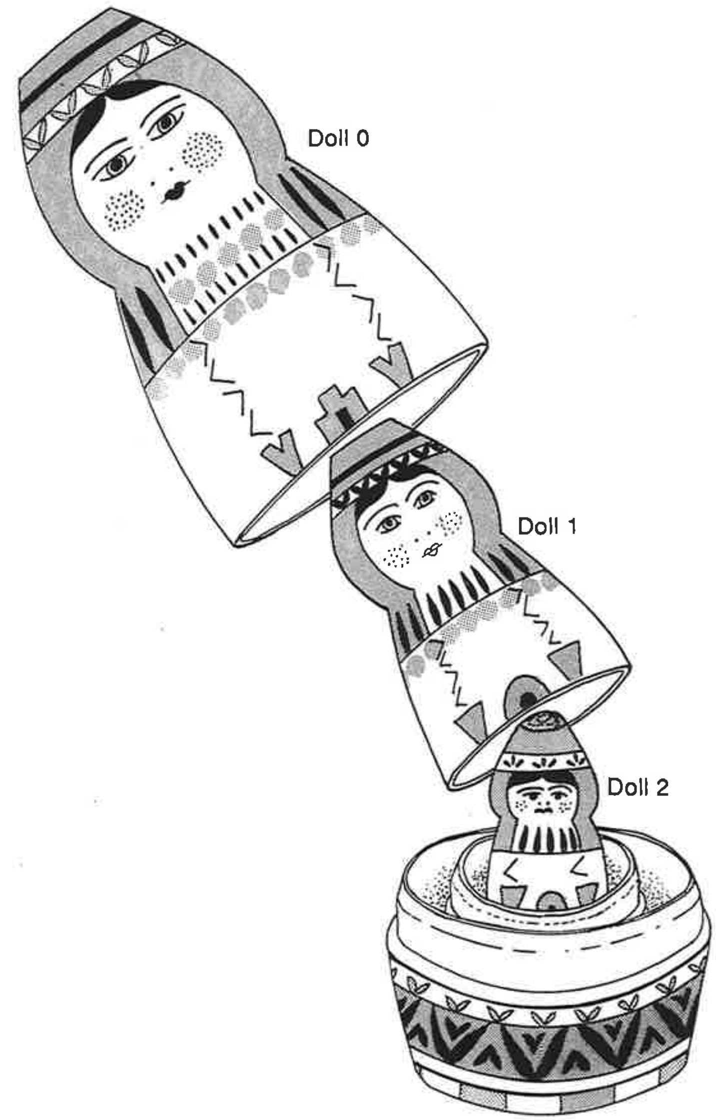

FIGURE 1.1. A Russian matryoshka doll.