servations that are mutually independent and identically distributed (IID), or X might be some general quantity. The set of possible values for X is the *sample space* and is often denoted as $\mathcal{X}$. The members P_θ of the parametric family will be distributions over this space $\mathcal{X}$. If X is continuous or discrete, then densities or probability mass functions[1] exist. We will denote the density or mass function for P_θ by $f_{X|\Theta}(\cdot|\theta)$. For example, if X is a single random variable with continuous distribution, then*

$$P_\theta(a < X \leq b) = \int_a^b f_{X|\Theta}(x|\theta)dx.$$

If $X = (X_1, \ldots, X_n)$, where the X_i are IID each with density (or mass function) $f_{X_1|\Theta}(\cdot|\theta)$ when $\Theta = \theta$, then

$$f_{X|\Theta}(x|\theta) = \prod_{i=1}^{n} f_{X_1|\Theta}(x_i|\theta), \tag{1.1}$$

where $x = (x_1, \ldots, x_n)$. After observing the data $X_1 = x_1, \ldots, X_n = x_n$, the function in (1.1), as a function of θ for fixed x, is called the *likelihood function*, denoted by $L(\theta)$. Section 1.3 is devoted to a motivation of the above structure based on the concept of *exchangeability* and DeFinetti's representation theorem 1.49. Exchangeability is discussed in detail in Section 1.2, and DeFinetti's theorem is the subject of Section 1.4.

1.1.2 Classical Statistics

Classical inferential techniques include tests of hypotheses, unbiased estimates, maximum likelihood estimates, confidence intervals and many other things. These will be covered in great detail in the text, but we remind the reader of a few of them here. Suppose that we are interested in whether or not the parameter lies in one portion Ω_H of the parameter space. We could then set up a *hypothesis* $H : \Theta \in \Omega_H$ with the corresponding *alternative* $A : \Theta \notin \Omega_H$. The simplest sort of *test* of this hypothesis would be to choose a subset $R \subseteq \mathcal{X}$, and then reject H if $x \in R$ is observed. The set R would be called the *rejection region* for the test. If $x \notin R$, we would say that we do not reject H. Tests are compared based on their power functions. The *power function* of a test with rejection region R is $\beta(\theta) = P_\theta(X \in R)$. The *size* of a test is $\sup_{\theta \in \Omega_H} \beta(\theta)$. Chapter 4 covers hypothesis testing in depth.

Example 1.2. Suppose that $X = (X_1, \ldots, X_n)$ and the X_i are IID with $N(\theta, 1)$ distribution under P_θ. The usual size α test of $H : \Theta = \theta_0$ versus $A : \Theta \neq \theta_0$ is

---

[1]Using the theory of measures (see Appendix A) we will be able to dispense with the distinction between densities and probability mass functions. They will both be special cases of a more general type of "density."

*Later in this chapter we will distinguish between the probability P_θ on $\mathcal{X}$ and the corresponding probability on the underlying space $(S, \mathcal{A})$ as discussed in Appendix B. In this section, we will use P_θ to stand for both probabilities.