of the distribution rather than the moments of a specific random variable. If a distribution possesses a k^{th} moment, it also possesses all moments of order less than k.

The higher moments just defined are called the **uncentered moments** of a distribution, because, in general, X does not have mean zero. It is often more useful to work with the **central moments**, which are defined as the ordinary moments of the difference between the random variable and its expectation. Thus the k^{th} central moment of the distribution of a continuous r.v. X is

$$\mu_k \equiv \mathrm{E}\big(X - \mathrm{E}(X)\big)^k = \int_{-\infty}^{\infty} (x-\mu)^k f(x)\,dx,$$

where $\mu \equiv \mathrm{E}(X)$. For a discrete X, the k^{th} central moment is

$$\mu_k \equiv \mathrm{E}\big(X - \mathrm{E}(X)\big)^k = \sum_{i=1}^{m} p(x_i)(x_i-\mu)^k.$$

By far the most important central moment is the second. It is called the **variance** of the random variable and is frequently written as $\mathrm{Var}(X)$. Another common notation for a variance is σ^2. This notation underlines the important fact that a variance cannot be negative. The square root of the variance, σ, is called the **standard deviation** of the distribution. Estimates of standard deviations are often referred to as **standard errors**, especially when the random variable in question is an estimated parameter.

Multivariate Distributions

A **vector-valued random variable** takes on values that are vectors. It can be thought of as several scalar random variables that have a single, joint distribution. For simplicity, we will focus on the case of **bivariate random variables**, where the vector has two elements. A continuous, bivariate random variable (X_1, X_2) has a distribution function

$$F(x_1, x_2) = \Pr\big((X_1 \le x_1) \cap (X_2 \le x_2)\big),$$

where $\cap$ is the symbol for set intersection. Thus $F(x_1, x_2)$ is the joint probability that both $X_1 \le x_1$ and $X_2 \le x_2$. For continuous variables, the PDF, if it exists, is the **joint density function**[2]

$$f(x_1, x_2) = \frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2}. \tag{1.09}$$

[2] Here we are using what computer scientists would call "overloaded function" notation. This means that $F(\cdot)$ and $f(\cdot)$ denote, respectively, the CDF and the PDF of whatever their argument(s) happen to be. This practice is harmless provided there is no ambiguity.