21-256: Applications of integration to probability

Clive Newstead, Monday 23rd June 2014

Intuitively, a (real, continuous) *random variable* is a real number quantity whose precise value is unknown until it is observed. The *probability distribution function* (pdf) of a random random variable X is a function f_X which allows us to compute the probability that the outcome will be in a given range.

Specifically, if X is a random variable, its pdf is a function f_X such that

$$f_X(x) \geqslant 0 \text{ for all } x \qquad \text{and} \qquad \int_{-\infty}^{\infty} f(x)\, dx = 1$$

The *probability* that X takes a value between real numbers a and b, where $a < b$, is defined to be

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)\, dx$$

We consider what happens when, instead of just one random variable, there are more, which may be related in some way.

Suppose X and Y are random variables. Their *joint distribution function* (jdf) $f_{X,Y}$ is a bivariate function such that

$$f_{X,Y}(x,y) \geqslant 0 \text{ for all } x, y \qquad \text{and} \qquad \int_{\mathbb{R}^2} f(x,y)\, dA = 1$$

The *probability* that the pair (X,Y) takes its value in some subset U of $\mathbb{R}^2$ is defined to be

$$\mathbb{P}((X,Y) \in U) = \int_U f(x,y)\, dA$$

For instance, if U is the rectangle defined by $a \leqslant x \leqslant b$ and $c \leqslant y \leqslant d$ then

$$\mathbb{P}(a \leqslant X \leqslant b \text{ and } c \leqslant Y \leqslant d) = \int_a^b \int_c^d f_{X,Y}(x,y)\, dy\, dx$$

We say random variables X and Y are *independent* if $f_{X,Y}(x,y) = f_X(x) f_Y(y)$, where $f_{X,Y}$ is the jdf of X and Y, f_X is the pdf of X and f_Y is the pdf of Y. This makes computations much easier, since when X and Y are independent it follows that

$$\int_a^b \int_c^d f_{X,Y}(x,y)\, dy\, dx = \left(\int_a^b f_X(x)\, dx \right) \left(\int_c^d f_Y(y)\, dy \right)$$