We give a proof of the positive definiteness of convolution kernels. Convolution kernels (Haussler '99) are defined by

$$K(x,y) = \sum_{\vec{x}\in R^{-1}(x)} \sum_{\vec{y}\in R^{-1}(y)} k(\vec{x},\vec{y}). \tag{1}$$

Here $\vec{x},\vec{y} \in X$ are considered as parts of x and y; $k(\vec{x},\vec{y})$ is a positive definite kernel defined on $X \times X$; R is a relation which associates a *finite* subset of X with each x in the object space $\mathcal{X}$, and this subset is denoted by $R^{-1}(x)$.

To simplify notation, let us write $R^{-1}(x)$ as A_x. Thus $K(x,y)$ assigns a number to two finite subsets $A_x, A_y \subseteq X$. We can *extend* this definition to all pairs of finite subsets of X by defining a kernel $K^s : \mathcal{U} \times \mathcal{U} \to \Re$, where $\mathcal{U} = \{A \mid A \text{ finite}, A \subseteq X\}$,

$$K^s(A,B) = \sum_{\vec{x}\in A} \sum_{\vec{y}\in B} k(\vec{x},\vec{y}).$$

Clearly, if the extension K^s is a positive definite kernel, then K is positive definite.

We shall prove this using the random variable "trick." For any vector-valued random variable Y, taking values in $\Re^n$, the matrix $E\{YY^\top\}$ is clearly always positive definite, because $c^\top E\{YY^\top\}c = E\{(c^\top Y)^2\} \geq 0$ for any vector $c \in \Re^n$.

Lemma 1. *K^s is positive definite.*

Proof. We need to show for any n, $A_1, \ldots, A_n \subseteq X$, the matrix $\bar{K}$ with $\bar{K}_{ij} = K^s(A_i, A_j)$ is positive definite. We shall prove this by constructing an n-dimensional random variable $Y = (Y_1, \ldots, Y_n)^\top$, taking values in $\Re^n$, such that $E\{YY^\top\} = \bar{K}$. Let $A = \cup_{i=1}^n A_i$ and m be the size of A. Create an m-dimensional zero-mean Gaussian random variable Z, whose components are indexed by $a \in A$ and whose covariance is defined by

$$E\{Z_a\, Z_{\hat{a}}\} = k(a,\hat{a}), \quad a, \hat{a} \in A.$$

We now define $Y_i, i = 1, \ldots, n$ by

$$Y_i = \sum_{a\in A_i} Z_a.$$

Then,

$$E\{Y_iY_j\} = E\Big\{\Big(\sum_{a\in A_i} Z_a\Big)\Big(\sum_{\hat{a}\in A_j} Z_{\hat{a}}\Big)\Big\} = \sum_{a\in A_i}\sum_{\hat{a}\in A_j} E\{Z_a Z_{\hat{a}}\} = \sum_{a\in A_i}\sum_{\hat{a}\in A_j} k(a,\hat{a}) = K^s(A_i, A_j),$$

so $E\{YY^\top\} = \bar{K}$; this completes our proof. □