



## Why

We can identify probability distributions with vectors.

## Definition

Let  $p : \Omega \rightarrow \mathbf{R}$  be a probability distribution on a finite set  $\Omega = \{\omega_1, \dots, \omega_n\}$ . We can associate  $p$  with the vector  $x \in \mathbf{R}^n$  defined by  $x_i = p(\omega_i)$  for  $i = 1, \dots, n$ . We call this vector  $y$  the *probability vector* associated with  $p$ . The conditions on  $p$  mean that (1)  $\langle 1, y \rangle = 1$  and (2)  $y \geq 0$  (i.e.,  $y_i \geq 0$  for all  $i = 1, \dots, n$ ).

Conversely, suppose  $z \in \mathbf{R}^n$  satisfies (1) and (2). Then  $q : \Omega \rightarrow \mathbf{R}$  defined by  $q(\omega_i) = z_i$  for  $i = 1, \dots, n$  is a probability distribution. For this reason, we call  $z$  satisfying the conditions a *distribution vector*. Notice that implicit in this correspondence is a numbering  $\omega : \{1, \dots, n\} \rightarrow \Omega$  of the set of outcomes  $\Omega$ .

## Expectation

Suppose  $\rho \in \mathbf{R}^n$  is a distribution vector corresponding to  $p : \Omega \rightarrow \mathbf{R}$  its corresponding distribution. Let  $x : \Omega \rightarrow \mathbf{R}$  and (similar to  $\rho$ ) define  $\xi \in \mathbf{R}^n$  by  $\xi_i = x(\omega_i)$  for  $i = 1, \dots, n$ . Then  $\mathbf{E}x = \langle \rho, \xi \rangle$ .

## Example

For  $\rho = (-1, -1, 1, 1, 2)$  and  $\xi = (0.1, 0.15, 0.1, 0.25, 0.4)$

$$\mathbf{E}x = \langle \rho, \xi \rangle = -1 - 0.15 + 0.1 + 0.25 + 2(0.4) = 0.9.$$



