ENEE 627
SPRING 2011
INFORMATION THEORY

CONVEXITY

Convex sets

A subset K of $\mathbb{R}^d$ is said to be *convex* if for any elements $\boldsymbol{x}$ and $\boldsymbol{y}$ of K, and any λ in $]0,1]$, we have

$$\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y} \in K, \quad \lambda \in [0,1]$$

It is a simple exercise to show the following by induction:

Lemma 0.1 *A set K in $\mathbb{R}^d$ is convex if and only if for any integer $p = 2, 3, \ldots$, and any collection $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in K, we have*

$$\lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p \in K$$

for arbitrary $\lambda_1, \ldots, \lambda_p$ in $[0,1]$ such that

$$\lambda_1 + \ldots + \lambda_p = 1.$$

We refer to the linear combination $\lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p$ with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in $\mathbb{R}^d$ and $\lambda_1, \ldots, \lambda_p$ in $[0,1]$ such that

$$\lambda_1 + \ldots + \lambda_p = 1$$

as a *convex* combination.

Convex functions

Consider a convex set K in $\mathbb{R}^d$. A function $\varphi : K \to \mathbb{R}$ is said to be *convex* if for any elements $\boldsymbol{x}$ and $\boldsymbol{y}$ of K, and any λ in $]0,1]$, we have

$$\varphi(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) \leq \lambda\varphi(\boldsymbol{x}) + (1-\lambda)\varphi(\boldsymbol{y}), \quad \lambda \in [0,1].$$

A function $\varphi : K \to \mathbb{R}$ is said to be *concave* if $-\varphi$ is a convex function.

It is also a simple exercise to show the following by induction:

Lemma 0.2 *Consider a convex set K in $\mathbb{R}^d$. A function $\varphi : K \to \mathbb{R}$ is convex if and only if for any integer $p = 2, 3, \ldots$, and any collection $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in K, we have*

$$\varphi(\lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p) \leq \lambda_1 \varphi(\boldsymbol{x}_1) + \ldots + \lambda_p \varphi(\boldsymbol{x}_p)$$

for arbitrary $\lambda_1, \ldots, \lambda_p$ in $[0, 1]$ such that

$$\lambda_1 + \ldots + \lambda_p = 1.$$

Strictly convex functions

A function $\varphi : K \to \mathbb{R}$ is said to be *strictly convex* if it is convex and whenever the equality

$$\varphi(\lambda \boldsymbol{x} + (1-\lambda)\boldsymbol{y}) = \lambda \varphi(\boldsymbol{x}) + (1-\lambda)\varphi(\boldsymbol{y}), \qquad \begin{array}{l} \boldsymbol{x}, \boldsymbol{y} \in K \\ \lambda \in (0,1) \end{array}$$

holds, we necessarily have $\boldsymbol{x} = \boldsymbol{y}$. As expected, a function $\varphi : K \to \mathbb{R}$ is said to be *strictly concave* if $-\varphi$ is a strictly convex function.

Of great usefulness in many arguments is the following observation: Consider a strictly convex $\varphi : K \to \mathbb{R}$. Suppose that for some $p = 2, 3, \ldots$, with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$ in K, we have the equality

$$\varphi(\lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p) = \lambda_1 \varphi(\boldsymbol{x}_1) + \ldots + \lambda_p \varphi(\boldsymbol{x}_p) \tag{1}$$

with $\lambda_1, \ldots, \lambda_p$ in $(0, 1)$ such that

$$\lambda_1 + \ldots + \lambda_p = 1.$$

Under such circumstances, what can we say about $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_p$? We shall show that we must necessarily have

$$\boldsymbol{x}_1 = \ldots = \boldsymbol{x}_p. \tag{2}$$

If $p = 2$, since $0 < \lambda_1, \lambda_2 < 1$, by definition of strict convexity we automatically have the conclusion $\boldsymbol{x}_1 = \boldsymbol{x}_2$. If $p > 2$, the matter is more involved. To proceed, with any subset I of $\{1, \ldots, p\}$ such that $1 \leq |I| < p$ we define

$$\lambda_I = \sum_{i \in I} \lambda_i.$$

Under the foregoing assumptions we have $0 < \lambda_I < 1$, so that the definition

$$\boldsymbol{x}_I = \sum_{i \in I} \frac{\lambda_i}{\lambda_I} \boldsymbol{x}_i$$

is well posed and yields an element of K. We also note that

$$\lambda_I \boldsymbol{x}_I + \lambda_{I^c} \boldsymbol{x}_{I^c} = \lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p.$$

with

$$\lambda_I + \lambda_{I^c} = 1.$$

Using the convexity of φ twice we get

$$\begin{aligned}
&\varphi(\lambda_1 \boldsymbol{x}_1 + \ldots + \lambda_p \boldsymbol{x}_p) \\
&\quad = \varphi(\lambda_I \boldsymbol{x}_I + \lambda_{I^c} \boldsymbol{x}_{I^c}) \\
&\quad \leq \lambda_I \varphi(\boldsymbol{x}_I) + \lambda_{I^c} \varphi(\boldsymbol{x}_{I^c}) \\
&\quad \leq \lambda_I \left(\sum_{i \in I} \frac{\lambda_i}{\lambda_I} \varphi(\boldsymbol{x}_i) \right) + \lambda_{I^c} \left(\sum_{j \notin I} \frac{\lambda_j}{\lambda_{I^c}} \varphi(\boldsymbol{x}_j) \right) \\
&\quad = \lambda_1 \varphi(\boldsymbol{x}_1) + \ldots + \lambda_p \varphi(\boldsymbol{x}_p). \qquad (3)
\end{aligned}$$

Moreover, convexity again gives

$$\varphi(\boldsymbol{x}_I) \leq \sum_{i \in I} \frac{\lambda_i}{\lambda_I} \varphi(\boldsymbol{x}_i) \qquad (4)$$

and

$$\varphi(\boldsymbol{x}_{I^c}) \leq \sum_{j \notin I} \frac{\lambda_j}{\lambda_{I^c}} \varphi(\boldsymbol{x}_j). \qquad (5)$$

However, because of (1) the inequalities leading to (3) must necessarily hold as equalities, and this implies

$$\varphi(\lambda_I \boldsymbol{x}_I + \lambda_{I^c} \boldsymbol{x}_{I^c}) = \lambda_I \varphi(\boldsymbol{x}_I) + \lambda_{I^c} \varphi(\boldsymbol{x}_{I^c}), \qquad (6)$$

$$\varphi(\boldsymbol{x}_I) = \sum_{i \in I} \frac{\lambda_i}{\lambda_I} \varphi(\boldsymbol{x}_i) \qquad (7)$$

and

$$\varphi(\boldsymbol{x}_{I^c}) = \sum_{j \notin I} \frac{\lambda_j}{\lambda_{I^c}} \varphi(\boldsymbol{x}_j) \qquad (8)$$

as we make use of the fact that

$$0 < \lambda_I, \ \lambda_{I^c} < 1.$$

By strict convexity it follows from (6) that

$$\boldsymbol{x}_I = \boldsymbol{x}_{I^c}.$$

With (7) and (8) as point of departure, in lieu of (1), we can repeat the arguments above with I and I^c, respectively, instead of $\{1, \ldots, p\}$. Upon doing this as many times as needed we can eventually conclude that

$$\boldsymbol{x}_i = \boldsymbol{x}_j, \qquad \begin{array}{l} i, j = 1, \ldots, p \\ i \neq j \end{array}$$

and this completes the proof of (2). ■

Kullback-Leibler distance

Consider a set $\mathcal{X}$ of finite cardinality. With $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ pmfs on $\mathcal{X}$, define

$$D(\boldsymbol{\nu}||\boldsymbol{\mu}) = \sum_x \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right)$$

with the conventions

$$0 \log\left(\frac{0}{0}\right) = 0,$$

$$p \log\left(\frac{p}{0}\right) = \infty \quad \text{if } p > 0$$

and

$$0 \log\left(\frac{0}{q}\right) = 0 \quad \text{if } q > 0$$

The proof of Theorem 2.6.3 revisited: Thus,

$$\begin{aligned} -D(\boldsymbol{\nu}||\boldsymbol{\mu}) &= -\sum_x \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right) \\ &= -\sum_{x:\ \nu(x)>0} \nu(x) \log\left(\frac{\nu(x)}{\mu(x)}\right) \\ &= \sum_{x:\ \nu(x)>0} \nu(x) \log\left(\frac{\mu(x)}{\nu(x)}\right) \\ &\leq \log\left(\sum_{x:\ \nu(x)>0} \nu(x) \frac{\mu(x)}{\nu(x)}\right) \qquad (9) \\ &= \log\left(\sum_{x:\ \nu(x)>0} \mu(x)\right) \\ &\leq \log 1 = 0 \qquad (10) \end{aligned}$$

whence $-D(\boldsymbol{\nu}||\boldsymbol{\mu}) \leq 0$, or equivalently, $D(\boldsymbol{\nu}||\boldsymbol{\mu}) \geq 0$.

The equality $D(\boldsymbol{\nu}||\boldsymbol{\mu}) = 0$ occurs if and only if equality occurs at both (9) and (10). By the strict concavity of $t \rightarrow \log t$, equality occurs at (9) if and only terre exists $c > 0$ such that

$$\frac{\mu(x)}{\nu(x)} = c, \qquad \begin{array}{c} x \in \mathcal{X} \\ \nu(x) > 0 \end{array}$$

As a result,

$$\sum_{x:\ \nu(x)>0} \mu(x) = c \sum_{x:\ \nu(x)>0} \nu(x) = c$$

since

$$\sum_{x:\ \nu(x)>0} \nu(x) = \sum_{x} \nu(x) = 1.$$

On the other hand, (10) occurs if and only if

$$\sum_{x:\ \nu(x)>0} \mu(x) = 1$$

Consequently, $c = 1$ and

$$\sum_{x:\ \nu(x)=0} \mu(x) = 0,$$

whence $\mu(x) = 0$ if and only if $\nu(x) = 0$. In sum, $\mu(x) = \nu(x)$ for *all* x in $\mathcal{X}$.