In Lecture 1, we defined the following basic information measures:

- **Relative entropy:**
  \[
  D(p||q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = \mathbb{E} \left( \log \frac{p(X)}{q(X)} \right) \text{ where } X \sim p(x).
  \]

- **Mutual information:**
  \[
  I(X;Y) = D(p(x,y)||p(x)p(y))
  \]

- **Entropy:**
  \[
  H(X) = I(X;X) = \mathbb{E} \left( \log \frac{1}{p(X)} \right).
  \]

In Lecture 2, we showed several key properties of these information measures. These followed from using properties of convex functions. The following notes provide some additional detail about convex functions.

**Definition:** A real-valued function \( f \) is **convex** over an interval \([a, b]\) if and only if for all \( x_1, x_2 \in [a, b] \) and all \( \alpha \in [0, 1] \)

\[
  f(\alpha x_1 + (1 - \alpha)x_2) \leq \alpha f(x_1) + (1 - \alpha)f(x_2)
  \]

The right-hand side of this equation is a point on the line-segment connecting \( f(x_1) \) with \( f(x_2) \); thus this definition implies that this line-segment lies above the function. If \( f \) is twice differentiable then the above definition is equivalent to requiring that \( \frac{d^2}{dx^2} f(x) \geq 0 \) for all \( x \in [a, b] \) (see Theorem 2.6.1 in the text). If for any \( \alpha \in (a, b) \), the above inequality is strict, then \( f \) is said to be **strictly convex**.

**Definition:** \( f \) is (strictly) **concave** over an interval \([a, b]\) if \(-f\) is (strictly) convex.

Notice that a linear function will be both convex and concave (but not strictly convex or strictly concave).

Some useful properties for convex functions are:

1. If \( f \) and \( g \) are both convex on \([a, b]\) then so is \( \alpha f + \beta g \) for any \( \alpha \geq 0 \) and \( \beta \geq 0 \).
2. If $f$ is convex $[a, b]$ and $g$ is linear then $f(g(x))$ is convex.

3. If $f$ and $g$ are both convex on $[a, b]$ then so is $\max\{f(x), g(x)\}$.

4. If $f$ is convex over $[a, b]$ and differentiable at $x \in [a, b]$, then for all $y \in [a, b]$, $f(y) \geq f(x) + f'(x)(y - x)$.

You may want to try proving these using the above definitions. A useful bound that follows from the last property is that $\log(x) \leq x - 1$ for all $x \geq 0$ (with equality only at $x = 1$).

These definitions and properties extend naturally to multi-variable functions, e.g. let $x = (x_1, \ldots, x_n)$ denote a vector in $\mathbb{R}^n$, then a real-valued function $f(x)$ is convex over the set $\mathcal{X} = \{x = (x_1, \ldots, x_n) : x_i \in [a_i, b_i], i = 1, \ldots, n\}$ if for all $x, y \in \mathcal{X}$, and all $\alpha \in [0, 1]$,

$$f(\alpha x + (1 - \alpha)y) \leq \alpha f(x) + (1 - \alpha)f(y).$$

The definition of convexity can also be extended to functions defined on non-retangular sets, $\mathcal{Y}$; the key property needed is that the set contains the line segment connecting any two points in the set; such sets are also called convex. In other words, $\mathcal{Y}$ is a convex set if and only if for all $y_1, y_2 \in \mathcal{Y}$, and all $\alpha \in [0, 1]$ , then $\alpha y_1 + (1 - \alpha)y_2 \in \mathcal{Y}$.

Given a convex function $f$ over a convex set $\mathcal{X}$, then the set of points

$$\{(x, y) : y \geq f(x), x \in \mathcal{X}\}$$

will also be a convex set (this set is called the epigraph of the function).

A key property of convex functions is Jensen’s inequality:

**Jensen’s inequality:** Let $X$ be a random variable and $f$ a convex function. Then

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]).$$

Furthermore, if $f$ is strictly convex, then equality implies $X = \mathbb{E}[X]$ with probability 1.

If $X$ is a binary RV, inequality follows from the definition of convexity. For a general discrete RV, the inequality can be proved by induction on number of mass points (see text).

We give an alternate proof, assuming $f$ is differentiable.

**Proof:** If $f$ is differentiable then using property 4, above we have that for all $x, y$

$$f(x) \geq f(y) + f'(y)(x - y).$$

Taking expectations with respect of $x$ we have

$$\mathbb{E}[f(X)] \geq f(y) + f'(y)(\mathbb{E}[X] - y).$$

Setting $y = \mathbb{E}[X]$, Jensen’s ineq. follows.■

Jensen’s inequality is still true for continuous valued random variables and also generalizes directly to random vectors.