Fisher Information in a Statistic
Stat 543 Spring 2005

One would hope that "information" is defined in such a way that the "information" in a statistic $T(X)$ can be no larger than the information in X. Here is something in that direction.

Suppose that X is a discrete random quantity taking values in a finite set $\mathcal{X}$ and with probability mass function $f_\theta(x) > 0$ for $\theta \in \Theta$, an open interval in $\mathcal{R}$. Suppose that for every $x \in \mathcal{X}$ the function $f_\theta(x)$ is differentiable in θ. Then in this notation, denoting differentiation wrt θ by a prime,

$$I_X(\theta) = \mathrm{E}_\theta \left(\frac{f'_\theta(X)}{f_\theta(X)} \right)^2$$

For each t, let

$$g_\theta(t) = \sum_{x|T(x)=t} f_\theta(x)$$

Then

$$\begin{aligned}
I_X(\theta) &= \sum_x \left(\frac{f'_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) \\
&= \sum_t \sum_{x|T(x)=t} \left(\frac{f'_\theta(x)}{f_\theta(x)} \right)^2 f_\theta(x) \\
&= \sum_t g_\theta(t) \sum_{x|T(x)=t} \left(\frac{f'_\theta(x)}{f_\theta(x)} \right)^2 \frac{f_\theta(x)}{g_\theta(t)} \\
&\geq \sum_t g_\theta(t) \left(\sum_{x|T(x)=t} \frac{f'_\theta(x)}{f_\theta(x)} \cdot \frac{f_\theta(x)}{g_\theta(t)} \right)^2 \\
&= \sum_t \left(\frac{\sum_{x|T(x)=t} f'_\theta(x)}{g_\theta(t)} \right)^2 g_\theta(t) \\
&= \sum_t \left(\frac{g'_\theta(t)}{g_\theta(t)} \right)^2 g_\theta(t) \\
&= I_{T(X)}(\theta)
\end{aligned}$$

The inequality in the string above is an application of Jensen's Inequality applied to the conditional distributions of X given $T(X) = t$, one t at a time.