$$\mathcal{I}(\boldsymbol{\theta}; \mathbf{x}) = \mathcal{I}(\boldsymbol{\theta}; T(\mathbf{x})) \qquad \Leftrightarrow \qquad T(\mathbf{x}) \text{ sufficient for } \boldsymbol{\theta}$$

1. Sufficient data summary
Holds for simple models (e.g. Gaussian, Poisson). The factorization $p(\mathbf{x} \mid \boldsymbol{\theta}) = h(\mathbf{x})\, g(\boldsymbol{\theta}, T(\mathbf{x}))$ implies that $T(\mathbf{x})$ is sufficient for $\boldsymbol{\theta}$, and preserves *all* information for inference.

Insight into optimal summaries for idealized models (1) helps guide the design of effective summaries in general settings (2).

$$\mathcal{I}(\boldsymbol{\theta}; \mathbf{x}) = \mathcal{I}(\boldsymbol{\theta}; T(\mathbf{x})) + \mathbb{E}_{p(\mathbf{x})}\left[D_{\mathrm{KL}}\left(p(\boldsymbol{\theta} \mid \mathbf{x}) \;||\; p(\boldsymbol{\theta} \mid T(\mathbf{x}))\right)\right]$$

2. Lossy data summary
In general, data summaries do not preserve all information. The Kullback-Leibler (KL) divergence in the second term quantifies the loss of information *relevant for inference* due to summarization.