1. Sufficiency

A statistic $S = S(X)$ is called *sufficient* (for θ) in a model $X \sim f_n(x \mid \theta)$ if the conditional distribution of X, given S (and θ of course), does not depend on θ; this will be the case exactly when the density function factors as

$$f_n(x \mid \theta) = g(S, \theta) \cdot h(x).$$

Evidently all the evidence an observation of $X = x$ gives about θ is already given in $S = s \equiv S(x)$.
For example, in an Exponential Family the likelihood

$$f_n(x \mid \theta) = h(x) \exp\left[\eta(\theta) \cdot T(x) - c(\theta)\right]$$

evidently factors as the product of a function $g(T, \theta) \equiv e^{\eta(\theta) \cdot T(x) - c(\theta)}$ of T and θ and a function $h(x)$ of x that does not depend on θ, so the natural statistic in an exponential family is sufficient. This is the most important class of examples, but not the only ones; others include:

- The maximum $T(x) \equiv \max(x_1, ..., x_n)$, for $x_i \sim \mathsf{Un}(0, \theta)$;
- The range $T(x) \equiv [\min(x_1, ..., x_n), \max(x_1, ..., x_n)]$, for $x_i \sim \mathsf{Un}(\theta_1, \theta_2)$;
- The vector of order statistics $T(x) = (x_{[1]}, x_{[2]}, ..., x_{[n]})$, for any i.i.d. family $x_i \sim f(x \mid \theta)$;
- The likelihood function itself, $T(x) \equiv \{f_n(x \mid \theta) : \theta \in \Theta\}$;
- Any even function like $T_i = |X_i|$ or $T_i = (X_i)^2$, if the distribution of X_i is symmetric about 0 for all θ;
- Any statistic U for which a sufficient statistic $S(X)$ may be written $S(X) = g(U(X))$ for some function $g(\cdot)$.

For example, if $X_i \sim \mathsf{No}(0, \sigma^2)$ then $S_1(X) \equiv (x_1, ..., x_n)$ is sufficient, as are $S_2(X) = ({x_1}^2, ..., {x_n}^2)$, $S_3(X) = ({x_1}^2 + ... + {x_k}^2, {x_{k+1}}^2 + ... + {x_n}^2)$, and $S_4(X) = ({x_1}^2 + ... + {x_n}^2)$. A sufficient statistic S is called *minimal* sufficient if for every other sufficient statistic T there exists a function $g(\cdot)$ such that S may be written $S(X) = g(T(X))$; S_4 is minimal sufficient in this example while the others are not (we can show that the natural sufficient statistic $T(X)$ in an exponential family is always minimal sufficient, so long as the components of $\eta(\theta)$ are linearly independent and also those of $T(X)$; S_4 *is* the natural statistic for the $\mathsf{No}(0, \sigma^2)$ family of distributions).

The general principal is that it is enough to store any sufficient statistic, rather than the entire data set, for any possible inferential goal in an accepted model $x_i \sim f(x \mid \theta)$; for example, for normally distributed data $x_i \sim \mathsf{No}(\mu, \sigma^2)$ it is enough to know the Maximum likelihood estimators $T_n(x) = (\bar{x}_n, S_n^2)$ of μ and σ^2, for any possible inference problem concerning μ or σ^2 or both (can you verify that T_n is sufficient?).

For those familiar with measure theory, a *sigma algebra* $\mathcal{G}$ is sufficient if the conditional expectation $\mathsf{E}[X \mid \mathcal{G}]$ does not depend on θ; this generalizes the definition above (take $\mathcal{G} = \sigma(S)$), but also allows one to consider infinite collections of variables which together are sufficient, through $\mathcal{G} = \sigma(\{S_i\})$. A *minimal sufficient* sigma algebra $\mathcal{G}$ satisfies $\mathcal{G} \subset \mathcal{H}$ for every sufficient sigma algebra $\mathcal{H}$.

Indeed, it is not only "good enough" to base inference on a sufficient statitic, it is actually *better* to do so:

Theorem 1 (Rao-Blackwell) *Let T be any estimator of any quantity $\psi(\theta)$ and let S be sufficient. Define $T_S = \mathsf{E}[T(X) \mid S]$. Then $R(T_S, \theta) < R(T, \theta)$ for all $\theta \in \Theta$, unless $T = g(S)$ in which case $R(T_S, \theta) = R(T, \theta)$ $\forall \theta \in \Theta$.*

The proof is immediate; since $\mathsf{E}[(T - T_S)(T_S - \psi(\theta)) \mid S] = 0$:

$$\begin{aligned}
R(T, \theta) &= \mathsf{E}\left[\left(T(X) - \psi(\theta)\right)^2 \mid \theta\right] \\
&= \mathsf{E}\left[\left(\{T(X) - T_S(X)\} + \{T_S(X) - \psi(\theta)\}\right)^2 \mid \theta\right] \\
&= \mathsf{E}\left[\left(T(X) - T_S(X)\right)^2 \mid \theta\right] + \mathsf{E}\left[\left(T_S(X) - \psi(\theta)\right)^2 \mid \theta\right] \\
&> 0 + \mathsf{E}\left[\left(T_S(X) - \psi(\theta)\right)^2 \mid \theta\right] = R(T_S, \theta),
\end{aligned}$$

or $\mathsf{E}\left[(T - T_S)^2 \mid \theta\right] = 0$ if $T = g(S)$ and hence $R(T_S, \theta) = R(T, \theta)$. Thus it can only *improve* any estimator T to condition it on a sufficient statistic S.

From the factorization criterion for the sufficiency of a statistic S, $f_n(x \mid \theta) = g(S, \theta) \cdot h(x)$, it is clear that the MLE $\hat{\theta} \equiv \text{argmax}(f_n(x \mid \theta) = \text{argmax}(g(S, \theta)$ depends on x only through $S(x)$, and hence cannot be improved by "Rao-Blackwellization", i.e., taking the conditional expectation given S. Often $\hat{\theta}$ will itself be sufficient; if so, it is also *minimal* sufficient, since it is a function of any sufficient S.

WARNING: the concept of sufficiency is only meaningful *within an uncontested model*; if data X_i do *not* come from a $\mathsf{No}(\mu, \sigma^2)$ distribution but instead come from a thicker-tailed distribution like the Cauchy or $t_\nu(\mu, \sigma^2)$ double exponential or Tukey's ϵ-mixture of $\mathsf{No}(\mu, \sigma^2)$ and $\mathsf{No}(\mu, \tau^2 \sigma^2)$ for some $\tau > 0$, for example, then $(\bar{X}_n, S^2)$ is no longer sufficient and more of the data must be retained. In the (rather extreme) case of the Cauchy, for example, the smallest sufficient statistic is the n-dimensional vector of order statistics. Thus, if you intend to test the adequacy of a model or to compare models, it is not enough to keep only sufficient statistics for one particular model.