**Sufficient Statistic**

To find it (call it $\Psi$), expand $\prod_{i=1}^{n} f(X_i \mid \theta)$ into as many factors as possible, ignore those without $\theta$; if those *with* $\theta$ contain only a single combination of all the $X_i$s (usually a sample *sum* such as $\sum_{i=1}^{n} X_i$, $\sum_{i=1}^{n} X_i^2$, $\sum_{i=1}^{n} \ln X_i$, etc. in the regular case, an *order* statistic such as $X_{(n)}$ in the non-regular one). Such $\Psi$ may not always exist (try it with Cauchy!) - we are then out of luck.

Then, find $\mathbb{E}(\Psi)$ (easy in both cases - *univariate* integration required) which must be (usually a simple) function of $\theta$, say $g(\theta)$. Make it equal to $\Psi$ and solve for $\hat{\theta} = g^{-1}(\Psi)$, which is (in general only) an *asymptotically* unbiased (sufficient) *estimator*. In the regular case its variance must tend to (as $n \longrightarrow \infty$) RCV.

When $g$ is non-linear, computing this $\hat{\theta}$'s *exact* bias (let alone removing it) is always theoretically possible but practically very difficult (since it usually involves $n$-dimensional integration) and we will not even try it (being happy with its asymptotic behaviour).

The concept of sufficient estimator is superseded by **ML estimator**, which yields practically the same answer (when sufficient estimator exists), and results in MVaUE in *all* other cases (why would then anyone want to do anything else?). This time it's easier to work with $\sum_{i=1}^{n} \ln f(X_i \mid \theta)$, seen as a function of $\theta$ (the $X_i$s being 'fixed' at their observed values) called ln of the Likelihood function (LF for short), and simply *maximize* it by varying $\theta$. The value of $\theta$ which achieves this is the MLE. It is guaranteed to (uniquely) exist, be asymptotically unbiased, asymptotically RC efficient (in the regular case), and overall 'best' otherwise. The only slight (and only occasional) problem is that this $\hat{\theta}$ may not have a neat analytic form (can be computed only numerically - no problem in the age of computers!).

Estimators thus produced (by either technique) are then always of 'highest quality'. This is definitely *not* the case with Method-of-Moments estimator which is constructed by solving $\mathbb{E}(X) = \bar{X}$. Since $\mathbb{E}(X)$ is a function of $\theta$, say $h(\theta)$, this yields $\hat{\theta} = h^{-1}(\bar{X})$; the estimator is thus always a function of the sample mean $\bar{X}$. This is fine when $\bar{X}$ is sufficient (giving the same answer as the other two techniques), but inefficient in all other cases (*grossly* inefficient particularly in non-regular situations). Besides, it does not work when $\mathbb{E}(X)$ is indefinite (Cauchy). It's difficult to understand why would anyone bother to use it.