Suppose $X \sim P_\theta$, $\theta \in \Theta$.

Basu's Lemma: If $T(X)$ is complete and sufficient (for $\theta \in \Theta$), and $S(X)$ is ancillary, then $S(X)$ and $T(X)$ are independent for all $\theta \in \Theta$.

In other words, a complete sufficient statistic is independent of any ancillary statistic.

Preliminary remarks (prior to proof):

Let $S = S(X)$, $T = T(X)$.

Let E_θ denote expectation w.r.t. P_θ.

1. The joint distribution of (S, T) depends on θ, so in general it is possible for S and T to be independent for **some** values of θ, but **not** for others. (Basu's Lemma says this does **not** happen in this case.)

2. For any rv's Y and Z, we know that $E(Y \mid Z) = g(Z)$, i.e., the conditional expectation is a function of Z.

   If the joint distribution of (Y, Z) depends on a parameter θ, then $E_\theta(Y \mid Z) = g(Z, \theta)$, i.e., the conditional expectoration is a function of both Z and θ. (However, this function may turn out to be constant in one or both variables.)

3. In general, $E(Y) = EE(Y \mid Z)$ and $E_\theta(Y) = E_\theta E_\theta(Y \mid Z)$.

4. To show that Y and Z are independent, it suffices to show that $\mathcal{L}(Y \mid Z) = \mathcal{L}(Y)$ which means that $P(Y \in A \mid Z) = P(Y \in A)$ for all (Borel) sets A.

   Let $w(Y) = I(Y \in A)$. Then $P(Y \in A) = E\,w(Y)$ and $P(Y \in A \mid Z) = E(w(Y) \mid Z)$. The indicator function $w(Y)$ is a bounded (Borel measurable) function. Therefore we have:

   To show that Y and Z are independent, it suffices to show $E(w(Y) \mid Z) = E\,w(Y)$ for all bounded (Borel measurable) functions.

5. Thus, to show that S and T are independent for all θ, it suffices to show that $E_\theta(w(S) \mid T) = E_\theta\,w(S)$ for all θ and all bounded (B.m.) functions $w(S)$.

Proof: Let $w(S)$ be a given bounded function of S. Consider both sides of the identity:

$$E_\theta w(S) = E_\theta \left[E_\theta(w(S) \mid T) \right] \quad \text{for all } \theta \ .$$

Consider the LHS. Since S is ancillary, the distribution of $w(S)$ is the same for all θ so that the LHS is constant in θ. Call this constant c.

Now consider the RHS. We know that $E_\theta(w(S) \mid T)$ will be some function of θ and T. However, since T is a sufficient statistic, $\mathcal{L}(X \mid T)$ does not depend on θ. Since $S = S(X)$, this implies $\mathcal{L}(S \mid T)$ does not depend on θ so that in turn $\mathcal{L}(w(S) \mid T)$ does not depend on θ. Thus, by sufficiency, $E_\theta(w(S) \mid T)$ is constant in θ and must be a function of T only. Call this function $\psi(T)$. That is,

$$\psi(T) = E_\theta(w(S)\,|\,T)\,.$$

The original identity can now be written as

$$\begin{aligned} c &= E_\theta\,\psi(T) \quad \text{for all } \theta \\ & \quad\ \text{or equivalently} \\ 0 &= E_\theta\,(\psi(T) - c) \quad \text{for all } \theta \quad . \end{aligned}$$

Since T is complete, this implies

$$\begin{aligned} P(\psi(T) - c = 0) &= 1 \quad \text{for all } \theta \\ & \quad\ \text{or equivalently} \\ \psi(T) &= c \quad \text{with probability 1 for all } \theta. \end{aligned}$$

This means

$$E_\theta(w(S)\,|\,T) = E_\theta\,w(S) \quad \text{with probability 1 for all } \theta.$$

Since $w(S)$ is an arbitrary bounded function, by the earlier discussion this implies S and T are independent for all θ.

QED

Applications of Basu's Lemma

Example:

Let $X = (X_1, \ldots, X_n)$ iid Uniform$(0, \theta)$.

Recall:

> $T(X) = X_{(n)} = \max X_i$ is a CSS.
>
> Uniform$(0, \theta)$ is a scale family, so any scale invariant statistic $S(X)$ is ancillary.

Thus, by Basu's lemma, all of the following are independent of $X_{(n)}$ for all θ:

$$\frac{\bar{x}}{s}, \quad \frac{X_{(1)}}{X_{(n)}}, \quad \text{the vector} \left(\frac{X_{(1)}}{X_{(n)}}, \ldots, \frac{X_{(n-1)}}{X_{(n)}} \right).$$

Example: (Using Basu's Lemma to obtain an indirect proof of completeness.)

Let $X = (X_1, \ldots, X_n)$ iid Uniform$(\theta, \theta + 1)$.

Recall:

> $T(X) = (X_{(1)}, X_{(n)})$ is a MSS.
>
> $S(X) = X_{(n)} - X_{(1)}$ is ancillary.

Since S is a function of T, the rv's S and T can**not** be independent. Thus T cannot be complete (for then we would get a contradiction with Basu's Lemma).

Example:

Let $X = (X_1, \ldots, X_n)$ be iid $N(\mu, \sigma^2)$.

Let $\bar{x}, s^2, Z$ be the sample mean, sample variance, and standardized residuals (z-scores) of the data X.

Reminder: $Z = (Z_1, \ldots, Z_n)$ with $Z_i = (X_i - \bar{x})/s$.

Fact: $\bar{x}$, s^2, Z are mutually independent.

Proof: We first show that the pair $(\bar{x}, s^2)$ is independent of Z, and then show that $\bar{x}$ and s^2 are independent. Each stage uses Basu's Lemma.

Stage 1: $(\bar{x}, s^2)$ is independent of Z.

Consider the family of all $N(\mu, \sigma^2)$ distributions (with both parameters allowed to vary).

Recall:

> $(\bar{x}, s^2)$ is a CSS.
>
> This is a location-scale family so that any location-scale invariant statistic is ancillary.
>
> Z is location-scale invariant.

Thus, Z is ancillary so that (by Basu's Lemma) it must be independent of $(\bar{x}, s^2)$ for all (μ, σ^2).

Stage 2: $\bar{x}$ and s^2 are independent.

Fix σ^2 at an arbitrary value σ_0^2 and consider the family of $N(\mu, \sigma_0^2)$ distributions, μ unknown.

Recall:

This is a 1pef and the natural SS $\sum_i X_i$ is a CSS. $\bar{x}$ is a 1-1 function of this and so also a CSS.

This is a location family so that any location invariant statistic is ancillary.

s^2 is location invariant.

Thus, s^2 is ancillary so (by Basu's Lemma) it must be independent of $\bar{x}$ for all μ (and also for all σ^2 since σ_0^2 is arbitrary).

QED

Example: (an amusing calculation via Basu)

Let $X = (X_1, \ldots, X_n)$ iid $N(0, \sigma^2)$, $\sigma^2 > 0$.

Goal: Compute ES where $S = \dfrac{\left(\sum_i X_i\right)^2}{\sum_i X_i^2}$.

This is a 1pef with $\theta = \sigma^2$, $t(x) = x^2$ and $w(\theta) = \frac{-1}{2\sigma^2}$. (Check this.)

Therefore $T(X) = \sum_i X_i^2$ is CSS.

This is also a scale family so that scale invariant statistics are ancillary.

S is scale invariant $\Rightarrow$ S ancillary $\Rightarrow$ (by Basu) S independent of T.

Thus $E(ST) = (ES)(ET) \Rightarrow ES = \dfrac{E(ST)}{ET}$ which becomes

$$ES = \frac{E\left[\left(\sum X_i\right)^2\right]}{E\left(\sum X_i^2\right)} = \frac{n\sigma^2}{n\sigma^2} = 1 \quad \text{for all } n \text{ and } \sigma^2.$$