Theorem 1 (Factorization Theorem). *Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and let $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{A})$. Suppose there exists a σ-finite measure μ on $(\mathcal{X}, \mathcal{A})$ such that each P_θ is absolutely continuous with respect to μ, and let $p_\theta(x) = \frac{dP_\theta}{d\mu}(x)$ denote the Radon-Nikodym derivative.*

A statistic $T : \mathcal{X} \to \mathcal{Y}$ is sufficient for θ if and only if there exist measurable functions $u : \mathcal{Y} \times \Theta \to [0, \infty)$ and $v : \mathcal{X} \to [0, \infty)$ such that for all $x \in \mathcal{X}$ and all $\theta \in \Theta$,

$$p_\theta(x) = u(T(x), \theta)\, v(x).$$

Proof. We will prove both directions of the equivalence.

($\Rightarrow$) Sufficiency implies factorization:

Assume that T is sufficient for θ. By the definition of sufficiency, the conditional distribution of X given $T(X) = t$ does not depend on θ. Formally, for any measurable set $B \in \mathcal{A}$ and for all $\theta \in \Theta$,

$$P_\theta(X \in B \mid T(X) = t) = P(X \in B \mid T(X) = t),$$

where the right-hand side is independent of θ.

Consider the conditional density of X given $T(X) = t$, which we denote by $k(x \mid t)$. Since this density does not depend on θ, we can write

$$p_\theta(x) = P_\theta(X = x) = P_\theta(T(X) = t, X = x) = P_\theta(X = x \mid T(X) = t)\, P_\theta(T(X) = t).$$

However, in continuous settings, we need to be careful with densities. Instead, we can use the following approach.

Let $f_{T,\theta}(t)$ be the marginal density of $T(X)$ under P_θ, and let $k(x \mid t)$ be the conditional density of X given $T(X) = t$, which does not depend on θ due to sufficiency. Then,

$$p_\theta(x) = f_{T,\theta}(t)\, k(x \mid t).$$

Since $k(x \mid t)$ does not depend on θ, we can let

$$v(x) = k(x \mid t),$$

and since $t = T(x)$, $v(x)$ is a function of x alone.

Now, define

$$u(t, \theta) = f_{T,\theta}(t).$$

Therefore, we have

$$p_\theta(x) = u(T(x), \theta)\, v(x).$$

($\Leftarrow$) Factorization implies sufficiency:

Assume that there exist measurable functions u and v such that

$$p_\theta(x) = u(T(x), \theta)\, v(x).$$

We need to show that T is sufficient for θ.

Consider any measurable set $B \in \mathcal{A}$. We aim to show that the conditional distribution of X given $T(X) = t$ does not depend on θ.

First, compute the conditional probability density of X given $T(X) = t$ under P_θ:

$$P_\theta(X \in B \mid T(X) = t) = \frac{\int_{B \cap T^{-1}(t)} p_\theta(x)\, d\mu(x)}{\int_{T^{-1}(t)} p_\theta(x)\, d\mu(x)}.$$

Substitute $p_\theta(x) = u(T(x), \theta)\, v(x)$:

$$P_\theta(X \in B \mid T(X) = t) = \frac{\int_{B \cap T^{-1}(t)} u(t, \theta)\, v(x)\, d\mu(x)}{\int_{T^{-1}(t)} u(t, \theta)\, v(x)\, d\mu(x)}.$$

Since $u(t, \theta)$ is constant with respect to x over $T^{-1}(t)$, it can be factored out:

$$P_\theta(X \in B \mid T(X) = t) = \frac{u(t, \theta) \int_{B \cap T^{-1}(t)} v(x)\, d\mu(x)}{u(t, \theta) \int_{T^{-1}(t)} v(x)\, d\mu(x)}.$$

The $u(t, \theta)$ terms cancel out:

$$P_\theta(X \in B \mid T(X) = t) = \frac{\int_{B \cap T^{-1}(t)} v(x)\, d\mu(x)}{\int_{T^{-1}(t)} v(x)\, d\mu(x)}.$$

This expression is independent of θ, which means that the conditional distribution of X given $T(X) = t$ does not depend on θ. Therefore, T is sufficient for θ. $\square$