# factorization criterion

| | |
|---|---|
| Canonical name | FactorizationCriterion |
| Date of creation | 2013-03-22 15:02:48 |
| Last modified on | 2013-03-22 15:02:48 |
| Owner | CWoo (3771) |
| Last modified by | CWoo (3771) |
| Numerical id | 4 |
| Author | CWoo (3771) |
| Entry type | Theorem |
| Classification | msc 62B05 |
| Synonym | factorization theorem |
| Synonym | Fisher-Neyman factorization theorem |

Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random vector whose coordinates are observations, and whose probability (density) function is, $f(\boldsymbol{x} \mid \theta)$ where $\theta$ is an unknown parameter. Then a statistic $T(\boldsymbol{X})$ for $\theta$ is a sufficient statistic iff $f$ can be expressed as a product of (or *factored into*) two functions $g, h$, $f = gh$ where $g$ is a function of $T(\boldsymbol{X})$ and $\theta$, and $h$ is a function of $\boldsymbol{x}$. In symbol, we have

$$f(\boldsymbol{x} \mid \theta) = g(T(\boldsymbol{X}), \theta) h(\boldsymbol{x}).$$

**Applications**.

1. In view of the above statement, let's show that the sample mean $\overline{X}$ of $n$ independent observations from a normal distribution $N(\mu, \sigma^2)$ is a sufficient statistic for the unknown mean $\mu$. Since the $X_i$'s are independent random variables, then the probability density function $f(\boldsymbol{x} \mid \mu)$, being the joint probability density function of each of the $X_i$, is the product of the individual density functions $f(x \mid \mu)$:

$$\begin{aligned}
f(\boldsymbol{x} \mid \mu) &= \prod_{i=1}^{n} f(x \mid \mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] && (1)\\
&= \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} \exp\left[\sum_{i=1}^{n} -\frac{(x_i-\mu)^2}{2\sigma^2}\right] && (2)\\
&= \frac{1}{\sqrt{(2\pi)^n \sigma^{2n}}} \exp\left[\frac{-1}{2\sigma^2} \sum_{i=1}^{n} x_i^2\right] \exp\left[\frac{\mu}{\sigma^2} \sum_{i=1}^{n} x_i - \frac{n\mu^2}{2\sigma^2}\right] && (3)\\
&= h(\boldsymbol{x}) \exp\left[\frac{n\mu}{\sigma^2} T(\boldsymbol{x}) - \frac{n\mu^2}{2\sigma^2}\right] && (4)\\
&= h(\boldsymbol{x}) g(T(\boldsymbol{x}), \mu) && (5)
\end{aligned}$$

   where $g$ is the last exponential expression and $h$ is the rest of the expression in (3). By the factorization criterion, $T(\boldsymbol{X}) = \overline{X}$ is a sufficient statistic.

2. Similarly, the above shows that the sample variance $s^2$ is not a sufficient statistic for $\sigma^2$ if $\mu$ is unknown.

3. But, if $\mu$ is a known constant, then the statistic

$$T(X_1, \ldots, X_n) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \mu)^2$$

is sufficient for $\sigma^2$ by observing in (2) above, and letting $h(\boldsymbol{x}) = 1$ and $g(T, \sigma^2)$ be all of expression (2).