# Definition

Let the rv’s form a random sample from a discrete or a continuous distribution whose p.f. or p.d.f. is .

For every observed vector in the sample, the value of the **joint** p.f. or the joint p.d.f. is denoted as . Note that random variables from a random sample is considered to be independent. Thus we have

To simplify optimization works, we yield the **log-likelihood function**, which is also called cross-entropy

Our optimization goal is

借用某博客中的一句话。对于一个模型已定、样本已有、参数未知的模型（这就是statistical learning），MLE就是找出一组参数，使得模型产生出样本的概率最大。

# Examples

有待补充