先验概率，条件概率与后验概率

先验概率是基于背景常识或者历史数据的统计得出的预判概率，一般只包含一个变量，例如 $P(X)$， $P(Y)$ 。

条件概率是表示一个事件发生后另一个事件发生的概率，例如 $P(Y|X)$ 代表 X 事件发生后 Y 事件发生的概率。

后验概率是由果求因，也就是在知道结果的情况下求原因的概率，例如Y事件是X引起的，那么 $P(X|Y)$ 就是后验概率，也可以说它是事件发生后的反向条件概率。

似然函数

在数理统计学中，似然函数是一种关于统计模型中的参数的函数，表示模型参数中的似然性。似然函数可以理解为条件概率的逆反。

在已知某个参数 α 时，事件 A 会发生的条件概率可以写作 $P(A;\alpha)$ ，也就是 $P(A|\alpha)$ 。我们也可以构造似然性的方法来表示事件 A 发生后估计参数 α 的可能性，也就表示为 $L(\alpha|A)$ ，其中 $L(\alpha|A)=P(A|\alpha)$ 。

这里Wikipedia的解释比较全面详细，可以参见似然函数。

最大似然估计（MLE）与最大后验概率（MAP）

最大似然估计是似然函数最初也是最自然的应用。似然函数取得最大值表示相应的参数能够使得统计模型最为合理。从这样一个想法出发，最大似然估计的做法是：首先选取似然函数（一般是概率密度函数或概率质量函数），整理之后求最大值。实际应用中一般会取似然函数的对数作为求最大值的函数，这样求出的最大值和直接求最大值得到的结果是相同的。似然函数的最大值不一定唯一，也不一定存在。

这里简单的说一下最大后验概率（MAP），如下面的公式

$$P(\alpha|X)=\frac{P(X|\alpha)P(\alpha)}{P(X)}$$

其中等式左边 $P(\alpha|X)$ 表示的就是后验概率，优化目标即为 $argmax_\alpha P(\alpha|X)$ ，即给定了观测值 X 以后使模型参数 α 出现的概率最大。等式右边的分子式 $P(X|\alpha)$ 即为似然函数 $L(\alpha|X)$，MAP考虑了模型参数 α 出现的先验概率 $P(\alpha)$ 。即就算似然概率 $P(X|\alpha)$ 很大，但是 α 出现的可能性很小，也更倾向于不考虑模型参数为 α 。

生成式模型与判别式模型

最后简单说一下生成式模型与判别式模型。

判别式模型学习的目标是条件概率 $P(Y|X)$ 或者是决策函数 $Y=f(X)$ ，其实这两者本质上是相同的。例如KNN，Decision Tree，SVM，CRF等模型都是判别式模型。

生成式模型学习的是联合概率分布 $P(X,Y)$ ，从而求得条件概率分布 $P(Y|X)$ 。例如NB，HMM等模型都是生成式模型。

参考

- 似然函数
- 最大似然估计 （MLE） 最大后验概率（MAP）
- 《统计学习方法》李航