# Fisher打分法

尽管最大似然方法应用实例可能会给出不同的估计式，但通常无法给出最大似然估计量的显式公式。在这种情况下，我们需要使用数值近似方法。

对于给定的观测值 $x$，似然函数 $\theta \mapsto L(\theta;x)$ 是参数 $\theta$ 的一个"正常"函数，而我们需要找到使该函数达到最大值的 $\theta$ 值。我们可以使用牛顿-拉夫森（Newton-Raphson）方法，或者该方法的变体，在统计学中被称为 Fisher 打分法（Fisher's Scoring）。现在我们简要描述这些数值方法。

在大多数情况下，所需的值 $\hat{\theta}$ 是对数似然函数相对于 $\theta$ 的导数的驻点。因此，我们讨论找到函数 $\theta \mapsto \dot{\Lambda}(\theta;x)$ 的零点 $\hat{\theta}$，其中 $\dot{\Lambda}$ 是对数似然函数的偏导数向量 $\theta \mapsto \Lambda(\theta;x) = \log L(\theta;x)$。牛顿-拉夫森方法的思想是从 $\hat{\theta}$ 的"合理"初始估计值 $\tilde{\theta}_0$ 出发，并用线性近似替换函数 $\dot{\Lambda}$, 这里要注意的是初始值的选择非常重要, 否则将会造成比较严重的误差：

$$\dot{\Lambda}(\theta;x) \approx \dot{\Lambda}\left(\tilde{\theta}_0;x\right) + \ddot{\Lambda}\left(\tilde{\theta}_0;x\right)\left(\theta - \tilde{\theta}_0\right)$$

这里，$\ddot{\Lambda}(\theta;x)$ 是对数似然函数相对于参数的二阶导数矩阵。我们不再寻找 $\dot{\Lambda}(\theta;x) = 0$ 的解，而是转向求解方程 $\dot{\Lambda}\left(\tilde{\theta}_0;x\right) + \ddot{\Lambda}\left(\tilde{\theta}_0;x\right)\left(\theta - \tilde{\theta}_0\right) = 0$。该方程的零点为

$$\tilde{\theta}_1 = \tilde{\theta}_0 - \ddot{\Lambda}\left(\tilde{\theta}_0;x\right)^{-1}\dot{\Lambda}\left(\tilde{\theta}_0;x\right)$$

由于线性近似并不精确，$\tilde{\theta}_1$ 通常不是期望的零点 $\hat{\theta}$。然而，我们预计 $\tilde{\theta}_1$ 比初始值 $\tilde{\theta}_0$ 更接近 $\hat{\theta}$。然后我们将 $\tilde{\theta}_1$ 作为初始值，计算第三个值，依此类推。这就形成了一个近似序列 $\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2, \ldots$，在某些条件下，该序列收敛于零点 $\hat{\theta}$。如果初始值 $\tilde{\theta}_0$ 足够接近目标值 $\hat{\theta}$ 并且函数 $\dot{\Lambda}$ 足够光滑，收敛是可以保证的，但在实践中我们当然无法保证这一点。

对算法的各种不同修改可以使收敛更可靠。然而，如果这个对数似然函数有多个局部极大值和/或极小值，则需要小心，首先得一种坏情况就是收敛到 $\dot{\Lambda}$ 的另一个零点（对应于局部极大值或极小值）, 此外，序列 $\tilde{\theta}_0, \tilde{\theta}_1, \tilde{\theta}_2, \ldots$ 还可能发散。

在之后章节中，我们将看到，对数似然函数在最大似然估计量 $\hat{\theta}$ 处的二阶导数 $\ddot{\Lambda}(\hat{\theta};x)$ 具有特殊意义。该二阶导数称为观测信息，与 Fisher 信息近似相等。在牛顿-拉夫森算法中，有时使用另一种矩阵代替二阶导数。比如这里如果如果使用 Fisher 信息，则该算法称为 Fisher 打分法。值得一提的是,当利用 Fisher 信息解析计算时，这种方法特别有趣。