== 贝叶斯分析 ==

在德国坦克问题中，贝叶斯方法要考虑当观察到的坦克总数<math>\scriptstyle K</math>等于数<math>\scriptstyle k</math>、序列号最大值<math>\scriptstyle M</math>等于数<math>\scriptstyle m</math>时敌方坦克总数<math>\scriptstyle N</math>等于数<math>\scriptstyle n</math>的可信度<math>\scriptstyle (N=n\mid M=m, K=k)</math>

为了简便起见，以下将<math>\scriptstyle (N=n\mid M=m,K=k)</math>写成<math>\scriptstyle (n\mid m,k)</math>。

[[条件概率]]的法则会给出

:<math>(n\mid m,k) = (m\mid n,k)\frac {(n\mid k)}{(m\mid k)}</math>

表达式<math>\scriptstyle (m\mid n,k)=(M=m\mid N=n,K=k)</math>是当敌方坦克总数等于已知的<math>\scriptstyle n</math>、且观察到了<math>\scriptstyle k</math>辆敌方坦克时，所观察到的序列号最大值等于<math>\scriptstyle m</math>的条件概率。其为

:<math>

(m\mid n,k) =

\begin{cases}

\frac{\binom{m - 1}{k - 1}}{\binom{n}{k}} &\text{if } k \le m \le n\\

0 &\text{otherwise}

\end{cases}

</math>

其中[[二项式系数]]<math>\scriptstyle \binom n k</math>是从总数为<math>\scriptstyle n</math>的总体中取<math>\scriptstyle k</math>个样本的方法数。

表达式<math>\scriptstyle (m\mid k)=(M=m\mid K=k)</math>是在真正观测前，假设观察了''k''辆坦克后，最大序号等于''m''的概率。通过合并所有可能的<math>\scriptstyle n</math>，<math>\scriptstyle (m\mid k)</math>可以改写为其它形式。

:<math>\begin{align}

(m\mid k)

&=(m\mid k)\cdot 1 \\

&=(m\mid k){\sum\_{n=0}^\infty(n\mid m,k)} \\

&=(m\mid k){\sum\_{n=0}^\infty(m\mid n,k)\frac {(n\mid k)}{(m\mid k)}} \\

&=\sum\_{n=0}^\infty(m\mid n,k)(n\mid k)

\end{align}</math>

表达式<math>\scriptstyle (n\mid k)=(N=n\mid K=k)</math>为当已观察到''k''辆坦克、但序列号尚未统计前坦克总数等于''n''的可信度。假定它是某种[[离散均匀分布]]

:<math>

(n\mid k) =

\begin{cases}

\frac 1{\Omega - k} &\text{if } k \le n < \Omega \\

0 &\text{otherwise}

\end{cases}

</math>

上限<math>\Omega</math>必须是有限的，因为该函数

:<math>

f(n)=\lim\_{\Omega\rarr\infty}

\begin{cases}

\frac 1{\Omega - k} &\text{if } k \le n < \Omega \\

0 &\text{otherwise}

\end{cases}

</math>

的结果为:<math>f(n) = 0</math>，而这不是一个概率质量函数。

因而

:<math>

(n\mid m,k) =

\begin{cases}

\frac{(m|n,k)}{\sum\_{n=m}^{\Omega - 1} (m|n,k)} &\text{if } m \le n < \Omega \\

0 &\text{otherwise}

\end{cases}

</math>

如果<math>\scriptstyle \sum\_{n=m}^\infty(m|n,k)<\infty</math>，那么不受欢迎的变量<math>\scriptstyle \Omega</math>就能从表达式中消失。

:<math>

(n\mid m,k) =

\begin{cases}

0 &\text{if } n < m \\

\frac{(m|n,k)}{\sum\_{n=m}^\infty(m|n,k)} &\text{if } n\ge m

\end{cases}

</math>

当''k'' ≥ 1时，敌方坦克数量分布的[[众数 (数学)|众数]]为''m''。

当''k'' ≥ 2时，敌方坦克数量“等于”<math>n</math>的可信度为

:<math>

(N=n|M=m\ge k,K=k\ge 2) =

\begin{cases}

0 &\text{if } n < m \\

\frac {k - 1}{k }\frac {\binom{m - 1}{k - 1}}{\binom n k} &\text{if } n \ge m

\end{cases}

</math>

而敌方坦克数量<math>\scriptstyle N</math>“大于”<math>\scriptstyle n</math>的可信度为

:<math>

(N > n|M = m \ge k , K = k \ge 2) =

\begin{cases}

1 &\text{if } n < m \\

\frac {\binom{m - 1}{k - 1}}{\binom n {k - 1}} &\text{if } n \ge m

\end{cases}

</math>

当''k'' ≥ 3时，<math>N</math>的[[平均数|均值]]有限：

:<math>\frac{(m - 1)(k - 1)}{k - 2}</math>

当''k'' ≥ 4时，<math>\scriptstyle N</math>的[[标准差]]有限：

:<math>\sqrt{\frac{(m - 1)(k - 1)(m + 1 - k)}{(k - 2)^2(k - 3)}}</math>

这些公式将在下面推导。

=== 求和公式 ===

下面的等式用[[二项式系数#涉及二项式系数的恒等式|二项式系数表示法]]来简化德国坦克问题中的[[级数]]。

:<math>\sum\_{n=m}^\infty \frac 1 {\binom n k} = \frac k{k - 1}\frac 1 {\binom{m - 1}{k - 1}}</math>

这个求和公式有点类似于积分公式

:<math>\int\_{n=m}^\infty \frac {dn}{n^k} = \frac 1{k - 1}\frac 1{m^{k - 1}}</math>

这些公式在''k'' > 1时适用。

=== 一辆坦克 ===

从''n''辆坦克的总体中随机观察一辆坦克，当''m'' ≤ ''n''时，其序列号为''m''的概率为1/''n''，而当''m'' > ''n''时概率是零。用[[艾弗森括号]]表示法可写成

:<math>(M=m|N=n,K=1) = (m|n) = \frac{[m \le n]}{n}</math>

这是<math>\scriptstyle m</math>的条件概率质量分布函数。

当''m''为定值时，这是一个''n''的[[似然函数]]。

:<math>\mathcal{L}(n) = \frac{[n\ge m]}{n}</math>

[[最大似然估计]]的坦克总数为''N''<sub>0</sub> = ''m''。

总概率为[[无穷|无穷大]]，因为尾部为一个[[调和级数]]数列。

:<math>\sum\_n \mathcal{L}(n) = \sum\_{n=m}^\infty \frac{1}{n} = \infty</math>

但

:<math>\begin{align}

\sum\_n \mathcal{L}(n)[n < \Omega]

&= \sum\_{n=m}^{\Omega - 1} \frac{1}{n} \\

&= H\_{\Omega-1} - H\_{m - 1}

\end{align}</math>

其中<math>H\_n</math>为[[调和数]]。

可信度质量分布函数依赖于先前的限制<math>\scriptstyle \Omega</math>：

:<math>\begin{align}

&(N=n\mid M=m,K=1) \\

= {} &(n|m) = \frac{[m\le n]}{n} \frac{[n<\Omega]}{H\_{\Omega - 1} - H\_{m - 1}}

\end{align}</math>

<math>\scriptstyle N</math>的均值为

:<math>\begin{align}

\sum\_n n\cdot(n|m) &= \sum\_{n=m}^{\Omega - 1} \frac{1}{H\_{\Omega - 1} - H\_{m - 1}} \\

&= \frac{\Omega - m}{H\_{\Omega - 1} - H\_{m - 1}} \\

&\approx \frac{\Omega - m}{\log\left(\frac{\Omega - 1}{m - 1}\right)}

\end{align}</math>

=== 两辆坦克 ===

如果观察到了两辆坦克，而不是一辆，那么所观察到的两个序列号中较大值为''m''的概率为

:<math>(M=m\mid N=n,K=2) = (m|n) = [m \le n]\frac{m - 1}{\binom{n}{2}}</math>

当''m''为定值时，这是一个''n''的[[似然函数]]

:<math>\mathcal{L}(n) = [n \ge m]\frac{m - 1}{\binom{n}{2}}</math>

总概率为

:<math>\begin{align}

\sum\_{n}\mathcal{L}(n) &= \frac{m - 1}{1} \sum\_{n=m}^\infty \frac{1}{\binom n 2} \\

&= \frac{m - 1}{1} \cdot \frac{2}{2 - 1} \cdot \frac{1}{\binom {m - 1}{2 - 1}} \\

&= 2

\end{align}</math>

可信度质量分布函数为

:<math>\begin{align}

&(N=n\mid M=m,K=2) \\

= {} &(n\mid m) \\

= {} &\frac{\mathcal{L}(n)}{\sum\_n \mathcal{L}(n)} \\

= {} &[n \ge m]\frac{m - 1}{n(n - 1)}

\end{align}</math>

[[中位数]]<math>\scriptstyle \tilde{N}</math>满足

:<math>\sum\_n [n \ge \tilde{N}](n|m) = \frac{1}{2}</math>

所以

:<math>\frac{m - 1}{\tilde N - 1} = \frac{1}{2}</math>

因而中位数为

:<math>\tilde{N} = 2m - 1</math>

但''N''的均值为无穷大

:<math>\mu = \sum\_n n \cdot (n|m) = \frac{m - 1}1\sum\_{n=m}^\infty \frac{1}{n - 1} = \infty</math>

=== 多辆坦克 ===

==== 可信度质量分布函数 ====

在序列号{1,...,''n''}中观测到的最大值为''k''的条件概率''m''为

:<math>\begin{align}

&(M=m|N=n,K=k\ge 2) \\

= {} &(m\mid n,k) \\

= {} &[m\le n]\frac{\binom{m - 1}{k - 1}}{\binom{n}{k}}

\end{align}</math>

''n''的似然函数表达式与此相同

:<math>\mathcal{L}(n) = [n \ge m]\frac{\binom{m - 1}{k - 1}}{\binom{n}{k}}</math>

''k'' ≥ 2的总概率为一个有限值：

:<math>\begin{align}

\sum\_n \mathcal{L}(n)

&= \frac{\binom{m - 1}{k - 1}}{1} \sum\_{n=m}^\infty {1 \over \binom n k} \\

&= \frac{\binom{m - 1}{k - 1}}{1} \cdot \frac{k}{k-1} \cdot \frac{1}{\binom{m - 1}{k - 1}} \\

&= \frac k{k - 1}

\end{align}</math>

可信度质量分布函数为

:<math>\begin{align}

&(N=n|M=m,K=k \ge 2) = (n|m,k) \\

= {} &\frac{\mathcal{L}(n)}{\sum\_n \mathcal{L}(n)} \\

= {} &[n\ge m]\frac{k-1}{k} \frac{\binom{m - 1}{k - 1}}{\binom n k} \\

= {} &[n\ge m]\frac{m-1}{n} \frac{\binom{m - 2}{k - 2}}{\binom{n - 1}{k - 1}} \\

= {} &[n\ge m]\frac{m-1}{n} \frac{m - 2}{n - 1} \frac{k - 1}{k - 2} \frac{\binom{m - 3}{k - 3}}{\binom{n-2}{k-2}}

\end{align}</math>

[[互补累积分布函数]]是''N'' > ''x''的可信度

:<math>\begin{align}

&(N>x\mid M=m,K=k) \\

= {} &\begin{cases}

1 &\text{if }x < m \\

\sum\_{n=x+1}^\infty (n|m,k) &\text{if }x \ge m

\end{cases} \\

= {} &[x<m] + [x \ge m]\sum\_{n=x+1}^\infty \frac{k - 1}{k}\frac{\binom{m - 1}{k - 1}}{\binom{N}{k}} \\

= {} &[x<m] + [x \ge m]\frac{k - 1}{k} \frac{\binom{m - 1}{k - 1}}{1} \sum\_{n=x+1}^\infty \frac{1}{\binom{n}{k}} \\

= {} &[x<m] + [x \ge m]\frac{k - 1}{k} \frac{\binom{m - 1}{k - 1}}{1} \cdot \frac{k}{k - 1} \frac{1}{\binom{x}{k - 1}} \\

= {} &[x<m] + [x \ge m]\frac{\binom{m - 1}{k - 1}}{\binom{x}{k - 1}}

\end{align}</math>

[[累积分布函数]]是''N'' ≤ ''x''的可信度

:<math>\begin{align}

&(N\le x|M=m,K=k) \\

= {} &1 - (N>x\mid M=m,K=k) \\

= {} &[x \ge m]\left(1 - \frac{\binom{m - 1}{k - 1}}{\binom{x}{k - 1}}\right)

\end{align}</math>

==== 数量级 ====

敌方坦克数目的数量级为

:<math>\begin{align}

\mu &= \sum\_n n\cdot(N=n|M=m,K=k) \\&

= \sum\_n n [n\ge m]\frac {m-1}n \frac {\binom{m-2}{k-2}}{\binom{n-1}{k-1}} \\&

= \frac{m-1}1 \frac{\binom{m-2}{k-2}}1\sum\_{n=m}^\infty \frac 1{\binom{n-1}{k-1}}\\&

= \frac{m-1}1 \frac{\binom{m-2}{k-2}}1 \cdot \frac{k-1}{k-2}\frac {1}{\binom{m-2}{k-2}}\\&

= \frac{m-1}1 \frac{k-1}{k-2}

\end{align}</math>

==== 统计不确定度 ====

统计的不确定度用标准差''σ''表示，其满足等式

:<math>\sigma^2 + \mu^2 = \sum\_n n^2 \cdot (N=n\mid M=m,K=k)</math>

所以

:<math>\begin{align}

\sigma^2+\mu^2-\mu &= \sum\_n n(n-1)\cdot(N=n\mid M=m,K=k)\\&

= \sum\_{n=m}^\infty n(n-1)\frac{m-1}n \frac{m-2}{n-1} \frac{k-1}{k-2} \frac{\binom{m-3}{k-3}}{\binom{n-2}{k-2}}\\&

= \frac{m-1}1 \frac{m-2}1 \frac{k-1}{k-2} \cdot \frac{\binom{m-3}{k-3}}1 \sum\_{n=m}^\infty \frac 1{\binom{n-2}{k-2}}\\

& = \frac{m-1}1 \frac{m-2}1 \frac{k-1}{k-2} \frac{\binom{m-3}{k-3}}1 \frac{k-2}{k-3} \frac 1{\binom{m-3}{k-3}}\\

& = \frac{m-1}1 \frac{m-2}1 \frac{k-1}{k-3}\\&

\end{align}</math>

及

:<math>\begin{align}

\sigma &= \sqrt{\frac{m-1}1 \frac{m-2}1 \frac{k-1}{k-3}+\mu-\mu^2} \\&

= \sqrt{\frac{(k-1)(m-1)(m-k+1)}{(k-3)(k-2)^2}} \\&

\end{align}</math>

[[方差均值比]]则为

:<math>\frac{\sigma^2}\mu = \frac{m - k + 1}{(k - 3)(k - 2)}</math>