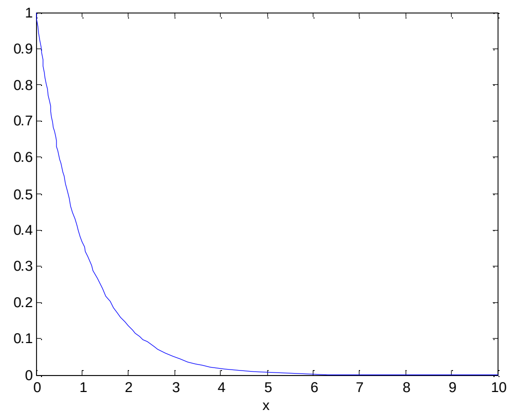
第三章 最大似然估计和贝叶斯参数估计

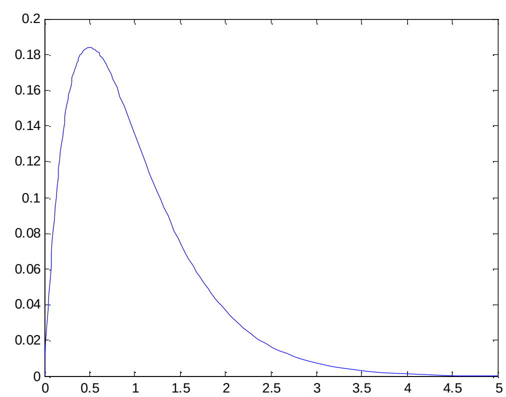
1、指数概率密度函数的分布：

$$p(x|q)=\begin{cases} q e^{-qx}, & x\ge 0 \\ 0, & others \end{cases}$$

a) $q=1$时，



$x=2$时，



b) $x_1,\cdots,x_n \sim p(x|q)$

定义对数似然函数：

$$l(q)=\sum_{i=1}^{n}\ln p(x_i|q)=\sum_{i=1}^{n}(\ln q-q x_i)$$

计算导数：

$$\frac{dl(q)}{dq}=\sum_{i=1}^{n}\left(\frac{1}{q}-x_i\right)=\frac{n}{q}-\sum_{i=1}^{n}x_i=0$$

因此：

$$\hat{q} = \frac{1}{n\sum_{i=1}^{n} x_i}$$

c) $q = 1$，当 n 非常大时，

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \int_{-\infty}^{+\infty} xp(x)\,dx$$

$$= \int_{0}^{+\infty} xe^{-x}dx = -(x+1)e^{-x}\Big|_{0}^{+\infty} = 1$$

因此：

$$\hat{q} = \frac{1}{\lim_{n\to\infty} n\sum_{i=1}^{n} x_i} = 1$$

2． x 服从均匀分布：

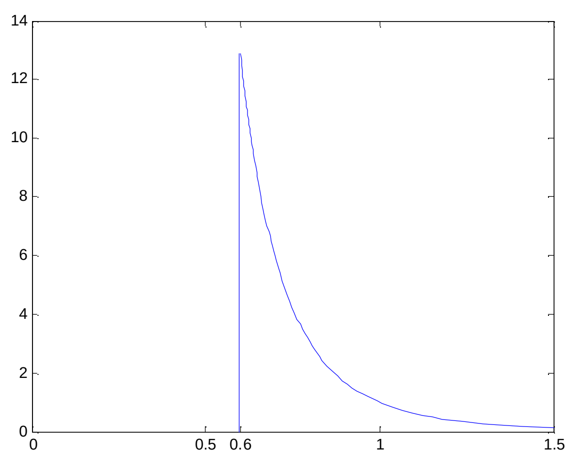
$$p(x|q) = \begin{cases} \frac{1}{q}, & 0 \le x \le q \\ 0, & others \end{cases}$$

a) 构造似然函数：

$$l(q) = \begin{cases} \prod_{i=1}^{n} \frac{1}{q}, & q \ge \max(D) \\ 0, & others \end{cases} = \begin{cases} \frac{1}{q^n}, & q \ge \max(D) \\ 0, & others \end{cases}$$

显然，当 $q \ge \max(D)$ 时，$l(q)$ 为 q 的单调下降函数，而 $q < \max(D)$ 时，$l(q) = 0$，因此 $l(q)$ 的最大值产生在 $\{q \,|\, \max(D)\}$ 的最小值处，即 $q = \max(D)$。

b)



4．$\mathbf{x}$ 服从 d 维的Bernoulli分布：

$$p(\mathbf{x}|\boldsymbol{\theta})=\prod_{i=1}^{d}q_i^{x_i}\left(1-q_i\right)^{1-x_i}$$

因此：

$$p\left(\mathbf{x}^1,\cdots,\mathbf{x}^n\middle|\boldsymbol{\theta}\right)=\prod_{k=1}^{n}\prod_{i=1}^{d}q_i^{x_i^k}\left(1-q_i\right)^{1-x_i^k}$$

对数似然函数：

$$l(\boldsymbol{\theta})=\sum_{k=1}^{n}\sum_{i=1}^{d}\left[x_i^k\ln q_i+\left(1-x_i^k\right)\ln\left(1-q_i\right)\right]$$

$$\frac{\partial l(\boldsymbol{\theta})}{\partial q_i}=\sum_{k=1}^{n}\left[\frac{x_i^k}{q_i}-\frac{1-x_i^k}{1-q_i}\right]=\sum_{k=1}^{n}\frac{x_i^k-q_i}{q_i\left(1-q_i\right)}=0$$

因此有：

$$\sum_{k=1}^{n}\left(x_i^k-q_i\right)=0$$

$$q_i=\frac{1}{n}\sum_{k=1}^{n}x_i^k$$

写成矢量形式：

$$\hat{\boldsymbol{\theta}}=\frac{1}{n}\sum_{k=1}^{n}\mathbf{x}_k$$

16．$\mathbf{A},\mathbf{B}$ 为两个同阶的非奇异矩阵。

a) $$\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)\left[\mathbf{A}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}\right]=\left(\mathbf{I}+\mathbf{B}^{-1}\mathbf{A}\right)\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}$$

$$=\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}+\mathbf{B}^{-1}\mathbf{A}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}$$

$$=\mathbf{B}^{-1}\mathbf{B}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}+\mathbf{B}^{-1}\mathbf{A}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}$$

$$=\mathbf{B}^{-1}\left(\mathbf{B}+\mathbf{A}\right)\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B}$$

$$=\mathbf{B}^{-1}\mathbf{B} \qquad \text{(因为}\left(\mathbf{B}+\mathbf{A}\right)=\left(\mathbf{A}+\mathbf{B}\right)\text{)}$$

$$=\mathbf{I}$$

上式两边左乘$\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}$，则有：

$$\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}=\mathbf{A}\left(\mathbf{A}+\mathbf{B}\right)^{-1}\mathbf{B} \qquad \text{(要求}\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)\text{为非奇异矩阵)}$$

同理：

$$\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)=\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\left(\mathbf{I}+\mathbf{A}\mathbf{B}^{-1}\right)$$

$$=\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}+\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\mathbf{B}^{-1}$$

$$=\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{B}\mathbf{B}^{-1}+\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}\mathbf{B}^{-1}$$

$$=\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}(\mathbf{B}+\mathbf{A})\mathbf{B}^{-1}$$

$$=\mathbf{B}\mathbf{B}^{-1}=\mathbf{I}$$

两边右乘$\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)^{-1}$，有：

$$\left(\mathbf{A}^{-1}+\mathbf{B}^{-1}\right)=\mathbf{B}(\mathbf{A}+\mathbf{B})^{-1}\mathbf{A}$$

b) 当$\mathbf{A},\mathbf{B}$非方阵时，$\mathbf{A},\mathbf{B}$均为奇异阵，所以$\mathbf{A}^{-1}$和$\mathbf{B}^{-1}$均不存在，因此以上恒等式部成立。

另外：

对于奇异阵$\mathbf{A}$和$\mathbf{B}$存在伪逆矩阵(广义逆矩阵)，伪逆矩阵定义为：

$\mathbf{A}$为$m\times n$的矩阵，当$\mathbf{A}$的秩为n时：$\operatorname{rank}(\mathbf{A})=n$，$\mathbf{A}$的伪逆矩阵定义为：

$$\mathbf{A}^{+}=\left(\mathbf{A}^{t}\mathbf{A}\right)^{-1}\mathbf{A}^{t}$$

当$\mathbf{A}$的秩为m时：$\operatorname{rank}(\mathbf{A})=m$，$\mathbf{A}$的伪逆矩阵定义为：

$$\mathbf{A}^{+}=\mathbf{A}^{t}\left(\mathbf{A}\mathbf{A}^{t}\right)^{-1}$$

伪逆矩阵的性质：

当$\operatorname{rank}(\mathbf{A})=n$时，有：$\mathbf{A}^{+}\mathbf{A}=\mathbf{I}_n$

当$\operatorname{rank}(\mathbf{A})=m$时，有：$\mathbf{A}\mathbf{A}^{+}=\mathbf{I}_m$

但可以验证当以伪逆矩阵代替上式中的逆阵时，以上恒等式不成立。（如令$\mathbf{A}=\begin{bmatrix}8 & 0 & 4\\ 9 & 2 & 8\end{bmatrix},\mathbf{B}=\begin{bmatrix}0 & 12 & 99\\ 9 & 7 & 8\end{bmatrix}$，可以验证$\left(\mathbf{A}^{+}+\mathbf{B}^{+}\right)^{+}\neq\mathbf{A}(\mathbf{A}+\mathbf{B})^{+}\mathbf{B}$）

c) 由(41)式：

$$\boldsymbol{\Sigma}_n^{-1}=n\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}_0^{-1}$$

等式两边求逆：

$$\boldsymbol{\Sigma}_n=\left(n\boldsymbol{\Sigma}^{-1}+\boldsymbol{\Sigma}_0^{-1}\right)^{-1}=\left(\boldsymbol{\Sigma}_0^{-1}+n\boldsymbol{\Sigma}^{-1}\right)^{-1}$$

等式右边利用(a)的恒等式，有：

$$\boldsymbol{\Sigma}_n = \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma} \right)^{-1} \left(\frac{1}{n}\boldsymbol{\Sigma} \right) \tag{*}$$

同理有：

$$\boldsymbol{\Sigma}_n = \left(n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\Sigma}_0^{-1} \right)^{-1} = \left(\frac{1}{n}\boldsymbol{\Sigma} \right)\left(\frac{1}{n}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_0 \right)^{-1} \boldsymbol{\Sigma}_0 \tag{**}$$

由(42)式：

$$\boldsymbol{\Sigma}_n^{-1}\boldsymbol{\mu}_n = n\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0$$

等式两边左乘 $\boldsymbol{\Sigma}_n$：

$$\boldsymbol{\mu}_n = \boldsymbol{\Sigma}_n \left(n\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_n + \boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 \right)$$

$$= \boldsymbol{\Sigma}_n n\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_n + \boldsymbol{\Sigma}_n\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0$$

$$= \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma} \right)^{-1} \left(\frac{1}{n}\boldsymbol{\Sigma} \right) n\boldsymbol{\Sigma}^{-1}\hat{\boldsymbol{\mu}}_n + \left(\frac{1}{n}\boldsymbol{\Sigma} \right)\left(\frac{1}{n}\boldsymbol{\Sigma} + \boldsymbol{\Sigma}_0 \right)^{-1} \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\mu}_0 \qquad \text{(代入(*)式)}$$

$$= \boldsymbol{\Sigma}_0 \left(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma} \right)^{-1} \hat{\boldsymbol{\mu}}_n + \frac{1}{n}\boldsymbol{\Sigma} \left(\boldsymbol{\Sigma}_0 + \frac{1}{n}\boldsymbol{\Sigma} \right)^{-1} \boldsymbol{\mu}_0 \qquad \text{（***）}$$

其中(*)式为(46)式，(***)式为(45)式。