

Parameter Estimation

赖宜贵 2019211249

Maximum Likelihood Estimation

1. 依题意得：泊松分布的函数表达式为 $P(k|\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}$ ，其中 $\lambda > 0$ ， $k \in \{0, 1, 2, \dots\}$

1.1 已知样本 $\{k_1, k_2, \dots, k_n\}$ 属于独立同分布，故 $P(k_i|\lambda) = \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}$

联合分布函数： $P(k_1, k_2, \dots, k_n|\lambda) = \prod_{i=1}^n P(k_i|\lambda) = \prod_{i=1}^n \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}$

似然函数： $P(\lambda|k_1, k_2, \dots, k_n) = P(k_1, k_2, \dots, k_n|\lambda) = \prod_{i=1}^n \frac{\lambda^{k_i} e^{-\lambda}}{k_i!}$

对数似然函数：
$$L(\lambda|k_1, k_2, \dots, k_n) = \ln P(k_1, k_2, \dots, k_n|\lambda) = \sum_{i=1}^n \ln\left(\frac{\lambda^{k_i} e^{-\lambda}}{k_i!}\right)$$
$$= \sum_{i=1}^n (k_i \ln \lambda - \lambda - \ln k_i!) = \ln \lambda \sum_{i=1}^n k_i - n\lambda - \sum_{i=1}^n \ln k_i!$$

对数函数对参数 λ 的一阶导： $\frac{dL(\lambda|k_1, k_2, \dots, k_n)}{d\lambda} = \frac{\sum_{i=1}^n k_i}{\lambda} - n = 0$ ， $\hat{\lambda}_{mle} = \frac{\sum_{i=1}^n k_i}{n}$

对数函数对参数 λ 的二阶导： $\frac{d^2 L(\lambda|k_1, k_2, \dots, k_n)}{d\lambda^2} = -\frac{\sum_{i=1}^n k_i}{\lambda^2} < 0$ ，其中 $\lambda > 0$

所以参数 λ 的最大似然估计值为： $\hat{\lambda}_{mle} = \frac{\sum_{i=1}^n k_i}{n}$

1.2 已知随机变量服从参数为 λ 的泊松分布，故 $P(K|\lambda) = \frac{\lambda^K e^{-\lambda}}{K!}$ ， $\lambda > 0$

$$E[K] = \sum_{k=0}^{+\infty} \frac{k e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda e^{\lambda} = \lambda$$

$$\begin{aligned}
E[K^2] &= \sum_{k=0}^{+\infty} \frac{k^2 e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} \lambda \sum_{k=1}^{+\infty} \frac{k \lambda^{k-1}}{(k-1)!} = e^{-\lambda} \lambda \left(\sum_{k=1}^{+\infty} \frac{(k-1+1) \lambda^{k-1}}{(k-1)!} \right) \\
&= e^{-\lambda} \lambda \left(\sum_{k=1}^{+\infty} \frac{(k-1) \lambda^{k-1}}{(k-1)!} + \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) = e^{-\lambda} \lambda \left(\lambda \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \right) \\
&= e^{-\lambda} \lambda (\lambda e^{\lambda} + e^{\lambda}) = \lambda(\lambda+1) = \lambda^2 + \lambda
\end{aligned}$$

$$Var[K] = E[K^2] - (E[K])^2 = \lambda(\lambda+1) - \lambda^2 = \lambda$$

由 1.1 可知: $\hat{\lambda}_{mle} = \frac{\sum_{i=1}^n k_i}{n}$, 同时样本 $\{k_1, k_2, \dots, k_n\}$ 属于独立同分布

$$E[\hat{\lambda}_{mle}] = E\left[\frac{\sum_{i=1}^n k_i}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^n k_i\right] = E[k_i] = \lambda$$

$$Var[\hat{\lambda}_{mle}] = Var\left[\frac{\sum_{i=1}^n k_i}{n}\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^n k_i\right] = \frac{\lambda}{n}$$

分析结果可知: $\hat{\lambda}_{mle} = \frac{\sum_{i=1}^n k_i}{n}$ 是参数的无偏估计

2、依题意得: $p_1(x)$ 是任意形式的概率分布

$p_2(x)$ 是一维正态分布, 其概率分布函数为: $p_2(x|u, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-u)^2}{2\sigma^2}\right]$

现需要求解: 当分布 $p_1(x)$ 与分布 $p_2(x)$ 间的 Kullback-Leibler 散度最小时, 求解

$p_2(x)$ 的参数 u, σ^2 的取值

$$\begin{aligned}
D_{KL}[p_1(x), p_2(x)] &= \int p_1(x) \ln \frac{p_1(x)}{p_2(x)} dx = \int p_1(x) \ln \frac{p_1(x)}{\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-u)^2}{2\sigma^2}\right]} dx \\
&= \int p_1(x) \ln p_1(x) dx - \int p_1(x) \ln \frac{1}{\sqrt{2\pi}\sigma} dx + \int p_1(x) \frac{(x-u)^2}{2\sigma^2} dx
\end{aligned}$$

当 $D_{KL}[p_1(x), p_2(x)]$ 最小时, 由于 $\int p_1(x) \ln p_1(x) dx$ 已知, 为固定值

因此 $\min D_{KL}[p_1(x), p_2(x)]$ 等价于 $\min \left[-\int p_1(x) \ln \frac{1}{\sqrt{2\pi}\sigma} dx + \int p_1(x) \frac{(x-u)^2}{2\sigma^2} dx \right]$

$$\text{等价于 } \max[\int p_1(x) \ln \frac{1}{\sqrt{2\pi}\sigma} dx - \int p_1(x) \frac{(x-u)^2}{2\sigma^2} dx]$$

由于概率分布 $p_1(x)$ 与 $p_2(x)$ 位于同一空间，假定样本集 $\{x_1, x_2, \dots, x_n\}$

故可将 $\max[\int p_1(x) \ln \frac{1}{\sqrt{2\pi}\sigma} dx - \int p_1(x) \frac{(x-u)^2}{2\sigma^2} dx]$ 转换为概率累加形式

$$\max[\sum_{i=1}^n p_1(x_i) \ln \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^n p_1(x_i) \frac{(x_i-u)^2}{2\sigma^2}]$$

$$\text{假定函数: } L(x|u, \sigma^2) = \sum_{i=1}^n p_1(x_i) \ln \frac{1}{\sqrt{2\pi}\sigma} - \sum_{i=1}^n p_1(x_i) \frac{(x_i-u)^2}{2\sigma^2}$$

对函数求偏导，可得：

$$\frac{\partial L(x|u, \sigma^2)}{\partial u} = \sum_{i=1}^n p_1(x_i) \frac{(x_i-u)}{\sigma^2}, \quad \frac{\partial L(x|u, \sigma^2)}{\partial \sigma^2} = -\sum_{i=1}^n \frac{p_1(x_i)}{2\sigma^2} + \sum_{i=1}^n \frac{p_1(x_i)(x_i-u)^2}{2(\sigma^2)^2}$$

$$\text{当函数 } L(x|u, \sigma^2) \text{ 取得最大值时, } \frac{\partial L(x|u, \sigma^2)}{\partial u} = 0, \quad \frac{\partial L(x|u, \sigma^2)}{\partial \sigma^2} = 0$$

$$\text{即 } \sum_{i=1}^n p_1(x_i) \frac{(x_i-u)}{\sigma^2} = 0, \quad \sum_{i=1}^n p_1(x_i) x_i = \sum_{i=1}^n p_1(x_i) u$$

$$u = \frac{\sum_{i=1}^n p_1(x_i) x_i}{\sum_{i=1}^n p_1(x_i)} = \sum_{i=1}^n \frac{p_1(x_i)}{\sum_{i=1}^n p_1(x_i)} x_i, \quad \text{根据概率期望的定义可知: } u = E_1[x]$$

$$-\sum_{i=1}^n \frac{p_1(x_i)}{2\sigma^2} + \sum_{i=1}^n \frac{p_1(x_i)(x_i-u)^2}{2(\sigma^2)^2} = 0, \quad \sum_{i=1}^n \frac{p_1(x_i)(x_i-u)^2}{\sigma^2} = \sum_{i=1}^n p_1(x_i)$$

$$\sigma^2 = \sum_{i=1}^n \frac{p_1(x_i)(x_i-u)^2}{\sum_{i=1}^n p_1(x_i)}, \quad \sigma^2 = \sum_{i=1}^n \frac{p_1(x_i)}{\sum_{i=1}^n p_1(x_i)} (x_i-u)^2, \quad \text{同理可得: } \sigma^2 = E_1[(x_i-u)^2]$$

综上所述：当同一空间的概率分布函数 $p_1(x)$ 与 $p_2(x)$ 之间的 Kullback-Leibler 散度达到最小时，一维正态分布函数 $p_2(x)$ 的参数将满足以下条件， $u = E_1[x]$ 和 $\sigma^2 = E_1[(x_i-u)^2]$ 。

Bayes Estimation

1. 已知 d 维多元伯努利分布函数为: $P(x|\theta) = \prod_{i=1}^d \theta_i^{x_i} (1-\theta_i)^{1-x_i}$

1.1 一维伯努利分布函数为: $P(x|\theta) = \theta^x (1-\theta)^{1-x}$

选取任意两个参数 θ_1, θ_2 , 则 $P(x|\theta_1) = \theta_1^x (1-\theta_1)^{1-x}$, $P(x|\theta_2) = \theta_2^x (1-\theta_2)^{1-x}$

假设对于任意的 x , 均有 $P(x|\theta_1) = P(x|\theta_2)$

$$\text{即 } \theta_1^x (1-\theta_1)^{1-x} = \theta_2^x (1-\theta_2)^{1-x}, \quad \frac{\theta_1^x}{\theta_2^x} = \frac{(1-\theta_2)^{1-x}}{(1-\theta_1)^{1-x}}, \quad \left(\frac{\theta_1}{\theta_2}\right)^x = \left[\frac{(1-\theta_2)}{(1-\theta_1)}\right]^{1-x}$$

上式当且仅当 $\theta_1 = \theta_2$ 时成立

故对于任意两个参数 θ_1, θ_2 , 当 $P(x|\theta_1) = P(x|\theta_2)$ 成立时, $\theta_1 = \theta_2$ 成立

所以一维伯努利分布是可识别的。

1.2 已知 $P(x|\theta) = \prod_{i=1}^d \theta_i^{x_i} (1-\theta_i)^{1-x_i}$, $\{x_1, x_2, \dots, x_n\}$ 独立同分布

联合分布函数: $P(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n \prod_{j=1}^d \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}}$

似然函数: $L(\theta | x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n \prod_{j=1}^d \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}}$

对数似然函数:

$$\begin{aligned} \ln L(\theta | x_1, x_2, \dots, x_n) &= \sum_{i=1}^n \ln \prod_{j=1}^d \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}} \\ &= \sum_{i=1}^n \sum_{j=1}^d \ln \theta_j^{x_{ij}} (1-\theta_j)^{1-x_{ij}} = \sum_{i=1}^n \sum_{j=1}^d [x_{ij} \ln \theta_j + (1-x_{ij}) \ln(1-\theta_j)] \end{aligned}$$

最大似然值为:

$$\frac{\partial \ln L(\theta | x_1, x_2, \dots, x_n)}{\partial \theta_j} = \sum_{i=1}^n \left[\frac{x_{ij}}{\theta_j} + \frac{(x_{ij}-1)}{(1-\theta_j)} \right] = \sum_{i=1}^n \left[\frac{x_{ij}(1-\theta_j)}{\theta_j} + \frac{\theta_j(x_{ij}-1)}{(1-\theta_j)} \right] = 0$$

此时 $\hat{\theta} = \frac{1}{n} \sum_{k=1}^n x_k$ 是参数 θ 的最大似然解

1.3 对于样本集 $D = \{x_1, x_2, \dots, x_n\}$, 已知各维度样本的和 $s = \{s_1, s_2, \dots, s_d\}$

$$P(x_i | \theta) = \prod_{i=1}^d \theta_i^{x_i} (1-\theta_i)^{1-x_i}, \quad \text{那么 } P(D | \theta) = \prod_{i=1}^d \theta_i^{\sum_{k=1}^n x_k^i} (1-\theta_i)^{\sum_{k=1}^n (1-x_k^i)} = \prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i}$$

1.4 由贝叶斯公式可得: $p(\theta|D) = \frac{p(\theta,D)}{p(D)} = \frac{p(D|\theta)p(\theta)}{p(D)} = \frac{p(D|\theta)p(\theta)}{\int p(D|\theta)p(\theta)d\theta}$

由参数 θ 满足均匀分布 $U(0,1)$, 当 $\theta \in [0,1]$ 时, $p(\theta)=1$, 可知

$$p(\theta|D) = \frac{\prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i}}{\int_0^1 \prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta} = \frac{\prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i}}{\prod_{i=1}^d \int_0^1 \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta}$$

利用 $\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!}$ 得 $\int_0^1 \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta = \frac{s_i!(n-s_i)!}{(s_i+n-s_i+1)!} = \frac{s_i!(n-s_i)!}{(n+1)!}$

代入式子中可得:

$$p(\theta|D) = \frac{\prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i}}{\int_0^1 \prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta} = \frac{\prod_{i=1}^d \theta_i^{s_i} (1-\theta_i)^{n-s_i}}{\prod_{i=1}^d \frac{s_i!(n-s_i)!}{(n+1)!}} = \prod_{i=1}^d \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$$

1.5 依题意得: $p(x|D) = \int p(x|\theta)p(\theta|D)d\theta$

将 $P(x|\theta) = \prod_{i=1}^d \theta_i^{x_i} (1-\theta_i)^{1-x_i}$ 与 $p(\theta|D) = \prod_{i=1}^d \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i}$ 代入其中, 得

$$\begin{aligned} p(x|D) &= \int \prod_{i=1}^d \theta_i^{x_i} (1-\theta_i)^{1-x_i} \prod_{i=1}^d \frac{(n+1)!}{s_i!(n-s_i)!} \theta_i^{s_i} (1-\theta_i)^{n-s_i} d\theta \\ &= \int \prod_{i=1}^d \theta_i^{x_i+s_i} (1-\theta_i)^{1-x_i+n-s_i} \frac{(n+1)!}{s_i!(n-s_i)!} d\theta = \prod_{i=1}^d \frac{(n+1)!}{s_i!(n-s_i)!} \int \theta_i^{x_i+s_i} (1-\theta_i)^{1-x_i+n-s_i} d\theta \end{aligned}$$

由于 $\theta \in [0,1]$, 所以再次利用等式 $\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!}$ 可得

$$p(x|D) = \prod_{i=1}^d \frac{(n+1)!}{s_i!(n-s_i)!} \frac{(x_i+s_i)!(1-x_i+n-s_i)!}{(n+2)!} = \prod_{i=1}^d \frac{(x_i+s_i)!(1-x_i+n-s_i)!}{s_i!(n-s_i)!(n+2)}$$

根据伯努利分布特性可得, $x_i \in [0,1]$

当 $x_i = 0$ 时, $\frac{(x_i+s_i)!(1-x_i+n-s_i)!}{s_i!(n-s_i)!(n+2)} = \frac{(n-s_i+1)!}{(n-s_i)!(n+2)} = \frac{n-s_i+1}{n+2} = 1 - \frac{s_i+1}{n+2}$

当 $x_i = 1$ 时, $\frac{(x_i+s_i)!(1-x_i+n-s_i)!}{s_i!(n-s_i)!(n+2)} = \frac{s_i+1}{n+2}$

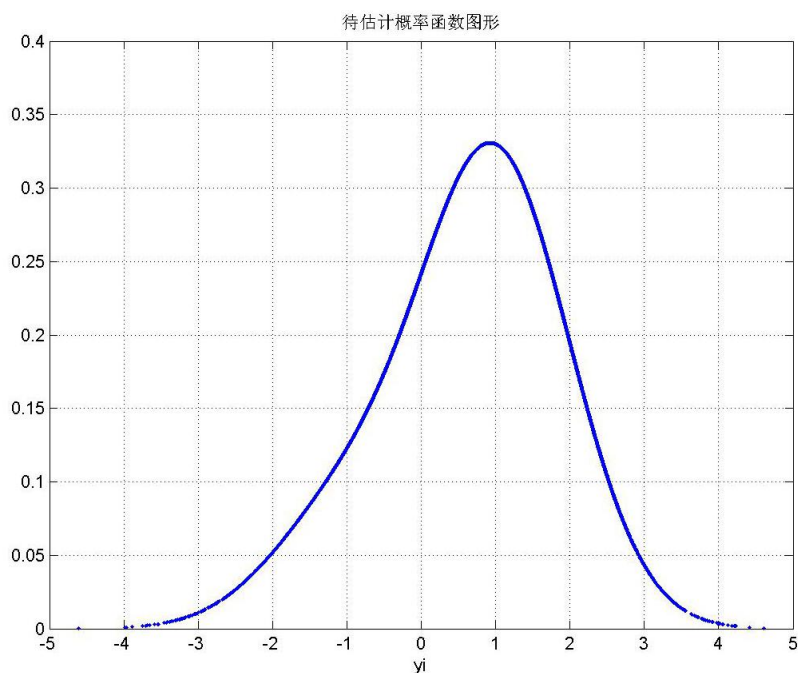
故综上可得: $p(x|D) = \prod_{i=1}^d \left(\frac{s_i+1}{n+2}\right)^{x_i} \left(1 - \frac{s_i+1}{n+2}\right)^{1-x_i}$

Programming

一、知识储备

1、原函数图像

依题意得， $p(x) \sim 0.2N(-1,1) + 0.8N(1,1)$ ，绘制图像，结果如下所示：



2、窗函数相关信息

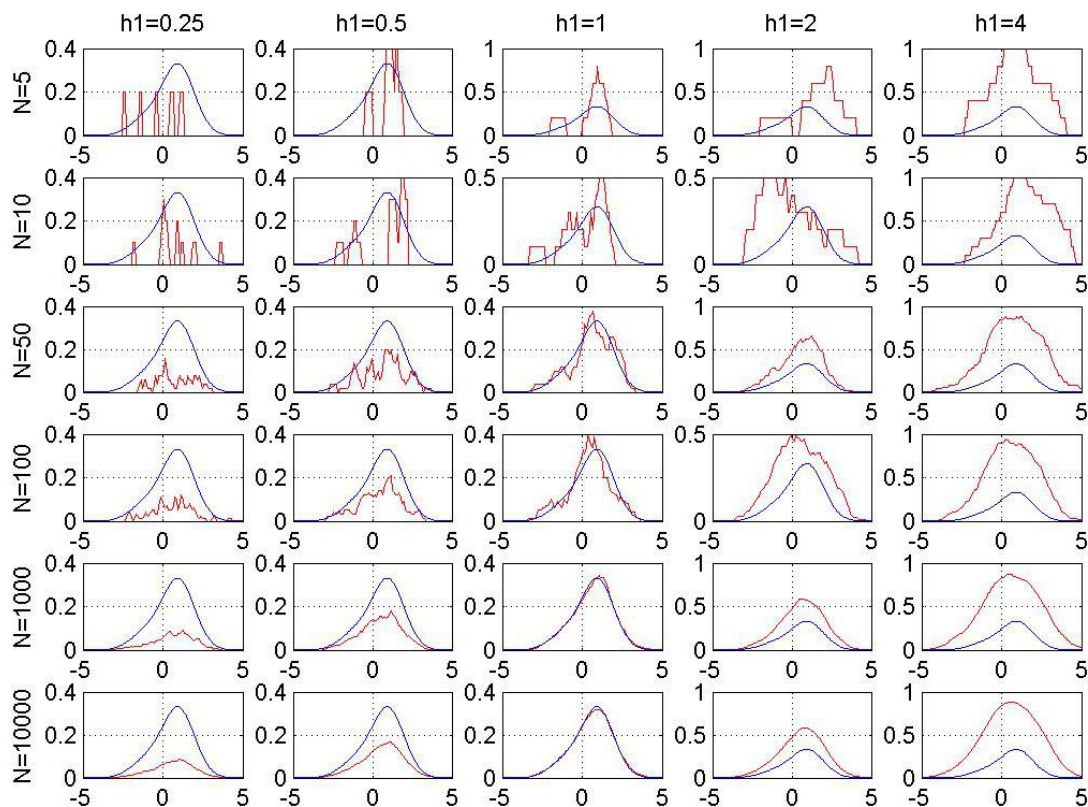
窗函数类型	窗函数 $\psi(u)$	参数	概率估计函数 $\hat{p}_N(x)$
方窗	$\psi(u) = \begin{cases} 1, & -\frac{1}{2} \leq u \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$	$h_N = a$ $V_N = 1$	$\hat{p}_N(x) = \frac{1}{NV_N} \sum_{i=1}^N \psi\left(\frac{x-x_i}{h_N}\right)$
高斯窗	$\psi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$	$h_N = \frac{\sigma}{\sqrt{N}}$	$\hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^N \psi\left(\frac{x-x_i}{h_N}\right)$
指数窗	$\psi(u) = \frac{1}{2} \exp\{- u \}$	$h_N = \frac{\lambda}{\sqrt{N}}$	$\hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^N \psi\left(\frac{x-x_i}{h_N}\right)$
三角窗	$\psi(u) = \begin{cases} 1- x , & -1 \leq u \leq 1 \\ 0, & \text{otherwise} \end{cases}$	$h_N = a$ $V_N = 1$	$\hat{p}_N(x) = \frac{1}{NV_N} \sum_{i=1}^N \psi\left(\frac{x-x_i}{h_N}\right)$

二、问题解答

(a) 取值情况如下： $a = \{0.25, 0.5, 1, 2, 4\}$ ， $N = \{5, 10, 50, 100, 1000, 10000\}$

利用 matlab, 可绘制出不同参数 a 以及不同样本数量 N 的估计分布以及原分布概率密度函数图线, 结果如下所示:

方窗估计



结论: ①当窗宽一定时, 增大样本数量时, 概率密度函数曲线由“多峰”变为“单峰”, 减少了毛刺;

②当样本数量一定时, 增大窗宽, 可以提高曲线的平滑性, 此外, 当窗宽增大至 1 时, 估计概率密度函数曲线最接近原分布曲线, 继续增大窗宽将减小函数的幅度值, 将远离真实分布;

③在研究参数区间内, 当窗宽选择 $a = 1$, 样本数量 $N = 10000$ 时, 估计分布最接近原分布。

(b) 取区间 $x \in [-5, 5]$, 假设在小区间 $dx = 0.02$ (足够小), 则估计分布函数值与原分布函数值间的差值 $\varepsilon(p_n) = \int [p_n(x) - p(x)]^2 dx$ 可转换为累加格式, 如下:

$$\varepsilon(p_n) = \int [p_n(x) - p(x)]^2 dx \approx \sum_{j=1}^N [p_n(x) - p(x)]^2 dx$$

$$\text{其中 } \hat{p}_N(x) = \frac{1}{NV_N} \sum_{i=1}^N \psi\left(\frac{x-x_i}{h_N}\right), \quad V_N = 1, \quad h_N = a, \quad \psi(u) = \begin{cases} 1, & -\frac{1}{2} \leq u \leq \frac{1}{2} \\ 0, & \text{otherwise} \end{cases}$$

因为 $p(x) \sim 0.2N(-1,1) + 0.8N(1,1)$

$$\text{所以 } p(x) = 0.2 * \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x+1)^2}{2}\right] + 0.8 * \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-1)^2}{2}\right]$$

(c) 根据 (b) 中的计算方法, 可计算出每一次估计分布函数值与原分布函数值间的差值 $\varepsilon(p_n)$, 先假设重复计算 20 次, 可以得到差值 $\varepsilon(p_n)$ 的期望和方差值随参数 a 以及样本数量 N 的变化关系表, 如下所示:

表 1 窗函数为方窗函数时, $\varepsilon(p_n)$ 的期望随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.2119	0.1688	0.1191	0.2698	1.3004
10	0.1191	0.1035	0.0227	0.3944	1.4122
50	0.1411	0.0863	0.0110	0.1898	1.3904
100	0.1304	0.0646	0.0064	0.2305	1.3893
1000	0.1282	0.0564	0.0011	0.1847	1.3785
10000	0.1268	0.0569	0.0004	0.1844	1.3385

表 2 窗函数为方窗函数时, $\varepsilon(p_n)$ 的方差随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.0008e-30	0.0032e-30	0	0	0.2076e-30
10	0.0002	0.0032e-30	0.0001e-30	0.0130e-30	0
50	0.0008	0.0008e-30	0	0.0032e-30	0.4671e-30
100	0	0.0008e-30	0	0.0032e-30	0.2076e-30
1000	0	0.0001e-30	0	0.0032e-30	0.4671e-30
10000	0.0032	0.0001e-30	0	0	0.0519e-30

(d) 分析表 1 和表 2 可得以下结论:

- ①当样本数量 N 一定时, 随着样本数量 N 增大, $\varepsilon(p_n)$ 的期望值减小后增大, 存在一个“峰谷”值, 反映在分布函数曲线上为曲线逐渐趋于平滑;
- ②当样本数量 N 一定时, 可通过计算 $\varepsilon(p_n)$ 的期望值随一组窗宽 a 的变化曲线图, 并从中寻找到该曲线的“峰谷”位置, 该位置所对应的窗宽 a 值即为最佳参数值, 估计出来的分布函数曲线最接近原分布函数曲线;
- ③在满足 $\lim_{n \rightarrow +\infty} \frac{a}{N} = 0$ 的条件下, 对于较大的样本数量 N , 应选取较小的窗宽 a 值, 对于较小的样本数量 N , 应选取较大的窗宽 a 值。

(e) 当窗函数为高斯窗函数时, 即 $\psi(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}u^2)$

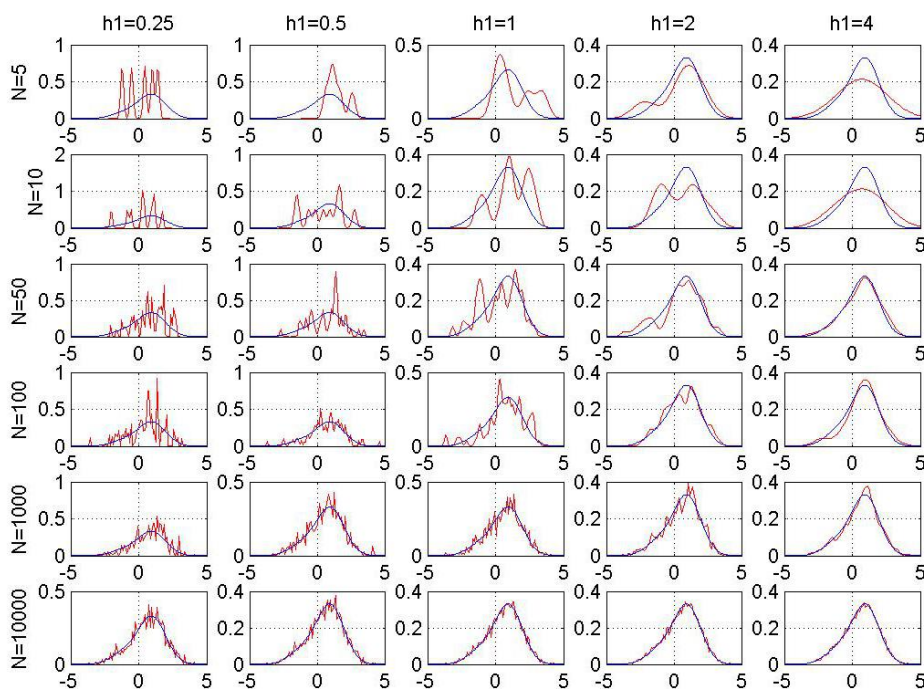
可知待调节参数为方差 σ , 窗宽为 $h_N = \frac{\sigma}{\sqrt{N}}$

概率估计函数为 $\hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^N \psi(\frac{x-x_i}{h_N})$, 按照 (a) - (d) 的步骤进行处理, 结果依次如下所示。

(e.a) 取值情况如下: $a = \{0.25, 0.5, 1, 2, 4\}$, $N = \{5, 10, 50, 100, 1000, 10000\}$

利用 matlab, 可绘制出不同参数 a 以及不同样本数量 N 的估计分布以及原分布概率密度函数图线, 结果如下所示:

高斯窗估计



结论：①当窗宽一定时，增大样本数量时，概率密度函数曲线由“多峰”变为“单峰”，减少了毛刺；

②当样本数量一定时，增大窗宽，可以提高曲线的平滑性，此外，当窗宽增大至 1 时，估计概率密度函数曲线最接近原分布曲线；

③在研究参数区间内，当方差选择 $\sigma = 1$ ，样本数量 $N = 10000$ 时，估计分布最接近原分布。

(e.b) 取区间 $x \in [-5, 5]$ ，假设在小区间 $dx = 0.02$ （足够小），则估计分布函数值与原分布函数值间的差值 $\varepsilon(p_n) = \int [p_n(x) - p(x)]^2 dx$ 可转换为累加格式，如下：

$$\varepsilon(p_n) = \int [p_n(x) - p(x)]^2 dx \approx \sum_{j=1}^N [p_n(x) - p(x)]^2 dx$$

$$\text{其中 } \hat{p}_N(x) = \frac{1}{Nh_N} \sum_{i=1}^N \psi\left(\frac{x - x_i}{h_N}\right), \quad h_N = \frac{\sigma}{\sqrt{N}}, \quad \psi(u) = \begin{cases} 1 \\ 0 \end{cases}, \quad \psi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}u^2\right)$$

因为 $p(x) \sim 0.2N(-1,1) + 0.8N(1,1)$

$$\text{所以 } p(x) = 0.2 * \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x+1)^2}{2}\right] + 0.8 * \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-1)^2}{2}\right]$$

(e.c) 根据 (b) 中的计算方法, 可计算出每一次估计分布函数值与原分布函数值间的差值 $\varepsilon(p_n)$, 先假设重复计算 20 次, 可以得到差值 $\varepsilon(p_n)$ 的期望和方差值随参数 a 以及样本数量 N 的变化关系表, 如下所示:

表 3 窗函数为高斯函数时, $\varepsilon(p_n)$ 的期望随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.4114	0.1090	0.0856	0.0063	0.0371
10	0.6629	0.2772	0.0844	0.0216	0.0233
50	0.1987	0.0642	0.0373	0.0133	0.0132
100	0.1211	0.0420	0.0291	0.0119	0.0041
1000	0.0347	0.0169	0.0080	0.0028	0.0016
10000	0.0113	0.0057	0.0025	0.0012	0.0006

表 4 窗函数为高斯函数时, $\varepsilon(p_n)$ 的方差随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.0324e-31	0.0182e-31	0.0081e-31	0	0.0005e-31
10	0.5190e-31	0.1297e-31	0.0020e-31	0.0011e-31	0.0001e-31
50	0.0730e-31	0	0.0020e-31	0	0
100	0.0182e-31	0.0020e-31	0.0011e-31	0.0001e-31	0
1000	0.0020e-31	0	0	0	0
10000	0	0	0	0	0

(e.d) 分析表 1 和表 2 可得以下结论:

- ①当样本数量 N 一定时, 随着样本数量 N 增大, $\varepsilon(p_n)$ 的期望值不断减小;
- ②当样本数量 N 一定时, 可通过计算 $\varepsilon(p_n)$ 的期望值随一组窗宽 a 的变化曲线图, 并从中寻找到该曲线的“最小值”位置, 该位置所对应的窗宽 a 值即为最佳参数值, 估计出来的分布函数曲线最接近原分布函数曲线;
- ③在满足 $\lim_{n \rightarrow +\infty} \frac{a}{N} = 0$ 的条件下, 选择较大的样本数量 N 与较大的窗宽 a 值, 此时估计出来的分布函数曲线最接近原分布函数曲线。

(f) 除了题给的方窗函数以及高斯窗函数之外, 本实验采用了指数窗以及三角窗进行研究, 取值情况如下: $a = \{0.25, 0.5, 1, 2, 4\}$, $N = \{5, 10, 50, 100, 1000, 10000\}$, 现将结果展示如下:

①指数窗

指数窗估计

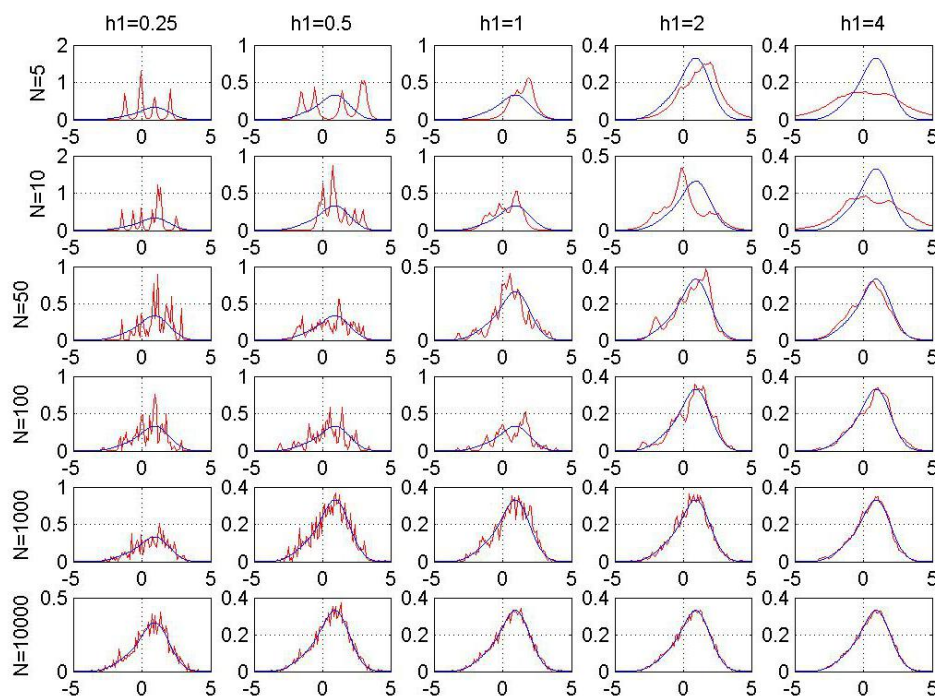


表 5 窗函数为指数函数时, $\varepsilon(p_n)$ 的期望随窗宽 a 以及样本数量 N 的关系表

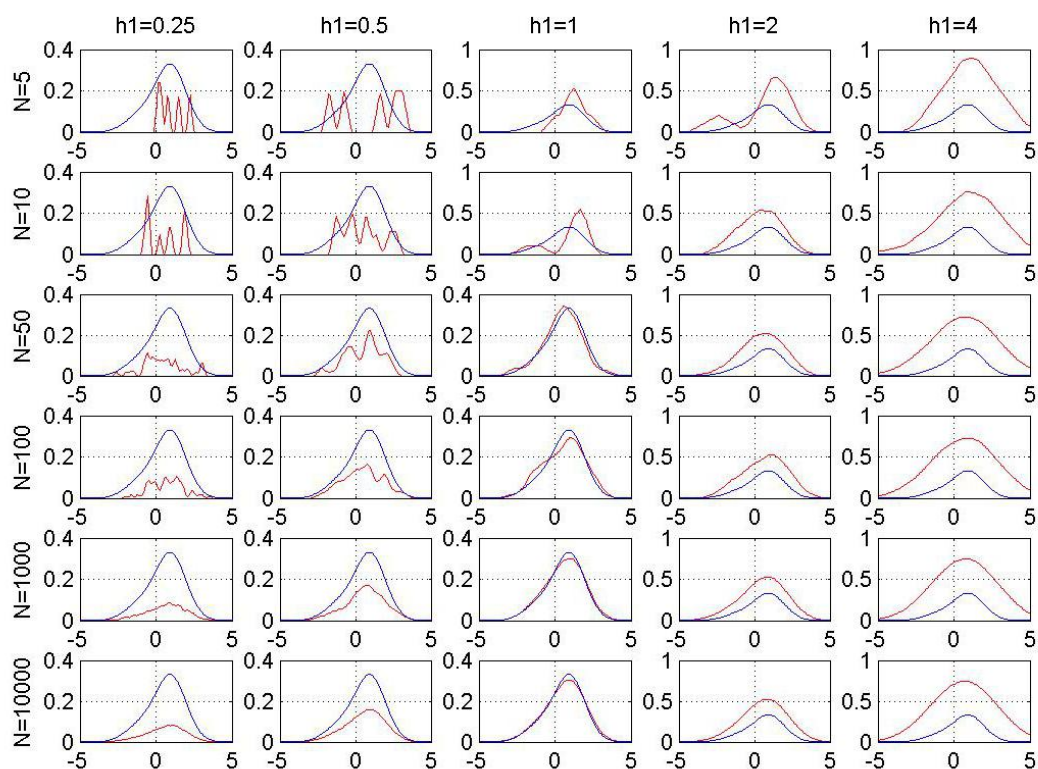
	0.25	0.5	1	2	4
5	0.3523	0.1561	0.1318	0.0640	0.0265
10	0.2563	0.0698	0.0363	0.0047	0.0092
50	0.1598	0.0646	0.0459	0.0164	0.0054
100	0.1121	0.0469	0.0112	0.0217	0.0040
1000	0.0309	0.0142	0.0079	0.0036	0.0026
10000	0.0099	0.0049	0.0023	0.0011	0.0005

表 6 窗函数为指数函数时, $\varepsilon(p_n)$ 的方差随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.2919e-31	0	0	0	0.0011e-31
10	0	0.0081e-31	0	0	0.0001e-31
50	0	0	0	0.0001e-31	0.0001e-31
100	0.0020e-31	0.0020e-31	0.0003e-31	0.0001e-31	0
1000	0	0.0001e-31	0	0	0
10000	0	0	0	0	0

②三角窗

三角窗估计

表 7 窗函数为三角窗函数时, $\varepsilon(p_n)$ 的期望随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.1408	0.0834	0.0651	0.2625	1.1749
10	0.1323	0.0957	0.0262	0.1790	1.0201
50	0.1314	0.0703	0.0083	0.1487	1.0417
100	0.1276	0.0583	0.0017	0.1685	1.0399
1000	0.1280	0.0578	0.0012	0.1541	1.0714
10000	0.1269	0.0573	0.0007	0.1602	1.0771

表 8 窗函数为三角窗函数时, $\varepsilon(p_n)$ 的方差随窗宽 a 以及样本数量 N 的关系表

	0.25	0.5	1	2	4
5	0.0008e-30	0.0008e-30	0	0.0032e-30	0.2076e-30
10	0.0032e-30	0.0008e-30	0.0001e-30	0.0073e-30	0
50	0	0.0002e-30	0	0.0008e-30	0.0519e-30
100	0	0.0005e-30	0	0.0032e-30	0.2076e-30
1000	0	0.0002e-30	0	0.0008e-30	0.2076e-30
10000	0.0032e-30	0.0001e-30	0	0.0032e-30	0

结论：综合以上分析，可以得到以下结论：

①高斯窗与指数窗具有相似性：由于高斯函数与指数函数具有开放性，因此根据其得到的估计函数曲线存在以下特性，当样本数量 N 一定时，随着样本数量 N 增大， $\varepsilon(p_n)$ 的期望值不断减小；

②方窗与三角窗具有相似性：由于高斯函数与指数函数具有开放性，因此根据其得到的估计函数曲线存在以下特性，当样本数量 N 一定时，随着样本数量 N 增大， $\varepsilon(p_n)$ 的期望值减小后增大，存在一个“峰谷”值，反映在分布函数曲线上为曲线逐渐趋于平滑；

③同等参数条件下，根据高斯窗得到的估计分布函数曲线与原分布函数曲线最接近，为得到 $a = \{0.25, 0.5, 1, 2, 4\}$ ， $N = \{5, 10, 50, 100, 1000, 10000\}$ 下的最优组合，分别列出在该参数空间下，不同窗函数得到的 $\varepsilon(p_n)$ 的期望与窗宽 a 以及样本数量 N 的关系表如下所示：

表 9 不同窗函数得到的 $\varepsilon(p_n)$ 的最优期望与窗宽 a 以及样本数量 N 的关系表

函数类型	最优期望值	窗宽 a	样本数量 N
方窗	0.0004	1	10000
高斯窗	0.0006	4	10000
指数窗	0.0005	4	10000
三角窗	0.0007	1	10000

分析表格可知，在 $a = \{0.25, 0.5, 1, 2, 4\}$ ， $N = \{5, 10, 50, 100, 1000, 10000\}$ 的条件下，窗宽 $a = 1$ ，样本数量 $N = 10000$ 的方窗结果最优。