

# ksdensity

单变量和单变量的核平滑函数估计 双变量数据

## 语法

```
[f,xi] = ksdensity(x)
[f,xi] = ksdensity(x,pts)
[f,xi] = ksdensity( ___,Name,Value)
[f,xi,bw] = ksdensity( ___)

ksdensity( ___)
ksdensity(ax, ___)

```

## 描述

[f, xi] = ksdensity (x) 返回一个概率 密度估计值, , 对于样本数据 矢量或两列矩阵。估计为 基于普通内核函数, 并以等间距计算 点, , 覆盖数据范围 在。估计 单变量数据的密度为 100 点, 双变量数据的密度为 900 点 数据。fxxixksdensity

ksdensity连续工作效果最佳 分布式样本。

[f, xi] = ksdensity (x, pts) 指定要计算的点 ( )。在这里, 并包含相同的 值。

ptsfxipts

[f,xi] = ksdensity( \_\_\_,Name,Value)使用由 中的一个或多个名称-值对参数指定的其他选项 添加到前面语法中的任何输入参数。例如 您可以定义函数类型的评估, 例如 如概率密度、累积概率、幸存者函数等。或 您可以指定平滑窗口的带宽。ksdensity

[f,xi,bw] = ksdensity( \_\_\_)还返回带宽 的内核平滑窗口 .默认 带宽是正常密度的最佳选择。bw

ksdensity( \_\_\_)情节 核平滑函数估计值。

ksdensity(ax, \_\_\_)情节 结果使用带有手柄的轴, 代替 的当前轴。axgca

## 例子

全部折叠

### 估计密度

从两个正态分布的混合生成样本数据集。

试试这个例子

复制命令

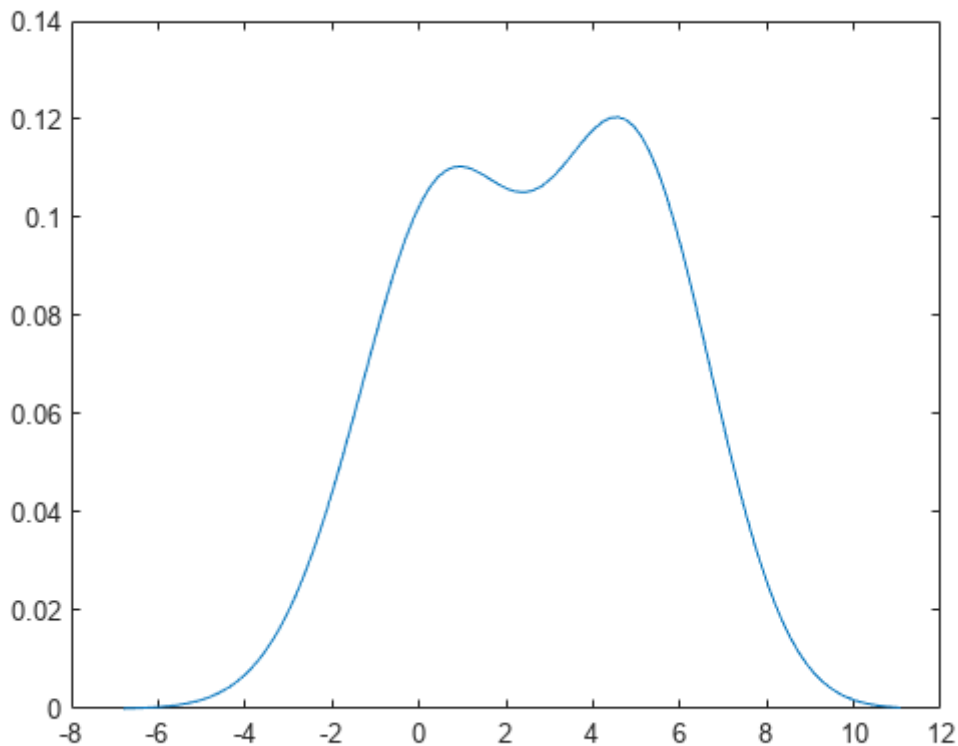
```
rng('default') % For reproducibility
x = [randn(30,1); 5+randn(30,1)];
```

获取

绘制估计密度。

```
[f,xi] = ksdensity(x);  
figure  
plot(xi,f);
```

获取



密度估计值显示了样品的双峰性。

### 使用边界校正估计密度

从半正态分布生成非负样本数据集。

试试这个例子

复制命令

```
rng('default') % For reproducibility  
pd = makedist('HalfNormal','mu',0,'sigma',1);  
x = random(pd,100,1);
```

获取

使用名称-值对参数，使用两种不同的边界校正方法（对数转换和反射）估计 pdf。 'BoundaryCorrection'

```
pts = linspace(0,5,1000); % points to evaluate the estimator  
[f1,xi1] = ksdensity(x,pts,'Support','positive');  
[f2,xi2] = ksdensity(x,pts,'Support','positive','BoundaryCorrection','re
```

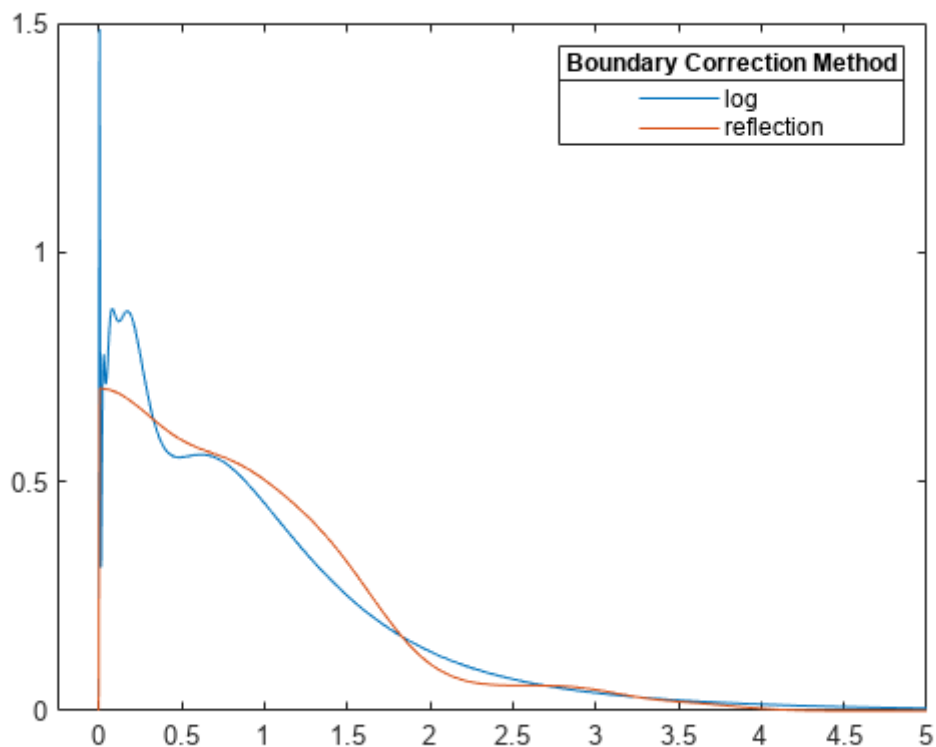
获取



绘制两个估计的 pdf。

```
plot(xi1,f1,xi2,f2)
lgd = legend('log','reflection');
title(lgd, 'Boundary Correction Method')
x1 = xlim;
xlim([x1(1)-0.25 x1(2)])
```

获取



ksdensity指定正支座或有界支座时使用边界校正方法。默认边界校正方法是对数转换。当转换回支持时，它会在核密度估计器中引入该术语。因此，估计值在附近有一个峰值。另一方面，反射方法不会在边界附近引起不希望的峰值。ksdensity1/xx = 0

## 估计指定值下的累积分布函数

加载示例数据。

试试这个例子

复制命令

```
load hospital
```

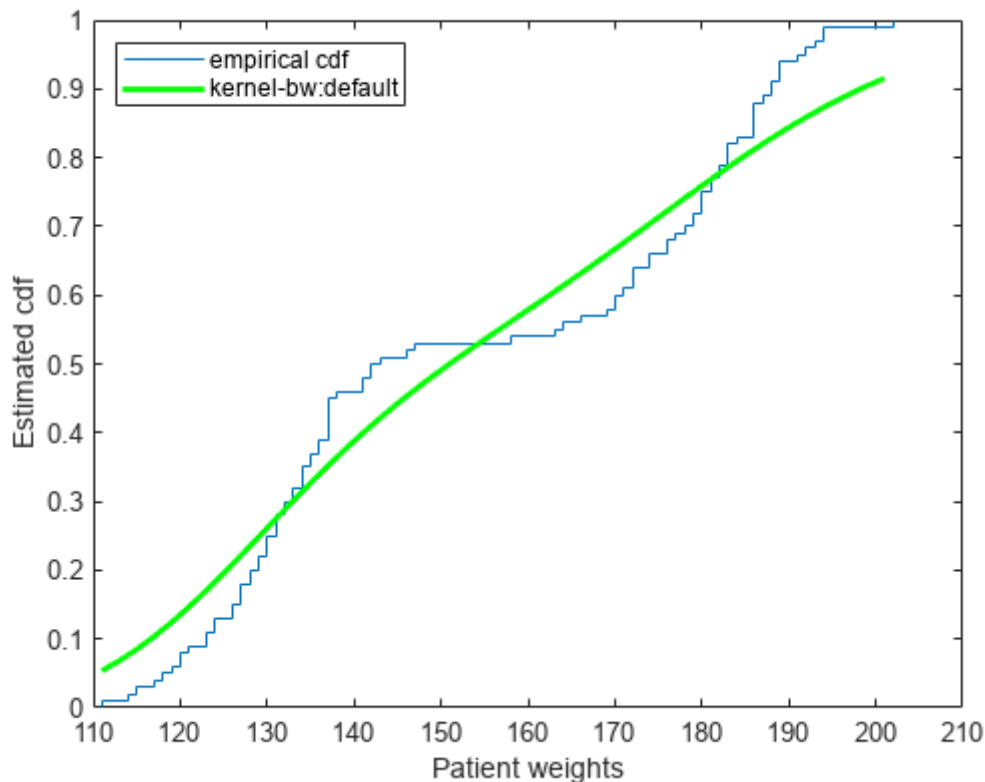
获取

计算并绘制在一组指定值下评估的估计 cdf。



```
pts = (min(hospital.Weight):2:max(hospital.Weight));
figure()
ecdf(hospital.Weight)
hold on
[f,xi,bw] = ksdensity(hospital.Weight,pts,'Support','positive',...
    'Function','cdf');
plot(xi,f,'-g','LineWidth',2)
legend('empirical cdf','kernel-bw:default','Location','northwest')
xlabel('Patient weights')
ylabel('Estimated cdf')
```

获取



ksdensity似乎平滑了累积分布函数估计太多。带宽较小的估计值可能会产生更接近经验累积分布函数的估计值。

返回平滑窗口的带宽。

```
bw
```

获取

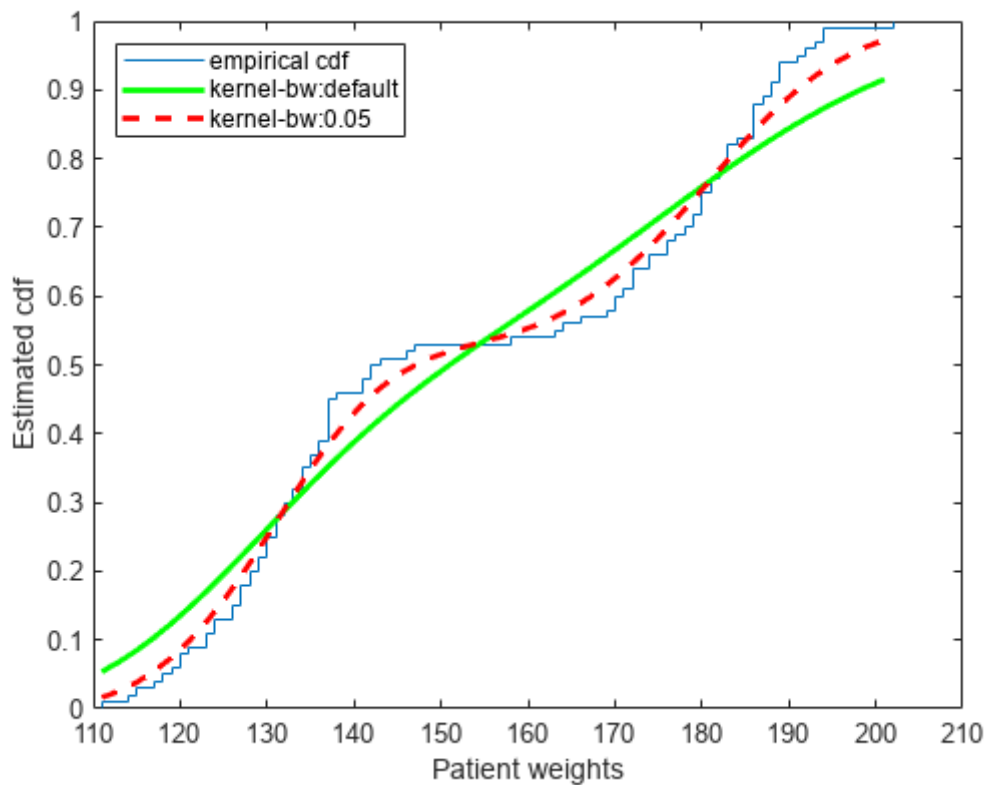
```
bw = 0.1070
```

使用较小的带宽绘制累积分布函数估计值。

```
[f,xi] = ksdensity(hospital.Weight,pts,'Support','positive',...
    'Function','cdf','Bandwidth',0.05);
plot(xi,f,'--r','LineWidth',2)
legend('empirical cdf','kernel-bw:default','kernel-bw:0.05',...
    'Location','northwest')
hold off
```

获取





带宽较小的估计值与经验累积分布函数匹配得更好。ksdensity

#### 绘制给定点数的估计累积分布函数

加载示例数据。

试试这个例子

复制命令

```
load hospital
```

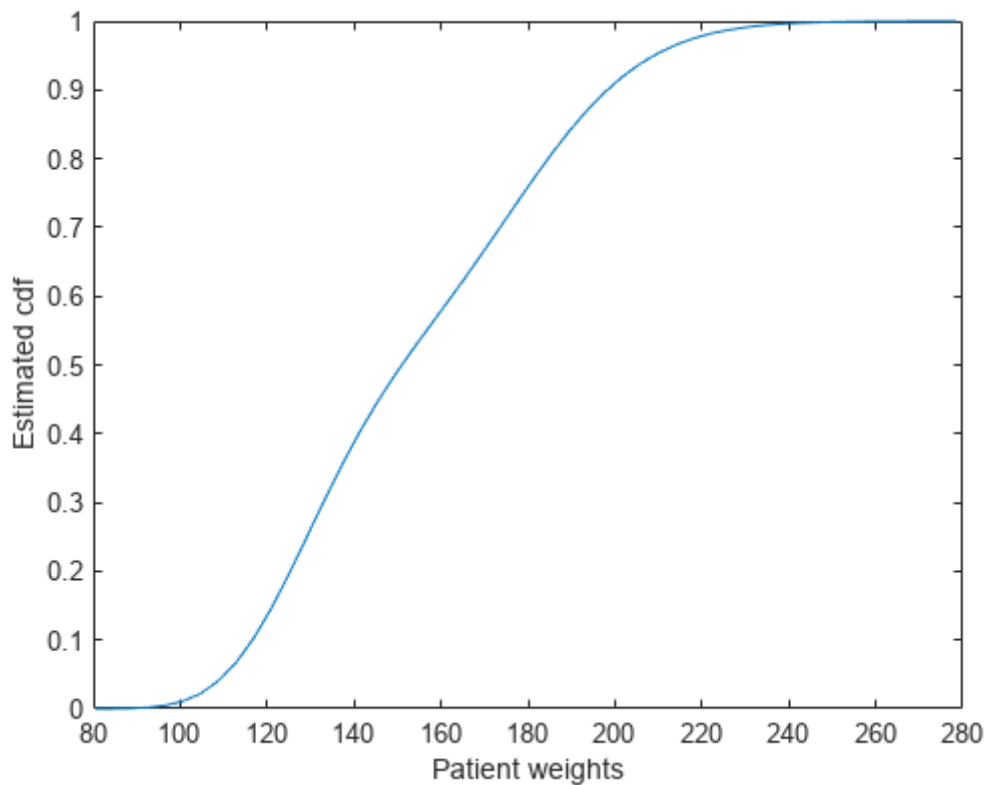
获取

绘制在 50 个等间距点处评估的估计 cdf。

```
figure()
ksdensity(hospital.Weight,'Support','positive','Function','cdf',...
'NumPoints',50)
xlabel('Patient weights')
ylabel('Estimated cdf')
```

获取





#### 估计删失故障数据的幸存者和累积危险

从均值为 3 的指数分布生成样本数据。

试试这个例子

复制命令

```
rng('default') % For reproducibility
x = random('exp',3,100,1);
```

获取

创建指示审查的逻辑向量。在这里，寿命超过10的观测被审查。

```
T = 10;
cens = (x>T);
```

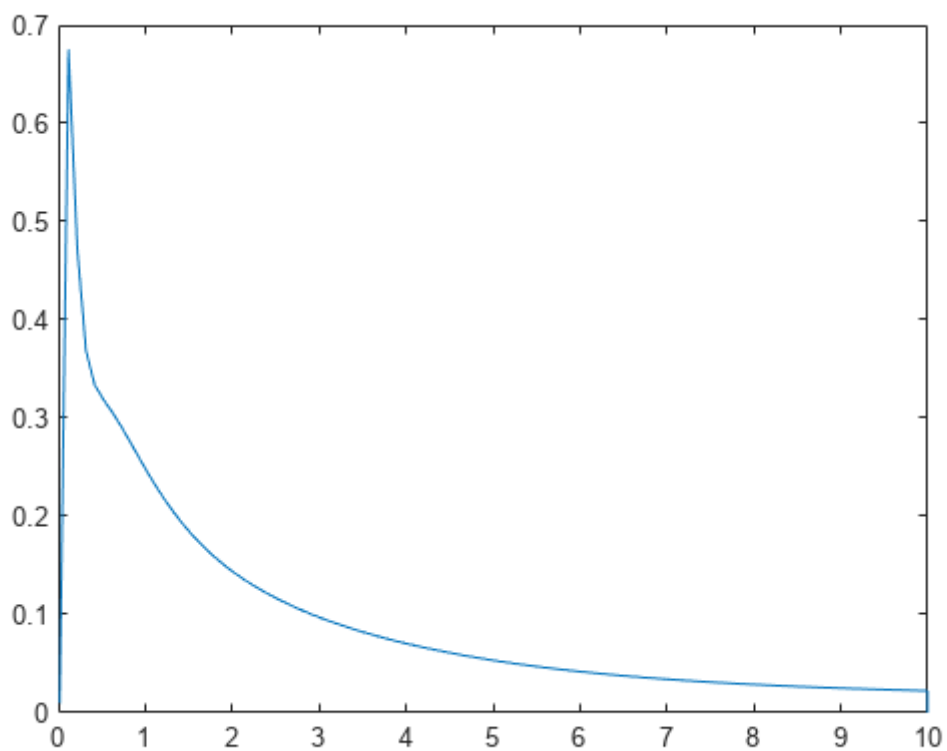
获取

计算并绘制估计的密度函数。

```
figure
ksdensity(x,'Support','positive','Censoring',cens);
```

获取

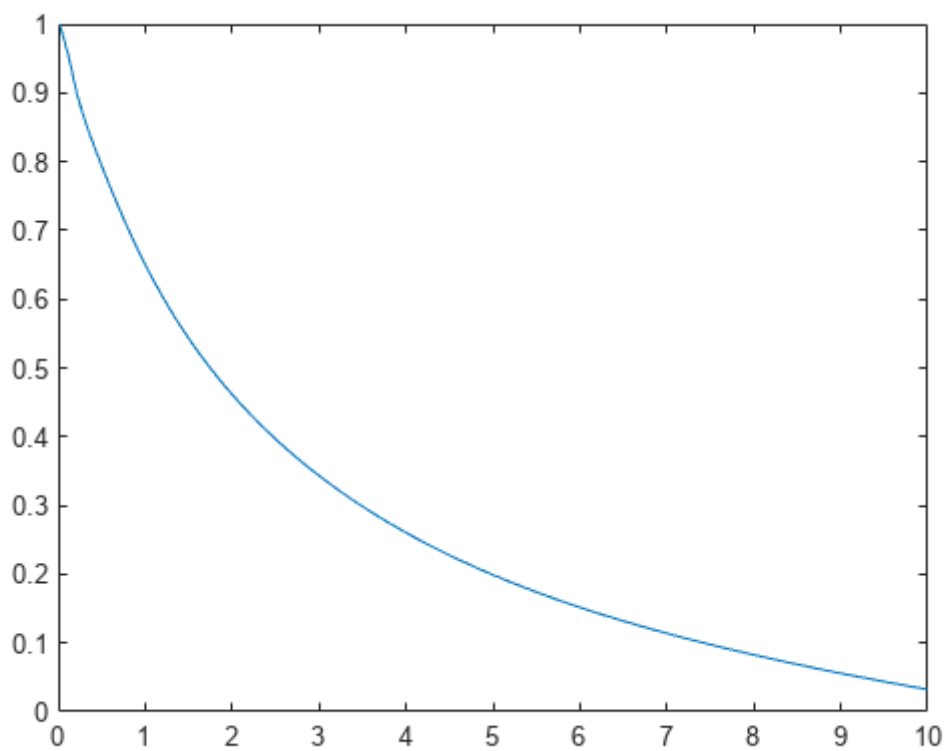




计算并绘制幸存者函数。

```
figure
ksdensity(x,'Support','positive','Censoring',cens,...
'Function','survivor');
```

获取

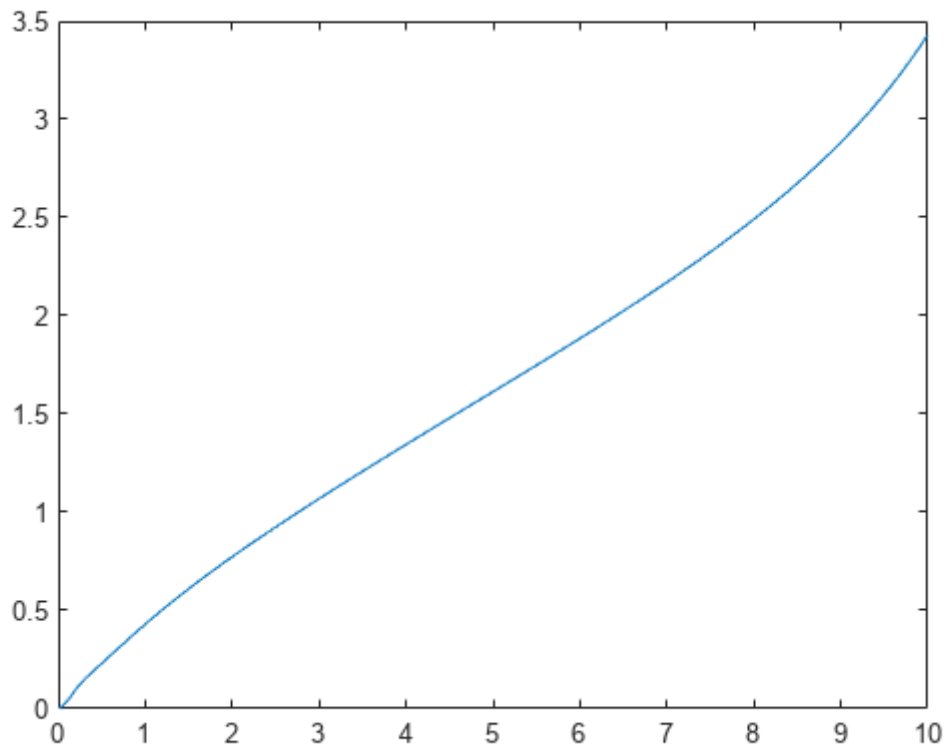


计算并绘制累积风险函数。



```
figure
ksdensity(x,'Support','positive','Censoring',cens,...
'Function','cumhazard');
```

获取



### 估计指定概率值的逆累积分布函数

生成两个正态分布的混合，并在一组指定的概率值处绘制估计的逆累积分布函数。

试试这个例子

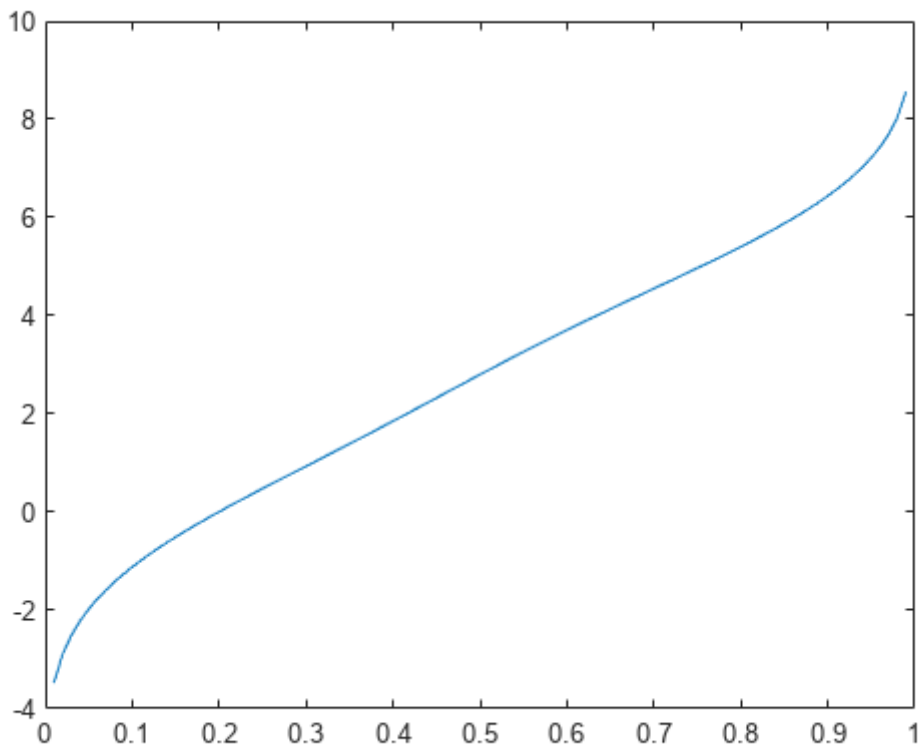
复制命令

```
rng('default') % For reproducibility
x = [randn(30,1); 5+randn(30,1)];
pi = linspace(.01,.99,99);
figure
ksdensity(x,pi,'Function','icdf');
```

获取







#### 返回平滑窗口的带宽

生成两个正态分布的混合。

试试这个例子

复制命令

```
rng('default') % For reproducibility
x = [randn(30,1); 5+randn(30,1)];
```

获取

返回概率密度估计的平滑窗口的带宽。

```
[f,xi,bw] = ksdensity(x);
bw
```

获取

```
bw = 1.5141
```

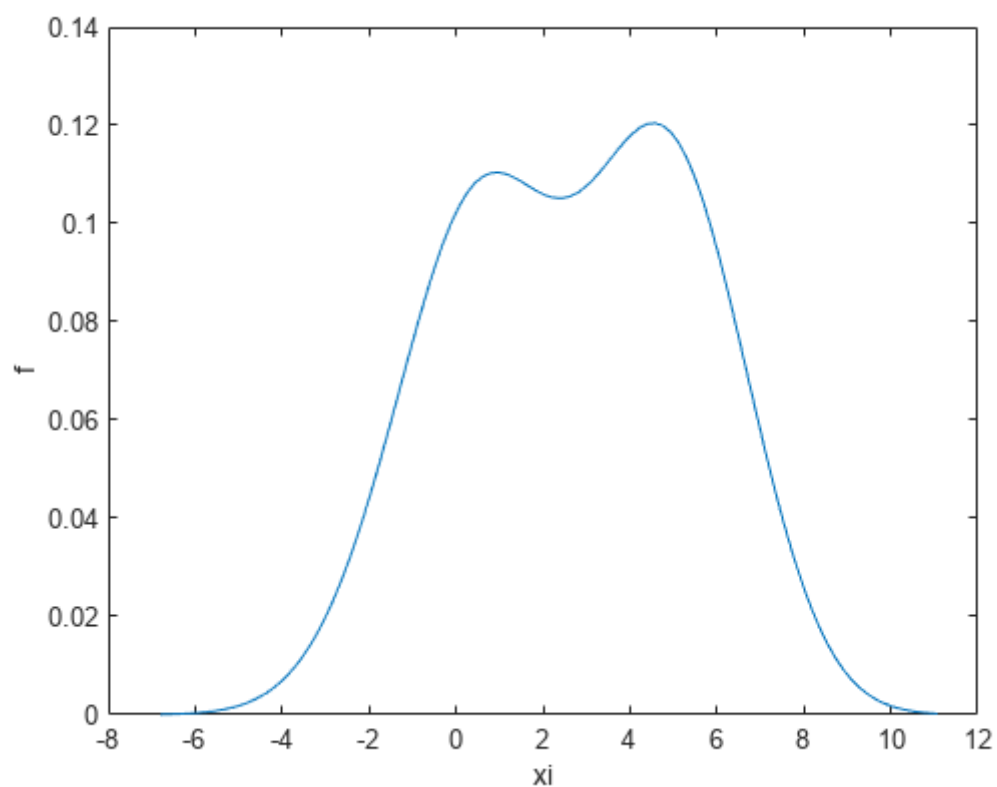
默认带宽最适合正常密度。

绘制估计密度。

```
figure
plot(xi,f);
xlabel('xi')
ylabel('f')
hold on
```

获取

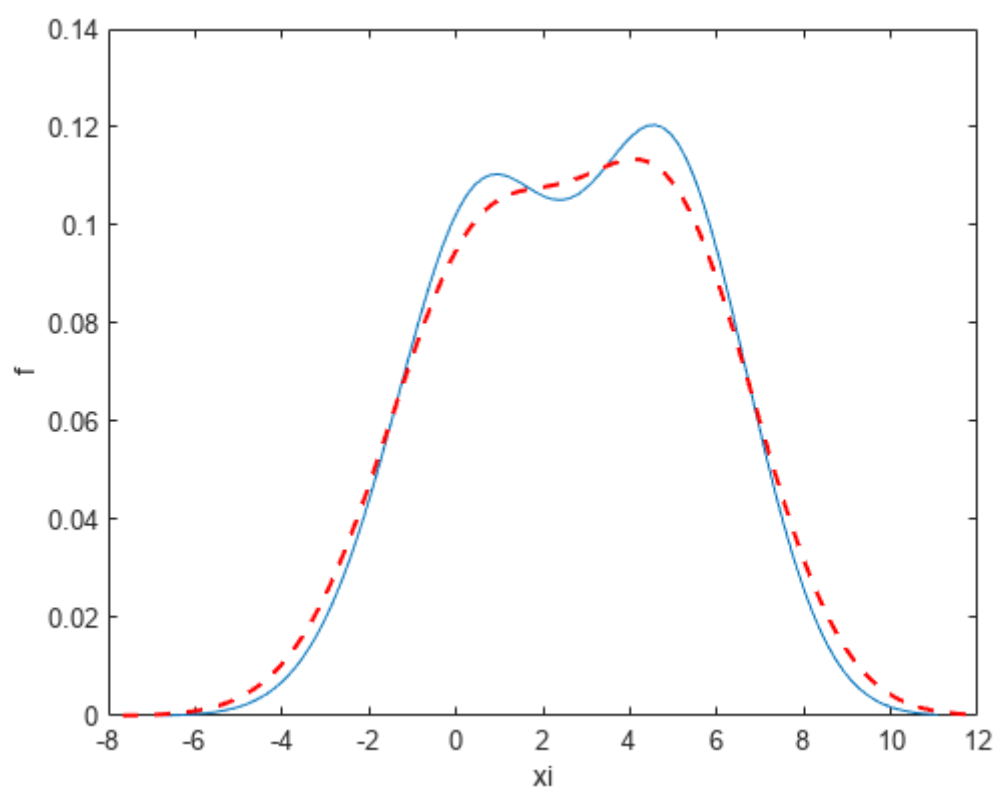




使用增加的带宽值绘制密度。

```
[f,xi] = ksdensity(x,'Bandwidth',1.8);  
plot(xi,f,'--r','LineWidth',1.5)
```

获取



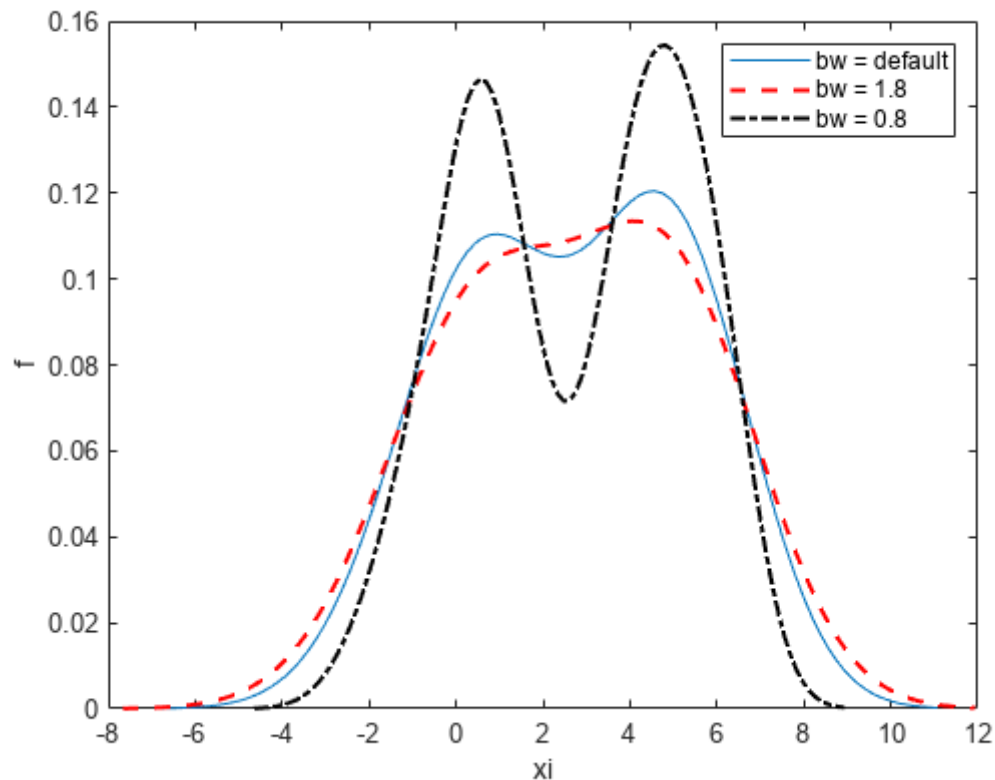
较高的带宽会进一步平滑密度估计值，这可能会掩盖分布的某些特征。

现在，使用减小的带宽值绘制密度。



```
[f,xi] = ksdensity(x,'Bandwidth',0.8);
plot(xi,f,'-.k','LineWidth',1.5)
legend('bw = default','bw = 1.8','bw = 0.8')
hold off
```

获取



较小的带宽使密度估计值平滑得越少，这会夸大样品的某些特征。

## 绘制二元数据的核密度估计

创建一个两列点向量，以评估密度。

试试这个例子

复制命令

```
gridx1 = -0.25:.05:1.25;
gridx2 = 0:.1:15;
[x1,x2] = meshgrid(gridx1, gridx2);
x1 = x1(:);
x2 = x2(:);
xi = [x1 x2];
```

获取

生成一个 30 x 2 矩阵，其中包含来自二元正态分布混合的随机数。

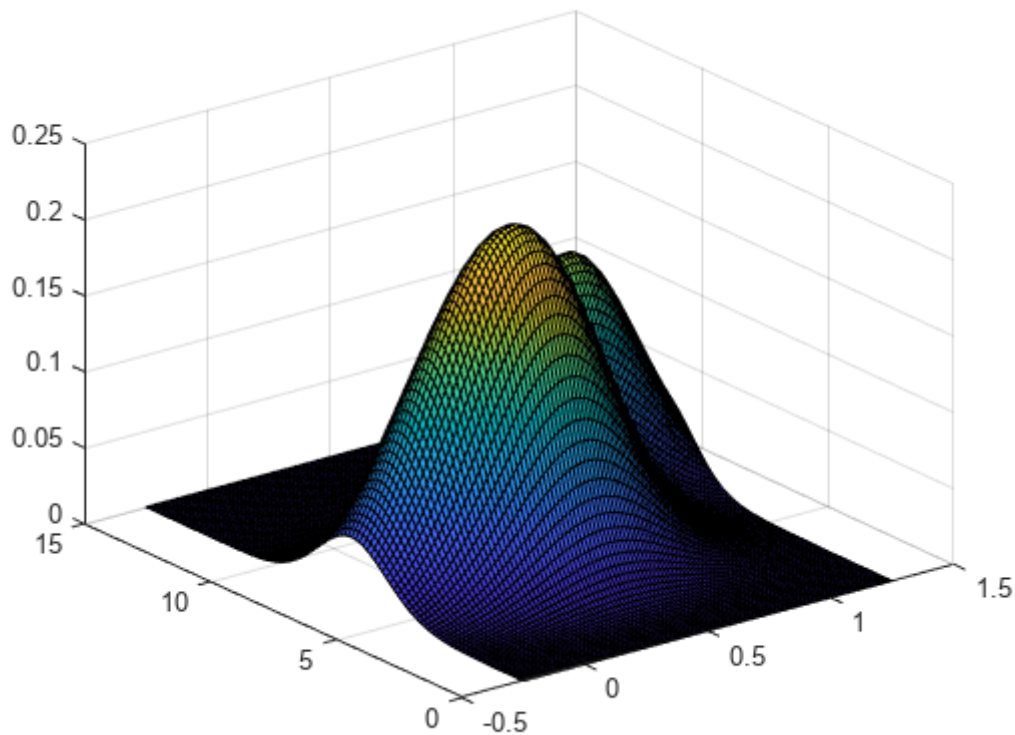
```
rng('default') % For reproducibility
x = [0+.5*rand(20,1) 5+2.5*rand(20,1);
     .75+.25*rand(10,1) 8.75+1.25*rand(10,1)];
```

获取

绘制样本数据的估计密度。

```
figure
ksdensity(x,xi);
```

获取



## 输入参数

全部折叠

- ✓ **x— 示例数据**  
列向量 | 两列矩阵

返回  $f$  值的示例数据，指定为列向量或两列矩阵。使用列向量用于单变量数据，以及用于双变量数据的两列矩阵。ksdensity

例: `[f,xi] = ksdensity(x)`

数据类型: |singledouble

- ✓ **pts— 计算  $f$**   
向量的点 | 两列矩阵

求值  $f$  的点，指定为向量或两列矩阵。对于单变量数据，可以是行或列向量。返回输出的长度等于中的点数。ptsfpts

例: `pts = (0:1:25); ksdensity(x,pts);`

数据类型: |singledouble



## ax— 轴手柄手柄

图形图的轴控点 to, 指定为句柄。ksdensity

例如, 如果 是图形的手柄, 然后可以绘制到该图, 如下所示。hksdensity

**例:** ksdensity(h,x)

## 名称-值参数

将可选的参数对指定为 , 其中 参数名称, 是相应的值。名称-值参数必须出现在其他参数之后, 但 成对无关紧要。Name1=Value1,...,NameN=ValueNNameValue

在 R2021a 之前, 使用逗号分隔每个名称和值, 并用引号括起来。Name

**示例:** “删失”, cens, “内核”, “三角形”, “NumPoints”, 20, “函数”, “cdf” 指定 ksdensity 通过在 20 处求值来估计 cdf 使用三角形核覆盖数据范围的等间距点 矢量CENS中删失数据信息的平滑函数和考虑。

## Bandwidth— 内核平滑窗口 的带宽 正常密度的最佳值 (默认值) |标量值 |双元素向量

内核平滑窗口的带宽, 它是 x 中的点数, 指定为 逗号分隔对由 和 组成 标量值。如果样本数据是二元的, 也可以是双元素向量。这 默认值是估计正常密度 [1] 的最佳选择, 但您可能希望选择 更大或更小的值以平滑或多或少。'Bandwidth'Bandwidth

如果将 “边界校正” 指定为 (默认), 将“支持”指定为向量或向量, 则会转换有界数据 使用日志转换不受限制。的值在转换的规模上 值。'log''positive'[L U]ksdensity'Bandwidth'

**例:** 'Bandwidth',0.8

**数据类型:** |singledouble

## BoundaryCorrection— 边界校正方法 “日志” (默认) | “反思”

边界校正方法, 指定为逗号分隔对 由 和 或 组成。'BoundaryCorrection''log''reflection'



价值	描述
'log'	<p>ksdensity有界转换 数据 x 不受 1 的约束 以下转换。然后，它 之后转换回原始有界比例 密度估计。</p> <ul style="list-style-type: none"> <li>对于单变量数据，如果指定，则应用 。 'Support','positive'ksdensitylog(x)</li> <li>对于单变量数据，如果指定，其中 和 数字标量和，然后应用 .'Support', [L U]LUL &lt; Uksdensitylog((x-L)/(U-x))</li> <li>对于二元数据，转换每个 列以同样的方式 与单变量数据。ksdensityx</li> </ul> <p>“带宽”的值和 带宽输出在刻度上 转换后的值。</p>
'reflection'	<p>ksdensity增强有界 通过在边界附近添加反射数据来获取数据，然后它返回对应于 原始支持。有关详细信息，请参阅反射方法。</p>

ksdensity仅在以下情况下应用边界校正 将“支持”指定为非 值。 'unbounded'

例: 'BoundaryCorrection','reflection'

✓ **Censoring— 0s 的逻辑向量**  
 (默认) |0 和 1s 的矢量

指示哪些条目被删失的逻辑向量，指定为 逗号分隔对由 和 组成 二进制值的向量。值为 0 表示没有审查，1 表示观察被审查。默认值为没有 审查。此名称-值对仅对单变量有效 数据。 'Censoring'

例: 'Censoring',censdata

数据类型: logical

✓ **Function— 估计**  
 “pdf”的功能 (默认) |“CDF” |“ICDF” |“幸存者” |“康哈尔”

要估计的函数，指定为逗号分隔对，包括 和以下之一。 'Function'

价值	描述
'pdf'	概率密度函数。
'cdf'	累积分布函数。
'icdf'	<p>逆累积分布函数。 计算 估计 x 中值的逆 CDF，并在 中指定的概率值。 ksdensitypi</p> <p>此值仅对单变量有效 数据。</p>
'survivor'	幸存者功能。



价值	描述
'cumhazard'	累积危险函数。  此值仅对单变量有效 数据。

**示例：** , 'Function''icdf'

Kernel— 内核类型更平滑“正常”（默认） |“盒子” |“三角形” |“叶帕内奇尼科夫” |功能句柄 |字符矢量 |字符串标量

内核平滑器的类型，指定为逗号分隔对 由和之一组成 以后。'Kernel'

- 'normal'（默认）
- 'box'
- 'triangle'
- 'epanechnikov'
- 作为自定义函数或内置函数的内核函数。将函数指定为函数句柄（例如，或）或字符向量或 字符串标量（例如，或）。该软件调用 指定函数，其中一个参数是 数据值与位置之间的距离，其中 评估密度。该函数必须返回一个数组 包含相应内核值的相同大小 功能。  
@myfunction@normpdf'myfunction''normpdf'

当为 时，内核函数返回 密度值。否则，它返回累积概率 值。'Function''pdf'

指定自定义内核时返回错误。'Function''icdf'

对于二元数据，应用相同的 内核到每个维度。ksdensity

**例：**'Kernel','box'

NumPoints— 等间距点数 100（默认） |标量值

xi 中等间距点的数量，指定 作为由 和 组成的逗号分隔对标量值。此名称-值对仅对单变量有效 数据。'NumPoints'

例如，对于指定函数的内核平滑估计 样本数据范围内的 80 个等间距点，输入：

**例：**'NumPoints',80

**数据类型：** |singledouble

Support— 支持密度“无界”（默认） |“积极” |双元素向量， [L U] |二乘二矩阵， [L1 L2;U1 U2]

支持密度，指定为逗号分隔对 由和之一组成 以后。'support'



价值	描述
'unbounded'	违约。允许密度扩展到整个 真正的线。
'positive'	将密度限制为正值。
双元素向量, [L U]	给出 支持密度。此选项仅对 单变量样本数据。
二乘二矩阵, [L1 L2; U1 U2]	给出 支持密度。第一行包含较低的 限制, 第二行包含上限。此选项仅对双变量样本有效 数据。

对于双变量数据, 可以是组合 指定为 或 的正变量、无界变量或有界变量。'Support'[0 -Inf; Inf Inf][0 L; Inf U]

例: 'Support','positive'

例: 'Support',[0 10]

数据类型: ||singledoublecharstring

PlotFcn— 用于创建核密度图

“surf”的函数（默认）|“轮廓”|“剧情3”|“冲浪”

用于创建核密度图的函数, 指定为 逗号分隔对由 和 1 组成 以下。'PlotFcn'

价值	描述
'surf'	3-D 着色曲面图, 使用冲浪创建
'contour'	等值线图, 使用等值线创建
'plot3'	3-D 线图, 使用 plot3 创建
'surfc'	创建的三维着色曲面图下的等值线图 使用冲浪

此名称/值对仅对双变量样本数据有效。

例: 'PlotFcn','contour'

Weights— 样本数据

向量的权重

样本数据的权重, 指定为逗号分隔对, 由长度和向量组成, 其中 x 是示例数据。'Weights'size(x,1)

例: 'Weights',xw

数据类型: |singledouble

输出参数

f— 估计函数值

向量





估计的函数值，以长度等于点数的向量形式返回 在 `xi` 或 `PTS` 中。

#### ✖ **xi— 评估点 100 个等距点 | 900 个等间距点 | 向量 | 两列矩阵**

计算 `f` 的评估点，以向量或两列矩阵的形式返回。对于单变量数据，默认值为 100 个等距点，覆盖 `x` 中的数据范围。对于二元数据，默认值为 900 使用网格创建的等间距点，从 30 开始 每个维度中间距相等的点。`ksdensity`

#### ✖ **bw— 平滑窗口 标量值的带宽**

平滑窗口的带宽，以标量值返回。

如果将“边界校正”指定为（默认），将“支持”指定为 要么 或向量，将有界数据转换为无界 使用日志转换。的值在 转换值的比例。`'log''positive'[L U]ksdensitybw`

### 更多关于

全部折叠

#### ✖ 内核分发

内核发行版 是随机概率密度函数（PDF）的非参数表示 变量。当参数分布无法正确使用时，可以使用 内核分布 描述数据，或者当您想要避免对 数据。内核分布由平滑函数和带宽值定义，它们 控制生成的 密度曲线的平滑度。

内核密度 估计器是随机变量的估计 PDF。对于  $x$  的任何实值，核密度估计器的公式由下式给出

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right),$$

哪里  $x_1, x_2, \dots, x_n$  是来自未知的随机样本 分布， $n$  是样本数量， $K(\cdot)$  是核平滑函数， $h$  是 带宽。


累积分布函数（cdf）的内核估计器，适用于任何实数  $x$  的值，由下式给出

$$\hat{F}_h(x) = \int_{-\infty}^x \hat{f}_h(t) dt = \frac{1}{n} \sum_{i=1}^n G\left(\frac{x - x_i}{h}\right),$$

哪里  $G(x) = \int_{-\infty}^x K(t) dt$ .

有关更多详细信息，请参阅内核分发。

#### ✖ 反射方法

反射法是一种边界校正方法，在随机变量有界时准确查找核密度估计器 支持。如果指定，则使用  方法。这种方法增强了 通过在边界附近添加反射数据来限定数据，并估计 PDF。然后，返回与 原始支持 与适当的规范化，以便估计的PDF的积分 超过原始支撑等于 1。`'BoundaryCorrection','reflection'ksdensityksdensity`

如果另外指定，则按如下方式查找内核估计器。`'Support',[L U]ksdensity`

- 如果是，则核密度估计器为'Function''pdf'

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n \left[ K\left(\frac{x - x_{我}^-}{h}\right) + K\left(\frac{x - x_{我}}{h}\right) + K\left(\frac{x - x_{我}^+}{h}\right) \right] \text{ 对于 } L \leq x \leq U,$$

哪里  $x_{我}^- = 2L - x_{我}$ ,  $x_{我}^+ = 2U - x_{我}$  和  $x_{我}$  是第 TH 个示例数据。i

- 如果是，则 CDF 的内核估计器是'Function''cdf'

$$\hat{F}_h(x) = \frac{1}{n} \sum_{i=1}^n \left[ G\left(\frac{x - x_{我}^-}{h}\right) + G\left(\frac{x - x_{我}}{h}\right) + G\left(\frac{x - x_{我}^+}{h}\right) \right] - \frac{1}{n} \sum_{i=1}^n \left[ G\left(\frac{L - x_{我}^-}{h}\right) + G\left(\frac{L - x_{我}}{h}\right) + G\left(\frac{L - x_{我}^+}{h}\right) \right]$$

对于  $L \leq x \leq U$ 。

- 为了获得逆 cdf（幸存者函数）的核估计器，或累积危险函数（当为、或时），使用双'Function''icdf''survivor''cumhazrd'ksdensity  $\hat{f}_h(x)$  和  $\hat{F}_h(x)$ 。

如果您另外指定为或，则通过在上面替换为方程。'Support''positive'[0 inf]ksdensity[L U][0 inf]

## 替代功能

您还可以使用 MATLAB kde 函数估计单变量数据的 pdf 或 cdf。与不同，不支持边界校正方法或数据审查。<sup>®</sup>ksdensitykde

## 引用

- [1] 鲍曼, A.W.和A.阿扎利尼。应用的用于数据分析的平滑技术。纽约：牛津大学出版社，1997年。
- [2] Hill, P. D. "Kernel estimation of a Distribution Function." 统计学通讯 - 理论与方法。第14卷，问题。3,1985年，第605-620页。
- [3] Jones, M. C. "核密度的简单边界校正估计。统计和计算。第3卷，第3期，1993年，第135-146页。
- [4] Silverman, B. W. *Density Estimation for Statistics and Data 分析*。查普曼和霍尔/CRC，1986年。

## 扩展功能

### 高数组

使用行数超过内存容量的数组进行计算。

### C/C++ 代码生成

使用 MATLAB® 编码器™生成 C 和C++代码。

### GPU 阵列

通过使用并行计算工具箱™在图形处理单元（GPU）上运行来加速代码。

## 版本历史

在 R2006a 之前推出



## 参见

---

直方图 | MVKS密度 | 凯德

## 主题

使用 ksdensity 拟合核分布

使用 ksdensity 将分布拟合到分组数据

使用概率分布

非参数和经验概率分布

支持的发行版



