Advanced Economics Statistics

FALL 2010

Second-Assignment Answer Sheet

by Freddy Rojas Cama

1. In the following function $f_X(x) = \frac{1}{2}e^{-|x|}, -\infty < x < \infty; y = |x|^3$ we find the pdf of Y and show that the pdf integrates to 1. The distribution of y is a weibull with parameters γ and β.

$$f_y(x) = \begin{cases} \sum_{i=1}^{k} f_x(g_i^{-1}(y)) \left| \frac{d}{d_y} g_i^{-1}(y) \right| & , y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

where $f_x(x) = \frac{1}{2}\exp(-|x|), -\infty < x < \infty; Y = |X|^3$. We have that $g(x) = |x|^3$ and it can turn out in three posible functions or parts as follows:

- $g_1(x) = (-x)^3 = -x^3$, if $x < 0$ and $x = -y^{1/3}$
- $g_2(x) = 0$, if $x = 0$
- $g_3(x) = x^3$, if $x > 0$ and $x = y^{1/3}$

And each derivatives to the inverse function respectively are given by:

- $\frac{d}{d_y} g_1^{-1}(-x^3) = \frac{1}{3}(-y)^{-2/3} = \frac{1}{3}(x^3)^{-2/3} = \frac{1}{3}x^{-2} = \frac{1}{3x^2}$
- $\frac{d}{d_y} g_2^{-1}(0) = 0$
- $\frac{d}{d_y} g_3^{-1}(x^3) = \frac{1}{3}(y)^{-2/3} = \frac{1}{3}(x^3)^{-2/3} = \frac{1}{3}x^{-2} = \frac{1}{3x^2}$

So we know that $(y)^{1/3} = g_i^{-1}(|x|^3)$. So we must replace them into $f_y(x)$

$$f_y(x) = f_x(g_1^{-1}(|x|^3)) \left| \frac{d}{d_y} g_1^{-1}(|x|^3) \right| + f_x(g_2^{-1}(|x|^3)) \left| \frac{d}{d_y} g_2^{-1}(|x|^3) \right| + f_x(g_3^{-1}(|x|^3)) \left| \frac{d}{d_y} g_3^{-1}(|x|^3) \right|$$

Where $f_x(g_i^{-1}(|x|^3)) = \frac{1}{2}\exp\left(-\left|g_i^{-1}(|x|^3)\right|\right)$ and $x = g_i^{-1}(y)$. And the derivatives are:

$$\begin{aligned} f_x(g_1^{-1}(-x^3)) \left| \frac{d}{d_y} g_1^{-1}(|x|^3) \right| &= \frac{1}{2}\exp(-|x|) \left| \frac{1}{3x^2} \right| \\ f_x(g_2^{-1}(0)) \left| \frac{d}{d_y} g_2^{-1}(|x|^3) \right| &= 0 \\ f_x(g_3^{-1}(x^3)) \left| \frac{d}{d_y} g_3^{-1}(|x|^3) \right| &= \frac{1}{2}\exp(-|x|) \left| \frac{1}{3x^2} \right| \end{aligned}$$

Including each term to the main expression:

$$\begin{aligned} f_y(x) &= \frac{1}{2}\exp(-|x|) \left| \frac{1}{3x^2} \right| + \frac{1}{2}\exp(-|x|) \left| \frac{1}{3x^2} \right| \\ &= \exp(-|x|) \left| \frac{1}{3x^2} \right|, \text{ or in terms of } y \\ f_y(x) &= \frac{1}{3}\exp(-y^{1/3})(y)^{-2/3}, 0 < y < \infty \end{aligned}$$

we show that this `pdf` integrates to 1.

$$
\begin{aligned}
\underset{Y\to\infty}{F_Y(y=Y)} &= \int_0^Y f_y(y).dy \\
\underset{Y\to\infty}{F_Y(y=Y)} &= \int_0^Y \frac{1}{3}\exp(-y^{1/3})(y)^{-2/3}dy \\
\underset{Y\to\infty}{F_Y(y=Y)} &= \lim_{Y\to\infty} -\exp(-y^{\frac{1}{3}})|_0^Y \\
\underset{Y\to\infty}{F_Y(y=Y)} &= 1
\end{aligned}
$$

The distribution of y is a weibull with parameters $\gamma = \frac{1}{3}$ and $\beta = 1.\square$

2. The `pdf` (probability density function) of the Weibull(γ, β) distribution is given in (3.3.12) on p.102 of **CB.**

$$f_Y(y|\,\gamma,\beta) = \tfrac{\gamma}{\beta} y^{\gamma-1}\exp(-y^{\gamma/\beta}), \quad 0 < y < \infty, \quad \gamma > 0, \ \beta > 0 \tag{1}$$

we obtain the cumulative density function (`cdf`) of y, $F_Y(y)$ using two methods we describe in the following lines:

(i) By using Theorem 2.1.3 on p.51 of CB and using the fact that the *Weibull* variable is the transformation of $X \sim exp(\beta)$ by $Y = X^{1/\gamma}$. Since we have that $X \sim exp(\beta)$, and $F_Y(y) = F_X(g^{-1}(y)) = F_X(X) = F_X(Y^\gamma)$

$$F_Y(y) = F_X(Y^\gamma) \tag{2}$$

And by integrating 1, we get, the `cdf` for the *weibull* distribution in terms of x:

$$F(x|\gamma,\beta) = 1 - \exp\left(-\frac{x}{\beta}\right)$$

By replacing X for Y^γ, by using 2, we have:

$$F(y|\gamma,\beta) = 1 - \exp\left(-\frac{y}{\beta^{\frac{1}{\gamma}}}\right)^\gamma$$

(ii) This problem calls for using the probability integral tranformation. The Weibull (k, λ^k) is $f_x = \left(\frac{k}{\lambda}\right)\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k}$, in terms of parameters of the problem we have that $k = \gamma$ and $\lambda^k = \beta$. In order to get the expression in the assignment we should perform the integral on the `pdf`, this would conduce us to have an expression por the `cdf`:

$$F_x(X) = \int_0^X f(x)\,dx$$

in terms of the *weibull*

$$F_y(Y) = \int_0^Y \left(\frac{k}{\lambda}\right)\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k}$$

then working on the integral

$$\begin{aligned} F_y(Y) &= -e^{-\left(\frac{y}{\lambda}\right)^k}\Big|_0^y \\ &= -e^{-\left(\frac{y}{\lambda}\right)^k}+1 \\ &= 1-e^{-\left(\frac{y}{\lambda}\right)^k} \end{aligned}$$

3. Again, this problem calls for using the probability integral tranformation. The Weibull (k,λ^k) is $f_x = \left(\frac{k}{\lambda}\right)\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k}$, in terms of parameters of the problem we have that $k=\gamma$ and $\lambda^k=\beta$. In order to get the expression in the assignment we should perform the integral of the `pdf`, this would conduce us to have an expression por the `cdf`:

$$F_x(X) = \int_0^X f(x)\,dx$$

in terms of the *weibull*

$$F_y(Y) = \int_0^Y \left(\frac{k}{\lambda}\right)\left(\frac{y}{\lambda}\right)^{k-1} e^{-\left(\frac{y}{\lambda}\right)^k}$$

then working on the integral

$$\begin{aligned} F_y(Y) &= -e^{-\left(\frac{y}{\lambda}\right)^k}\Big|_0^y \\ &= -e^{-\left(\frac{y}{\lambda}\right)^k}+1 \\ &= 1-e^{-\left(\frac{y}{\lambda}\right)^k} \end{aligned}$$

then $0 < F_y(Y) < 1$, so we can perform the probability integral transformation

$$u = 1-e^{-\left(\frac{y}{\lambda}\right)^k}$$

re-arranging the terms and putting the expresion in terms of random uniform variable $u \in]0,1[$

$$\begin{aligned} 1-e^{-\left(\frac{y}{\lambda}\right)^k} &= u \\ e^{-\left(\frac{y}{\lambda}\right)^k} &= 1-u \\ \left(\frac{y}{\lambda}\right)^k &= -\ln(1-u) \\ y &= \lambda\left(-\ln(1-u)\right)^{\frac{1}{k}} \end{aligned}$$

finally, we have that $k=\gamma$ and $\lambda^k=\beta$

$$y = \left(-\beta\ln(1-u)\right)^{\frac{1}{\gamma}}$$

4. The GAUSS program for drawing weibull random variables is attached in the appendix.This program is a user-friendly one and just you need to run it. We modified the

bandwith[1] for the following values $h = 0.85, 0.5, 0.15$ and 0.05 using a Gaussian kernel density. We also use the Epanechnikov kernel density with the optimal bandwith suggested by Silverman (1986). We report these kernel densities in this section as well as the estimates of the mean, variance, skewness and kurtosis (see graph below). We got the densities using the following kernel formula:

$$\hat{f}_h(x) = \frac{1}{nh} \sum\nolimits_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{3}$$

h is a smoothing parameter called the bandwith. Professor Tsurumi gave us a gaussian function K for performing the density of the random variables. We perform kernel densities using different values of the bandwith (see figure at the top and left side of figure 4.1).

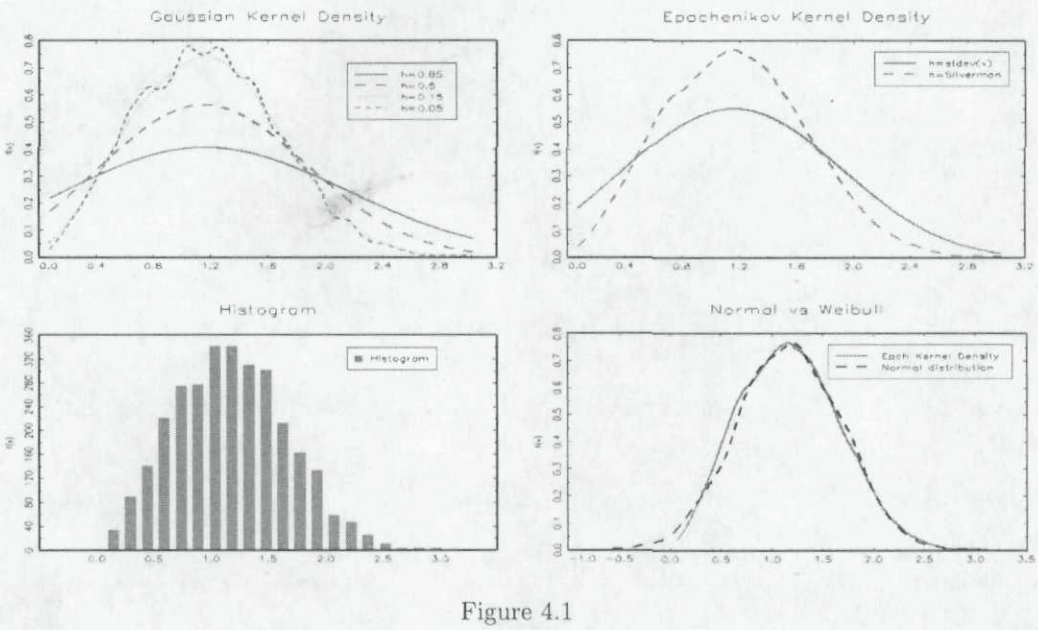

Figure 4.1

In order to leave the gaussian assumption we perform an Epanechnikov kernel:

$$f_h(x) = \frac{1}{nh} \sum\nolimits_{i=1}^{n} \widetilde{K}\left(\frac{x - x_i}{h}\right)$$

in this case

$$\widetilde{K}(z) = \frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}z^2\right)\left\{\mathbf{1}_{|z|<\sqrt{5}}\right\}$$

[1] The original code which was provided by Professor Tsurumi has a bandwith equal to 5.8074269e-008. This value is too small and we get a flat kernel density.

The moments of the distribution are shown in the following table, also we have statistics for the median, maximum and minimum values.

Table 4.1: Statistics

Statistics	Value
Mean	1.1782
Median	1.1598
Max	3.0360
Min	0.0717
Std	0.4930
Skewness	0.2716
Kurtosis	2.7795

The values of skewness and kurtosis are 0.27 and 2.77 respectively. According to `mathworld`, the skewness is a measure of the asymmetry of the probability function. The skewness can be positive or negative or even undefined. A value close to zero -as our calculations indicate for the exercise - indicates that the values are relatively evenly distributed on both sides of the mean[2]. Kurtosis indicates the *peakedness* of the probability function. The kurtosis is the results of extreme deviation respect to its mean. In the exercise we performed, we got a value of 2.77, a value of 3 would indicate the distribution is *mesokurtotic*[3]. In our case we got a negative excess of kurtosis (-0.23), in this case the distribution is called *platykurtic* (a flattened shape). Both statistics indicate that the distribution we are analyzing is similar to the normal distribution (in terms of moments), in other words this distribution has a few deviations from normal case (see figure at the bottom and right side of the figure 4.1).

5. Let X be *continuous, nonnegative random variable, the expected value of this variable is* $E(x) = \int_0^\infty x f(x)\,.dx$. We need to prove that $E(x) = \int_0^\infty [1 - F_X(x)]\,dx$. As requested, we must use in our solving procedure the integration-by-parts method. Considering the following expression: $u \cdot v|_0^\infty = \int_0^\infty u \cdot dv + \int_0^\infty v \cdot du$, so, in terms of our problem we have $u = x$ and $dv = f_X(x)\,dx$ we have the following:

$$
\begin{aligned}
-x\left(\int f_X(x)\,dx\right)\Big|_0^\infty &= \int_0^\infty x \cdot f_X(x)\,dx + \int_0^\infty \left(\int f_X(x)\,dx\right).dx \\
-x\,(1 - F_X)\,|_0^\infty &= \int_0^\infty x \cdot f_X(x)\,dx + \int_0^\infty \left(\int f_X(x)\,dx\right).dx \\
-x\,(1 - F_X)\,|_0^\infty &= \int_0^\infty x \cdot f_X(x)\,dx + \int_0^\infty -(1 - F_X)\,.dx
\end{aligned}
$$

we have used the fact that $\frac{\partial(-(1-F_X))}{\partial x} = f_X(x)$. So, the expected value of X is :

$$\int_0^\infty x \cdot f_X(x)\,dx = -x\,(1 - F_X)\,|_0^\infty + \int_0^\infty (1 - F_X)\,.dx \tag{4}$$

[2] Tipically means asymmetry, but this does not necessarily mean asymmetry (see `wikipedia`)

[3] Consider the normal distribution for comparison; in fact a value of 3 is the fourth moment for the normal distribution.

If $x \to \infty$ then $F_X \to 1$ and $-(1-F_X) \to 0$. Also if $x \to 0$ then $F_X \to 0$ and $-(1-F_X) \to -1$. We should take notice that first term on right side of the expression (4) should be evaluated on ∞ and zero. Because we have an expression which is not defined (∞ times zero) we need to do an arrangement and apply L'Hopital rule. Before this we replace the term $1 - F_X$ with a suitable expression with the same characteristics[4], a exponential function will work on this case because it is defined on the range of x $[0, \infty[$.We do that in the following lines:

$$\lim_{x\to\infty} -x\left(1-F_X\right)$$

$$\lim_{x\to\infty} -xe^{-x}$$

$$\lim_{x\to\infty} \frac{\frac{\partial(-x)}{\partial x}}{\frac{\partial e^x}{\partial x}}$$

$$\lim_{x\to\infty} -\frac{1}{e^x}$$

so

$$\lim_{x\to\infty} -e^{-x} = 0$$

Therefore,

$$\int_0^\infty x \cdot f_X(x)\,dx = \int_0^\infty (1-F_X)\,.dx$$

6. Let X be a discrete random variable whose range is the nonnegative integers. We show that:

$$EX = x_1 P(X = x_1) + x_2 P(X = x_2) + x_3 P(X = x_3) + ...$$

where $x_0 = 0 < x_1 < x_{2....}$ can be expressed as:

$$EX = \sum_{j=0}^{\infty} (x_{j+1} - x_j)\, P(X > x_j)$$

We have that:

$$\begin{aligned} a_1 &= P(X \le x_1) \\ a_2 &= P(X \le x_2) - P(X \le x_1) \\ a_3 &= P(X \le x_3) - P(X \le x_2) \\ &\ldots \\ &\ldots \\ &\ldots \\ a_n &= P(X \le x_n) - P(X \le x_{n-1}) \end{aligned}$$

[4] i.e. If $x \to \infty$ then $F_X \to 1$ and $-e^{-x} \to 0$. Also if $x \to 0$ then $F_X \to 0$ and $-e^{-x} \to -1$.

By performing a sum, we can write the above expression as follows:

$$\begin{aligned}\sum a_j x_j &= x_j \left(P(X \leq x_j) - P(X \leq x_{j-1})\right) \\ &= x_1 \left(P(X \leq x_1) - P(X \leq x_0)\right) ... \\ &\quad ... + x_2 \left(P(X \leq x_2) - P(X \leq x_1)\right) ... \\ &\quad ... + x_3 \left(P(X \leq x_3) - P(X \leq x_2)\right) ... \\ &\quad ... \\ &\quad + x_{j-1} \left(P(X \leq x_{j-1}) - P(X \leq x_{j-2})\right) \\ &\quad + x_j \left(P(X \leq x_j) - P(X \leq x_{j-1})\right)\end{aligned}$$

By re-grouping terms and gathering common expressions we have that:

$$\begin{aligned}\sum a_j x_j &= -(x_2 - x_1)P(X \leq x_1) - (x_3 - x_2)P(X \leq x_2) - ... \\ ...(x_4 - x_3)P(X &\leq x_3) ... - (x_j - x_{j-1})P(X \leq x_{j-1}) + x_j P(X \leq x_j)\end{aligned}$$

Using that $P(X \leq x_j) = 1 - P(X > x_j)$

$$\begin{aligned}\sum a_j x_j &= -(x_2 - x_1)(1 - P(X > x_1)) - (x_3 - x_2)(1 - P(X > x_2)) - ... \\ ...(x_4 - x_3)(1 - P(X &> x_3)) ... - (x_j - x_{j-1})(1 - P(X > x_{j-1})) + x_j \left(1 - P(X > x_j)\right)\end{aligned}$$

$$\begin{aligned}\sum a_j x_j &= (x_2 - x_1)(P(X > x_1) + (x_3 - x_2)(P(X > x_2) + ... \\ ...(x_4 - x_3)(P(X &> x_3) + ...(x_j - x_{j-1})(P(X > x_{j-1}) - x_j P(X > x_j) ... \\ &.... - (x_2 - x_1) - (x_3 - x_2) - (x_4 - x_3) ... - (x_j - x_{j-1}) + x_j\end{aligned}$$

then a initial condition in this case is $x_1 = 0^5$. Also, the ending condition is $P(X > x_j) = 0$, thus we have that

$$\begin{aligned}\sum a_j x_j &= (x_2 - x_1)(P(X > x_1) + (x_3 - x_2)(P(X > x_2) + \\ &= (x_4 - x_3)(P(X > x_3) + ...(x_j - x_{j-1})(P(X > x_{j-1}) - x_j P(X > x_j)\end{aligned}$$

$$EX = \sum_{j=1}^{\infty} (x_{j+1} - x_j) P(X > x_j) \tag{5}$$

The part (b) of the problem states that the expectation of X can be defined as follows:

$$EX = \sum_{j=1}^{\infty} (1 - F_X(j)) \tag{6}$$

as we mentioned at the beginning, X is a discrete variable whose range is the nonnegative integers, we have that $(x_{j+1} - x_j) = 1$ for all j. Furthermore, we use the fact that $P(X \leq x_j) = 1 - P(X > x_j)$, thus (5) is equivalent to (6). If we compare (6) with the expectation of a continuous non-negative variable, we got the same procedure for reckoning the expectation of X.

[5] We began the index with a subscript equal to 1.

References

[1] Casella, G and R. Berger. 2002. Statistical Inference. Second Edition, Duxbury Advanced Studies.

[2] Mendenhall and Scheaffer. 1973. Mathematical Statistics with Applications. Duxbury Press. North Scituate, Massachusetts.

[3] Mathworld website. http://mathworld.wolfram.com/

[4] http://www.math.ust.hk/excalibur/v11_n3.pdf

[5] GAUSS kernel density library. GAUSS.

[6] M.P Wand & M.C Jones. 1995. Kernel Smoothing. Monographs on Statistics and Applied Probability. Chapman & Hall, 1995.

[7] http://compdiag.molgen.mpg.de/docs/talk_05_01_04_stefanie.pdf

1 Appendix

1.1 Gauss Program

```
/*===============================================================================================
   Two parameter Weibull(gamma,beta) file name:  asign02 10.pro (code sent by Hiraki
Tsurumi)
   ===============================================================================================*,
   new;
   library pgraph;
   pqgwin auto;
   graphset;
   #include kernel.src;
   #include density.src;
   " ";
   " ";
   " Professor Tsurumi's Second assignment - Rutgers ";
   " PhD student:  Freddy Rojas Cama ";
   " Last modification:  Sept 19th ";
   /*====Drawing Weibull(2.5,1)======*/
   gam=2.5;
   beta=2;
   n=3000;
   print "n" n;
   seed=123;
   u1=rndus(n,1,seed);
   x= (-beta*ln(1-u1))^(1/gam);
   cls;
   /*====Saving Results======*/
   chdir C:\Users\Fred\Freddy\Personal\Rutgers 2010\Semester I\adv_Statistics\second
assignment\;
   output file=out1.txt reset;
   x;
   output off;
   cls;
   /*****************************************************/
   /* mean, median, max, min, and std of Weibull draws */
   /*****************************************************/
   result=meanc(x)~median(x)~maxc(x)~minc(x)~stdc(x);
   print;
   print "mean median max min std of Weibull random variables";
   print result;
   m=meanc(x);
   sd=stdc(x);
   skew=(x-m)^3/sd^3;
   skew=meanc(skew);
```

```
kurtos=(x-m)^4/sd^4;
kurtos=meanc(kurtos);
print "skewness and kurtosis" skew~kurtos;
/****************************************************************************/
/*===kernel density of Weibull random variables, multiple bandwiths =====*/
/****************************************************************************/
pos=seqa(1,1,n);
x=x[pos];
h=0.85;
he=stdc(x);
{x1,den1}=kden(x,h);
{xe,dene}=kernele(x,he);
h_05=0.5;
he_s=bandw1(x);
{x1_05,den1_05}=kden(x,h_05);
{xe_s,dene_s}=kernele(x,he_s);
h_005=0.05;
{x1_005,den1_005}=kden(x,h_005);
h_015=0.15;
{x1_015,den1_015}=kden(x,h_015);
z=rndn(rows(x),1)*0.5+meanc(x);
{x_gz,den_gz}=kden(z,he_s);
graphset;
begwind;
window(2,2,1);
 _pcolor = { 9 }; /* Colors for series */
 _pmcolor = { 1, 8, 2, 8, 8, 8, 8, 8, 15 };
 /*Colors for axes, title, x and y labels, date, box, and background */
 _plwidth={6.5 }; /*Controls line thickness for main curves*/
 _paxht=0.05; /*Controls size of axes labels*/
 _ptitlht = 0.125; /*Controls main title size */
 _plegstr = " h=0.85 \000 h=0.5 \000 h=0.15 \000 h=0.05";
 _plegctl = { 2 5 6 4.5 };
 title("Gaussian Kernel Density");
xy(x1~x1_05~x1_015~x1_005,den1~den1_05~den1_015~den1_005);
nextwind;
 _pcolor = { 9 3 }; /* Colors for series */
 _plegstr = " h=stdev(x) \000 h=Silverman";
 _plegctl = { 2 5 6 4.8 };
 title("Epachenikov Kernel Density");
xy(xe~xe_s,dene~dene_s);
nextwind;
 _plegstr = "Histogram";
 _plegctl = { 2 5 6 5 };
 title("Histogram");
```

```
hist(x,20);
nextwind;
 title("Normal vs Weibull");
 _pcolor = { 9 8}; /* Colors for series */
 _plegstr = " Epch Kernel Density \000 Normal distribution";
 _plegctl = { 2 5 5.2 4.8 };
xy(xe_s~x_gz,dene_s~den_gz);
endwind;
end;
/* ==================================================*/
/* kernel density estimation; Tsurumi's original code*/
/* but modified by Freddy Rojas */
/* ==================================================*/
proc(2)=kden(v,h);
local g,j,nn,res;
nn=rows(v);
" ";
" ";
@print "h ";@
@h;@
g=0;
j=1;
do while j <= nn;
g=g|meanc(pdfn((v[j]-v)./h))./h;
j=j+1;
endo;
res=sortc(v~g[2:nn+1],1);
retp(res[.,1],res[.,2]);
endp;
/* ================================================*/
/* Epachenikov kernel density estimation; */
/* ================================================*/

proc(2)= kernele(z,h);
local a,res,t,g,z_v,zv,zv_,j;
j=1;
g=0;
do while j <= rows(z);
 zv=(z[j]-z)./h;

 t=(abs(zv).<sqrt(5));
 a=code(t,sqrt(5)|1);
 zv_=t.*((3/4)*(1-(1/5).*(zv.^2))./a);
 g=g|meanc(zv_)./h;
j=j+1;
```

```
endo;
 res=sortc(z~g[2:rows(z)+1],1);
 retp(res[.,1],res[.,2]);
endp;
```