EXAM C QUESTIONS OF THE WEEK

S. Broverman, 2005

Question 1 - Week of July 25

You are given the following random sample of observations:

3 , 4 , 6 , 6 , 7 , 8 , 8 , 8 , 9 , 12

(a) Find the smoothed empirical estimate of the median.

(b) Find the estimate of the median using the kernel smoothed distribution function based on a uniform kernel with a bandwidth of 2.

Solution can be found below.

Question 1 - Week of July 25 - Solution

(a) With $n = 10$ data points, the smoothed percentiles of the given data points are

x	Smoothed Percentile
3	$\frac{1}{11} = 0.0909$
4	$\frac{2}{11} = 0.1818$
6	$\frac{3}{11} = 0.3636$
6	$\frac{4}{11} = 0.3636$
7	$\frac{5}{11} = 0.4545$
8	$\frac{6}{11} = 0.5455$
8	$\frac{7}{11} = 0.6364$
8	$\frac{8}{11} = 0.7273$
9	$\frac{9}{11} = 0.8182$
12	$\frac{10}{11} = 0.9091$

Since 0.5 is between $\frac{5}{11} = 0.4545$ and $\frac{6}{11} = 0.5455$, the smoothed empirical estimate of the 50th percentile will be between $x = 7$ and $x = 8$. The smoothed empirical 50th percentile, say m, is found by linear interpolation. The proportion of the way that m is between $x = 7$ and $x = 8$ is the same proportion of the way that 0.5 is between $\frac{5}{11} = 0.4545$ and $\frac{6}{11} = 0.5455$, which is the same proportion of the way that 5.5 is between 5 and 6 (we get these by multiplying 0.5 , 0.4545 and 0.5455 by 11). We see that 5.5 is $\frac{1}{2}$ of the way from 5 to 6, and therefore m is $\frac{1}{2}$ of the way from $x = 7$ to $x = 8$. Therefore, $m = 7\frac{1}{2} = 7.5$ is the smoothed empirical estimate of the 50th percentile.

(b) The empirical probability function is
$p(3) = .1$, $p(4) = .1$, $p(6) = .2$, $p(7) = .1$, $p(8) = .3$, $p(9) = .1$, $p(12) = .1$.

With a bandwidth of 2, the bands are $[1,5]$ for $y = 3$, $[2,6]$ for $y = 4$, $[4,8]$ for $y = 6$, $[5,9]$ for $y = 7$, $[6,10]$ for $y = 8$, $[7,11]$ for $y = 9$ and $[10,14]$ for $y = 12$

The cumulative kernel functions are
$K_3(x) = .25(x-1)$ for $1 \le x \le 5, K_3(x) = 0$ for $x < 1$ and $K_3(x) = 1$ for $x > 5$,
$K_4(x) = .25(x-2)$ for $2 \le x \le 6, K_4(x) = 0$ for $x < 2$ and $K_4(x) = 1$ for $x > 6$,
$K_6(x) = .25(x-4)$ for $4 \le x \le 8, K_6(x) = 0$ for $x < 4$ and $K_6(x) = 1$ for $x > 8$,
$K_7(x) = .25(x-5)$ for $5 \le x \le 9, K_7(x) = 0$ for $x < 5$ and $K_7(x) = 1$ for $x > 9$,
$K_8(x) = .25(x-6)$ for $6 \le x \le 10, K_8(x) = 0$ for $x < 6$ and $K_8(x) = 1$ for $x > 10$,
$K_9(x) = .25(x-7)$ for $7 \le x \le 11, K_9(x) = 0$ for $x < 7$ and $K_9(x) = 1$ for $x > 11$, and
$K_{12}(x) = .25(x-10)$ for $10 \le x \le 14, K_{12}(x) = 0$ for $x < 10$ and $K_{12}(x) = 1$ for $x > 14$.

The kernel smoothed distribution function is
$\widehat{F}(x) = \sum_y p(y) \cdot K_y(x)$.

For $x < 1$, $\widehat{F}(x) = 0$.

For $1 \leq x < 2$, $\widehat{F}(x) = (.1)[.25(x-1)]$, which ranges from 0 to .025 .

For $2 \leq x < 4$, $\widehat{F}(x) = (.1)[.25(x-1)] + (.1)[.25(x-2)]$, which ranges from .025 to .125 .

For $4 \leq x < 5$, $\widehat{F}(x) = (.1)[.25(x-1)] + (.1)[.25(x-2)] + (.2)[.25(x-4)]$,
which ranges from .125 . to .225 .

For $5 \leq x < 6$, $\widehat{F}(x) = (.1)(1) + (.1)[.25(x-2)] + (.2)[.25(x-4)] + (.1)[.25(x-5)]$,
which ranges from .225 . to .325 .

For $6 \leq x < 7$,
$\widehat{F}(x) = (.1)(1) + (.1)(1) + (.2)[.25(x-4)] + (.1)[.25(x-5)] + (.3)[.25(x-6)]$,
which ranges from .325 . to .475 .

For $7 \leq x < 8$,
$$\widehat{F}(x) = (.1)(1) + (.1)(1) + (.2)[.25(x-4)] + (.1)[.25(x-5)] + (.3)[.25(x-6)] + (.1)[.25(x-7)] ,$$
which ranges from .475 . to .650 .

The estimated median based on the kernel smoothed distribution function would occur at $x = m$, where $\widehat{F}(m) = .50$. Since $\widehat{F}(x)$ is a linear function, and since $\widehat{F}(7) = .475$ and $\widehat{F}(8) = .650$, we see that m must be $\frac{.025}{.175} = \frac{1}{7} = .143$ of the way from $x = 7$ to $x = 8$. Thus, $m = 7.14$.