$$f(x,y) = \frac{1}{h^2 n} \sum_{i=1}^{n} W\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)$$

(In practice, as mentioned above, we will usually be applying this formula to x^*, y^*, x_i^* and y_i^*, instead of x, y, x_i and y_i.)

As in the univariate case, we can think of $f(x, y)$ as the result of smearing the data points. Each (x_i, y_i) is replaced by a function of (x, y), namely $h^{-1}W\left(\frac{x-x_i}{h}, \frac{y-y_i}{h}\right)$, which is the smeared contribution of the observation (x_i, y_i) to the density at (x, y). Furthermore, since our cylinder smearing function has a discontinuity at all points on the unit circle $u^2 + v^2 = 1$, we might decide to replace the function with one that also has total integral one but decreases smoothly from a maximum at (0,0) to zero on the unit circle. This will remove the discontinuity. For example, we could use

$$W(u,v) = \begin{cases} \dfrac{1 + \cos \pi(\sqrt{u^2+v^2})}{\pi} & \text{if } u^2 + v^2 \leqslant 1 \\ 0 & \text{otherwise} \end{cases}$$

This function would give smoother two-dimensional local densities just as the cosine function in Section 2.9 gave smoother one-dimensional densities.

*4.11 MATHEMATICAL DETAILS OF LOWESS

In the lowess procedure the user chooses f, which is approximately the fraction of points to be used in the computation of each fitted value. Let q be fn rounded to the nearest integer. First, we will describe the computation of the neighborhood weight functions shown in the lower left panels of Figures 4.14 and 4.15. Let d_i be the distance from x_i to its qth nearest neighbor along the x axis. (x_i is counted as a neighbor of itself.) Let $T(u)$ be the *tricube weight function*:

$$T(u) = \begin{cases} (1 - |u|^3)^3 & \text{for } |u| < 1 \\ 0 & \text{otherwise} \end{cases} .$$

Then the weight given to the point (x_k, y_k) when computing a smoothed value at x_i is defined to be

$$t_i(x_k) = T\left(\frac{x_i - x_k}{d_i}\right).$$

If d_i is 0, meaning that the q nearest neighbors of x_i all have abscissas equal to x_i, then points whose abscissas are equal to x_i are given weight 1 and all other points are given weight 0. In this case, since the slope of a fitted line cannot be estimated, a constant is fit instead of a line.

To compute a fitted value at x_i in the first stage of lowess, a line (or constant if $d_i = 0$) is fitted to the points of the scatter plot using weighted least squares with weight $t_i(x_k)$ at the point (x_k, y_k). That is, values of a and b are found which minimize

$$\sum_{k=1}^{n} t_i(x_k)(y_k - a - bx_k)^2.$$

If $\hat{a}$ and $\hat{b}$ are the values that achieve the minimum, then the initial fitted value at x_i is defined to be

$$\hat{y}_i = \hat{a} + \hat{b}x_i.$$

After the computation of initial fitted values for all x_i, residuals are computed,

$$r_i = y_i - \hat{y}_i,$$

and robustness weights are computed from them. Let $B(u)$ be the bisquare weight function:

$$B(u) = \begin{cases} (1-u^2)^2 & \text{for } |u| < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let m be the median of the absolute values of the residuals, that is,

$$m = \text{median}|r_k|.$$

The robustness weight for the point (x_k, y_k) is defined to be

$$w(x_k) = B\left(\frac{r_k}{6m}\right).$$

The median absolute residual, m, is a measure of how spread out the residuals are. If a residual is small compared with $6m$, the corresponding robustness weight will be close to 1; if a residual is greater than $6m$, the corresponding weight is 0. Suppose the r_i behave very much like a sample from a normal distribution. Then m nearly estimates $2\sigma/3$ where σ is the population standard deviation, and so $6m$ nearly estimates 4σ. Thus for well-behaved normal residuals we would very seldom have a weight as small as 0.

The next stage is to get updated fitted values, by fitting lines again, but this time incorporating the robustness weights. In the weighted linear regression for refitting $\hat{y}_i$, the point (x_k, y_k) is given weight $w(x_k)t_i(x_k)$. If (x_k, y_k) is a peculiar point with a large residual it will play a small role, or no role at all, in any of the fitted lines in this stage of the computation.

4.12 SUMMARY AND DISCUSSION

Scatter plots are a powerful tool for helping us to understand the relationship between two measured or observed variables, x and y. When x is a factor and y is a response, scatter plots can show us how the empirical distribution of y depends on x. In the exchangeable case, when neither x nor y is regarded as a factor or a response, scatter plots can tell us much about the bivariate empirical distribution of x and y.

Strip medians and strip box plots show how the empirical distribution of y in vertical strips varies from strip to strip on the scatter plot. They are an easy-to-make summary but the price paid is to display no information about the change in the distribution of y within strips.

Smoothing procedures require more computation but provide a more detailed look at how various aspects of the local distribution of y depend on x. Smoothed values from lowess show the middle of the distribution of y as it changes with x. If we compute residuals from these smoothed values, and plot the absolute values of the residuals against the x_i, along with smoothed absolute residuals, we can see how the spread of the distribution of y depends on x.

Repeated points on scatter plots are a problem that we cannot ignore. Both jittering and sunflowers can be used to address the problem. Sunflowers together with cellulation provide counts in square regions even when points do not exactly coincide.

Two-dimensional local densities can be computed in a manner analogous to one-dimensional local densities. Portraying the local density values using several sharpened scatter plots with different cut-