I. Explanation of Code

To understand the coding of standard errors for `DCdensity` requires a derivation not included in McCrary (2007). Let $Y_j \equiv \frac{1}{nb}\sum_{i=1}^{n} \mathbf{1}(g(R_i) = X_j)$ and $S_{n,k} = h^k \sum_{j=1}^{J} K(t_j)\mathbf{1}(t_j > \frac{c-r}{h})t^k$ where $t_j = (X_j - r)/h$. Fix $r > c$. Then the density estimator is

$$\widehat{f}(r) = \sum_{j=1}^{J} K(t_j)\mathbf{1}\left(t_j > \frac{c-r}{h}\right)\frac{S_{n,2} - S_{n,1}ht_j}{S_{n,2}S_{n,0} - S_{n,1}^2}Y_j \equiv \frac{1}{n}\sum_{i=1}^{n} Z_{in} \tag{1}$$

Define $m = \max\{-1, \frac{c-r}{h}\}$ and note that $-1 \leq m \leq 0$ since we are considering the case where $r > c$. For $m = 0$, we obtain the boundary case, and for $m = -1$ we obtain the interior case. Intermediary points are those for which there is still some data to the left of r which can be used to estimate $f(r)$, but not as much as if r were far away from c.

By Riemann approximation,

$$\begin{aligned} S_{n,k} &= \frac{h}{b}h^k \sum_{j=1}^{J} \frac{h}{b}K(t_j)\mathbf{1}\left(t_j > \frac{c-r}{h}\right) = \frac{h^{k+1}}{b}\int_m^1 K(t)t^k dt + O\left(\frac{h^{k+1}}{b}\frac{b^2}{h^2}\right) \\ &= \frac{h^{k+1}}{b}\frac{1}{(k+1)(k+2)}\left(1 - m^{k+1}\left\{k(1+m) + 2 + m\right\}\right) + O\left(bh^{k-1}\right) \end{aligned} \tag{2}$$

Using this result for $k = 0, 1, 2$ shows that

$$\frac{S_{n,2} - S_{n,1}ht_j}{S_{n,2}S_{n,0} - S_{n,1}^2} = 6\,\frac{b}{h}\,\frac{A_2 - 2A_1t_j}{3A_2A_0 - 2A_1^2} + O\left(\frac{b^2}{h^2}\right) \tag{3}$$

where $A_0 = 1 - m(2+m)$, $A_1 = 1 - m^2(3+2m)$, and $A_2 = 1 - m^3(4+3m)$. Since $g(R_i) = X_j$ if and only if $X_j - b/2 < R_i \leq X_j + b/2$, we have $E[\frac{1}{b}\mathbf{1}(g(R_i) = X_j)] = f(X_j) + O(b^2)$. One can use these facts to show that to second order, the expectation of $\widehat{f}(r)$ is $f(r)$. Our focus here is on the variance. Following the calculations in McCrary (2007), we have

$$\begin{aligned} V[\widehat{f}(r)] &= \frac{1}{nh}f(r)\;36\int_m^1 K^2(t)\left(\frac{A_2 - 2A_1t}{3A_2A_0 - 2A_1^2}\right)^2 dt + O\left(\frac{1}{n}\right) \\ &\approx \frac{12f(r)}{5nh}\frac{2 - 3m^{11} - 24m^{10} - 83m^9 - 72m^8 + 42m^7 + 18m^6 - 18m^5 + 18m^4 - 3m^3 + 18m^2 - 15m}{(1 + m^6 + 6m^5 - 3m^4 - 4m^3 + 9m^2 - 6m)^2} \end{aligned} \tag{4}$$

One can verify that when $m = 0$, this formula reduces to the boundary case variance $\frac{1}{nh}\frac{24}{5}f(r)$, and that when $m = -1$, this reduces to the interior case variance $\frac{1}{nh}\frac{2}{3}f(r)$.

Now consider the case with $r < c$. The density estimator is then

$$\widehat{f}(r) = \sum_{j=1}^{J} K(t_j)\mathbf{1}\left(t_j < \frac{c-r}{h}\right)\frac{S_{n,2} - S_{n,1}ht_j}{S_{n,2}S_{n,0} - S_{n,1}^2}Y_j = \frac{1}{n}\sum_{i=1}^{n} Z_{in} \tag{5}$$

with $S_{n,k} = h^k \sum_{j=1}^{J} K(t_j)\mathbf{1}\left(t_j < \frac{c-r}{h}\right)t_j^k$. A similar analysis to that above shows that

$$S_{n,k} = \frac{h^{k+1}}{b}\frac{1}{(k+1)(k+2)}\left((-1)^k + m^{k+1}\left\{k(1-m) + 2 - m\right\}\right) + O\left(bh^{k-1}\right) \tag{6}$$

where now $m = \min\left\{\frac{c-r}{h}, 1\right\}$, with $0 \leq m \leq 1$. We can use this result to show that

$$\frac{S_{n,2} - S_{n,1}ht_j}{S_{n,2}S_{n,0} - S_{n,1}^2} = 6\,\frac{b}{h}\,\frac{A_2 - 2A_1t_j}{3A_2A_0 - 2A_1^2} + O\left(\frac{b^2}{h^2}\right) \tag{7}$$

where $A_0 = 1 + m(2 - m)$, $A_1 = -1 + m^2(3 - 2m)$, and $A_2 = 1 + m^3(4 - 3m)$. This leads to the expression

$$V[\widehat{f}(r)] = \frac{1}{nh} f(r)\ 36 \int_m^1 K^2(t) \left(\frac{A_2 - 2A_1 t}{3A_2 A_0 - 2A_1^2} \right)^2 dt + O\left(\frac{1}{n}\right) \tag{8}$$

$$\approx \quad \frac{12 f(r)}{5nh} \frac{2 + 3m^{11} - 24m^{10} + 83m^9 - 72m^8 - 42m^7 + 18m^6 + 18m^5 + 18m^4 + 3m^3 + 18m^2 + 15m}{(1 + m^6 - 6m^5 - 3m^4 + 4m^3 + 9m^2 + 6m)^2}$$

One can verify that when $m = 0$, this reduces to $\frac{1}{nh}\frac{24}{5} f(r)$, and that when $m = 1$, this reduces to $\frac{1}{nh}\frac{2}{3} f(r)$.

References

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, forthcoming 2007.