FIGURE 1. “Whisk in 1 tsp white wine vinegar.”

PAVA AND PAVLOVA: AN R VINAIGRETTE

ROGER KOENKER

1. Introduction

Isotonic regression has a long and illustrious history beginning with work in the 1950s by van Eeden and the four B’s as elegantly sketched in de Leeuw et al. (2009). It was from this paper and the associated R package **isotone** that I first learned about the generalization of the classical “pool adjacent violators algorithm” (PAVA) method of estimation for isotonic mean regression to isotonic median, and quantile regression. In Figure 1 I’ve illustrated an example where because the conditional distribution of the response, $y|x$ is right skewed, the fitted mean function lies above the fitted median function over most of the domain.

```
require(isotone)
set.seed(1729)
n <- 100
x <- sort(rchisq(n,4)) + 2
y <- log(x)+ .1*(log(x))^2 + log(x)*rchisq(n,4)/4
plot(x, y, cex = 0.5, col = "blue")
plotpava <- function(x, col = 1){
```

February 15, 2021. A genre manifesto for R Vinaigrettes is available at http://davoidofmeaning.blogspot.com/2016/12/r-vinaigrettes.html .

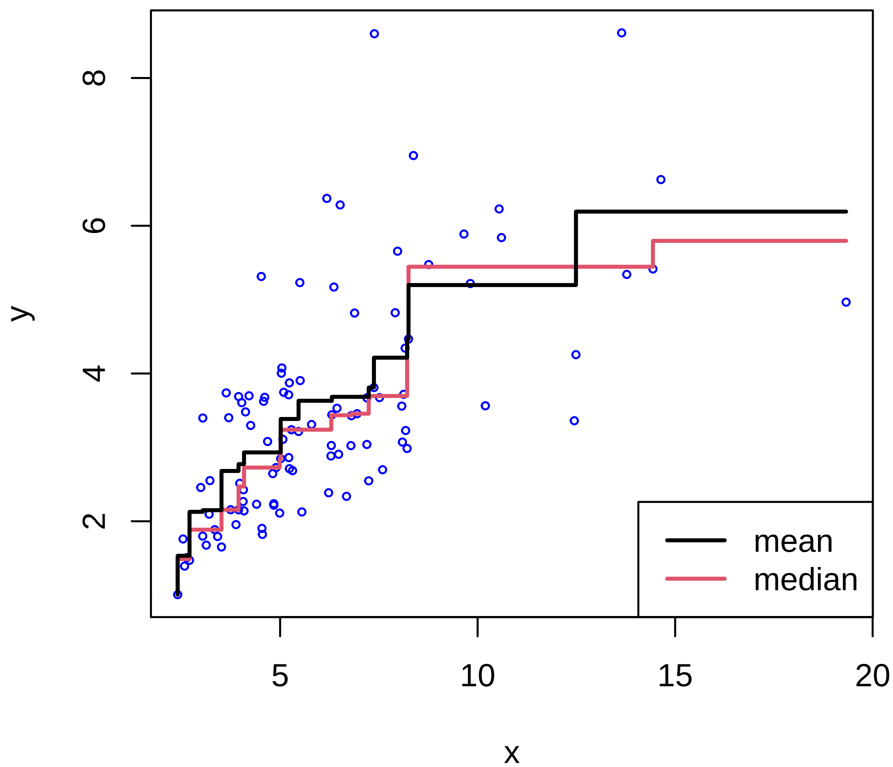

FIGURE 2. Comparison of mean and median isotonic regression: The response is conditionally rescaled χ^2 and consequently the isotonic mean fit is somewhat above the conditional median fit for this data.

```
    o <- order(x$z)
    xval <- x$z[o]
    yval <- x$x[o]
    lines(xval, yval, type = "S", col = col, lwd = 2)
}
g1 <- gpava(x,y,solver = weighted.fractile, p = 0.5)
g2 <- gpava(x,y)
plotpava(g1, col = 2)
plotpava(g2)
legend("bottomright", c("mean","median"), lwd = 2, col = 1:2)
```

The piecewise constant form of these estimates raised the question: Is there an equivalent formulation of the underlying optimization problem based on total variation penalization? In my experience TV penalization is usually imposed on the derivatiive, or gradient, of the fitted function, but there is a substantial literature on penalization of the fitted function itself, this leads naturally to fitted functions that are piecewise constant, rather than piecewise linear. Thus, to take the simplest example we might consider the penalized quantile regression problems,

$$\hat{g} = \text{argmin}_{g \in \mathcal{G}} \sum_{i=1}^{n} \rho_\tau(y_i - g(x_i)) + \lambda TV(g).$$

This is closely related to Davies and Kovac (2001) "taut-string" methods except that there the emphasis is on estimation of mean functions, and they have a much more sophisticated approach to a local rather than global penalty parameter, λ. The penalty enforces the piecewise constant form of the solutions since any departure from this class of solutions would inflate the penalty without improving the fidelity. From here, it is tempting to consider imposing the further constraint that $\hat{g}$ be monotone. Since we only need worry about the fitted function values at the observed design points, this entails only the requirement that the (ordered) $\hat{y}_i$'s be increasing, or decreasing.

Fortuitously, this formalism is easily implemented in the `rqss` framework of my **quantreg** package in R, so it is simple to compare solutions with those from the **isotone** package. The only fly in the ointment seems to be what to do with λ? We don't really want to do any smoothing, so we can just set $\lambda = 0$. I was doubtful that this would work, and assumed at first that I'd have to use some bogus "small" value, but to my surprise, zero seems to be fine.

```
set.seed(1729)
n <- 100
x <- sort(rchisq(n,4)) + 2
y <- log(x)+ .1*(log(x))^2 + log(x)*rchisq(n,4)/4
plot(x, y, cex = 0.5, col = "blue")
f  <- rqss(y ~ qss(x, constraint= "I", lambda = 0, Dorder = 0), tau = .501)
g1 <- gpava(x,y,solver = weighted.fractile, p = .501)
plotpava(g1)
plot(f, add = TRUE, col = 2)
Rho <- function(u, tau = 0.5) sum(u * (tau - (u < 0)))
R0 <- f$fidelity
R1 <- Rho(g1$y - g1$x, tau = .501)
legend("bottomright", c("isotone","TV Penalty"), lwd = 2:1, col = 1:2)
```

Note that in the `qss` term of the `rqss` fitting function I've specified `Dorder = 0` so that the TV penalty acts on g rather than g'. I've also specified `tau = 0.501` which may seem a bit odd, but serves the useful purpose that it avoids the multiple solutions that arise when `tau = 0.5` is specified. It should be noted that the same solutions can also be obtained by specifying `Dorder = 1` since the monotonicity enforces the same piecewise constant solutions.

In retrospect it is hardly surprising that the two problems produce the same solution, although I still find it remarkable that the two rather different algorithms, the active-set algorithm from the **isotone** package and the sparse Frisch-Newton interior point algorithm from **quantreg** produce solutions whose objective functions agree to eleven decimal digits.

2. Conclusion

Arguably, an advantage of the penalized version is that it fits nicely into the additive modeling framework of the `rqss` function, so one could easily incorporate additional covariate

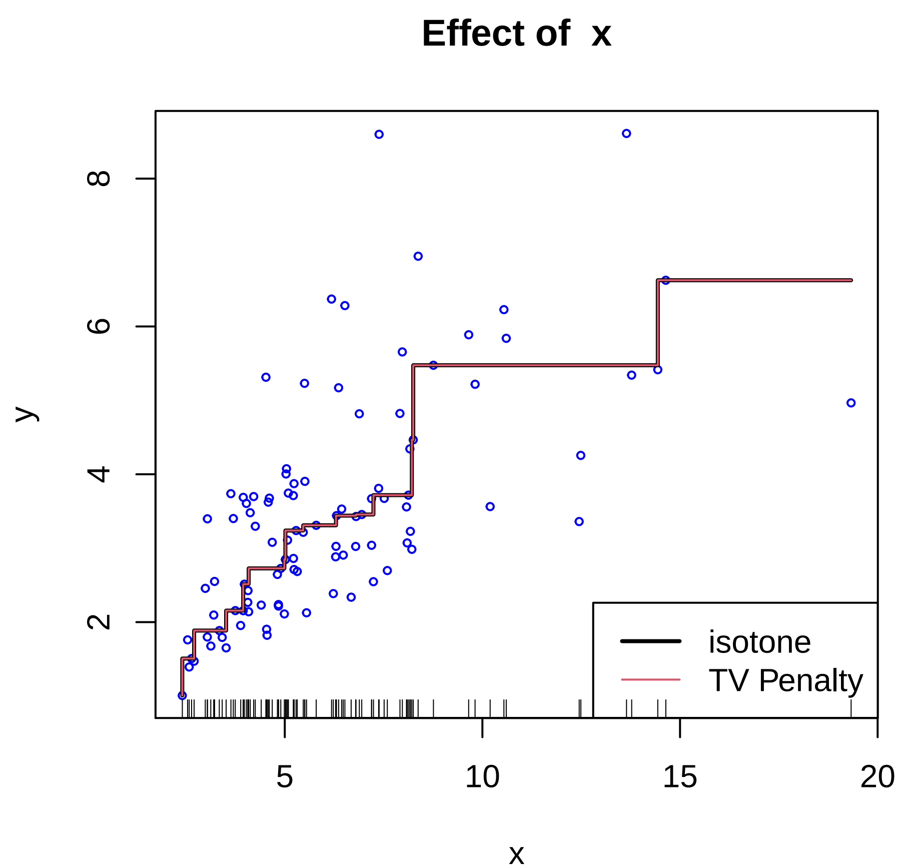

FIGURE 3. Comparison of median isotonic regression with penalized median regression: The response is the same conditionally rescaled χ^2. The TV penalized solution overplots the isotonic solution and the realized values of the objective function agree to 11 decimal digits.

effects. As always with shape constrained estimation, inference is a challenging subject. One might hope that having yet another perspective on the formulation of such problems might eventually shed some new light on these questions as well.

References

Davies, P. L. and Kovac, A. (2001), 'Local extremes, runs, strings and multiresolution', *Annals of Statistics* **29**, 1–65.

de Leeuw, J., Hornik, K. and Mair, P. (2009), 'Isotone optimization in r: Pool-adjacent-violators algorithm (pava) and active set methods', *Journal of Statistical Software, Articles* **32**, 1–24.