

ROBUST ESTIMATION OF THE GAUSSIAN SEQUENCE MODEL: AN R VINAIGRETTE

ROGER KOENKER

ABSTRACT. Sensitivity of posterior mean denoising based on the NPMLE of Gaussian mixture models is explored by replacing Gaussian likelihood with the classical Huber base density. Some protection against heavier tailed noise distributions is thereby achieved at modest cost in the Gaussian setting.

1. INTRODUCTION

Recent work on nonparametric empirical Bayes methods has emphasized the central role of the Gaussian sequence model, $Y_i \sim \mathcal{N}(\theta_i, 1)$, $i = 1, \dots, n$. Exchangeability of the Y_i 's implies that they have marginal, mixture density,

$$f_G(y) = \int \phi(y - \theta) dG(\theta).$$

For quadratic loss, optimal denoising of the Y_i is achieved by the ‘‘Tweedie formula’’ for the posterior mean,

$$\delta(y_i) = \mathbb{E}(\theta|Y = y_i) = y_i + f'_G(y_i)/f_G(y_i).$$

Of course, the mixing distribution, G , is typically unknown to the investigator, but it can be estimated as suggested in Robbins (1950), Kiefer and Wolfowitz (1956) and Koenker and Mizera (2014), which provides a convenient plug-in procedure for constructing a feasible $\delta_{\hat{G}}$.

A natural question that arises from this line of inquiry is: What justification do we have for the Gaussian noise assumption? Aside from predicting baseball batting averages where a binomial approximation argument can be relied upon, there seems to be no good general argument for the Gaussian assumption. When the base density is heavier tailed than the Gaussian, it becomes more difficult to distinguish signal – due to G , from noise.

2. A SIMULATION EXPERIMENT

Optimality, or near optimality, at the Gaussian model does not ensure good performance under modest departures from the Gaussian model as the seminal work of Tukey (1960) and Huber (1964) showed. Fortunately, the least favorable alternative to the Gaussian location model derived in Huber (1964) preserves the convex structure of the NPMLE problem. So it seems worthwhile to explore its performance under alternative noise assumptions. As an initial step, I generate data from the contaminated Gaussian model

$$F(u) = (1 - \alpha)\Phi(u) + \alpha\Phi(u/\sigma),$$

June 18, 2024. A genre manifesto for R Vinaigrettes is available at <http://davidofmeaning.blogspot.com/2016/12/r-vinaigrettes.html>.

	MLE	$\epsilon = 0$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.2$	Oracle
n= 500	1.343	0.984	0.930	0.887	0.837	0.837	0.707
n= 1000	1.342	0.974	0.915	0.869	0.813	0.818	0.707

TABLE 1. Root mean squared error for several alternative Huber/NPML estimators

	MLE	$\epsilon = 0$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.2$	Oracle
n= 500	1	0.591	0.593	0.607	0.653	0.808	0.579
n= 1000	1	0.585	0.588	0.602	0.648	0.804	0.578

TABLE 2. Root mean squared error for several alternative Huber/NPML estimators

with $\alpha = 0.1$ and $\sigma = 3$ which is heavier tailed than ϕ , but fairly mildly so. I took the mixing distribution, G to be the two-point distribution, $(1 - \beta)\delta_0 + \beta\delta_\theta$, with $\beta = 0.2$ and $\theta = 3$. Two sample sizes are considered $n = 500$ and $n = 1000$.

Four variants of the Huber density model,

$$\varphi(u) = \begin{cases} (1 - \epsilon)\phi(k) \exp(-k(u - k)) & u > k \\ (1 - \epsilon)\phi(u) & |u| \leq k \\ (1 - \epsilon)\phi(k) \exp(k(u + k)) & u < -k \end{cases}$$

are contrasted with Gaussian denoising. Here, k and ϵ are linked by the expression, $2\phi(k)/k - 2\Phi(-k) = \epsilon/(1 - \epsilon)$. The choice of ϵ can be interpreted as a proportion of contamination, while $\pm k$ are the corresponding values of u at which the influence function flips from being linear in the center to being constant in the tails. I consider 5 values of ϵ , $\{0, 0.025, 0.05, 0.10, 0.20\}$. Results based on 1000 replications are reported in Table 1. The column labeled “MLE” reports the naive MLE estimate performance, i.e. $\hat{\delta}(y) = y$, while the “Oracle” column reports performance of the Tweedie estimator based the true base distribution, F , and using the empirical distribution, $\hat{G} = G_n$, of the latent θ_i 's. Clearly, there is some advantage in the Huber procedures when the base distribution is heavier tailed than the Gaussian.

To evaluate the performance of the Huber procedures when the base distribution is Gaussian, we report in Table 2 results for the same experiment when data is generated with $F = \Phi$. As long as the Huber ϵ is small the “insurance premium” for using the Huber procedure is modest.

APPENDIX A. CODE FOR SIMULATION

```
# An NPML for G for the Huber likelihood model
hubereps <- function(k){
  v <- function(z, k)
    2*dnorm(k)/k - 2*pnorm(-k) - z/(1-z)
  uniroot(v, c(0.01, 0.3), k = k)$root
}
dhuber <- function(x, mu = 0, sigma = 1, k = 1.642, heps = hubereps(k)){
  (x > k)*(1-heps)*dnorm(k)*exp(-k*(x-k)) +
  (abs(x) <= k)*(1-heps)*dnorm(x) +
  (x < -k)*(1-heps)*dnorm(k)*exp(k*(x+k))
}
```

```

}
HLmix <- function(x, v = 300, sigma = 1, k = 1.345, heps = hubereps(k), ...){
  n = length(x)
  eps <- 1e-04
  if (length(v) == 1)
    v <- seq(min(x) - eps, max(x) + eps, length = v)
  m <- length(v)
  w <- rep(1,n)/n
  d <- rep(1, length(v))
  A <- dhuber(outer(x, v, "-"), sigma = sigma, k = k, heps = heps)
  f <- KWDual(A, d, w, ...)
  y <- f$f
  g <- f$g
  logLik <- n * sum(w * log(g))
  dy <- as.vector((A %*% (y * d * v))/g)
  z <- list(x = v, y = y, g = g, logLik = logLik, sigma = sigma,
    dx = x, dy = dy, k = k, heps = heps, status = f$status)
  class(z) <- c("HLmix", "density")
  return(z)
}

# Simulation to test HLmix -- Results saved in file sim.Rda

DGP <- function(n, alpha = 0.1, sigma = 3, beta = 0.2, theta = 3){
  mu <- sample(c(0, theta), n, replace = TRUE, prob = c(1-beta,beta))
  sd <- sample(c(1, sigma), n, replace = TRUE, prob = c(1-alpha,alpha))
  u <- rnorm(n, 0, sd = sd)
  y <- mu + u
  list(y = y, mu = mu, alpha = alpha, sigma = sigma, beta = beta)
}

Odelta <- function(D){ # Oracle: knows G_n, edf of the mu's
  # Oracle knows true base distribution
  y <- D$y
  mu <- D$mu
  T <- table(mu)
  G <- list(x = as.numeric(dimnames(T)$mu), y = as.numeric(T)/length(y))
  dgmix <- function(u, alpha, sigma)
    (1-alpha)*dnorm(u) + alpha * dnorm(u, sd = sigma)
  A <- dgmix(outer(y, G$x, "-"), alpha = D$alpha, sigma = D$sigma)
  as.vector((A %*% (G$y * G$x))/(A %*% G$y))
}

source("HLmix.R")
require(REBayes)
set.seed(1729)
sessionInfo()
R <- 1000

```

```

ns <- c(500,1000)
huberks <- c(0.862, 1.135, 1.398, 1.684) # approximately (0.2, 0.1, 0.05, 0.025)
J <- length(huberks)
huberepss <- rep(NA, J)
for(j in 1:J)
  huberepss[j] <- hubereps(huberks[j])
A <- array(NA, c(J+3, length(ns), R))
for(i in 1:length(ns)){
  n <- ns[i]
  for(r in 1:R){
    D <- DGP(n)
    mu <- D$mu
    y <- D$y
    a <- rep(NA, J+3)
    for(j in 1:J){
      k <- huberks[j]
      heps <- huberepss[j]
      a[j] <- mean((HLmix(y, k = k, heps = heps)$dy - mu)^2)
    }
    a[J+1] <- mean((GLmix(y)$dy - mu)^2)
    a[J+2] <- mean((Odelta(D) - mu)^2)
    a[J+3] <- mean((y - mu)^2)
    A[,i, r] <- a
  }
}

# Make latex table from sim.R results in sim.Rda
load("sim.Rda")
require(Hmisc)
A <- sqrt(apply(A,1:2,mean))
A <- t(A)
A <- A[, (J+3):1]
A <- round(A,3)
A <- A[,c(1, 3:7, 2)]
cap <- "Root mean squared error for several alternative Huber/NPMLE estimators"
Cn <- c("MLE", paste("$\\epsilon = $", c(0, round(huberepss[J:1],2))), "Oracle")
dimnames(A) <- list(paste("n=",ns), Cn)
latex(A, file = "sim.tex",rowlabel = "", caption = cap,
      caption.loc = "bottom", label = "tab.sim")

```

REFERENCES

Huber, P. J. (1964), ‘Robust estimation of a location parameter’, *Annals of Mathematical Statistics* **35**, 73–101.

- Kiefer, J. and Wolfowitz, J. (1956), ‘Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters’, *Annals of Mathematical Statistics* **27**, 887–906.
- Koenker, R. and Mizera, I. (2014), ‘Convex optimization, shape constraints, compound decisions, and empirical Bayes rules’, *Journal of the American Statistical Association* **109**, 674–685.
- Robbins, H. (1950), ‘A generalization of the method of maximum likelihood: Estimating a mixing distribution (abstract)’, *Annals of Mathematical Statistics* **21**, 314–315.
- Tukey, J. W. (1960), A survey of sampling from contaminated distributions, in I. Olkin, ed., ‘Contributions to Probability and Statistics’, Stanford University Press.