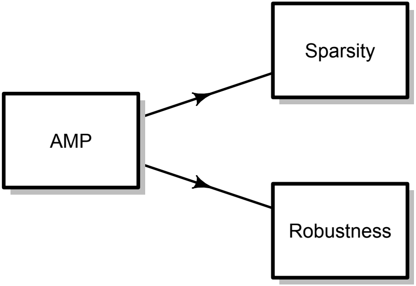
Comments on: "Big Data" Asymptotics via Approximate Message Passing

Cun-Hui Zhang, Rutgers University

May 2, 2014

Rutgers Statistics Symposium

Kill two birds with one stone:



- HD M-estimation: $\psi \rightarrow \widetilde{\Psi}$, $F \rightarrow \widetilde{F}$

- HD M-estimation: $\psi \to \widetilde{\Psi}$, $F \to \widetilde{F}$
- Robustness:

$$\widetilde{\Psi}(z) = \Psi(z; b_*) = b_* \psi\big(\mathrm{Prox}(z; b_*, \psi)\big)$$

- Extra Gaussian noise:

$$\widetilde{F} = F * N\big(0, \tau_\infty^2\big)$$

- HD M-estimation: $\psi \to \widetilde{\Psi}$, $F \to \widetilde{F}$
- Robustness:

$$\widetilde{\Psi}(z) = \Psi(z; b_*) = b_* \psi\big(\mathrm{Prox}(z; b_*, \psi)\big)$$

- Extra Gaussian noise:

$$\widetilde{F} = F * N\big(0, \tau_\infty^2\big)$$

- How much of is the extra noise intrinsic?
  - The smallest τ_∞^2 in HD M-estimation (Bean et al, 13)
  - HD Hájek-Le Cam-type convolution theorem for location-scale invariant $\widehat{\theta}$ and/or dense θ?

- Sparse θ (Donoho-Johnstone, 94): The extra noise is removable

- Sparse θ (Donoho-Johnstone, 94): The extra noise is removable
  - For "nice" F and $\mathbf{X}$, "regular" inference attains

$$\left\langle a, \left\{\widehat{\theta}^{(init)} + \mathbf{Z}^\top (Y - \mathbf{X}\widehat{\theta}^{(init)})\right\} - \theta \right\rangle \approx N\left(0, \left\langle a, \mathbf{\Sigma}^{-1} a \right\rangle \sigma_F^2\right)$$

when $\|\mathbf{Z}^\top \mathbf{X} - \mathbf{I}_{p\times p}\|_\infty$ is small, where $\mathbf{\Sigma} = \mathbb{E}\,\mathbf{X}^\top \mathbf{X}$

- Sparse θ (Donoho-Johnstone, 94): The extra noise is removable
    - For "nice" F and $\mathbf{X}$, "regular" inference attains

$$\left\langle a, \left\{\widehat{\theta}^{(init)} + \mathbf{Z}^\top (Y - \mathbf{X}\widehat{\theta}^{(init)})\right\} - \theta\right\rangle \approx N\left(0, \left\langle a, \mathbf{\Sigma}^{-1} a\right\rangle \sigma_F^2\right)$$

when $\|\mathbf{Z}^\top \mathbf{X} - \mathbf{I}_{p\times p}\|_\infty$ is small, where $\mathbf{\Sigma} = \mathbb{E}\,\mathbf{X}^\top \mathbf{X}$

- Dense θ: Empirical Bayes?

- Sparse θ (Donoho-Johnstone, 94): The extra noise is removable
  - For "nice" F and $\mathbf{X}$, "regular" inference attains

$$\left\langle a, \left\{\widehat{\theta}^{(init)} + \mathbf{Z}^\top (Y - \mathbf{X}\widehat{\theta}^{(init)})\right\} - \theta \right\rangle \approx N\left(0, \left\langle a, \mathbf{\Sigma}^{-1} a \right\rangle \sigma_F^2\right)$$

when $\|\mathbf{Z}^\top \mathbf{X} - \mathbf{I}_{p\times p}\|_\infty$ is small, where $\mathbf{\Sigma} = \mathbb{E}\,\mathbf{X}^\top \mathbf{X}$

- Dense θ: Empirical Bayes?
  - Let $G_n = p^{-1}\sum_{j=1}^p \delta_{\theta_j}$. In compressed sensing,

$$\begin{aligned} \frac{1}{p}\sum_{j=1}^p \mathbb{E}\left(t(\widehat{\theta}_j^M) - \theta_j\right)^2 &\approx \mathrm{BayesRisk}\left(t(\cdot), V(\widetilde{\Psi}, \widetilde{F}), G_n\right) \\ &= \int \mathbb{E}\left(t(\theta + \widetilde{\varepsilon}) - \theta\right)^2 G_n(d\theta) \end{aligned}$$

- Sparse θ (Donoho-Johnstone, 94): The extra noise is removable
  - For "nice" F and $\mathbf{X}$, "regular" inference attains

$$\left\langle a, \left\{\widehat{\theta}^{(init)} + \mathbf{Z}^\top(Y - \mathbf{X}\widehat{\theta}^{(init)})\right\} - \theta\right\rangle \approx N\Big(0, \left\langle a, \mathbf{\Sigma}^{-1} a\right\rangle \sigma_F^2\Big)$$

when $\|\mathbf{Z}^\top \mathbf{X} - \mathbf{I}_{p\times p}\|_\infty$ is small, where $\mathbf{\Sigma} = \mathbb{E}\,\mathbf{X}^\top \mathbf{X}$

- Dense θ: Empirical Bayes?
  - Let $G_n = p^{-1}\sum_{j=1}^p \delta_{\theta_j}$. In compressed sensing,

$$\frac{1}{p}\sum_{j=1}^p \mathbb{E}\Big(t\big(\widehat{\theta}_j^M\big) - \theta_j\Big)^2 \approx \text{BayesRisk}\Big(t(\cdot), V(\widetilde{\Psi}, \widetilde{F}), G_n\Big)$$

$$= \int \mathbb{E}\Big(t(\theta + \widetilde{\varepsilon}) - \theta\Big)^2 G_n(d\theta)$$

  - General $t(\cdot)$ (Robbins, 56); Linear $t(\cdot)$ (James-Stein, 61; Efron-Morris, 72)