Log-Concave Graph Estimation Using Convex-Plus-Concave Regression

Suppose we take $(X, Y) \in (0,1)^d$ with log-concave density $p(x,y) = e^{f(x,y)}$. Then the conditional mean satisfies

$$\mathbb{E}(Y \mid x) = \frac{\displaystyle\int y e^{f(x,y)}\, dy}{\displaystyle\int e^{f(x,y)}\, dy} = e^{g(x)-h(x)} \tag{1.1}$$

where $g(x)$ and $h(x)$ are concave. Thus $\log \mathbb{E}(Y \mid x)$ is convex plus concave.

If $\{(X_i, Y_i)\}_{i=1}^n$ are samples from the model then we can write

$$Y_i \mid X_i \;\sim\; \mathbb{E}(Y \mid X_i)\, W_i = e^{g(X_i)-h(X_i)}\, W_i \tag{1.2}$$

where $W_i \geq 0$ and $\mathbb{E}(W_i) = 1$. Thus, we have that

$$\log Y_i = g(X_i) - h(X_i) + \eta_i \tag{1.3}$$

where $\eta_i = \log W_i$.

This suggests neighborhood selection using a sparse concave-plus-convex regression of $\log Y_i$ on the remaining variables.

I believe that the first part (Lemma 6.1) of Min's additive faithfulness analysis goes through to the convex-plus-concave setting. The second part (Prop. 6.1) is more involved, and will need some modification. For example, we don't want to assume independence in this setting.

Minhua and I realized there is better, more natural formulation. Assuming only a log-concave density we can write

$$\mathbb{E}(e^Y \mid x) = \frac{\displaystyle\int e^y e^{f(x,y)}\, dy}{\displaystyle\int e^{f(x,y)}\, dy} = e^{g(x)-h(x)} \tag{1.4}$$

where $g(x)$ and $h(x)$ are concave. Thus $\log \mathbb{E}(e^Y \mid x)$ is convex plus concave.

Now we write

$$e^Y \mid X = x \;\sim\; \mathbb{E}(e^Y \mid x)\, W(x) = e^{g(x)-h(x)}\, W(x) \tag{1.5}$$

where $W(x) > 0$ and $\mathbb{E}[W(x)] = 1$. Then

$$Y = g(x) - h(x) + \eta \tag{1.6}$$

where $\eta = \log W[x]$.

But η may not be mean zero. Consider the Gaussian case, where W is log-normal, $\log W \sim N(\mu, \sigma^2)$. Then $\mathbb{E}W = e^{\mu + \frac{1}{2}\sigma^2}$ and $\mathbb{E}W = 1$ implies $\mu = -\frac{1}{2}\sigma^2$. We can then write

$$\begin{aligned} e^Y &= e^{g(x) - h(x) - \frac{1}{2}\sigma^2}\, W(x) e^{\frac{1}{2}\sigma^2} && (1.7) \\ &= e^{\widetilde{g}(x) - h(x)}\, \widetilde{W}(x) && (1.8) \end{aligned}$$

where $\widetilde{g}(x)$ is concave and $\mathbb{E}\big[\log \widetilde{W}(x)\big] = 0$.

In this way, if $\mathbb{E}\big[\log W(x)\big]$ is constant, we can subtract out this constant from $g(x)$ or $h(x)$, and thus assume that $\mathbb{E}\big[\log W(x)\big] = 0$. (It would be good to have an understanding of when this assumption on $\mathbb{E}\big[\log W(x)\big]$ is reasonable, or to come up with an alternative.)

We are then led to the sparse covex+plus+concave regression problem

$$Y = g(x) - h(x) + \eta \tag{1.9}$$

to carry out neighborhood selection. This gives a direct, natural extension of the Gaussian case with the Meinshausen-Bühlmann method.