Inverse Problems in Semiparametric Statistical Models

Guang Cheng

Department of Statistics, Purdue University

Applied Inverse Problems Conference
March 24th, 2011

Outline

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;
  - η: an infinite dimensional *nuisance* parameter, e.g. some function.

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;
  - η: an infinite dimensional *nuisance* parameter, e.g. some function.

- Given the above observations, we intend to make inferences on the Euclidean parameter θ in the presence of an infinite dimensional nuisance parameter η:

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;
  - η: an infinite dimensional *nuisance* parameter, e.g. some function.
- Given the above observations, we intend to make inferences on the Euclidean parameter θ in the presence of an infinite dimensional nuisance parameter η:
  - give a consistent estimate $\widehat{\theta}$;

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$
  - θ: Euclidean parameter of interest;
  - η: an infinite dimensional *nuisance* parameter, e.g. some function.

- Given the above observations, we intend to make inferences on the Euclidean parameter θ in the presence of an infinite dimensional nuisance parameter η:
  - give a consistent estimate $\widehat{\theta}$;
  - give a confidence interval/credible set (hypothesis testing) for θ

Semiparametric Models

- We observe i.i.d. data $\{X_i\}_{i=1}^n \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$
  - θ: Euclidean parameter of interest;
  - η: an infinite dimensional *nuisance* parameter, e.g. some function.
- Given the above observations, we intend to make inferences on the Euclidean parameter θ in the presence of an infinite dimensional nuisance parameter η:
  - give a consistent estimate $\widehat{\theta}$;
  - give a confidence interval/credible set (hypothesis testing) for θ
- Even we are only interested in θ, the estimation of η is usually unavoidable.

Model I. Cox Regression Model with Current Status Data

The hazard function of the survival time T of a subject with covariate Z is modelled as:

$$\lambda(t|z) \equiv \lim_{\Delta \to 0} \frac{1}{\Delta} Pr(t \leq T < t + \Delta | T \geq t, Z = z) = \lambda(t) \exp(\theta' z),$$

where λ is an unspecified baseline hazard function.

Consider the current status data in which the event time T is unobservable but we know whether the event has occurred at the examination time C or not. Thus, we observe $X = (C, \delta, Z)$, where $\delta = I\{T \leq C\}$.

Based on the above proportional hazard assumption, we can write down the log-likelihood as follows

$$\begin{aligned} & \log lik(\theta,\eta)(X) \\ = \; & \delta \log \left[1 - \exp(-\exp(\theta' Z)\eta(C))\right] - (1-\delta)\exp(\theta' Z)\eta(C), \end{aligned}$$

where the nuisance (monotone) function $\eta(y) \equiv \int_0^y \lambda(t)dt$, also called as cumulative hazard function.

Example II: Conditionally Normal Model

We assume that $Y|(W=w, Z=z) \sim N(\theta' w, \eta(z))$. The log-likelihood can be easily written as

$$\log lik(\theta, \eta)(X) = -\frac{1}{2}\log \eta(Z) - \frac{(Y-\theta' W)^2}{2\eta(Z)},$$

where $\eta(z)$ is positive.

Model III. Partly Linear Model

We assume that

$$Y = \theta' W + \eta(Z) + \epsilon,$$

where ϵ is independent of (W, Z) and η is an unknown smooth function belonging to second order Sobolev space. We assume that ϵ is normally distributed (can be relaxed to some tail conditions).

Model IV. Semiparametric Copula Model

We observe random vector $X = (X_1, \ldots, X_d)$ with multivariate distribution function $F(x_1, \ldots, x_d)$, and want to estimate the dependence structure in X. To avoid the curse of dimensionality, we will apply the following Copula approach.

According to Sklar (1959), there exists a unique Copula function $C_0(\cdot)$ such that

$$F(x_1, \ldots, x_d) = C_0(F_1(x_1), \ldots, F_d(x_d)),$$

where $F_j(\cdot)$ is the marginal distribution for X_j.

To model the dependence within X, we use the parametric Copula $C_\theta(\cdot)$, i.e., $C_{\theta_0} = C_0$. Thus, the log-likelihood is written as

$$\log lik(\theta, F_1, \ldots, F_d)(X) = \log c_\theta(F_1(X_1), \ldots, F_d(X_d)) + \sum_{j=1}^{d} \log f_j(X_j),$$

where f_j is the marginal density function and

$$c_\theta(t_1, \ldots, t_d) = \frac{\partial^d}{\partial t_1 \cdots \partial t_d} C_\theta(t_1, \ldots, t_d).$$

Semiparametric Efficiency Bound

- We hope to obtain the semiparametric efficient estimate $\widehat{\theta}$, which achieves the minimal asymptotic variance bound in the sense that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, V^*),$$

where V^* is the minimal one over all the regular semiparametric estimators.

Semiparametric Efficiency Bound

- We hope to obtain the semiparametric efficient estimate $\widehat{\theta}$, which achieves the minimal asymptotic variance bound in the sense that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N(0, V^*),$$

  where V^* is the minimal one over all the regular semiparametric estimators.
- IDEA: The minimal V^* actually corresponds to the largest asymptotic variance over all the parametric submodels $\{t \mapsto \log lik(t, \eta_t) : t \in \Theta\}$ of the semiparametric model in consideration. The parametric submodel achieving V^* is called as the *least favorable submodel (LFS)*, see Bickel et al (1996).

Intuition I

- Now, let us turn our attention to LFS defined as

$$t \mapsto \log lik(t, \eta_t^*),$$

where η_t^* is called as the least favorable curve.

Intuition I

- Now, let us turn our attention to LFS defined as

$$t \mapsto \log lik(t, \eta_t^*),$$

where η_t^* is called as the least favorable curve.

- The LFS needs to pass the true value (θ_0, η_0), i.e., $\eta_{\theta_0}^* = \eta_0$, and has the corresponding information matrix as

$$\widetilde{I}_0 = E\widetilde{\ell}_0\widetilde{\ell}_0', \text{ where } \widetilde{\ell}_0 \equiv \frac{\partial}{\partial t}|_{t=\theta_0} \log(t, \eta_t^*).$$

(This is just the usual way to calculate the information in parametric models)

Intuition I

- Now, let us turn our attention to LFS defined as

$$t \mapsto \log lik(t, \eta_t^*),$$

where η_t^* is called as the least favorable curve.

- The LFS needs to pass the true value (θ_0, η_0), i.e., $\eta_{\theta_0}^* = \eta_0$, and has the corresponding information matrix as

$$\widetilde{I}_0 = E\widetilde{\ell}_0\widetilde{\ell}_0', \quad \text{where} \quad \widetilde{\ell}_0 \equiv \frac{\partial}{\partial t}|_{t=\theta_0} \log(t, \eta_t^*).$$

(This is just the usual way to calculate the information in parametric models)

- Obviously, $V^* = \widetilde{I}_0^{-1}$.

Intuition II

What is the mysterious η_t^*?

Intuition II

What is the mysterious η_t^*?

- In fact, Severini and Wong (1992) discovered that

$$\eta_t^* = \arg\sup_{\eta \in \mathcal{H}} E \log lik(t, \eta) \quad \text{for any fixed } t \in \Theta$$

after some simple derivations! This is not surprising since η_t^* behaves like the true value for η at each fixed θ.

Intuition III

- According to our discussions on LFS in the above, we expect to obtain an efficient estimate of θ if we can estimate the abstract LFS, i.e., η_t^*, accurately.

Intuition III

- According to our discussions on LFS in the above, we expect to obtain an efficient estimate of θ if we can estimate the abstract LFS, i.e., η_t^*, accurately.
- Let $S_n(\theta) \equiv \sum_{i=1}^n \log lik(\theta, \widehat{\eta}_\theta)(X_i)$.

Intuition III

- According to our discussions on LFS in the above, we expect to obtain an efficient estimate of θ if we can estimate the abstract LFS, i.e., η_t^*, accurately.
- Let $S_n(\theta) \equiv \sum_{i=1}^n \log lik(\theta, \widehat{\eta}_\theta)(X_i)$.
- In fact, we can easily show that

$$\widehat{\theta} \equiv \arg\max_{\theta \in \Theta} S_n(\theta),$$

is semiparametric efficient if $\widehat{\eta}_t$ is a consistent estimate for η_t^*.

Intuition III

- According to our discussions on LFS in the above, we expect to obtain an efficient estimate of θ if we can estimate the abstract LFS, i.e., η_t^*, accurately.
- Let $S_n(\theta) \equiv \sum_{i=1}^n \log lik(\theta, \widehat{\eta}_\theta)(X_i)$.
- In fact, we can easily show that

$$\widehat{\theta} \equiv \arg\max_{\theta \in \Theta} S_n(\theta),$$

is semiparametric efficient if $\widehat{\eta}_t$ is a consistent estimate for η_t^*.
- Therefore, we can claim that the efficient estimation of θ boils down to the estimation of the least favorable curve η_t^*.

Summary

Efficient estimation of θ in presence of an infinite dimensional η

$$\Downarrow$$

Least favorable submodel: $t \mapsto \log lik(t, \eta_t^*)$

$$\Downarrow$$

Consistent estimation of η_t^*

Summary

Efficient estimation of θ in presence of an infinite dimensional η

$$\Downarrow$$

Least favorable submodel: $t \mapsto \log lik(t, \eta_t^*)$

$$\Downarrow$$

Consistent estimation of η_t^*

- The estimation accuracy of η_t^*, i.e., convergence rate, determines the second order efficiency of $\widehat{\theta}$ (Cheng and Kosorok, 2008);

Summary

Efficient estimation of θ in presence of an infinite dimensional η

$$\Downarrow$$

Least favorable submodel: $t \mapsto \log lik(t, \eta_t^*)$

$$\Downarrow$$

Consistent estimation of η_t^*

- The estimation accuracy of η_t^*, i.e., convergence rate, determines the second order efficiency of $\widehat{\theta}$ (Cheng and Kosorok, 2008);
- How we estimate η_t^* depends on the parameter space $\mathcal{H}$, and different regularizations on η_t^* gives different forms of $\widehat{\theta}$, see four examples to be presented.

Rigorous Statement

- In semiparametric literature,
  $\widetilde{\ell}_0$ is called as Efficient Score Function;
  $\widetilde{I}_0$ is called as Efficient Information Matrix.

Rigorous Statement

- In semiparametric literature,
  $\widetilde{\ell}_0$ is called as Efficient Score Function;
  $\widetilde{I}_0$ is called as Efficient Information Matrix.

- In fact, the efficient score function can be understood as the residual of the projection of the score function for θ onto the tangent space, which is defined as the closed linear span of the tangent set generated by the score function for η.

Rigorous Statement

- In semiparametric literature,
  $\widetilde{\ell}_0$ is called as Efficient Score Function;
  $\widetilde{I}_0$ is called as Efficient Information Matrix.
- In fact, the efficient score function can be understood as the residual of the projection of the score function for θ onto the tangent space, which is defined as the closed linear span of the tangent set generated by the score function for η.
- The LFS exists if the tangent set is closed. This is true for all of our examples in this talk.

Semiparametric Efficient Estimation

- As discussed above, we need to estimate η_t^* consistently in order to obtain the efficient $\widehat{\theta}$. Recall that

$$\eta_t^* = \arg\max_{\eta \in \mathcal{H}} E \log lik(t, \eta).$$

Semiparametric Efficient Estimation

- As discussed above, we need to estimate η_t^* consistently in order to obtain the efficient $\widehat{\theta}$. Recall that

$$\eta_t^* = \arg\max_{\eta\in\mathcal{H}} E \log lik(t, \eta).$$

- Therefore, a natural estimate for η_θ^* is

$$\widehat{\eta}_\theta = \arg\max_{\eta\in\mathcal{H}} \sum_{i=1}^{n} \log lik(\theta, \eta)(X_i) \tag{1}$$

for any fixed $\theta \in \Theta$.

- In the above, $\widehat{\eta}_\theta$ is the NPMLE, $S_n(\theta)$ is just the profile likelihood $\log pl_n(\theta)$, and $\widehat{\theta}$ becomes the semiparametric MLE.
- The above maximum likelihood estimation works for our example I, i.e., Cox model, due to monotone constraints (see the work by Jon Wellner and his coauthors). However, the NPMLE is not always well defined. Thus, some form of regularization is needed especially when η needs to be estimated smoothly.

- Kernel estimation: This is particularly useful when η_θ^* has an explicit form. In our example II, i.e., conditionally normal model, we have

$$\widehat{\eta}_{\theta,b_n}(z)=\frac{\sum_{i=1}^n(Y-\theta' W)^2K((z-Z_i)/b_n)}{\sum_{i=1}^n K((z-Z_i)/b_n)}>0, \qquad (2)$$

where $K(\cdot)$ is some kernel function and b_n is the related bandwidth.

- Penalized estimation: In our example III, i.e., partly linear model, we have

$$\widehat{\eta}_{\theta,\lambda_n} = \arg\max_{\eta\in\mathcal{H}} \left\{ \sum_{i=1}^{n} \log lik(\theta,\eta)(X_i) - \lambda_n \int_{\mathcal{Z}} [\eta^{(2)}(z)]^2 dz \right\}, \quad (3)$$

where λ_n is some smoothing parameter.

- Penalized estimation: In our example III, i.e., partly linear model, we have

$$\widehat{\eta}_{\theta,\lambda_n} = \arg\max_{\eta\in\mathcal{H}} \left\{ \sum_{i=1}^{n} \log lik(\theta,\eta)(X_i) - \lambda_n \int_{\mathcal{Z}} [\eta^{(2)}(z)]^2 dz \right\}, \quad (3)$$

where λ_n is some smoothing parameter.

- In the penalized estimation, we need to construct the penalized LFS, see Cheng and Kosorok (2009).

- Penalized estimation: In our example III, i.e., partly linear model, we have

$$\widehat{\eta}_{\theta,\lambda_n} = \arg\max_{\eta\in\mathcal{H}} \left\{ \sum_{i=1}^{n} \log lik(\theta,\eta)(X_i) - \lambda_n \int_{\mathcal{Z}} [\eta^{(2)}(z)]^2 dz \right\}, \quad (3)$$

where λ_n is some smoothing parameter.

- In the penalized estimation, we need to construct the penalized LFS, see Cheng and Kosorok (2009).

- In this example, $\widehat{\theta}$ is just the partial smoothing spline estimate.

- Sieve estimation: Here, we perform similar maximum likelihood estimation but replace the infinite dimensional parameter space $\mathcal{H}$ by its sieve approximation $\mathcal{H}_n$, e.g., B-spline space.

- Sieve estimation: Here, we perform similar maximum likelihood estimation but replace the infinite dimensional parameter space $\mathcal{H}$ by its sieve approximation $\mathcal{H}_n$, e.g., B-spline space.
- In our example IV, i.e., semiparametric copula model, we have

$$\widehat{\eta}_{\theta,s_n} = \arg\max_{\eta\in\mathcal{H}_n} \sum_{i=1}^{n} \log lik(\theta,\eta)(X_i), \qquad (4)$$

where $\mathcal{H}_n = \{\eta(\cdot) = \sum_{s=1}^{s_n} \gamma_s B_s(\cdot)\}$ is the B-spline space.

- Sieve estimation: Here, we perform similar maximum likelihood estimation but replace the infinite dimensional parameter space $\mathcal{H}$ by its sieve approximation $\mathcal{H}_n$, e.g., B-spline space.

- In our example IV, i.e., semiparametric copula model, we have

$$\widehat{\eta}_{\theta,s_n} = \arg\max_{\eta\in\mathcal{H}_n} \sum_{i=1}^{n} \log lik(\theta,\eta)(X_i), \tag{4}$$

  where $\mathcal{H}_n = \{\eta(\cdot) = \sum_{s=1}^{s_n} \gamma_s B_s(\cdot)\}$ is the B-spline space.

- An advantage of B-spline estimation is that we can transform the semiparametric estimation into the parametric estimation with increasing dimension as sample size.

Remark

- Under regularity conditions, all the above four estimation approaches yield the semiparametric efficient $\widehat{\theta}$, see Cheng (2011) for more details.

Remark

- Under regularity conditions, all the above four estimation approaches yield the semiparametric efficient $\widehat{\theta}$, see Cheng (2011) for more details.
- Cheng and Kosorok (2008) show that the second order semiparametric efficiency of $\widehat{\theta}$ is determined by the smoothing parameters, i.e., b_n, λ_n and s_n, and the size of $\mathcal{H}$ (in terms of entropy number).

Remark

- Under regularity conditions, all the above four estimation approaches yield the semiparametric efficient $\widehat{\theta}$, see Cheng (2011) for more details.
- Cheng and Kosorok (2008) show that the second order semiparametric efficiency of $\widehat{\theta}$ is determined by the smoothing parameters, i.e., b_n, λ_n and s_n, and the size of $\mathcal{H}$ (in terms of entropy number).
- In some situations, it might be more proper to use other criterion function than the likelihood function, e.g., use the least square criterion function in the partly linear model (replace $\epsilon \sim N(0, \sigma^2)$ by the sub-exponential tail condition).

In the end, I describe three (almost) automatic semiparametric inferential tools for obtaining the semiparametric efficient estimate and constructing the confidence interval/credible set in the literature.

In the end, I describe three (almost) automatic semiparametric inferential tools for obtaining the semiparametric efficient estimate and constructing the confidence interval/credible set in the literature.

- Bootstrap Inferences [Cheng and Huang (2010)]

In the end, I describe three (almost) automatic semiparametric inferential tools for obtaining the semiparametric efficient estimate and constructing the confidence interval/credible set in the literature.

- Bootstrap Inferences [Cheng and Huang (2010)]
- Profile Sampler [Lee, Kosorok and Fine (2005)]

In the end, I describe three (almost) automatic semiparametric inferential tools for obtaining the semiparametric efficient estimate and constructing the confidence interval/credible set in the literature.

- Bootstrap Inferences [Cheng and Huang (2010)]
- Profile Sampler [Lee, Kosorok and Fine (2005)]
- Sieve Estimation [Chen (2007)]

Bootstrap Inferences

The bootstrap resampling approach has the following well known advantages:

Bootstrap Inferences

The bootstrap resampling approach has the following well known advantages:

- Automatic procedure;

Bootstrap Inferences

The bootstrap resampling approach has the following well known advantages:

- Automatic procedure;
- Small sample advantages;

Bootstrap Inferences

The bootstrap resampling approach has the following well known advantages:

- Automatic procedure;
- Small sample advantages;
- Replace the tedious theoretical derivations in semiparametric inferences with routine simulations of bootstrap samples, e.g., the bootstrap confidence interval.

- The bootstrap estimator is defined as

$$(\widehat{\theta}^*, \hat{\eta}^*) = \arg \sup_{\theta\in\Theta, \eta\in\mathcal{H}} \sum_{i=1}^{n} \log lik(\theta, \eta)(X_i^*), \qquad (5)$$

where $(X_1^*, \ldots, X_n^*)$ is the bootstrap sample.

- The bootstrap estimator is defined as

$$(\widehat{\theta}^*, \hat{\eta}^*) = \arg \sup_{\theta \in \Theta, \eta \in \mathcal{H}} \sum_{i=1}^{n} \log lik(\theta, \eta)(X_i^*), \tag{5}$$

where $(X_1^*, \ldots, X_n^*)$ is the bootstrap sample.

- Recently, Cheng and Huang (2010) showed that (i) $\widehat{\theta}^*$ has the same asymptotic distribution as the semiparametric efficient $\widehat{\theta}$; (ii) the bootstrap confidence interval is theoretically valid, for a general class of exchangeably weighted bootstrap resampling schemes, e.g., Efron's bootstrap and Bayesian bootstrap.

Profile Sampler

Profile Sampler

- We assign some prior $\rho(\theta)$ on the profile likelihood $\log pl_n(\theta)$. MCMC is used for sampling from the posterior of the profile likelihood. This resulting MCMC chain is called as the profile sampler.

Profile Sampler

- We assign some prior $\rho(\theta)$ on the profile likelihood $\log pl_n(\theta)$. MCMC is used for sampling from the posterior of the profile likelihood. This resulting MCMC chain is called as the profile sampler.
- The inferences of θ are based on the profile sampler. Lee, Kosorok and Fine (2005) showed that chain mean (the inverse of chain variance) approximates the semiparametric efficient $\widehat{\theta}$ ($\tilde{I}_0$), and the credible set for θ has the desired asymptotic coverage probability.

Sieve Estimation

- Translate the semiparametric estimation into the parametric estimation with increasing dimension:

$$(\widehat{\theta}, \widehat{\gamma}) = \arg \max_{\theta \in \Theta, \gamma \in \Gamma} \sum_{i=1}^{n} \log lik(\theta, \gamma' B)(X_i).$$

Sieve Estimation

- Translate the semiparametric estimation into the parametric estimation with increasing dimension:

$$(\widehat{\theta}, \widehat{\gamma}) = \arg \max_{\theta \in \Theta, \gamma \in \Gamma} \sum_{i=1}^{n} \log lik(\theta, \gamma' B)(X_i).$$

- An advantage of B-spline estimation is that we are able to give an explicit B-spline estimate for the asymptotic variance V^* as a byproduct of the establishment of semiparametric efficiency of $\widehat{\theta}$. Indeed, it is simply the observed information matrix if we treat the semiparametric model as a parametric one after the B-spline approximation, i.e., $\mathcal{H} = \mathcal{H}_n$.

Future (Theoretical) Directions

- Limiting distribution of $\widehat{\eta}$ (expected to be nonstandard, e.g., Chernoff's distribution);

Future (Theoretical) Directions

- Limiting distribution of $\widehat{\eta}$ (expected to be nonstandard, e.g., Chernoff's distribution);
- Bootstrap Inferences for η (parametric bootstrap or *m* out of *n* bootstrap for nonstandard asymptotics);

Future (Theoretical) Directions

- Limiting distribution of $\widehat{\eta}$ (expected to be nonstandard, e.g., Chernoff's distribution);
- Bootstrap Inferences for η (parametric bootstrap or *m* out of *n* bootstrap for nonstandard asymptotics);
- Joint inferences for (θ, η) (extremely difficult.....);

Thanks for your attention....

Assistant Professor Guang Cheng
Department of Statistics, Purdue University
chengg@purdue.edu