Constrained Dirichlet Process and Functional Condition Model

Jaeyong Lee

Department of Statistics
Seoul National University

June 26th 2024

Jointly with

Jaequi Lee
SK Innovation, South Korea.

Content

Content

Dirichlet Process on $\mathcal{X}$ (Ferguson 1973)

- α: a finite nonnull measure on $\mathcal{X}$, a Euclidean space.
- (definition of DP)

$$P \sim \mathcal{D}_\alpha$$

$\iff$ for every finite partition $B_1, \ldots, B_k$ of $\mathcal{X}$

$$(P(B_1), \ldots, P(B_k)) \sim \mathit{Dirichlet}(\alpha(B_1), \ldots, \alpha(B_k)).$$

- The Dirichlet process $\mathcal{D}_\alpha$ is a probability measure on $M(\mathcal{X})$ and is used as a prior for F, where $M(\mathcal{X})$ = space of all probability measures on $\mathcal{X}$.

Properties of Dirichlet process

Conjugacy

Suppose

$$P \sim \mathcal{D}_\alpha$$
$$X_1, \ldots, X_n | P \sim P.$$

Then,

$$P | X_1, \ldots, X_n \sim \mathcal{D}(\alpha + \sum_{i=1}^{n} \delta_{X_i}).$$

Sethuraman's Representation

$P \sim \mathcal{D}_\alpha$ can be represented as a random discrete probability.

Marginalization Property

Suppose

$$P \sim \mathcal{D}_\alpha$$
$$X_1, X_2, \ldots | P \sim P.$$

Then, marginally $(X_1, X_2, \ldots)$ forms a Polya urn sequence:

$$X_1 \sim \bar{\alpha} := \alpha / \alpha(\mathcal{X})$$
$$X_{n+1} | X_1, \ldots, X_n \sim \frac{\alpha + \sum_{i=1}^{n} \delta_{X_i}}{\alpha(\mathcal{X}) + n}, \ n \geq 1.$$

Content

Motivation

- (Model)

$$X_1, \ldots, X_n | F, g(F) = \theta \sim F,$$

where g is a functional on $\mathcal{M}(\mathcal{X})$.

- (Prior) We often write as follows.

$$\theta \sim \pi$$
$$F | g(F) = \theta \sim \mathcal{D}_\alpha(dF) I(g(F) = \theta).$$

What is $\mathcal{D}_\alpha(dF) I(g(F) = \theta)$?

We mean it to be the $\mathcal{D}_\alpha$ restricted to the fiber

$$g^{-1}(\theta) = \{F \in \mathcal{M} : g(F) = \theta\}.$$

It is not well-defined because in most cases

$$\mathcal{D}_\alpha(F : g(F) = \theta) = 0, \ \forall \theta.$$

It is a conditioinal distribution, but the definition of the conditional distribution is not intuitive as the $\mathcal{D}_\alpha$ restricted to the fiber.

Disintegration

- $(\mathcal{X}, \mathcal{A})$, $(\mathcal{T}, \mathcal{B})$: measurable spaces
- λ: a σ-finite measure on $\mathcal{X}$
- μ: a σ-finite measure on $\mathcal{T}$
- $T : \mathcal{X} \to \mathcal{T}$: measurable function

- $\{\lambda_t, t \in \mathcal{T}\}$: disintegrating measures
- μ: mixing measure

Definition

λ is said to have (T, μ)-disintegration $\{\lambda_t, t \in \mathcal{T}\}$ if

(i) (concentration property) λ_t: a σ-finite measure on $\mathcal{X}$.

$$\lambda_t\{T \neq t\} = 0, \ \mu - a.a\ t.$$

(ii) (measurability)$t \mapsto \lambda_t f$: measurable, $\forall f \geq 0$ on $\mathcal{X}$.

(iii) (iterative integration) $\forall f \geq 0$ on $\mathcal{X}$,

$$\int_{\mathcal{X}} f(x)\lambda(dx) = \int_{\mathcal{T}} \int_{[T=t]} f(x)\lambda_t(dx)\mu(dt) \text{ or } \left(\lambda f = \mu^t(\lambda_t f) \right)$$

Existence Theorem of Disintegration

- $\mathcal{X}$: metric space
- λ: a σ-finite Radon measure on $\mathcal{X}$. The measure of a set can be approximated by that of a compact set from inside.
- $T\lambda := \lambda \circ T^{-1} << \mu$
- $\mathcal{B}$ is countably generated and contains all the singletons $\{t\}$.

Then, the following hold.

(a) λ has (T, μ)-disintegration $\{\lambda_t\}$.

(b) $\{\lambda_t\}$ is unique almost surely,
i.e., for another (T, μ)-disintegration $\{\lambda_t^*\}$,

$$\mu\{t \in \mathcal{T} : \lambda_t \neq \lambda_t^*\} = 0.$$

- If $\mathcal{X}$ is a Polish space, $\mathcal{T}$ is an Euclidean space, μ is the Lebesgue measure on $\mathcal{T}$ and $\lambda \circ T^{-1} << \mu$. Then, λ has the (T, μ)-disintegration.

Remarks

Suppose λ is a σ-finite measure and (λ_t) is the (T, μ)-disintegration of λ.

- If $X \sim \lambda$ and λ is a probability, $X|T(X) = t \sim \bar{\lambda}_t$.
- The conditional distributions defined by the disintegration is more intuitive than the usual definition.
- λ does not need to be a probability for the disintegration.
- A finite Borel measure on a Polish space is a Radon measure.
- The iterative integration condition can be written simply as

$$\lambda(dx) = \lambda_t(dx)\mu(dt).$$

Content

Disintegration of Dirichlet Process

- Suppose $F \sim \mathcal{D}_\alpha$, $g : \mathcal{M}(\mathcal{X}) \to \mathbb{R}$, $g(F) \sim h(\xi)\mu(d\xi)$, where μ is a σ-finite measure on $\mathbb{R}$. Then, $\mathcal{D}_\alpha$ has the (g, μ)-disintegration $\{\mathcal{D}_{\alpha,\xi}, \xi \in \mathbb{R}\}$.
- (iterative integation)

$$\begin{aligned}\mathcal{D}_\alpha(dF) &= \mathcal{D}_{\alpha,\xi}(dF)\mu(d\xi)\\ &= \frac{\mathcal{D}_{\alpha,\xi}(dF)}{\mathcal{D}_{\alpha,\xi}(\mathcal{M})}\mathcal{D}_{\alpha,\xi}(\mathcal{M})\mu(d\xi)\\ &= \mathcal{D}_\alpha(dF|g(F) = \xi)h(\xi)\mu(d\xi).\end{aligned}$$

- (simple identities)

$$\begin{aligned}\mathcal{D}_{\alpha,\xi}(dF) &= \mathcal{D}_\alpha(dF)I(g(F) = \xi)\\ h(\xi) &= \mathcal{D}_{\alpha,\xi}(\mathcal{M})\\ \mathcal{D}_\alpha(dF|g(F) = \xi) &= \frac{\mathcal{D}_{\alpha,\xi}(dF)}{\mathcal{D}_{\alpha,\xi}(\mathcal{M})}.\end{aligned}$$

Posterior with Constrained Dirichlet Process

$$F \sim \mathcal{D}_\alpha(\cdot | g(F) = \xi)$$

$$X_1, \ldots, X_n | F \overset{iid}{\sim} F$$

Then,

(a) (posterior) $F | X_1, \ldots, X_n \sim \mathcal{D}_{\alpha + nF_n}(dF | g(F) = \xi)$;

(b) (marginal distribution) $X_1, \ldots, X_n \sim \dfrac{h(\xi : g, \alpha + nF_n)}{h(\xi : g, \alpha)} Polya(d\mathbf{x}_n)$,
where $h(\cdot : g, \alpha)$ is the density of $g(F)$ when $F \sim \mathcal{D}_\alpha$,
$\mathbf{x}_n = (x_1, \ldots, x_n)$ and $Polya(d\mathbf{x}_n)$ is the distribution of the first n elements of the Polya sequence.

A1 If $F \sim \mathcal{D}_\alpha$, $g(F)$ has a density $h(\xi : g, \alpha)$ with respect to the Lebesgue measure μ, and $h(\xi : g, \alpha)$ is positive for each possible value in the range of g.

Example

$$F \sim \mathcal{D}_\alpha(\cdot|F(A) = \frac{1}{2}),\ A \subset \mathcal{X}$$

$$X_1, \ldots, X_n | F \overset{iid}{\sim} F$$

Then,

(a) (posterior) $F|X_1, \ldots, X_n \sim \mathcal{D}_{\alpha + nF_n}(dF|F(A) = \frac{1}{2})$;

(b) (marginal distribution)

$$\begin{aligned} X_1, \ldots, X_n &\sim \frac{Beta(\frac{1}{2} : \alpha(A) + n_A(\mathbf{x}_n), \alpha(A^c) + n_{A^c}(\mathbf{x}_n))}{Beta(\frac{1}{2} : \alpha(A), \alpha(A^c))} Polya(d\mathbf{x}_n) \\ &= \frac{[\alpha(\mathcal{X})]_n}{[\alpha(A)]_{n(A)}[\alpha(A^c)]_{n(A^c)}} (\frac{1}{2})^n Polya_n(d\mathbf{x}_n), \end{aligned}$$

where $Beta(x : a, b)$ is the density of $Beta(a, b),\ a, b > 0$ at x,

$$n_A = n(A) = n_A(\mathbf{x}_n) = \sum_{i=1}^{n} I(x_i \in A),$$

and $[a]_k := (a)(a+1) \times \ldots \times (a + k - 1)$.

Example: continued

- When $n = 1$, from the tail free property of the Dirichlet process, we know that
$$X_1 \sim \frac{1}{2}\overline{\alpha|_A}(dx) + \frac{1}{2}\overline{\alpha|_{A^c}}(dx).$$
- We get the same result from the above theorem.

$$\begin{aligned}
&\frac{[\alpha(\mathcal{X})]_1}{[\alpha(A)]_{n(A)}[\alpha(A^c)]_{n(A^c)}}\frac{1}{2}\bar{\alpha}(dx) \\
&= \frac{\alpha(\mathcal{X})}{\alpha(A)^{n(A)}\alpha(A^c)^{n(A^c)}}\frac{1}{2}\bar{\alpha}(dx) \\
&= \frac{1}{2}\alpha(\mathcal{X})\Big(\frac{1}{\alpha(A)}I(x \in A) + \frac{1}{\alpha(A^c)}I(x \in A^c)\Big)\frac{\alpha(dx)}{\alpha(\mathcal{X})} \\
&= \frac{1}{2}\Big(\frac{\alpha(dx)}{\alpha(A)}I(x \in A) + \frac{\alpha(dx)}{\alpha(A^c)}I(x \in A^c)\Big) \\
&= \frac{1}{2}\overline{\alpha|_A}(dx) + \frac{1}{2}\overline{\alpha|_{A^c}}(dx).
\end{aligned}$$

Content

Functional Condition Model

Model

$$\begin{aligned}\epsilon_i = t(y_i, x_i, \xi) &\overset{i.i.d.}{\sim} F, \;\; i = 1, 2, \dots, n \\ g(F, \xi) &= 0, \\ \nu &= 0.\end{aligned} \tag{1}$$

- y_i's are observations and x_i's are fixed covariates.
- F is a distribution satisfying a functional condition $g(F, \xi) = 0$.
- $\xi = (\theta, \nu)$.

Prior

$$\begin{aligned}\pi(dF, d\xi) &= \mathcal{D}_\alpha(dF | g(F, \xi) = 0)\pi(\theta)d\theta\delta_0(d\nu) \\ &= \frac{\mathcal{D}_{\alpha, g(F,\xi)=0}(dF)}{h(0 : g(F, \xi), \alpha)}\pi(\theta)d\theta\delta_0(d\nu).\end{aligned}$$

Estimation of Moments

$$y_1, \ldots, y_n \overset{i.i.d.}{\sim} F.$$

To estimate the four moments of F, $\xi = \theta = (\mu, \sigma, \gamma, \kappa)$ of F.

$$\begin{aligned}
\mu &= \mathbb{E}(y), \;\; y \sim F \\
\sigma &= sd(y) \\
\gamma &= \mathbb{E}\left(\frac{y-\mu}{\sigma}\right)^3 \\
\kappa &= \mathbb{E}\left(\frac{y-\mu}{\sigma}\right)^4.
\end{aligned}$$

Here

$$t(y_i, x_i, \xi) = y_i$$

$$g(F, \xi) = \begin{pmatrix} \mathbb{E}(y) - \mu \\ sd(y) - \sigma \\ \mathbb{E}\left(\frac{y-\mu}{\sigma}\right)^3 - \gamma \\ \mathbb{E}\left(\frac{y-\mu}{\sigma}\right)^4 - \kappa \end{pmatrix} = 0.$$

The parameter ν and covariate x_i's do not exist in this model.

Estimation of Quantiles

$$y_1, \ldots, y_n \overset{i.i.d.}{\sim} F.$$

To estimate the quantiles of F,

$$\theta = (F_{p_1}, F_{p_2}, \ldots, F_{p_k}),\ 0 < p_1 < \ldots p_k < 1,$$

where F_p is the pth-quantile of F.

$$t(y_i, x_i, \xi) = y_i$$

$$g(F, \xi) = \begin{pmatrix} F_{p_1} - \theta_1 \\ F_{p_2} - \theta_2 \\ \vdots \\ F_{p_k} - \theta_k \end{pmatrix} = 0.$$

The parameter ν and covariate x_i's do not exist in this model.

Regression with random covariates

- To estimate β which satisfies the relation
$$y = x'\beta + \epsilon \tag{2}$$
where $(y, x) \sim F$, F is a distribution on $\mathbb{R}^{p+1}$, $y, \epsilon \in \mathbb{R}$ and $x \in \mathbb{R}^p$.
- We observe
$$(y_1, x_1), \ldots, (y_n, x_n) \overset{i.i.d.}{\sim} F.$$
- Assume
$$\mathbb{E}(x\epsilon) = \mathbb{E}(\epsilon) = 0.$$
- $\xi = \beta$, $\epsilon_i = t(y_i, x_i, \beta) = y_i - \beta^T x_i$ and
$$g(F, \xi) = \mathbb{E}\begin{pmatrix} y - \beta^T x \\ x(y - \beta^T x) \end{pmatrix} = 0.$$
- If we can assume the homoscadescity, the functional condition can be modified as
$$g(F, \xi) = \mathbb{E}\begin{pmatrix} y - \beta^T x \\ x(y - \beta^T x) \\ (y - \beta^T x)^2 - \sigma^2 \end{pmatrix} = 0$$
with additional parameter $\sigma^2 > 0$.

Regression with fixed covariates

- (model)
$$y_i = x_i'\beta + \epsilon_i,\ \epsilon \sim F, i = 1, 2, \dots, n \tag{3}$$
where $y_i \in \mathbb{R}$, covariates $x_i \in \mathbb{R}^p$ are not random, and F is a distribution on $\mathbb{R}$ with mean 0.
- $t(y_i, x_i, \xi) = y_i - x_i'\beta$,
$$g(F) = \int \epsilon dF(\epsilon) = 0.$$
- For the fixed covariate case, the ETEL (exponentially tilted empirical likelihood) approach needs complicated conditional moment models.

Quantile regression

- (model)
$$y_i = x_i'\beta + \epsilon_i,\ \epsilon \sim F, i = 1, 2, \ldots, n \tag{4}$$
where $y_i \in \mathbb{R}$, covariates $x_i \in \mathbb{R}^p$ are not random, and F is a distribution on $\mathbb{R}$ with
$$g(F) = F_\alpha = 0,$$
where $\alpha \in (0, 1)$.
- $t(y_i, x_i, \xi) = y_i - x_i'\beta$,
$$g(F) = F_\alpha = 0.$$

Instrument variable regression

- (Model)
$$y = \alpha + x\beta + w\delta + \epsilon,\ \mathbb{E}(\epsilon) = 0 \tag{5}$$
where y is the response variable, x and w are random covariates, and ϵ is the error.
- $\mathbb{E}(w\epsilon) = 0$, but $\mathbb{E}(x\epsilon) \neq 0$.
- There are nstrumental variables z_1 and z_2. They are correlated with x but not with ϵ.
- We observe
$$(y_i, x_i, w_i, z_{1i}, z_{2i}) \overset{i.i.d.}{\sim} F,\ i = 1, 2, \ldots, n$$
- $$g(F, \xi) = \begin{pmatrix} \mathbb{E}(y - \alpha - x\beta - w\delta) \\ \mathbb{E}\Big((y - \alpha - x\beta - w\delta)z_1\Big) \\ \mathbb{E}\Big((y - \alpha - x\beta - w\delta)z_2\Big) \\ \mathbb{E}\Big((y - \alpha - x\beta - w\delta)w\Big) \end{pmatrix} = 0$$
where $\xi = (\alpha, \beta, \delta)$.

Posterior of Functional Condtion Model

- $\mathbf{x}_n = (x_1, \ldots, x_n)$, $\mathbf{y}_n = (y_1, \ldots, y_n)$.

$$\boldsymbol{\epsilon}_n := (\epsilon_1, \ldots, \epsilon_n) = t(\mathbf{y}_n, \mathbf{x}_n, \xi) = (t(y_1, x_1, \xi), \ldots, t(y_n, x_n, \xi)).$$

- $\boldsymbol{\epsilon}^* = (\epsilon_1^*, \ldots, \epsilon_{k_\epsilon}^*)$ be the distinct values of $\boldsymbol{\epsilon}_n$.
- $\Pi(\boldsymbol{\epsilon}_n)$ is the partition of $[n] = \{1, 2, \ldots, n\}$ defined by the equivalence class

$$i \sim j \iff \epsilon_i = \epsilon_j, \ \forall i, j \in [n].$$

- Let $a(\cdot)$ be the density of $\bar{\alpha}$ and $b(y : \xi, x)$ be the density of y when $\epsilon = t(y, x, \xi) \sim a(\cdot)$.
- **A2** $\Pi(\boldsymbol{\epsilon}_n) = \Pi(\mathbf{y}_n)$, for all ξ and $(\mathbf{x}_n, \mathbf{y}_n)$.

Posterior of Functional Condtion Model

Assume **A1** and **A2**. Under the functional condition model and the prior, the following hold.

(a) (joint posterior)

$$\pi(dF, d\theta|\mathbf{y}_n) \propto \mathcal{D}_{\alpha+nF_{n,(\theta,0)}}(dF|g(F,(\theta,0))=0)\prod_{i=1}^{k} b(y_i^* : (\theta,0), x_i^*)$$
$$\times \frac{h(0 : g(F,(\theta,0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F,(\theta,0)), \alpha)}\pi(\theta)d\theta$$
$$= \mathcal{D}_{\alpha+nF_{n,(\theta,0)},g(F,(\theta,0))=0}(dF)\prod_{i=1}^{k} b(y_i^* : (\theta,0), x_i^*)\frac{\pi(\theta)}{h(0 : g(F,(\theta,0)), \alpha)}d\theta$$

(b) (posterior of F)

$$\pi(dF|\mathbf{y}_n, \theta) = \mathcal{D}_{\alpha+nF_{n,(\theta,0)}}(dF|g(F,(\theta,0))=0)$$

(c) (posterior of θ)

$$\pi(d\theta|\mathbf{y}_n) \propto \prod_{i=1}^{k} b(y_i^* : (\theta,0), x_i^*)\frac{h(0 : g(F,(\theta,0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F,(\theta,0)), \alpha)}\pi(\theta)d\theta,$$

where $F_{n,\xi} = \frac{1}{n}\sum_{i=1}^{n}\delta_{t(y_i,x_i,\xi)}$.

Remarks

$$\pi(d\theta|\mathbf{y}_n) \propto \prod_{i=1}^{k} b(y_i^* : (\theta, 0), x_i^*) \frac{h(0 : g(F, (\theta, 0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F, (\theta, 0)), \alpha)} \pi(\theta) d\theta.$$

- If $\epsilon_i = t(y_i, x_i)$, the factor in the posteior

  $$\prod_{i=1}^{k} b(y_i^* : (\theta, 0), x_i^*)$$

  disappears.
- If $\epsilon_i = t(y_i, x_i, \xi)$, the factor does not disappear even as $n \to \infty$. Thus, we recommend the posterior

  $$\pi(d\theta|\mathbf{y}_n) \propto \frac{h(0 : g(F, (\theta, 0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F, (\theta, 0)), \alpha)} \pi(\theta) d\theta.$$

Posterior of Functional Condtion Model

If $\xi = \theta$ and $\epsilon_i = t(y_i, x_i)$, we have the following.

1. (joint posterior)

$$\pi(dF, d\theta|\mathbf{y}_n) \propto \mathcal{D}_{\alpha+nF_n}(dF|g(F,\theta)=0)\frac{h(0:g(F,\theta),\alpha+nF_n)}{h(0:g(F,\theta),\alpha)}\pi(\theta)d\theta$$

$$= \mathcal{D}_{\alpha+nF_n,g(F,\theta)=0}(dF)\frac{\pi(\theta)}{h(0:g(F,\theta),\alpha)}d\theta$$

2. (posterior of F)

$$\pi(dF|\mathbf{y}_n,\theta) = \mathcal{D}_{\alpha+nF_n}(dF|g(F,\theta)=0)$$

3. (posterior of θ)

$$\pi(d\theta|\mathbf{y}_n) \propto \frac{h(0:g(F,\theta),\alpha+nF_n)}{h(0:g(F,\theta),\alpha)}\pi(\theta)d\theta.$$

Here $F_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{t(y_i,x_i)}$.

As $A := \alpha(\mathcal{X}) \to 0$

Theorem

(i) As $A \to 0$,

$$\frac{h(0:g(F,(\theta,0)),\alpha + nF_{n,(\theta,0)})}{h(0:g(F,(\theta,0)),\alpha)} \to \frac{h(0:g(F,(\theta,0)),nF_{n,(\theta,0)})}{\tilde{h}(0:(\theta,0))}, \ \forall a.a-\theta.$$

(ii) There exists a funtion $u(\theta)$ satisfying

$$\left| \prod_{i=1}^{k} b(y_i^* : (\theta,0),x_i^*) \frac{h(0:g(F,(\theta,0)),\alpha + nF_{n,(\theta,0)})}{h(0:g(F,(\theta,0)),\alpha)} \right| \le u(\theta), \ \forall\theta$$

and

$$\int u(\theta)\pi(\theta)d\theta < \infty.$$

Suppose (i) and (ii) hold. Then, as $A \to 0$,

$$\pi(d\theta|\mathbf{y}_n) \rightsquigarrow \pi_0(d\theta|\mathbf{y}_n) \propto \prod_{i=1}^{k} b(y_i^* : (\theta,0),x_i^*) \frac{h(0:g(F,(\theta,0)),nF_{n,(\theta,0)})}{\tilde{h}(0:(\theta,0))}\pi(\theta)d\theta.$$

When $A := \alpha(\mathcal{X}) \to 0$

Consider the model

$$X_1, \ldots, X_n | F \sim F$$
$$g(F) = \theta$$
$$\pi(dF, d\theta) = \mathcal{D}_\alpha(dF | g(F) = \theta)\pi(\theta)d\theta.$$

The posterior is

$$\pi(d\theta | \mathbf{x}_n) \propto \frac{h(\theta : g(F), \alpha + nF_n)}{h(\theta : g(F), \alpha)} \pi(\theta)d\theta.$$

Under a strict set of conditions, as $A \to 0$,

$$\frac{h(\theta : g(F), \alpha + nF_n)}{h(\theta : g(F), \alpha)} \pi(\theta) \to \frac{h(\theta : g(F), nF_n)}{\tilde{h}(\theta, g(F))} \pi(\theta).$$

Note $h(\theta : g(F), \alpha)$ is the Bayesian bootstrap posterior of θ.

Posterior Computation

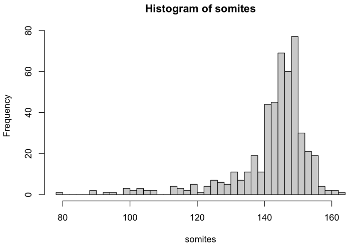
$$\pi(d\theta|\mathbf{y}_n) \propto \frac{h(0 : g(F,(\theta,0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F,(\theta,0)), \alpha)} \pi(\theta) d\theta.$$

Three proposals

1. When the closed form of $\frac{h(0 : g(F,(\theta,0)), \alpha + nF_{n,(\theta,0)})}{h(0 : g(F,(\theta,0)), \alpha)}$ is known, the posterior can be obtained using by an MCMC algorithm.
2. Bayesian bootstrap with reweighting. The posterior of each parameter is separately computed.
3. Constrained Hamiltonian Monte Carlo: Shake Algorithm

Content

Numbers of Somites of Garden Earthworms



$$X_1, \ldots, X_n \overset{i.i.d.}{\sim} F$$

- The data set consists of the number of somites of 487 earthworms (Pearl and Fuller, 1905).
- We estimate the mean (μ), standard deviation (σ), skewness (γ), kurtosis (κ) through the moment condition model.

moment functions

- $g_1(X, \theta) = \begin{pmatrix} X - \theta_1 \\ e^{-2\theta_2}(X - \theta_1)^2 - 1 \\ e^{-3\theta_2}(X - \theta_1)^3 - \theta_3 \\ e^{-4\theta_2}(X - \theta_1)^4 - e^{\theta_4} \end{pmatrix}$,
  where $\theta = (\mu, \log(\sigma), \gamma, \log(\kappa + 3))$
- $g_2(X, \theta) =$
  $\begin{pmatrix} X - \theta_1 \\ ((X - \theta_1)/\theta_2)^2 - 1 \\ ((X - \theta_1)/\theta_2)^3 - \theta_3 \\ ((X - \theta_1)/\theta_2)^4 - (\theta_4 + 3) \end{pmatrix}$,
  where $\theta = (\mu, \sigma, \gamma, \kappa)$

Estimators

Estimators for Earthworm data

	μ	σ	γ	κ
Sample	142.715	11.853	-2.179	5.857
GMM	142.715	11.853	-2.179	5.857
EL	142.715	11.853	-2.179	5.857
Bootstrap	142.756	11.935	-2.149	5.604
BB	142.726	11.840	-2.105	5.490
DP	142.661	12.109	-2.228	6.481

Table: Estimators using g_1

	μ	σ	γ	κ
Sample	142.715	11.853	-2.179	5.857
GMM	142.715	11.853	-2.179	5.857
EL	142.715	11.853	-2.179	5.857
Bootstrap	142.756	11.825	-2.149	5.527
BB	142.716	11.766	-2.151	5.528
DP	142.674	12.234	-2.288	6.442

Table: Estimators using g_2

- The transformation of parameters does not have significant effects on the estimators.
- Estimators of GMM and EL are the same as sample moments.
- All models have similar estimators except κ.

IV regression

- **(Model)**

$$y = \alpha + \beta x + \delta w + \epsilon \quad \textit{where} \;\; \mathbb{E}_P[\epsilon] = 0, \;\; Cor(x, \epsilon) \neq 0.$$

  z_1, z_2 : Instrumental variables where $Cor(z_j, x) \neq 0, \;\; j = 1, 2.$

- **(Moment function)**

$$g(X, \theta) = \begin{pmatrix} y - \alpha - \beta x - \delta w \\ (y - \alpha - \beta x - \delta w) z_1 \\ (y - \alpha - \beta x - \delta w) z_2 \\ (y - \alpha - \beta x - \delta w) w \end{pmatrix}$$

  where $X = (x, y, w, z_1, z_2)$, $\theta = (\alpha, \beta, \delta)$.

- **Data**
  - The number of observations are $n \in \{10, 25, 50, 100, 250, 500, 1000, 2000\}$.
  - 100 simulation data sets are generated from the true model for each n.
  - Prediction data set which has 500 observations is generated from the true model.

- **Analysis**
  - BB, DP, BETEL, GMM, EL, and Bootstrap are used for moment condition models.
  - The performance of the models is compared using MSE and distributions for each n.

MSE

MSE for $n = 25$

	α	β	δ	predict
GMM	0.283	0.023	0.850	1.158
EL	0.297	0.023	0.779	1.189
Bootstrap	4.339	3.757	2.827	16.242
BETEL	0.368	0.065	0.959	1.470
BB	0.188	0.018	0.415	1.094
DP	0.142	0.017	0.300	1.066

MSE for $n = 50$

	α	β	δ	predict
GMM	0.107	0.013	0.341	1.069
EL	0.103	0.014	0.317	1.079
Bootstrap	0.619	0.252	0.400	1.913
BETEL	0.108	0.016	0.337	1.108
BB	0.088	0.012	0.233	1.045
DP	0.072	0.012	0.197	1.039

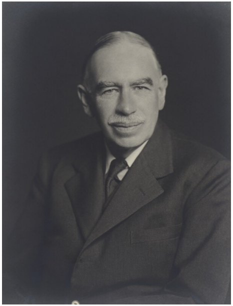
- $n = 25$
  - All models except bootstrap have small MSE.
  - BETEL failed to estimate the parameters from 1 dataset.
  - The size of MSE is $DP < BB < GMM \approx EL < BETEL < Bootstrap$.
- $n = 50$
  - All models have small MSE.
  - MSE of all models except bootstrap are similar.
  - MSE of bootstrap is slightly larger than other models.

References

- Chang, J. T., & Pollard, D. (1997). Conditioning as disintegration. Statistica Neerlandica, 51(3), 287-317.
- Chib, S., Shin, M., & Simoni, A. (2018). Bayesian estimation and comparison of moment condition models. JASA, 113(524), 1656-1668.
- Efron, B., & Tibshirani, R. J. (1994). An introduction to the bootstrap. Chapman and Hall/CRC.
- Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica: Journal of the econometric society, 1029-1054.
- Owen, A. B. (2001). Empirical likelihood. Chapman and Hall/CRC.
- Rubin, D. B. (1981). The bayesian bootstrap. The annals of statistics, 130-134.
- Schennach, S. M. (2005). Bayesian exponentially tilted empirical likelihood. Biometrika, 92(1), 31-46.

When the facts change, I change my opinion. What do you do, Sir?

- John Maynard Keynes



Thank you.