



研究興味領域:

(**Step1**): **Dependent Dirichlet Process**と高頻度取引データを用いたマーケット関与者の数および状態の動的ベイジアンノンパラメトリック推論

(**Step2**): オプションのダイナミックヘッジと現物価格動態の関係？

2021年6月10日 | 10 June. 2021

馬 煜茗 | Yuming MA

東京工業大学 環境・社会理工学院 技術経営専門職学位課程
Department of Technology and Innovation Management
School of Environment and Society
Tokyo Institute of Technology

この資料は、上記に示す作成者により管理され使用されるものとします。

この資料のいかなる部分についても、本作成者の事前の書面による承諾を得ずに、回覧・引用・複製、あるいは配布してはならないものとします。

**時代背景:**

時代背景: 米国の金融市場を先駆けにして、金融工学の応用はここ20年間に於いて飛躍的に発展を成し遂げた。現物、先物、デリバティブの売買双方を含め自動化されたアルゴリズム取引が市場の重要成分となっている。

研究目的:

隠れ変数となるマーケットプレイヤの状態を取引データから推論するモデルを開発すること

研究手法:

Dependent Dirichlet Processを用いたベイジアンノンパラメトリック推論手法、これで時系列確率過程のクラスタリング問題を解く

研究意義と新規性:

この研究はベイジアン統計学習手法の側面に重点を置き、該当手法の発展に工学の意味合いを持ち、工学の論文となる時間依存性を持つDependent Dirichlet Processを金融市場の時系列データに適用した研究はまだない

RQ1:

単一品種の現物金融商品の日中売買データに適用して、そこで得たクラスタデータになんの物理的な統計意味を持つのか？

RQ2:

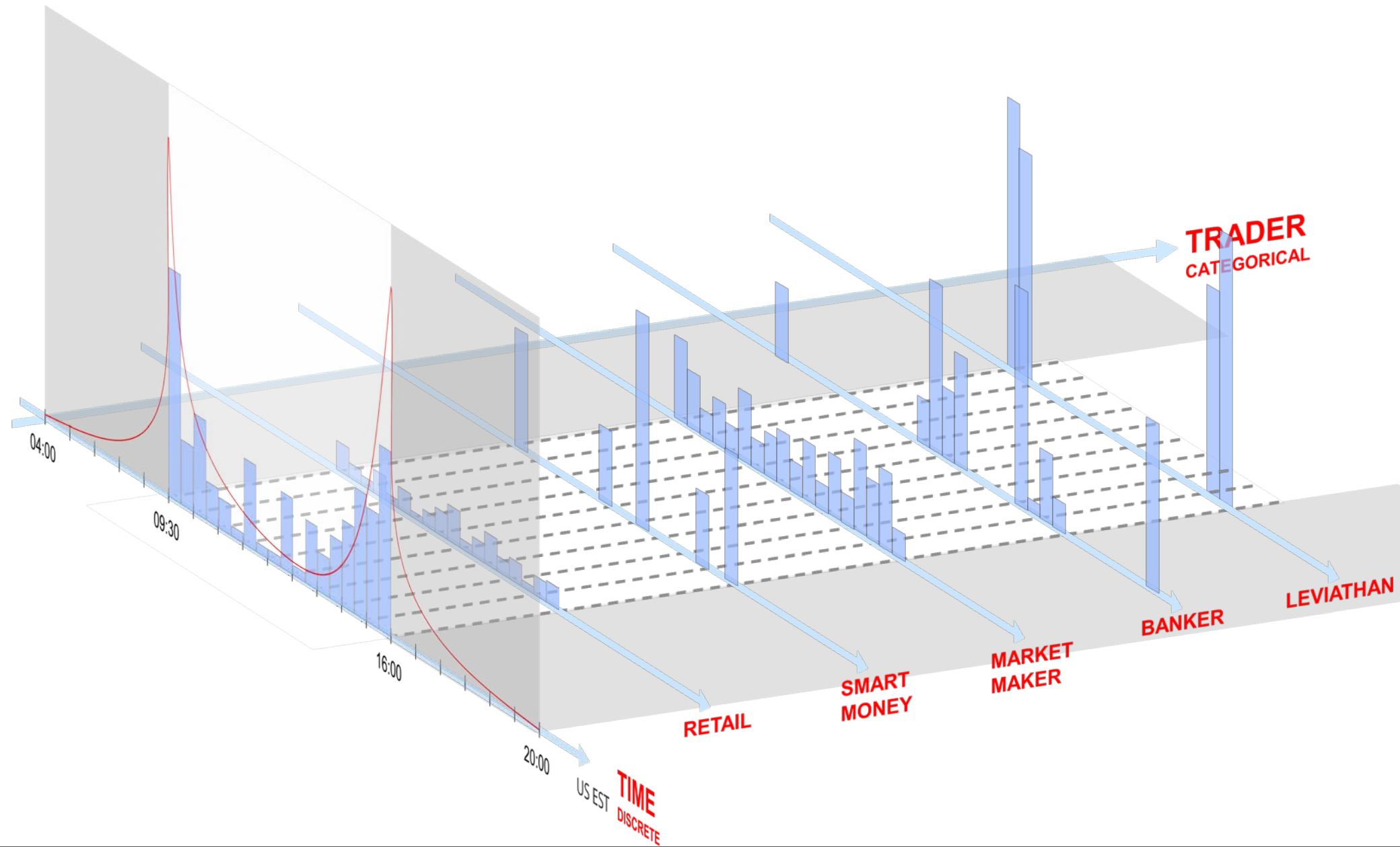
多品種の現物金融商品の日中売買データに適用に適用して、そこで得たクラスタデータになんの物理的な統計意味を持つのか？

RQ3:

現物先物、デリバティブの時系列データに適用に適用して、そこで得たクラスタデータになんの物理的な統計意味を持つのか？

研究成果:

時間依存性を持つランダム確率測度の上で定義するDirichlet Processの実装とそれが現実世界の生データに応用した一例が得られて、該当ベイジアン統計学習手法の工学発展に寄与する





マーケットプレイスは各々の持つマーケットエクスポージャー X_t に従って $\Pr(vol_t | X_{l,t}, X_{s,t})$ で注文を出し成約されると仮定し、マーケットエクスポージャーが似た同士を一つのクラスタに集約する

Market Exposure

Def: ある成分においてヘッジされていない資金の割合

Formulation:

今、ポートフォリオがあってその価値が $V_t(X_t) = \omega_t^\top S_t$
 ω_t はポートフォリオの組み入れ比率で S_t は標的資産の市場価格
 これに対してTaylor展開(伊藤の補題)するが、扱いたい変数によって式が変わる

$$dV_t(X_t) = \nabla V_t^\top (dX_t) + \frac{1}{2} dX_t^\top \text{Hessian}(V_t(X_t)) dX_t + o$$

Black-Scholes-MertonのMannerに倣って $V_t(\omega_t, S_t; \sigma_t, r_{f_t}, \dots)$

$$\begin{aligned} dV_t &= V_{t+s} - V_t \\ &= \frac{\partial V_t}{\partial S_t} (S_{t+s} - S_t) + \frac{1}{2} \frac{\partial^2 V_t}{\partial S_t^2} (S_{t+s} - S_t)^2 + \dots + \frac{\partial V_t}{\partial \omega_t} (\omega_{t+s} - \omega_t) + \dots \\ &\quad + \frac{\partial V_t}{\partial \sigma_t} (\sigma_{t+s} - \sigma_t) + \dots + \frac{\partial V_t}{\partial r_{f_t}} (r_{f_{t+s}} - r_{f_t}) + \dots \end{aligned}$$

ここで一般に総称Greeksとの記号を使って表現を図る

$$\text{Delta: } \frac{\partial V_t}{\partial S_t}, \text{Gamma: } \frac{\partial^2 V_t}{\partial S_t^2}, \text{Theta: } \frac{\partial V_t}{\partial t}, \text{Vega: } \frac{\partial V_t}{\partial \sigma_t}, \text{Pho: } \frac{\partial V_t}{\partial r_{f_t}}$$

Multi-Factor Model Manner

$$\begin{aligned} V_t &= \omega_t^\top S_t = f(r_t) = f(X_t b_t + u_t) \\ dV_t &= V_{t+s} - V_t \\ &= \frac{\partial f}{\partial X_t} (X_{t+s} - X_t) + \dots + \frac{\partial f}{\partial u_t} (u_{t+s} - u_t) + \dots \end{aligned}$$

Hedger (Variety with Greeks Positions and Hedge Strategy)

管理目的に応じて裁量的に影響を消したい因子を選び、その因子の導関数を参考に相反なポジションを取ることでヘッジ取引を行う
 各HedgerのGreeksは自分の持つポジションですぐ計算ができるが、ヘッジ目的で売買を行う量は各Hedgerの戦略に依存して確率的に揺らぐ

しかし、たとえエクスポージャーを持ち、それが即座に全部ヘッジできるのは運用資産残高の高いマーケットに対し大きな影響力を示す主要プレイヤーについてとても考えにくい、これに相反する場合はマーケットが激動する場合であろう

従って、**Hedgerの時刻 t と時刻 $t+s$ における各エクスポージャーの差分はベル型の分布に従うと仮定**する

残したいエクスポージャーの時間差分は狭い分布、残したくないエクスポージャーは歪度を持った分布のように揺らぐ(エクスポージャーは売買立場で正負を持つため)

歪度尖度など階数の高いモーメントを持つ確率分布をシミュレーションするにはパラメータの次元数から無理があると考え

従って**Dependent Dirichlet Process - Gaussian Mixture Model**で近似

Directional Investor (Delta Position Only No Hedge)

リスクのマネジメント方法はポートフォリオの(数と重み)最適化で行われ、下記状態の変化に応じてリバランス操作を行う

$$r_t = X_t b_t + u_t$$

r_t : 超過リターン X_t : ファクタエクスポージャー b_t : ファクタリターン u_t : 個別因子

$$\begin{pmatrix} r_{1,t} \\ \vdots \\ r_{n,t} \\ \vdots \\ r_{N,t} \end{pmatrix} = \begin{pmatrix} x_{11,t} & \cdots & x_{1k,t} & \cdots & x_{1K,t} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{n1,t} & \cdots & x_{nk,t} & \cdots & x_{nK,t} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N1,t} & \cdots & x_{Nk,t} & \cdots & x_{NK,t} \end{pmatrix} \begin{pmatrix} b_{1,t} \\ \vdots \\ b_{k,t} \\ \vdots \\ b_{K,t} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ \vdots \\ u_{n,t} \\ \vdots \\ u_{N,t} \end{pmatrix}$$

$V_t = X_t F_t X_t^\top + \Delta_t$ V_t : 超過リターンのリスク(分散共分散)行列

$$F_t = \text{cov}(b_t, b_t) = \mathbb{E}[(b_t - \mathbb{E}[b_t])(b_t - \mathbb{E}[b_t])^\top]$$

$$\Delta_t = \text{diag}(\text{cov}(u_t, u_t))$$

X_t は往々にして統計量変換で正規化されるので、**方向性取引の状態の確率過程は概ねGaussian Processで表現できるのであろう**



Raw Data: SPY 2021-02-03 Tick-by-Tick

	time	exchange	price	size	conditions	id	tape
990	2021-02-03 12:05:22.268400+00:00	P	382.80	104	[' ', 'F', 'T']	52983525034830	B
991	2021-02-03 12:05:53.756600+00:00	P	382.84	100	[' ', 'T']	52983525034831	B
992	2021-02-03 12:06:08.811700+00:00	P	382.80	196	[' ', 'T']	52983525034832	B
993	2021-02-03 12:06:08.811700+00:00	P	382.80	1	[' ', 'T', 'I']	52983525034833	B
994	2021-02-03 12:06:08.811700+00:00	P	382.80	300	[' ', 'T']	52983525034834	B
995	2021-02-03 12:06:08.821600+00:00	P	382.80	3	[' ', 'T', 'I']	52983525034835	B
996	2021-02-03 12:06:44.574600+00:00	P	382.81	261	[' ', 'T']	52983525034836	B
997	2021-02-03 12:06:51.025400+00:00	P	382.80	7	[' ', 'T', 'I']	52983525034837	B
998	2021-02-03 12:06:51.025400+00:00	P	382.80	5	[' ', 'T', 'I']	52983525034838	B
999	2021-02-03 12:06:51.025400+00:00	P	382.80	10	[' ', 'T', 'I']	52983525035056	B
1000	2021-02-03 12:06:51.025400+00:00	T	382.80	30	[' ', 'T', 'I']	62879132118885	B
1001	2021-02-03 12:06:51.857400+00:00	K	382.80	1	[' ', 'T', 'I']	52983525029221	B
1002	2021-02-03 12:06:57.169700+00:00	Y	382.79	100	[' ', 'F', 'T']	52983525027907	B
1003	2021-02-03 12:07:03.646500+00:00	P	382.75	12	[' ', 'T', 'I']	52983525035057	B
1004	2021-02-03 12:07:16.928100+00:00	T	382.71	30	[' ', 'T', 'I']	62879132170602	B
1005	2021-02-03 12:07:20.288200+00:00	P	382.71	188	[' ', 'T']	52983525035058	B
1006	2021-02-03 12:07:25.560700+00:00	T	382.71	500	[' ', 'F', 'T']	62879132171373	B
1007	2021-02-03 12:07:41.171300+00:00	P	382.72	1	[' ', 'T', 'I']	52983525035059	B
1008	2021-02-03 12:07:43.020900+00:00	P	382.74	100	[' ', 'T']	52983525035060	B
1009	2021-02-03 12:07:46.508500+00:00	P	382.72	400	[' ', 'F', 'T']	52983525035061	B

同じ時刻で記録されている箇所が存在
1対多数の取引が成立された場合が記録される

データの処理

等時間間隔でリサンプリング

補完:

1成約は売り手と買い手の双方が存在して成す
従って、損益のことを一旦無視して以下のような追加を行う

size_long:=size

買い手が自分のエクスポージを正の方向に修正したことを表す

size_short:=-size

売り手が自分のエクスポージャを負の方向に修正したことを表す

t 時刻において観測データ $\mathbf{Z}_t \in \mathbb{R}^{M_t \times 2 \times D}$

$$\mathbf{Z}_{l,t}^m = \{\text{timestamp}', 'price', 'size_long', \dots\}$$

$$\mathbf{Z}_{s,t}^m = \{\text{timestamp}', 'price', 'size_short', \dots\}$$

$$\mathbf{Z}_t = \{(\mathbf{Z}_{l,t}^1, \mathbf{Z}_{s,t}^1), \dots, (\mathbf{Z}_{l,t}^m, \mathbf{Z}_{s,t}^m), \dots, (\mathbf{Z}_{l,t}^{M_t}, \mathbf{Z}_{s,t}^{M_t})\}$$

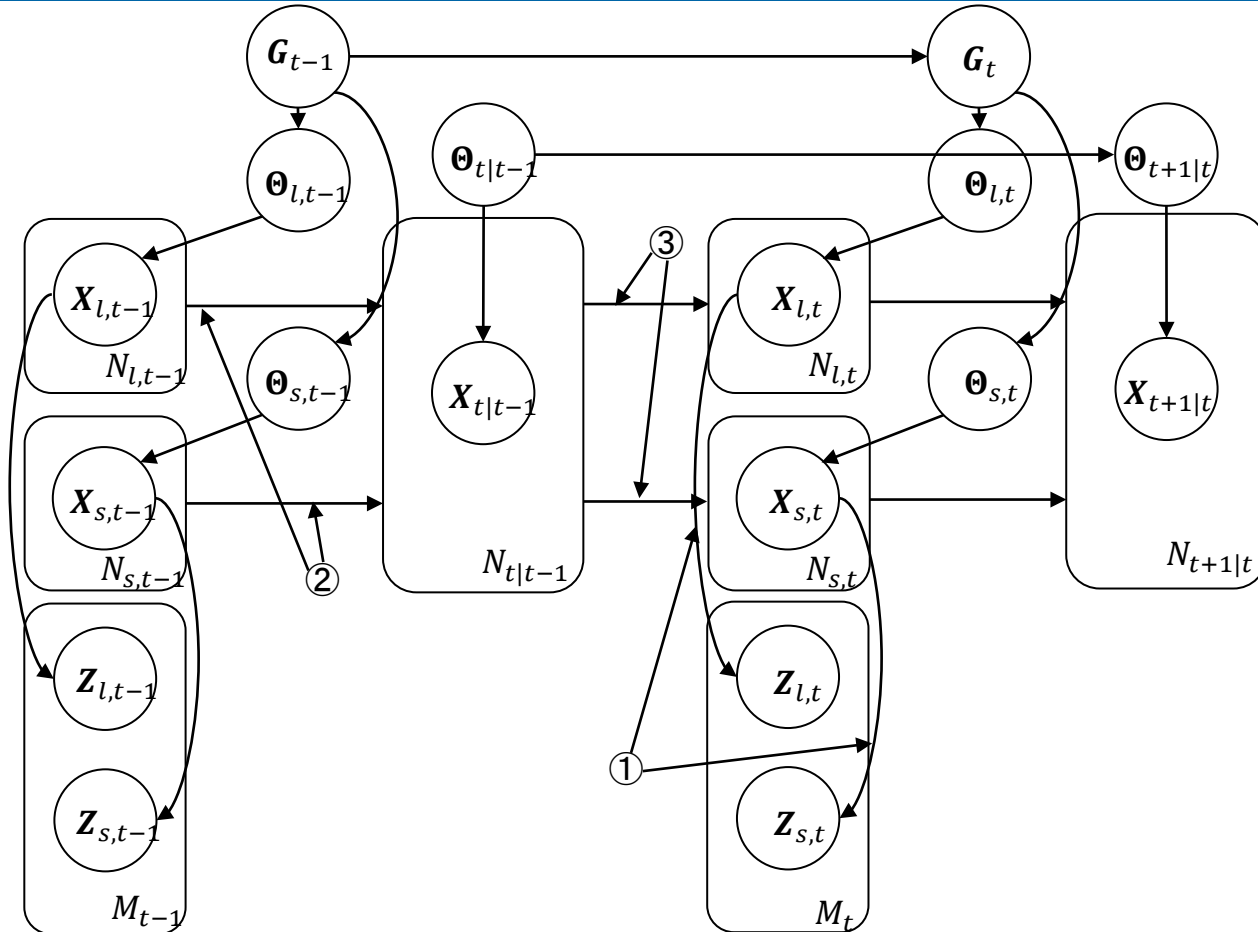
t 時刻において確率的に存在するプレイヤーの集合はそれぞれの状態ベクトルで表現する $\mathbf{X}_t \in \mathbb{R}^{N_t \times L}$

$$\mathbf{X}_t = \{\mathbf{x}_t^1, \dots, \mathbf{x}_t^n, \dots, \mathbf{x}_t^{N_t}\}$$

$$\mathbf{x}_t^n = \{p_1(\cdot | \boldsymbol{\theta}_1), \dots, p_l(\cdot | \boldsymbol{\theta}_l), \dots, p_L(\cdot | \boldsymbol{\theta}_L)\}$$

\mathbf{x}_t^n の各成分は \mathbf{x}_t^n の指すプレイヤーのマーケット
エクスポージャ

$\boldsymbol{\theta}_l$ によって \mathbf{x}_t^n の各成分が独立にサンプリングされる



Graphical Model

t 時刻において観測データ $\mathbf{Z}_t \in \mathbb{R}^{M_t \times 2 \times D}$

$$\mathbf{Z}_{l,t}^m = \{ 'time', 'price', 'size_long', \dots \}$$

$$\mathbf{Z}_{s,t}^m = \{ 'time', 'price', 'size_short', \dots \}$$

$$\mathbf{Z}_t = \{ (\mathbf{Z}_{l,t}^1, \mathbf{Z}_{s,t}^1), \dots, (\mathbf{Z}_{l,t}^m, \mathbf{Z}_{s,t}^m), \dots, (\mathbf{Z}_{l,t}^{M_t}, \mathbf{Z}_{s,t}^{M_t}) \}$$

t 時刻において確率的に存在するプレイヤーの集合はそれぞれの状態ベクトルで表現する $\mathbf{X}_t \in \mathbb{R}^{N_t \times L}$

$$\mathbf{X}_t = \{ \mathbf{x}_t^1, \dots, \mathbf{x}_t^n, \dots, \mathbf{x}_t^{N_t} \}$$

$$\mathbf{x}_t^n = \{ p_1(\cdot | \boldsymbol{\theta}_1^n), \dots, p_l(\cdot | \boldsymbol{\theta}_l^n), \dots, p_L(\cdot | \boldsymbol{\theta}_{L_n}^n) \}$$

\mathbf{x}_t^n の各成分は \mathbf{x}_t^n の指すプレイヤーのマーケットエクスポージャ成分、全体を隠れ変数と扱う
 $\boldsymbol{\theta}_l^n$ によって \mathbf{x}_t^n の各成分が独立にサンプリングされる

$$\boldsymbol{\theta}_t = \left\{ \begin{pmatrix} \boldsymbol{\theta}_1^1 \\ \vdots \\ \boldsymbol{\theta}_l^1 \\ \vdots \\ \boldsymbol{\theta}_{L_1}^1 \end{pmatrix}_t, \dots, \begin{pmatrix} \boldsymbol{\theta}_1^n \\ \vdots \\ \boldsymbol{\theta}_l^n \\ \vdots \\ \boldsymbol{\theta}_{L_n}^n \end{pmatrix}_t, \dots, \begin{pmatrix} \boldsymbol{\theta}_1^{N_t} \\ \vdots \\ \boldsymbol{\theta}_l^{N_t} \\ \vdots \\ \boldsymbol{\theta}_{L_{N_t}}^{N_t} \end{pmatrix}_t \right\}$$

ここで、クラスタ(類似するプレイヤー同士)の中であれば同じパラメータベクトルで表現することができるため、同じクラスタだと思われる $\boldsymbol{\theta}^n$ 同士の期待値を取りユニックなクラスタのパラメータ空間を作る

$$\boldsymbol{\theta}_t^* = \{ \boldsymbol{\theta}_t^{*1}, \dots, \boldsymbol{\theta}_t^{*u}, \dots, \boldsymbol{\theta}_t^{*U_t} \}$$

① how much money it would be traded?

Select a \mathbf{Z}_t and then the amount of money $price_{m,t} \times size_{m,t}$
 $m \sim \text{Categorical}(p = \boldsymbol{\pi}), \boldsymbol{\pi} \sim \text{GEM}(\alpha), \alpha \sim \text{InverseGamma}(1,1)$

② survival, ,death, or birth?

$$1_{\{x_t^n \text{ survives}\}} \sim \text{Bernolli}(p_{n,t|t-1})$$

③ do long or short or wait?

$$decision_{t|t-1} \sim \text{Categorical}(p(A|\mathbf{x}_{t-1}^n)), A \in \{-1,0,1\}$$



Dependent Dirichlet Process – Gaussian Mixture Model

Dirichlet Process on 1-dimension define a partition A on params axis

$$G(A) = \sum_{j=1}^{\infty} \pi_j \delta_{\theta_j}(A) \sim \text{DirichletProcess}(\alpha, H)$$

$\pi_j \sim GEM(\alpha), \alpha \sim \text{InverseGamma}(1,1), \theta_j \sim H, H := \text{base measure}$

$$\delta_{\theta_j}(A) = \begin{cases} 1, & \theta_j \in A \\ 0, & \theta_j \notin A \end{cases}$$

ランダム確率測度 G と任意確率密度関数 $f(\cdot | \theta)$ が混合確率測度を作る

$$F(\cdot) = \int_{\theta} f(\cdot | \theta) dG(\theta)$$

観測データは分布 F で生成されると仮定

$$\mathbf{Z}_t = \{(\mathbf{z}_{l,t}^1, \mathbf{z}_{s,t}^1), \dots, (\mathbf{z}_{l,t}^m, \mathbf{z}_{s,t}^m), \dots, (\mathbf{z}_{l,t}^{M_t}, \mathbf{z}_{s,t}^{M_t})\}$$

隠れ状態変数の数 N_t は不明なため、一旦最大数を置き、後に学習する

$$\mathbf{X}_t = \{\mathbf{x}_t^1, \dots, \mathbf{x}_t^n, \dots, \mathbf{x}_t^{N_t}\} \quad N_t \leq 2M_t$$

\mathbf{x}_t^n の状態空間が多次元に及ぶ可能性はある、ここで一旦多次元正規分布だと仮定

$$\mathbf{x}_t^n \sim \text{MultiNormal}(\boldsymbol{\mu}_t^n, \boldsymbol{\sigma}_t^n)$$

この場合 \mathbf{x}_t^n の状態ベクトルのパラメータベクトルが

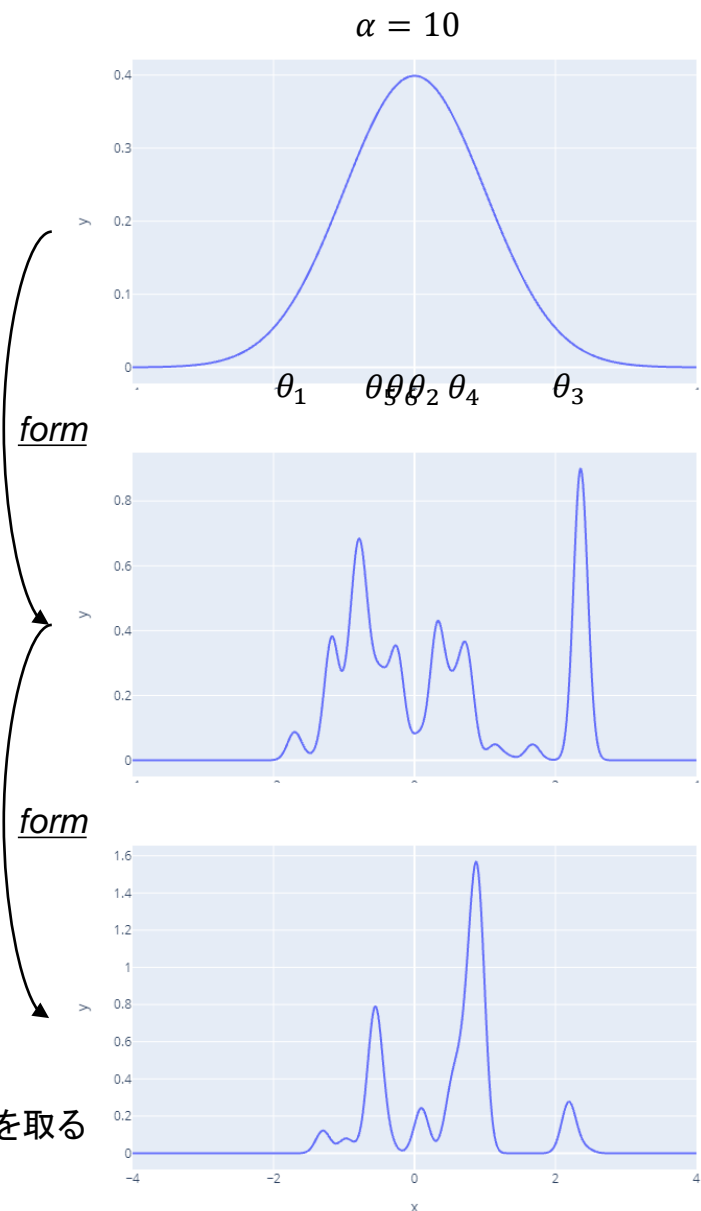
$$\boldsymbol{\theta}_t^n = (\boldsymbol{\mu}_t^n, \boldsymbol{\sigma}_t^n)$$

パラメータベクトル $\boldsymbol{\theta}_t^n$ によって \mathbf{x}_t^n の従う確率分布が定まるので

複数の \mathbf{x} があると、DP で得た $\boldsymbol{\pi}$ (確率測度 = 重み) で DP で得た候補 $\boldsymbol{\theta}$ の中で選ぶ

$i(\boldsymbol{\theta}_t^n) \sim \mathbb{E}[\text{Cartegorical}(\boldsymbol{\pi})]$ ← 多次元の場合はサンプリングしたインデックスの期待値を取る

$$i(\boldsymbol{\theta}_t^{*u}), c^*(\mathbf{x}_t^n), n(\boldsymbol{\theta}_t^{*u}), \sim \text{unique_with_count}(\mathbb{E}[\text{Cartegorical}(\boldsymbol{\pi})])$$





Dependent Dirichlet Process Prior Construction

Corresponded Parameters

Implied Entities State

Observations

Clusters Parameters

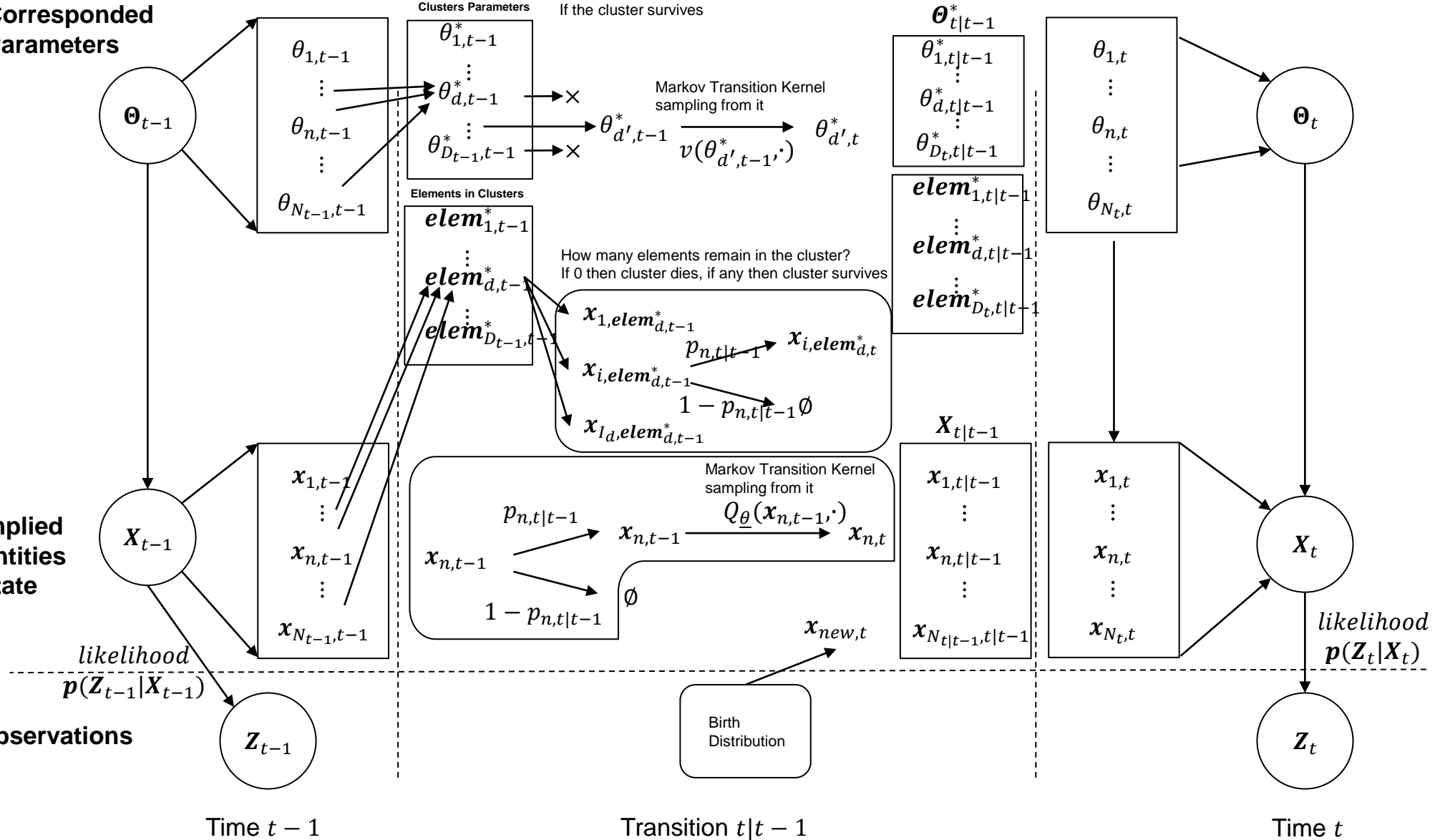
If the cluster survives

Elements in Clusters

How many elements remain in the cluster?
If 0 then cluster dies, if any then cluster survives

Markov Transition Kernel sampling from it

Birth Distribution





$x_{n,t}$ → the object survived and transited and assigned into an existing cluster

$$\Pi_{j,t}^1(\text{assign into } j\text{th cluster} | \theta_{1,t}, \dots, \theta_{n-1,t}) = \frac{\#elem_{j,t|t-1}^* + \sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t})}{(n-1) + \sum_j^{n-1} \sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t}) + \alpha}$$

$$p(x_{n,t} | x_{1,t}, \dots, x_{n-1,t}, \mathbf{X}_{t|t-1}, \boldsymbol{\theta}_{t|t-1}^*, \boldsymbol{\theta}_t) = Q_{\underline{\theta}}(x_{n,t-1}, x_{n,t}) f(x_{n,t} | \theta_{n,t}^*)$$

the object survived and assigned into an existing cluster but without elements assigned yet

$$\Pi_{j,t}^2(\text{assign into } j\text{th cluster without elem} | \theta_{1,t}, \dots, \theta_{n-1,t}) = \frac{\sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t})}{(n-1) + \sum_j^{n-1} \sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t}) + \alpha}$$

$$p(x_{n,t} | x_{1,t}, \dots, x_{n-1,t}, \mathbf{X}_{t|t-1}, \boldsymbol{\theta}_{t|t-1}^*, \boldsymbol{\theta}_t) = Q_{\underline{\theta}}(x_{n,t-1}, x_{n,t}) v(\theta_{n,t-1}^*, \theta_{n,t}) f(x_{n,t} | \theta_{n,t}^*)$$

the object survived but assigned into a brand-new cluster

$$\Pi_t^3(\text{creating a new cluster} | \theta_{1,t}, \dots, \theta_{n-1,t}) = \frac{\sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t})}{(n-1) + \sum_j^{n-1} \sum_{i=1}^{D_{t-1}} \#elem_{i,t|t-1}^* CluSurvInd_{i,t|t-1} \delta_i(c_{j,t}) + \alpha}$$

$$p(x_{n,t} | x_{1,t}, \dots, x_{n-1,t}, \mathbf{X}_{t|t-1}, \boldsymbol{\theta}_{t|t-1}^*, \boldsymbol{\theta}_t) = \int_{\theta} f(x_{n,t} | \theta) dH(\theta)$$

The new base distribution $\sum_{\boldsymbol{\theta}_t} \Pi_{j,t}^1 \delta_{\theta_{n,t}} + \sum_{\boldsymbol{\theta}_{t|t-1}^* \setminus \boldsymbol{\theta}_t} \Pi_{j,t}^2 v(\theta_{n,t-1}^*, \theta_{n,t}) \delta_{\theta_{n,t}} + \Pi_t^3 H$