Example 5.26 (Exponential frailty model). 生存時間のペアのサンプル $(X_1, Y_1), ..., (X_n, Y_n)$ を考える．例えば X_i は" 父親" の生存時間で Y_i は" 息子" の生存時間などである．観測されない値 z_i が与えられた時に，X_i と Y_i は独立で，それぞれ指数分布 $Exp(z_i)$, $Exp(\theta z_i)$ に従うと仮定する．パラメータ θ を推定する問題を考える．

まず，$z_1, ..., z_n$ はある未知の分布に従う確率変数 $Z_1, ..., Z_n$ の実現値であるとする．

一つ目のアプローチは θ が既知のときの，z_i が与えられた時の $X_i + \theta Y_i$ の十分性に基づいた方法である．

$Z_i = z$ が与えられた時．統計量 $X_i + \theta Y_i$ はガンマ分布 $Gamma(2, z)$ に従う．

$\because$ $X_i \sim Exp(z) = Gamma(1, z)$. $\theta Y_i \sim \theta Exp(\theta z) = \theta Gamma(1, \theta z) = Gamma(1, z)$. z が与えられたとき，X_i, Y_i は独立と仮定していたので，ガンマ分布の再生性から $X_i + \theta Y_i = Gamma(2, z)$

さて，$X_i + \theta Y_i$ は z_i に対する十分統計量より，フィッシャーの因子分解定理から，(X, Y) の同時分布 $p_\theta(x, y|z)$ は次のように分解できる．

$$p_\theta(x, y|z) = h_\theta(x, y) g_\theta(x + \theta y|z)$$

ただし

$$\begin{aligned} p_\theta(x, y|z) &= ze^{-zx}\theta ze^{-\theta zy} \\ &= \frac{\theta}{x + \theta y} z^2 (x + \theta y) e^{-z(x+\theta y)} \end{aligned}$$

より，

$$h_\theta(x, y) = \frac{\theta}{x + \theta y}, \quad g_\theta(x + \theta y|z) = z^2 s e^{-zs}, \quad \text{where } s = x + \theta y.$$

$g_\theta(s|z)$ は Gamma(2,z) の密度関数である．

$X_i + \theta Y_i$ は z_i に依存するが，z_i は観測できないので，$g_\theta(x + \theta y|z)$ を尤度から無視して，$h_\theta(x, y)$ だけを使うことを考える．これは，部分尤度法と呼ばれる生存時間解析などの分野でセミパラメトリックモデルを扱う際に用いられる方法と類似している．しかし，この" 条件付き尤度" $h_\theta(x, y)$ は尤度とは異なる振る舞いをする．なぜなら，h_θ に対応する，θ を求めるための条件付き尤度方程式

$$\frac{\dot{h}_\theta}{h_\theta}(x, y) = \frac{\partial}{\partial \theta} \log h_\theta(x, y)$$

は真の分布で期待値をとったとき，$\theta > 0$ をどのようにとっても 0 とならないからである．実際

$$\frac{\dot{h}_\theta}{h_\theta}(x, y) = \frac{x}{\theta(x + \theta y)} > 0$$

である．

このバイアスは十分統計量によって条件付けることで修正できる．

$$\psi_\theta(X, Y) = 2\theta \frac{\dot{h}_\theta}{h_\theta}(X, Y) - 2\theta E_\theta \left(\frac{\dot{h}_\theta}{h_\theta}(X, Y) | X + \theta Y \right) = \frac{X - \theta Y}{X + \theta Y}.$$

とする．

$\because$ 2 つ目の等号が成り立つことを示す．

$$2\theta \frac{\dot{h}_\theta}{h_\theta}(X, Y) = \frac{2X}{X + \theta Y}.$$

よって

$$2\theta E_\theta \left(\frac{\dot{h}_\theta}{h_\theta}(X, Y) | X + \theta Y \right) = 1$$

を示せばよい．ここで，$T = T(X,Y) = X + \theta Y$ と書くことにする．すると，

$$\begin{aligned} p_\theta(x,y|T=t) &= p_\theta(x,y|Z=z,T=t) \\ &= \frac{p_\theta(x,y|Z=z)}{Pr(T=t|Z=z)} \\ &= h_\theta(x,y). \end{aligned}$$

よって，

$$\begin{aligned} 2\theta E_\theta\left(\frac{\dot{h}_\theta}{h_\theta}(X,Y)|T=t\right) &= 2\theta\int_0^{t/\theta}\frac{t-\theta y}{\theta t}\frac{\theta}{t}dy \\ &= \frac{2\theta}{t^2}\left[ty - \frac{\theta y^2}{2}\right]_0^{t/\theta} \\ &= \frac{2\theta}{t^2}\left(\frac{t^2}{\theta} - \frac{t^2}{2\theta^2}\right) \\ &= 2\theta\left(\frac{1}{\theta} - \frac{1}{2\theta}\right) = 2\theta\frac{1}{2\theta} = 1. \end{aligned}$$

たしかに，

$$\psi_\theta(X,Y) = \frac{X - \theta Y}{X + \theta Y}.$$

$\hat{\theta}_n$ を $\mathbb{P}_n\psi_\theta = 0$ の解として得る．

ここで，$\dot{\psi}_\theta = \frac{-xy-x}{(x+\theta y)^2} < 0$ であり，$\psi_\theta \to -1.(\theta \to \infty).\psi_\theta \to 1(\theta \to 0)$. より，任意の $x, y > 0$ で，ψ_θ は 1 から -1 に単調減少する．よって $\mathbb{P}_n\psi_\theta = 0$ は唯一解を持つ．このとき，

Lemma 5.10. Θ を実数直線の部分集合とする．Ψ_n をランダム関数，Ψ を，任意の θ で $\Psi_n(\theta) \xrightarrow{\mathrm{P}} \Psi(\theta)$ となるような固定した関数とする．このとき Ψ_n が連続かつ Ψ_n が 0 になるような $\hat{\theta}_n$ が 1 つであると仮定，あるいは Ψ_n が非減少で，$\Psi_n(\hat{\theta}_n) = o_p(1)$ であると仮定する．また，θ_0 を任意の $\forall\epsilon > 0$ に対して $\Psi(\theta_0 - \epsilon) < 0 < \Psi(\theta_0 + \epsilon)$ となるような点であるとするとき，$\hat{\theta}_n$ は θ_0 に確率収束する．

より，$\mathbb{P}_n\psi_\theta = 0$ を解いて求めた $\hat{\theta}_n$ は一致推定量である．さらに

$$P_{\theta_0}\psi_\theta = -\frac{\theta+\theta_0}{\theta-\theta_0} - \frac{2\theta\theta_0}{(\theta-\theta_0)^2}\log\frac{\theta_0}{\theta} = \frac{1}{3\theta_0}(\theta_0-\theta) + o(\theta_0-\theta).$$

となる．2 つ目の等号を示す．まず，$\log\frac{\theta_0}{\theta}$ を 3 次まで θ_0 回りでテーラー展開すると

$$\begin{aligned} \log\frac{\theta_0}{\theta} &= \log\frac{\theta_0}{\theta_0} - \frac{1}{\theta_0}(\theta-\theta_0) + \frac{1}{2\theta_0^2}(\theta-\theta_0)^2 - \frac{1}{3\theta_0^3}(\theta-\theta_0)^3 + o(\theta-\theta_0)^3 \\ &= -\frac{1}{\theta_0}(\theta-\theta_0) + \frac{1}{2\theta_0^2}(\theta-\theta_0)^2 - \frac{1}{3\theta_0^3}(\theta-\theta_0)^3 + o(\theta-\theta_0)^3. \end{aligned}$$

よって，

$$\begin{aligned} P_{\theta_0}\psi_\theta &= -\frac{\theta+\theta_0}{\theta-\theta_0} - \frac{2\theta\theta_0}{(\theta-\theta_0)^2}\log\frac{\theta_0}{\theta} \\ &= -\frac{\theta+\theta_0}{\theta-\theta_0} + \frac{2\theta}{\theta-\theta_0} - \frac{\theta}{\theta_0} + \frac{2\theta}{3\theta_0^2}(\theta-\theta_0) + o(\theta-\theta_0) \\ &= 1 - \frac{\theta}{\theta_0} + \frac{2\theta}{3\theta_0^2}(\theta-\theta_0) + o(\theta-\theta_0) \\ &= \frac{(\theta_0-\theta)(\theta_0+2(\theta_0-\theta))}{3\theta_0^2} + o(\theta-\theta_0) \\ &= \frac{1}{3\theta_0}(\theta_0-\theta) + o(\theta_0-\theta) \end{aligned}$$

となり示せた．このように，$\theta \mapsto P_{\theta_0}\psi_\theta$ は $\theta = \theta_0$ でユニークにゼロとなる．次に，$\sqrt{n}(\hat{\theta}_n - \theta_0)$ の漸近正規性を Theorem 5.21

Theorem 5.21. ユークリッド空間の開部分集合 Θ でノルム $||\cdot||$ が定義されているとする．以下を満たすベクトル値可測関数 $x \mapsto \psi_\theta(x), \theta \in \Theta$ を定義する．

1 可測関数 $\dot{\psi}, P\dot{\psi}^2 < \infty$ が存在し，

$$\|\psi_{\theta_1}(x) - \psi_{\theta_2}(x)\| \le \dot{\psi}(x)\|\theta_1 - \theta_2\|.$$

2 $P\|\psi_{\theta_0}\|^2 < \infty$.

3 $\theta \mapsto P\psi_\theta$ が θ_0 において微分可能であり，θ_0 におけるヤコビアン $V_{\theta_0}^{-1}$ は非特異である．

このとき，$\mathbb{P}_n\psi_{\hat{\theta}_n} = o_P(n^{-1/2})$ かつ $\hat{\theta}_n \xrightarrow{\mathrm{P}} \theta_0$ ならば，次式が成り立つ．

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\psi_{\theta_0}(X_i) + o_P(1)$$

に基づき示す．

まず，$\dot{\psi}_\theta(x,y) = \frac{-x(1+y)}{(x+\theta y)^2}$ は任意の $x, y > 0$ と $(0,\infty)$ 上のコンパクト集合上の θ について一様有界より，平均値の定理から定理の条件 1 の $\dot{\psi}$ を定数としてとれる．他の条件は自明に成り立つ．以上により，Theorem 5.21. から漸近正規性が示せた．

この推定量は計算は容易であるが，第 25 章において漸近的に最適でないことが示せる．つまり，efficiency bounds を達成しない．第 25 章のセミパラメトリックモデルにおいてより分散の小さい推定量について議論する．