Chapter 1

Vogt(2015)의 첫번째 리뷰

1.1 Model setting

(정의) $\{X_{t,T} : t = 1, \ldots, T\}_{T=1}^{\infty}$가 locally stationary라는 것은 모든 $u \in [0,1]$에 대하여 아래식을 만족하는 strictly stationary process $\{X_t(u) : t \in \mathbb{Z}\}$를 찾을 수 있다는 것이다.

$$\|X_{t,T} - X_t(u)\| \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) U_{t,T}(u) \;\; a.s.$$

- 이때 u는 rescaled time 이다. [1]

- 여기에서 t가 고정되었을때 $X_{t,T}$는 d차원 r.v.이고 따라서 $\|\cdot\|$는 $\mathbb{R}^d$에서의 norm이다.

- 또한 $\{U_{t,T} : t = 1, \ldots, T\}_{T=1}^{\infty}$는 $E[U_{t,T}^{\rho}(u)] \leq C < \infty$를 만족하는 어떠한 양의 확률변수이다.

(정의) $\lambda_{t,T}$는 $E(X_{t,T})$혹은 $V(X_{t,T})$와 같이 시간에 따라서 변화하는 locally stationary process $X_{t,T}$ 의 특징이다.

- 여기에서 $\lambda_{t,T}$는 아래와 같은 성질을 만족한다.

[1] 보통 $u = \frac{t}{T}$로 정의하는데 이 정의는 모호한 편이다. 왜냐하면 $T = \infty$인 경우를 따지기 힘들기 때문임.

(1) $\lambda_{t,T}$는 $\{X_{t,T}\}$에서 어떠한 measurable함수 $f : \mathbb{R}^d \to \mathbb{R}$를 취한것의 평균으로 표현된다. 즉

$$\lambda_{t,T} = E(f(X_{t,T}))$$

이다. 여기에서 f들의 집합을 $\mathcal{F}$라고 표현한다.

example: $X_{t,T} = (X_{t,T,1}, \ldots, X_{t,T,d})'$라고 하자. $\lambda_{t,T} = cov(X_{t,T,i}, X_{t,T,j})$이면 $\mathcal{F} := \{f_{ij} : 1 \leq i \leq j \leq d\}$ $where\ f_{ij}(x) = x_i x_j$

(2) $\lambda_{t,T} = E(f(X_{t,T}))$는 아래를 만족한다.

$$\sup_{f \in \mathcal{F}} |E[f(X_{t,T})] - E[f(X_t(u))]| \leq C \left(\left| \frac{1}{T} - u \right| + \frac{1}{T} \right).$$

즉 $\lambda_{t,T}$가 locally하게는 $\lambda(u)$에 수렴한다는 의미이다. 여기에서 $\lambda(u) = E(f(X_t(u)))$

(가정) 우리는 $\lambda(u)$가 $[0, u_0]$에서는 변화하지 않고, u_0이후에서만 변화한다고 가정할 것이다.

(목표) 그리고 우리의 목적은 u_0를 찾는것, 즉 $\lambda(u)$가 시간에 따라 변화하기 시작하는 시점을 찾는 것이다.

1.2 A measure of time-variation

(정의) $D(u, v, f)$는 아래와 같이 정의된다.

$$D(u, v, f) = \int_0^v E(f(X_t(w)))dw - \left(\frac{v}{u}\right) \int_0^u E(f(X_t(w)))dw$$

- 따라서 고정된 f와 u에 대하여

$$\sup_{v\in[0,u]} |D(u,v,f)$$

의 값이 0이라면, $E(f(X_t(w)))$는 구간 $[0,u]$에서 변화가 없다는 의미가 된다.

(정의) 함수 $\mathcal{D} : [0,1] \to \mathbb{R}_{\geq 0}$을 아래와 같이 정의하자.

$$\mathcal{D}(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} |D(u,v,f)|.$$

- 만약에 $X_{t,T}$가 strong stationary process라고 하면, 어떠한 함수 f에 대하여서도

$$\sup_{v\in[0,u]} |D(u,v,f)| = 0.$$

이다. 따라서 이때는 $\mathcal{D}(u) = 0$이 된다.

- 여기에서 $\int_0^v E[f(X_t(w))]dw$는 아래와 같이 추정할 수 있다.

$$\frac{1}{T}\sum_{t=1}^{[vT]} f(X_{t,T}).$$

이때 앞에 $\frac{1}{T}$를 곱해준 이유는 시간축이 $[0,1]$로 scaling되어 있기 때문이다.

- 따라서 $\hat{D}_T(u,v,f)$와 $\hat{\mathcal{D}}_T(u)$는 각각 아래와 같이 추정하는 것이 바람직 하다.

$$\hat{D}_T(u,v,f) = \frac{1}{T}\sum_{t=1}^{[vT]} f(X_{t,T}) - \left(\frac{v}{u}\right)\frac{1}{T}\sum_{t=1}^{[uT]} f(X_{t,T}).$$

$$\hat{\mathcal{D}}_T(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} |\hat{D}(u,v,f)|.$$

1.3 Estimating the gradual change point u_0.

- 일반적으로

$$\hat{\mathcal{D}}_T(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} |\hat{D}(u,v,f)|.$$

는 $u \le u_0$에서는 0에 가까운 값을 가지고 $u > u_0$에서는 어떠한 bounded된 값을 가지게 된다. 즉 $\hat{\mathcal{D}}_T(u)$는 아래와 같은 느낌이다.

$$\hat{\mathcal{D}}_T(u) = \begin{cases} \xrightarrow{p} 0, & u \le u_0, \\ \le C < \infty, & u > u_0. \end{cases}$$

- 또한 $\sqrt{T}\hat{\mathcal{D}}_T(u)$는 대충 아래와 같은 느낌을 가지게 된다.

$$\sqrt{T}\hat{\mathcal{D}}_T(u) = \begin{cases} \xrightarrow{d} \mathcal{H}(u), & u \le u_0, \\ \xrightarrow{p} \infty, & u > u_0. \end{cases}$$

- $\sqrt{T}\hat{\mathcal{D}}_T(u)$가 어떠한 분포를 가지지 않고 ∞값을 가지게 되는 시점을 찾기 위해서 아래와 같은 성질을 만족하는 τ_T를 도입하자.

$$\begin{cases} P(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) \to 1, & u \le u_0 \\ P(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) \to 0, & u > u_0 \end{cases}$$

note: 이를 위해서는 적당히 τ_T를 $\mathcal{H}(u)$의 분포에서 나올수 있는 최대값으로 정의하는것이 바람직하다. 즉 $h_1, \dots, h_T$가 분포 $\mathcal{H}(u)$에서 나온 샘플이라고 한다면 대충

$$\tau_T \approx \max\{h_1, \dots, h_T\}$$

와 같이 정의하는 것이 바람직하다. 이렇게 정의하면 자연스럽게 τ_T는 T가 커질수록 무한대로 (천천히) 발산하게 됨을 이해할 수 있다.

- 위의 식을 다시 쓰면

$$\begin{cases} P(1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) = 1) \to 1, & u \le u_0 \\ P(1(\sqrt{T}\hat{\mathcal{D}}_T(u) \le \tau_T) = 1) \to 0, & u > u_0 \end{cases}$$

와 같이 쓸 수 있고 여기에서 $\hat{r}_T(u) := 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \leq \tau_T)$라고 정의하면

$$\hat{r}_T(u) \xrightarrow{p} \begin{cases} 1, & u \leq u_0 \\ 0, & u > u_0 \end{cases}$$

와 같이 된다. 따라서 $\int_0^1 \hat{r}_T(u)du$를 이용하여 u_0를 추정할 수 있다. 즉

$$\hat{u}_0(\tau_T) = \int_0^1 \hat{r}_T(u)du$$

와 같이 쓸 수 있다.

1.4 Asymptotic properties.

(약속) $\ell_\infty(S)$를 supremum norm이 bound된 $f : S \to \mathbb{R}$들의 집합으로 정의하자.

(약속) 아래를 가정하자.

$$\frac{\mathcal{D}(u)}{(u-u_0)^\kappa} \longrightarrow c_\kappa > 0 \ \ as\ u \downarrow u_0$$

여기에서 κ는 자연수이고 c_κ는 양수이다.

- 여기에서 큰 수의 κ는 $\mathcal{D}$가 u_0시점 이후로 좀 더 천천히 0으로부터 멀어짐을 의미한다. 즉 κ는 $\mathcal{D}$의 smoothness의 정도를 알려주는 파라메터이다.

note: $u > u_0$에서 $u - u_0$는 1보다 작은 양수이므로, κ가 클수록 $(u-u_0)^\kappa$는 매우 천천히 변화한다.

1.4.1 Assumption.

- 내용이없어서..

1.4.2 Weak convergence of the measure of time-variation.

- 아래와 같은 통계량을 정의하자.

$$\hat{H}_T(u,v,f) = \sqrt{T}\left(\hat{D}_T(u,v,f) - D(u,v,f)\right)$$

여기에서 $\hat{H}$, $\hat{D}_T$, D는 모두 공간 $\Delta \times \mathcal{F}$에서 $\mathbb{R}$로 가는 함수이다. 이때 $\Delta = \{(u,v) \in [0,1]^2 : v \leq u\}$로 정의한다.

- 추가적으로 $\hat{H}_T$는 $\ell_\infty(\Delta \times \mathcal{F})$의 원소라고 가정하자. 즉 $\hat{H}_T$의 sup이 bound되어있다고 가정하자.

note: 생각해보면 이 가정은 매우 특이하다. $X_{t,T}$가 Ω에서 정의된 확률과정이듯이 $\hat{H}_T$는 $\Omega \times \Delta \times \mathcal{F}$에서 정의되는 일종의 확률변수로 볼 수 있는데, 이 확률변수의 최대값(=sup)이 bound되어 있다는 의미이기 때문이다.

(정리) 적당한 가정하에서 통계량 $\hat{H}_T(u,v,f)$의 분포가 평균이 0이고 분산이

$$Cov(H(u,v,f), H(u',v',f'))$$

인 Gaussian process H로 수렴한다. 즉,

$$\hat{H}_T \xrightarrow{d} H$$

이다.

- $\hat{H}_T(u,v,f)$와 $\hat{H}_T(u',v',f')$는 모두 $X_{t,T}$로부터 유도된 확률변수이지만 엄연히 서로 다른 확률변수이다.

note: X가 r.v.라고 할때 $Y_1 = f_1(X)$, $Y_2 = f_2(X)$과 같은 느낌

- 따라서 두 확률변수의 분산을 구할수 있는데 이를 아래와 같이 정의한다.

$$Cov(H(u,v,f), H(u',v',f')).$$

(정의) 확률변수 $\hat{\mathcal{D}}_T(u)$, $\hat{\mathcal{H}}_T(u)$를 아래와 같이 정의하자.

$$\hat{\mathcal{D}}_T(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} |\hat{D}_T(u,v,f)|.$$

$$\hat{\mathcal{H}}_T(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} |\hat{H}_T(u,v,f)|.$$

즉 $\hat{\mathcal{D}}_T(u)$는 고정된 u에 대하여 확률변수 $\hat{D}_T(u,v,f)$가 취할수 있는 값 중의 가장 큰 값을 의미하고, $\hat{\mathcal{H}}_T(u)$역시 그러하다.

note: $\hat{\mathcal{D}}_T(u)$는 $\Omega \times [0,1]$에서 정의된 확률변수라고 볼 수 있고 $\hat{\mathcal{H}}_T(u)$역시 그러하다.

note: 만약에 $u = 0.6$으로 이라고 한다면 $\hat{\mathcal{D}}_T(0.6)$은 아래와 같은 확률변수들 중에서 최대값이다.

$$\hat{D}_T(0.6, 0.5, f_1)$$

$$\hat{D}_T(0.6, 0.4, f_1)$$

$$\hat{D}_T(0.6, 0.3, f_1)$$
$$\hat{D}_T(0.6, 0.2, f_1)$$
$$\dots$$
$$\hat{D}_T(0.6, 0.5, f_2)$$
$$\hat{D}_T(0.6, 0.4, f_2)$$
$$\hat{D}_T(0.6, 0.3, f_2)$$
$$\hat{D}_T(0.6, 0.2, f_2)$$
$$\dots$$

(정의) 확률변수 $\hat{\mathbb{D}}_T(u)$, $\hat{\mathbb{H}}_T(u)$를 아래와 같이 정의하자.

$$\hat{\mathbb{D}}_T(u) = \sup_{f \in \mathcal{F}} \sup_{0 \le w \le v \le u} |\hat{D}_T(v, w, f)|.$$

$$\hat{\mathbb{H}}_T(u) = \sup_{f \in \mathcal{F}} \sup_{0 \le w \le v \le u} |\hat{H}_T(v, w, f)|.$$

note: 만약에 $u = 0.6$으로 이라고 한다면 $\hat{\mathbb{D}}_T(0.6)$은 아래와 같은 확률변수들 중에서 최대값이다.

$$\hat{D}_T(0.6, 0.5, f_1)$$
$$\hat{D}_T(0.6, 0.4, f_1)$$
$$\hat{D}_T(0.6, 0.3, f_1)$$
$$\hat{D}_T(0.6, 0.2, f_1)$$
$$\dots$$
$$\hat{D}_T(0.4, 0.2, f_1)$$
$$\hat{D}_T(0.4, 0.1, f_1)$$

$$\cdots$$
$$\hat{D}_T(0.3, 0.2, f_2)$$
$$\hat{D}_T(0.3, 0.1, f_2)$$
$$\cdots$$

note: 따라서 자연스럽게 $\hat{\mathbb{D}}_T(u) \geq \hat{\mathcal{D}}_T(u)$임을 알 수 있다.

- 아래와 같이 정의하기도 한다.

$$\hat{\mathbb{D}}_T(u) = \sup_{v \in [0,u]} |\hat{\mathcal{D}}_T(v)|.$$

$$\hat{\mathbb{H}}_T(u) = \sup_{v \in [0,u]} |\hat{\mathcal{H}}_T(v)|.$$

(정리) 적당한 조건하에서 H_T가 H로 분포수렴한다는 정리를 이용하면 쉽게 아래를 보일 수 있다.

$$\sup_{f \in \mathcal{F}} \sup_{v \in [0,u]} |\hat{H}_T(u, v, f)| \xrightarrow{d} \sup_{f \in \mathcal{F}} \sup_{v \in [0,u]} |H(u, v, f)|.$$

$$\sup_{f \in \mathcal{F}} \sup_{0 \leq w \leq v \leq u} |\hat{H}_T(v, w, f)| \xrightarrow{d} \sup_{f \in \mathcal{F}} \sup_{0 \leq w \leq v \leq u} |H(v, w, f)|.$$

1.4.3 Convergence of the estimator $\hat{u}_0(\tau_T)$.

- 앞서서 $\hat{\mathcal{D}}$는 아래와 같은 느낌이라고 하였다.

$$\hat{\mathcal{D}}_T(u) = \begin{cases} \xrightarrow{p} 0, & u \leq u_0, \\ \leq C < \infty, & u > u_0. \end{cases}$$

그런데

$$\begin{cases} P(\sqrt{T}\hat{\mathcal{D}}_T(u) \leq \tau_T) \to 1, & u \leq u_0 \\ P(\sqrt{T}\hat{\mathcal{D}}_T(u) \leq \tau_T) \to 0, & u > u_0 \end{cases}$$

이므로 아래가 성립한다.

$$\begin{cases} P(\hat{\mathcal{D}}_T(u) \leq \frac{\tau_T}{\sqrt{T}}) \to 1, & u \leq u_0 \\ P(\hat{\mathcal{D}}_T(u) \leq \frac{\tau_T}{\sqrt{T}}) \to 0, & u > u_0 \end{cases}$$

(약속) 이런 사실을 종합하면 τ_T가 아래의 성질을 만족해야 함을 유추할 수 있다.

(i) $\tau_T \longrightarrow \infty$

(ii) $\tau_T/\sqrt{T} \longrightarrow 0$

따라서 앞으로는 특별한 언급이 없어도 τ_T가 위의 두 성질을 만족한다는 것으로 생각하자.

(정의) γ_T를 아래와 같이 정의하자.

$$\gamma_T = \left(\frac{\tau_T}{\sqrt{T}}\right)^{1/\kappa}$$

note: 이때 κ는 앞에서 약속한것 처럼 다음식을 만족시켜야 한다.

$$\frac{\mathcal{D}(u)}{(u-u_0)^\kappa} \longrightarrow c_\kappa > 0 \;\; as \; u \downarrow u_0$$

- $(\gamma_T)^\kappa$는 T가 커짐에 따라 0으로 수렴한다. 따라서 γ_T역시 T가 커짐에 따라서 0으로 수렴한다. 동일한 κ일경우 굳이 따지면 $(\gamma_T)^\kappa$이 γ_T보다 빠르게 수렴한다.

- 그런데 $P(\hat{\mathcal{D}}_T(u) \leq \frac{\tau_T}{\sqrt{T}}) \to 1$임을 생각하면 대충 T가 무한대로 감에 따라 $\mathcal{D}(u) \leq (\gamma_T)^\kappa$와 같은 식이 성립한다고 생각할 수 있다. 또한

$$\frac{\mathcal{D}(u)}{(u-u_0)^\kappa} \longrightarrow c_\kappa > 0 \ \ as\ u \downarrow u_0$$

임을 감안하면, 아래식이 성립하는 느낌이 든다.

$$\frac{(\gamma_T)^\kappa}{(u-u_0)^\kappa} \longrightarrow c_\kappa > 0 \ \ as\ u \downarrow u_0,\ T \to \infty.$$

즉 $T \to \infty$일때 $(\gamma_T)^\kappa$가 0으로 가는 속도는 $u \downarrow u_0$일때 $(u-u_0)^\kappa$가 0으로 가는 속도와 비슷하다. 즉 γ_T는 적당하게 $u-u_0$가 0으로 가는 속도와 비슷하도록 κ와 τ_T를 조절하여 얻을 수 있다. 기본적으로 γ_T의 값은 κ와 τ_T에 따라서 아래와 같은 규칙으로 변화한다.

(1) κ가 커질수록, 즉 $\mathcal{D}$의 부드러움의 정도가 클수록 γ_T는 천천히 0으로 수렴한다.

(2) τ_T가 천천히 무한대로 갈수록, 즉 $\sqrt{T}\hat{\mathcal{D}}_T$의 극한분포 $\mathcal{H}$에서 나오는 최대값을 좀더 타이트하게 잡을수록 γ_T는 빠르게 0으로 수렴한다.

(정리) 적당한 조건하에서 아래가 성립한다.

$$\hat{u}_0(\tau_T) - u_0 = O_p(\gamma_T).$$

- γ_T는 기본적으로 $\mathcal{D}$의 부드러움의 정도가 클수록 천천히 0으로 수렴하기 때문에 $\hat{u}_0(\tau_T)$이 보다 정확한 추정량이 되기 위해서는 u_0에서 파라메터가 서서히 변화하는 모델보다 파라메터가 급격하게 변화하는 모델이 유리함을 알 수 있다.

- 또한 τ_T가 좀 더 천천히 발산할수록 γ_T를 보다 빠르게 0으로 수렴시킬수 있기 때문에 $\sqrt{T}\hat{\mathcal{D}}_T$의 극한분포 $\mathcal{H}$에서 나오는 최대값을 좀더 타이트 하게 잡을수록 u_0를 정확하게 추정할 수 있음을 알 수 있다.

1.4.4 Some mean squared error considerations.

1.4.5 Choice of the threshold level τ_T

(약속) $q_\alpha(u)$를 $\mathbb{H}(u)$의 $1-\alpha$-분위수라고 약속하자.

(약속) 모든 $u \in [0,1]$에 대하여 $\mathbb{H}(u)$의 분포를 알고있다고 약속하자.

note: 실제로는 분포를 몰라서 적당한 방법을 통하여 추론을 해야한다. 이런 방법은 다음장(Implementation)에서 자세히 다룸.)

- 참고로 $u < u_0$인곳에서 $\mathbb{H}(u) = \sqrt{T}\mathbb{D}(u)$임을 기억하자.

- 또한 $\mathbb{H}(u)$의 분포는 고정된 u에 대해서 정규확률과정 $H(u,v,f)$에서 나오는 확률변수들의 최대값들의 분포를 의미한다.

Preliminary choice of τ_T

- H_0를 "변화점이 없다."로 설정하자. 우선 우리의 첫번째 목표는 귀무가설하에서

$$P(\hat{u}_0(\tau_T) < 1) = \alpha$$

가 되도록 하는 어떠한 τ_T를 구하는 것이다.

- 우선 τ_T가 T에 depend하지 않는 어떠한 상수 τ라고 하자. 즉 모든 T에 대하여 $\tau_T = \tau$이다.

- 아래식이 성립한다.

$$P(\hat{u}_0(\tau) < 1) \leq P(\sqrt{T}\hat{\mathcal{D}}_T(u) > \tau, \text{ for some } u \in [0,1]) = P(\sqrt{T}\mathbb{D}_T(1) > \tau)$$

그런데 changing point가 없다고 하였으므로

$$P(\sqrt{T}\mathbb{D}_T(1) > \tau) \longrightarrow P(\mathbb{H}(1) > \tau)$$

가 된다. $\mathbb{H}(1)$의 분포를 완전히 알고 있다고 가정하였으므로 그것의 $1-\alpha$-분위수 $q_\alpha(1)$를 정확하게 알 수 있다. 이것을 τ로 잡자. 즉 $\tau = q_\alpha(1)$로 잡자. 그러면,

$$P(\hat{u}_0(\tau) < 1) \leq \alpha + o(1)$$

이 된다. 이때의 τ를 τ_α°이라고 하자. 즉 $\tau_\alpha^\circ = q_\alpha(1)$. 그러면 위의식을 아래와 같이 다시 쓸 수 있다.

$$P(\hat{u}_0(\tau_\alpha^\circ) < 1) \leq \alpha + o(1)$$

- 하지만 τ_α°는 앞에서 가정한것 처럼 T가 커짐에 따라서 (천천히) 무한대로 발산하는 수가 아니라 T에 영향을 받지 않는 어떠한 상수이다. 따라서 $\hat{u}_0(\tau_\alpha^\circ)$는 consistent estimator가 될 수 없다. $\hat{u}_0(\tau_\alpha^\circ)$를 consitent estimator가 되게 하려면 τ_α°를 무한대로 천천히 발산시켜야 한다. 그 방법은 T가 커짐에 따라서 α를 0으로 보내야 한다. 즉 아래와 같은 수열을 정의해야 한다.

$$\alpha_T \to 0, \ as \ T \to \infty$$

여기에서 α_T를 충분히 천천히 0으로 보내야 함을 이해해야 한다. α_T를 빠르게 0으로 보내면 어떤일이 생길까? 그렇다면 τ_T가 $\sqrt{T}$보다 빠르게 무한대로 발산할 수도 있을 것이고 이것은 적절한 τ_T에 대한 선택이 아니다.

- 얼마나 천천히 α_T를 0으로 보내야 하는 것일까? 만족시켜야 할것은

$$\frac{\tau_{\alpha_T}^\circ}{T} \to 0$$

이어야 한다는 것이다. 그런데 이것은 적당한 c와 r에 대하여 $\alpha_T \leq cT^{-r}$를 만족하는 α_T만 잡으면 된다고 논문에 나와있다.

- 실제로 T는 고정되어 있으므로, 천천히 수렴하는 수열 $\{\alpha_T\}$를 잡는다느니 하는 말은 다 이론용이다. 실제로는 하나의 α_T만 정해주면 된다.

Refined choice of τ_T

- 귀무가설을 살짝 수정하여 u_0이전까지는 변화점이 없다고 하자.

- 편의상 u_0를 알고 있다고 생각하자.

- 아래식이 성립한다.

$$P(\hat{u}_0(\tau) < u_0) \leq P(\sqrt{T}\hat{\mathcal{D}}_T(u) > \tau, \text{ for some } u \in [0, u_0]) = P(\sqrt{T}\mathbb{D}_T(u_0) > \tau)$$

그런데 u_0까진 changing point가 없다고 하였으므로

$$P(\sqrt{T}\mathbb{D}_T(u_0) > \tau) \longrightarrow P(\mathbb{H}(u_0) > \tau)$$

가 된다. $\mathbb{H}(u_0)$의 분포를 완전히 알고 있다고 가정하였으므로 그것의 $1-\alpha$분위수 $q_\alpha(u_0)$를 정확하게 알 수 있다. 이것을 τ_α로 잡자. 그러면 아래가 성립한다.

$$P(\hat{u}_0(\tau_\alpha) < u_0) \leq \alpha + o(1)$$

- 추가적으로 아래도 성립함을 논문에서 증명하였다.

$$P(\hat{u}_0(\tau_\alpha) > u_0 + C\gamma_T) = o(1)$$

(정리) 적당한 가정하에서 아래식들이 성립한다.

$$P(\hat{u}_0(\tau_\alpha) < u_0) \leq \alpha + o(1)$$

$$P(\hat{u}_0(\tau_\alpha) > u_0 + C\gamma_T) = o(1)$$

여기에서 $C > 0$이고, $\gamma_T = \left(\frac{\tau_T}{\sqrt{T}}\right)^{1/\kappa}$이다.

- 실제로는 u_0를 모르기 때문에 τ_α, 혹은 $\{\tau_{\alpha_T}\}$를 잡을 수 없다. 따라서 아래의 과정을 통하여 u_0의 추정치혹은 추정치열을 구하고 그것을 이용하여 τ_α혹은 $\{\tau_{\alpha_T}\}$를 잡아야 한다.

(i) 0으로 천천히 수렴하는 적당한 수열 $\{\alpha_T\}$를 잡는다. 여기에서 α_T는 적당한 c와 r에 대하여 $\alpha_T \leq cT^{-r}$를 만족하기만 하면 된다.

(ii) $\mathbb{H}(1)$의 분포를 알고 있으므로 $\mathbb{H}(1)$의 $1-\alpha_T$분위수에 해당하는 값들을 각각 구하고 이를 $\{\tau^\circ_{\alpha_T}\}$로 정의한다. (1차선택!!)

(iii) 각각의 $\{\tau^\circ_{\alpha_T}\}$에 대하여 $\hat{\mu}_0(\tau^\circ_{\alpha_T})$를 계산한다. 아래식을 써야한다.

$$\hat{\mu}_0(\tau^\circ_{\alpha_T}) = \int_0^1 1(\sqrt{T}\hat{\mathcal{D}}_T(u) \leq \tau^\circ_{\alpha_T})du.$$

여기에서

$$\hat{\mathcal{D}}_T(u) = \sup_{f\in\mathcal{F}} \sup_{v\in[0,u]} \left| \frac{1}{T}\sum_{t=1}^{[vT]} f(X_{t,T}) - \left(\frac{v}{u}\right)\frac{1}{T}\sum_{t=1}^{[uT]} f(X_{t,T}) \right|$$

이다. 따라서 이 과정을 계산할때 주어진 샘플들, 즉 $\{X_{t,T}\}$이 사용된다.

(iv) $\hat{\mu}_0(\tau^\circ_{\alpha_T})$를 값을 넣어서 $\mathbb{H}(\hat{\mu}_0(\tau^\circ_{\alpha_T}))$의 분포를 구하고 또 그것의 $1-\alpha_T$분위수에 해당하는 값들을 각각 구하고 이를 $\{\hat{\tau}^\circ_{\alpha_T}\}$로 정의한다. 이것이 $\{\tau_T\}$이다. (2차선택!!)

(정리) τ_α대신에 $\hat{\tau}_\alpha$를 넣어도 위의 정리가 성립한다. 즉 적당한 가정하에서 아래식들이 성립한다.

$$P(\hat{u}_0(\hat{\tau}_\alpha) < u_0) \leq \alpha + o(1)$$

$$P(\hat{u}_0(\hat{\tau}_\alpha) > u_0 + C\gamma_T) = o(1)$$

여기에서 $C > 0$이고, $\gamma_T = \left(\frac{\tau_T}{\sqrt{T}}\right)^{1/\kappa}$이다.

Chapter 2

Vogt(2015)의 두번째 리뷰

2.1 Introduction

- goal

2.2 Model setting

- 시계열 $\{X_t\}$ 가 locally stationary process 라고 하자.
- 우리는 $\{X_t\}$ 에서 생성된 하나의 realization $\{x_t\}$ 를 관측했다고 하자.
- Let λ_t be some time-varying feature of X_t.

example: $\lambda_t = E(X_t)$.

example: $\lambda_t = V(X_t)$.

note: 만약에 $\{X_t\}$가 정상이라면 λ_t는 항상 상수이어야 할것이다.

- λ_t는 구체적으로 (1) 시계열 $\{X_t\}$ 에서 (2) 임의의 measurable function $f \in \mathcal{F}$를 취한 후 (3) 그것의 평균으로 정의된다. 즉 아래와 같은 형태를 가져야한다.

$$\lambda_t = E(f(X_t)), \quad f \in \mathcal{F}$$

example: $\lambda_t = E(X_t)$인 경우 $\mathcal{F} = \{\text{id}\}$.

- 관측한 시계열이 univariate timeseries 이지만 λ_t가 벡터일수도 있다. (이때는 univariate 를 관측했지만 multivariate를 관측했다고 가정한다) 아래는 그 예제이다.

example: $\{Y_t\}$가 univariate timeseries라고 하자. 우리는 아래에 관심이 있다고 하자.

$$\boldsymbol{\gamma} = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \dots \\ \gamma_p \end{bmatrix}$$

이럴때는

$$\boldsymbol{\lambda}_t = \boldsymbol{\gamma}$$

로 설정한다. 여기에서 $\gamma_\ell = Cov(X_t, X_{t-\ell})$을 의미함.

Bibliography

[1] Vogt, M., & Dette, H. (2015). Detecting gradual changes in locally stationary processes. The Annals of Statistics, 43(2), 713-740.