Edgeworth expansion for the pre-averaging estimator

Bezirgen Veliyev (Aarhus)
jointly with Mark Podolskij (Aarhus), Nakahiro Yoshida (Tokyo)

Dynstoch 2014, Warwick, September 11, 2014

Table of Contents

What is Edgeworth expansion?

- Let $(X_i)_{i=1}^{\infty}$ be i.i.d. rv with mean μ and variance σ^2.

What is Edgeworth expansion?

- Let $(X_i)_{i=1}^{\infty}$ be i.i.d. rv with mean μ and variance σ^2. Recall the CLT:

$$\text{If } S_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu), \text{ then } F_n(x) = \mathbb{P}[S_n \leq x] \to \Phi(x).$$

What is Edgeworth expansion?

- Let $(X_i)_{i=1}^{\infty}$ be i.i.d. rv with mean μ and variance σ^2.
  Recall the CLT:

$$\text{If } S_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu), \text{ then } F_n(x) = \mathbb{P}[S_n \leq x] \to \Phi(x).$$

- Edgeworth expansion for distributions:

$$F_n(x) = \Phi(x) + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(1 - x^2)\phi(x) + o\left(\frac{1}{\sqrt{n}}\right).$$

What is Edgeworth expansion?

- Let $(X_i)_{i=1}^{\infty}$ be i.i.d. rv with mean μ and variance σ^2. Recall the CLT:

$$\text{If } S_n = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{n}(X_i - \mu), \text{ then } F_n(x) = \mathbb{P}[S_n \leq x] \to \Phi(x).$$

- Edgeworth expansion for distributions:

$$F_n(x) = \Phi(x) + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(1 - x^2)\phi(x) + o\left(\frac{1}{\sqrt{n}}\right).$$

- Edgeworth expansion for densities:

$$f_n(x) = \phi(x) + \frac{\kappa_3}{6\sigma^3\sqrt{n}}(x^3 - 3x)\phi(x) + o\left(\frac{1}{\sqrt{n}}\right).$$

Volatility Estimation

- On $[0, 1]$, we consider a continuous semimartingale

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

where μ_t, σ_t are continuous processes.

Volatility Estimation

- On $[0,1]$, we consider a continuous semimartingale

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

  where μ_t, σ_t are continuous processes.
- Observations at points $i/n,\ 0 \leq i \leq n$.

Volatility Estimation

- On $[0, 1]$, we consider a continuous semimartingale

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

  where μ_t, σ_t are continuous processes.
- Observations at points $i/n,\ 0 \leq i \leq n$.
- High frequency (infill asymptotics): $n \to \infty$.

Volatility Estimation

- On $[0,1]$, we consider a continuous semimartingale

$$dX_t = \mu_t dt + \sigma_t dW_t.$$

  where μ_t, σ_t are continuous processes.
- Observations at points $i/n,\ 0 \leq i \leq n$.
- High frequency (infill asymptotics): $n \to \infty$.
- Object of interest: quadratic variation (integrated volatility)

$$V = \int_0^1 \sigma_t^2 dt.$$

Realized Volatility

- Estimator: Realized volatility (realized variance)

$$RV_n = \sum_{i=1}^{n} \left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right)^2 \xrightarrow{\mathbb{P}} V.$$

Realized Volatility

- Estimator: Realized volatility (realized variance)

$$RV_n = \sum_{i=1}^{n} \left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right)^2 \xrightarrow{\mathbb{P}} V.$$

- Stable CLT:

$$Z_n = \sqrt{n}(RV_n - V) \xrightarrow{d_{st}} M \sim MN\left(0, 2\int_0^1 \sigma_t^4 dt\right).$$

Realized Volatility

- Estimator: Realized volatility (realized variance)

$$RV_n = \sum_{i=1}^{n} \left(X_{\frac{i}{n}} - X_{\frac{i-1}{n}} \right)^2 \xrightarrow{\mathbb{P}} V.$$

- Stable CLT:

$$Z_n = \sqrt{n}(RV_n - V) \xrightarrow{d_{st}} M \sim MN\left(0, 2\int_0^1 \sigma_t^4 dt\right).$$

- Empirical studies show that when applied to real data RV_n diverges as $n \to \infty$.

Microstructure noise model

- Observations

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}},\ 0 \leq i \leq n$$

where $(\varepsilon_{\frac{i}{n}})_{i\geq 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0,\ \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.

Microstructure noise model

- Observations

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}},\ 0 \le i \le n$$

where $(\varepsilon_{\frac{i}{n}})_{i\ge 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0,\ \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.

- Balancing the noise with blocks:

Microstructure noise model

- Observations

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}},\ 0 \le i \le n$$

where $(\varepsilon_{\frac{i}{n}})_{i\geq 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0,\ \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.

- Balancing the noise with blocks:
  - Size of blocks $k_n \sim \theta\sqrt{n}$, where $\theta > 0$ is a tuning parameter.

Microstructure noise model

- Observations
$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}},\ 0 \le i \le n$$
where $(\varepsilon_{\frac{i}{n}})_{i\geq 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0,\ \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.
- Balancing the noise with blocks:
  - Size of blocks $k_n \sim \theta\sqrt{n}$, where $\theta > 0$ is a tuning parameter.
  - Number of blocks $d_n = \frac{n}{k_n}$

Microstructure noise model

- Observations
$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}}, \ 0 \le i \le n$$
where $(\varepsilon_{\frac{i}{n}})_{i \ge 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0, \ \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.
- Balancing the noise with blocks:
  - Size of blocks $k_n \sim \theta\sqrt{n}$, where $\theta > 0$ is a tuning parameter.
  - Number of blocks $d_n = \frac{n}{k_n}$
  - g is a weight function, e.g. $g(x) = \min(x, 1-x)$ on $[0, 1]$.

Microstructure noise model

- Observations

$$Y_{\frac{i}{n}} = X_{\frac{i}{n}} + \varepsilon_{\frac{i}{n}}, \; 0 \le i \le n$$

where $(\varepsilon_{\frac{i}{n}})_{i\geq 1}$ is i.i.d., $\mathbb{E}[\varepsilon_0] = 0, \; \mathbb{E}[\varepsilon_0^2] = \omega^2$ and $\varepsilon \perp X$.

- Balancing the noise with blocks:
  - Size of blocks $k_n \sim \theta\sqrt{n}$, where $\theta > 0$ is a tuning parameter.
  - Number of blocks $d_n = \frac{n}{k_n}$
  - g is a weight function, e.g. $g(x) = \min(x, 1-x)$ on $[0, 1]$.
- Pre-averaging:

$$\bar{Y}_{\frac{ik_n}{n}} = \sum_{j=1}^{k_n-1} g\left(\frac{j}{k_n}\right)\left(Y_{\frac{ik_n+j}{n}} - Y_{\frac{ik_n+j-1}{n}}\right).$$

Pre-averaging estimator

- Define

$$V_n = \frac{1}{\psi_2^n} \sum_{i=0}^{d_n-1} \left(\bar{Y}_{\frac{ik_n}{n}} \right)^2 - \frac{\psi_1^n}{2(k_n)^2 \psi_2^n} \sum_{i=1}^{n} \left(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \right)^2,$$

where ψ_1^n and ψ_2^n depend on g. Non-overlappping blocks.

Pre-averaging estimator

- Define

$$V_n = \frac{1}{\psi_2^n} \sum_{i=0}^{d_n - 1} \left(\bar{Y}_{\frac{ik_n}{n}} \right)^2 - \frac{\psi_1^n}{2(k_n)^2 \psi_2^n} \sum_{i=1}^{n} \left(Y_{\frac{i}{n}} - Y_{\frac{i-1}{n}} \right)^2,$$

  where ψ_1^n and ψ_2^n depend on g. Non-overlappping blocks.

- Stable CLT

$$Z_n = n^{1/4}(V_n - V) \xrightarrow{d_{st}} M \sim MN(0, C),$$

$$C = \int_0^1 2\theta \left(\sigma_t^2 + \frac{\omega^2 \psi_1}{\theta^2 \psi_2} \right)^2 dt.$$

Feasible CLT

- Consistent and positive estimator for C :

$$F_n = \frac{2\sqrt{n}}{3(\psi_2^n)^2} \sum_{i=0}^{d_n-1} (\bar{Y}_{\frac{ik_n}{n}})^4 \xrightarrow{\mathbb{P}} C.$$

Feasible CLT

- Consistent and positive estimator for C :

$$F_n = \frac{2\sqrt{n}}{3(\psi_2^n)^2} \sum_{i=0}^{d_n-1} (\bar{Y}_{\frac{ik_n}{n}})^4 \xrightarrow{\mathbb{P}} C.$$

- Hence, we obtain

$$Z_n/\sqrt{F_n} \xrightarrow{d} N(0,1).$$

Feasible CLT

- Consistent and positive estimator for C :

$$F_n = \frac{2\sqrt{n}}{3(\psi_2^n)^2} \sum_{i=0}^{d_n-1} (\bar{Y}_{\frac{ik_n}{n}})^4 \xrightarrow{\mathbb{P}} C.$$

- Hence, we obtain

$$Z_n/\sqrt{F_n} \xrightarrow{d} N(0,1).$$

- Our aim: Edgeworth expansion of

$$Z_n/\sqrt{F_n}.$$

Decomposition

- Suppose that

$$Z_n = M_n + r_n N_n,$$

where N_n is tight, $r_n \to 0$ and M_n is a terminal value of a continuous martingale $(M_t^n)_{t\in[0,1]}$ with $M_0^n = 0$.

Decomposition

- Suppose that

$$Z_n = M_n + r_n N_n,$$

  where N_n is tight, $r_n \to 0$ and M_n is a terminal value of a continuous martingale $(M_t^n)_{t\in[0,1]}$ with $M_0^n = 0$.
- Asymptotic expansion of (Z_n, F_n).

Decomposition

- Suppose that

$$Z_n = M_n + r_n N_n,$$

  where N_n is tight, $r_n \to 0$ and M_n is a terminal value of a continuous martingale $(M_t^n)_{t\in[0,1]}$ with $M_0^n = 0$.
- Asymptotic expansion of (Z_n, F_n).
- Characteristic function of (Z_n, F_n) :

$$\mathbb{E}\left[e^{\mathrm{i}uZ_n+\mathrm{i}vF_n}\right].$$

Random Symbol $\underline{\sigma}$

- Let $C_n = \langle M^n \rangle_1$. Suppose that $C_n \xrightarrow{\mathbb{P}} C$ and $F_n \xrightarrow{\mathbb{P}} C$ hold. Denote $\widehat{C}_n = r_n^{-1}(C_n - C)$ and $\widehat{F}_n = r_n^{-1}(F_n - C)$.

Random Symbol $\underline{\sigma}$

- Let $C_n = \langle M^n \rangle_1$. Suppose that $C_n \xrightarrow{\mathbb{P}} C$ and $F_n \xrightarrow{\mathbb{P}} C$ hold. Denote $\widehat{C}_n = r_n^{-1}(C_n - C)$ and $\widehat{F}_n = r_n^{-1}(F_n - C)$.
- **Assumption 1:**

$$(M_t^n, N_n, \widehat{C}_n, \widehat{F}_n) \xrightarrow{d_{st}} (M_t, N, \widehat{C}, \widehat{F}) \sim MN(\mu, \Sigma).$$

Random Symbol $\underline{\sigma}$

- Let $C_n = \langle M^n \rangle_1$. Suppose that $C_n \xrightarrow{\mathbb{P}} C$ and $F_n \xrightarrow{\mathbb{P}} C$ hold. Denote $\widehat{C}_n = r_n^{-1}(C_n - C)$ and $\widehat{F}_n = r_n^{-1}(F_n - C)$.
- **Assumption 1:**

$$(M_t^n, N_n, \widehat{C}_n, \widehat{F}_n) \xrightarrow{d_{st}} (M_t, N, \widehat{C}, \widehat{F}) \sim MN(\mu, \Sigma).$$

- Define

$$\underline{\sigma}(z, \mathrm{i}u, \mathrm{i}v) = \frac{(\mathrm{i}u)^2}{2}\widetilde{C}(z) + \mathrm{i}u\widetilde{N}(z) + \mathrm{i}v\widetilde{F}(z)$$

It is a random polynomial in $(\mathrm{i}u, \mathrm{i}v)$. It was already in (Yos97).

Random Symbol $\overline{\sigma}$

- Let $e_t^n = e^{\mathrm{i}uM_t^n + \frac{u^2}{2}C_t^n}$

Random Symbol $\overline{\sigma}$

- Let $e_t^n = e^{\mathrm{i}uM_t^n + \frac{u^2}{2}C_t^n}$
- Define

$$\Phi_n(u,v) = \mathbb{E}[\Psi(u,v)(e_1^n(u) - 1)\psi_n]$$

with a truncation functional $\psi_n \sim 1$.

Random Symbol $\overline{\sigma}$

- Let $e_t^n = e^{\mathrm{i}uM_t^n + \frac{u^2}{2}C_t^n}$
- Define

$$\Phi_n(u,v) = \mathbb{E}[\Psi(u,v)(e_1^n(u) - 1)\psi_n]$$

with a truncation functional $\psi_n \sim 1$.

- **Assumption 2:**

$$\lim_{n\to\infty} \frac{\Phi_n(u,v)}{r_n} = \mathbb{E}[\Psi(u,v)\overline{\sigma}(\mathrm{i}u, \mathrm{i}v)]$$

where

$$\overline{\sigma} = \sum_j \overline{c}_j(z)(\mathrm{i}u)^{m_j}(\mathrm{i}v)^{n_j}.$$

Main Result

- Full random symbol:
  $\sigma(z, \mathrm{i}u, \mathrm{i}v) = \underline{\sigma} + \overline{\sigma} = \sum_j c_j(z)(\mathrm{i}u)^{m_j}(\mathrm{i}v)^{n_j}$.

Main Result

- Full random symbol:
  $\sigma(z, \mathrm{i}u, \mathrm{i}v) = \underline{\sigma} + \overline{\sigma} = \sum_j c_j(z)(\mathrm{i}u)^{m_j}(\mathrm{i}v)^{n_j}$.
- Using information in random symbol σ, we construct the approximated density $p_n(z, x)$.

Main Result

- Full random symbol: $\sigma(z, \mathrm{i}u, \mathrm{i}v) = \underline{\sigma} + \overline{\sigma} = \sum_j c_j(z)(\mathrm{i}u)^{m_j}(\mathrm{i}v)^{n_j}$.
- Using information in random symbol σ, we construct the approximated density $p_n(z, x)$.

Theorem (Yoshida13)

Under some integrability conditions, we obtain

$$\sup_h \left| \mathbb{E}[h(Z_n, F_n)] - \int h(z, x) p_n(z, x) dz dx \right| = o(r_n).$$

Stochastic Decomposition

- Let $b^{[1]}, b^{[2]} \in C^\infty$ and $dX_t = b^{[1]}(X_t)dW_t + b^{[2]}(X_t)dt$.

Stochastic Decomposition

- Let $b^{[1]}, b^{[2]} \in C^\infty$ and $dX_t = b^{[1]}(X_t)dW_t + b^{[2]}(X_t)dt$.
- Using Ito's formula, we obtain for $k = 1, 2$

$$db^{[k]}(X_t) = b^{[k.1]}(X_t)dW_t + b^{[k.2]}(X_t)dt.$$

Stochastic Decomposition

- Let $b^{[1]}, b^{[2]} \in C^\infty$ and $dX_t = b^{[1]}(X_t)dW_t + b^{[2]}(X_t)dt$.
- Using Ito's formula, we obtain for $k = 1, 2$

$$db^{[k]}(X_t) = b^{[k.1]}(X_t)dW_t + b^{[k.2]}(X_t)dt.$$

Lemma

Let $\alpha_{\frac{ik_n}{n}} = b^{[1]}_{\frac{ik_n}{n}} \bar{W}_{\frac{ik_n}{n}} + \bar{\varepsilon}_{\frac{ik_n}{n}}$. *Then, we obtain*

$$Z_n = n^{1/4}(V_n - V) = M_n + \frac{1}{n^{1/4}} N_n + o_{\mathbb{P}}(\frac{1}{n^{1/4}}),$$

where

$$M_n = \frac{n^{1/4}}{\psi_2^n} \sum_{i=0}^{d_n - 1} \alpha^2_{t_{ik_n}} - \mathbb{E}[\alpha^2_{t_{ik_n}} | \mathcal{F}_{t_{ik_n}}] \text{ and } N_n = \sum_{k=1}^{6} N_{n,k}$$

Some Terms

$$N_{n,3} = \frac{k_n^2(\psi_3^n)^2}{n^{3/2}\psi_2^n} \sum_{i=0}^{d_n-1} (b_{t_{ik_n}}^{[2]})^2,$$

$$N_{n,4} = \frac{k_n(2k_n\psi_4^n - (k_n+1)\psi_2^n)}{2n^{3/2}\psi_2^n} \sum_{i=0}^{d_n-1} 2b_{t_{ik_n}}^{[1]} b_{t_{ik_n}}^{[1.2]} + (b_{t_{ik_n}}^{[1.1]})^2$$

$$N_{n,5} = -2n^{1/2} \sum_{i=0}^{d_n-1} b_{t_{ik_n}}^{[1]} b_{t_{ik_n}}^{[1.1]} \int_{t_{ik_n}}^{t_{(i+1)k_n}} \int_{t_{ik_n}}^{u} dW_s du,$$

$$N_{n,6} = \frac{n^{3/2}\psi_1^n}{\psi_2^n(k_n)^2} \left[\omega^2 - \frac{1}{2n} \sum_{i=1}^{n} (\Delta\varepsilon_i^n)^2\right].$$

Computation of σ

- Computation of $\underline{\sigma}$ is via

$$\left(M_n, N_n, \widehat{C}_n, \widehat{F}_n\right) \xrightarrow{d_{st}} (M, N, \widehat{C}, \widehat{F}) \sim MN\left(\mu, \int_0^1 \Sigma_s ds\right).$$

Computation of σ

- Computation of $\underline{\sigma}$ is via

$$\left(M_n, N_n, \widehat{C}_n, \widehat{F}_n\right) \xrightarrow{d_{st}} (M, N, \widehat{C}, \widehat{F}) \sim MN\left(\mu, \int_0^1 \Sigma_s ds\right).$$

- Coefficients in $\overline{\sigma}$ involve Malliavin derivatives.

Computation of σ

- Computation of $\underline{\sigma}$ is via

$$\left(M_n, N_n, \widehat{C}_n, \widehat{F}_n\right) \xrightarrow{d_{st}} (M, N, \widehat{C}, \widehat{F}) \sim MN\left(\mu, \int_0^1 \Sigma_s ds\right).$$

- Coefficients in $\overline{\sigma}$ involve Malliavin derivatives.
- $\sigma = \underline{\sigma} + \overline{\sigma}$

Expansion for pre-averaging

Approximated density has 8 second order terms:

$$p_n(z,x) = \phi(z;0,x)p^C(x) + \frac{1}{n^{1/4}} \sum_{j=1}^{8} A_j.$$

Expansion for pre-averaging

Approximated density has 8 second order terms:

$$p_n(z,x) = \phi(z;0,x)p^C(x) + \frac{1}{n^{1/4}}\sum_{j=1}^{8} A_j.$$

Our main result is:

Theorem

Under $b^{[1]}, b^{[2]} \in C^\infty$ and some more conditions, we obtain

$$\sup_h \left| \mathbb{E}[h(Z_n, F_n)] - \int h(z,x)p_n(z,x)dzdx \right| = o\left(\frac{1}{n^{1/4}}\right)$$

Studentization

Edgeworth expansion of studentized statistic $Z_n/\sqrt{F_n}$:

Corollary

Under some smoothness conditions on $b^{[1]}, b^{[2]}$, we get

$$p^{Z_n/\sqrt{F_n}}(y) = \phi(y) + \frac{1}{n^{1/4}}\phi(y)\Big[y\Big(c_1\mathbb{E}[DC^{-3/2}] + \mathbb{E}[\mu_2 C^{-1/2}]\dots\Big) + y^3\Big(c_2\mathbb{E}[DC^{-3/2}] + \dots\Big)\Big]$$

where

$$C = 2\theta\int_0^1 \left[(b^{[1]})^2(X_t) + \frac{\omega^2\psi_1}{\theta^2\psi_2}\right]^2 dt,$$

$$D = \int_0^1 \left[(b^{[1]})^2(X_t) + \frac{\omega^2\psi_1}{\theta^2\psi_2}\right]^3 dt,$$

Thanks for your attention!

References

[PVY14] M. Podolskij, B. Veliyev, and N. Yoshida. Edgeworth expansion for the pre-averaging estimator. 2014. Preprint.

[PY13] M. Podolskij and N. Yoshida. Edgeworth expansion for functionals of continuous diffusion processes. 2013. Preprint.

[Yos97] N. Yoshida. Malliavin calculus and asymptotic expansion for martingales. *Probab. Theory Related Fields*, 109(3):301–342, 1997.

[Yos13] N. Yoshida. Martingale expansion in mixed normal limit. *Stochastic Process. Appl.*, 123(3):887–933, 2013.