Gaussian Approximations for Random Functions

March, 17th 2021
Séminaire des Doctorants, EDMH

Etienne Lasalle

Office : 3A19
etienne.lasalle@universite-paris-saclay.fr

Outline

1. Reminders and Introduction
    - Classical Stats Results
    - The problem of Gaussian Approximation

2. Historical Methods
    - Skorokhod Embedding
    - Komlós, Major and Tusnády (KMT)
    - Generalizations

3. My Thesis Project
    - Introduction
    - Results and Questions

Classical Framework for Statisticians

$$X_1, \ldots, X_N \underset{i.i.d}{\sim} P$$

Classical Question : Can we infer $\mathbb{E}\left[X\right]$?

$$\mathbb{E}\left[X\right] := \sum_{x \in \mathcal{X}} xP(X = x) \quad \left(\text{or} \quad \int_{\mathbb{R}} xdP(x)\right)$$

Classical Framework for Statisticians

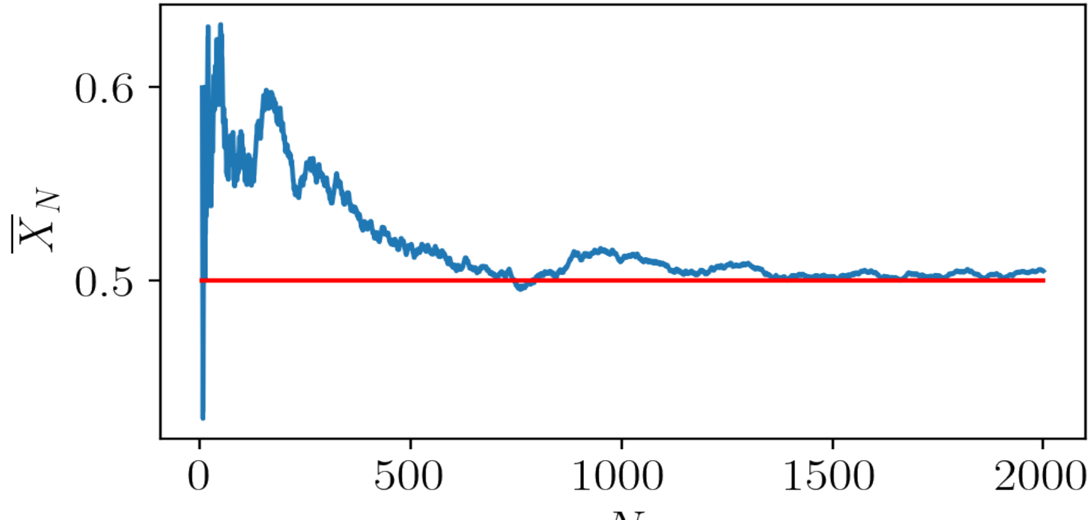
$$X_1, \dots, X_N \underset{i.i.d}{\sim} P$$

Classical Question : Can we infer $\mathbb{E}\left[X\right]$?

$$\mathbb{E}\left[X\right] := \sum_{x \in \mathcal{X}} x P(X = x) \quad \left(\text{or} \quad \int_{\mathbb{R}} x dP(x) \right)$$

Strong Law of Large Numbers : If $X_1, X_2, \dots$ is an i.i.d sequence and $\mathbb{E}\left[|X_1|\right] < \infty$ then

$$\overline{X}_N := \frac{1}{N} \sum_{i=1}^{N} X_i \xrightarrow[N \to \infty]{a.s.} \mathbb{E}\left[X\right]$$

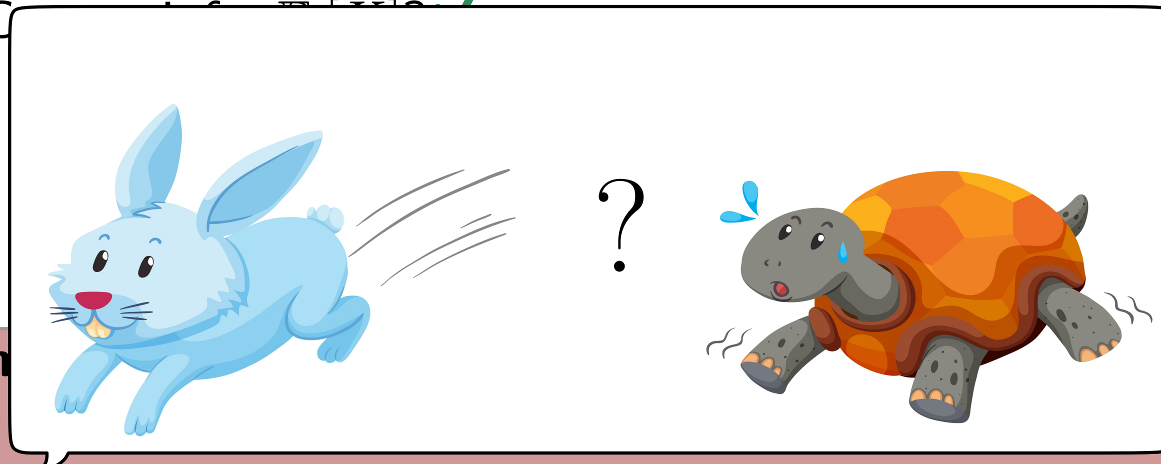


Classical Framework for Statisticians

$$X_1, \dots, X_N \underset{i.i.d}{\sim} P$$

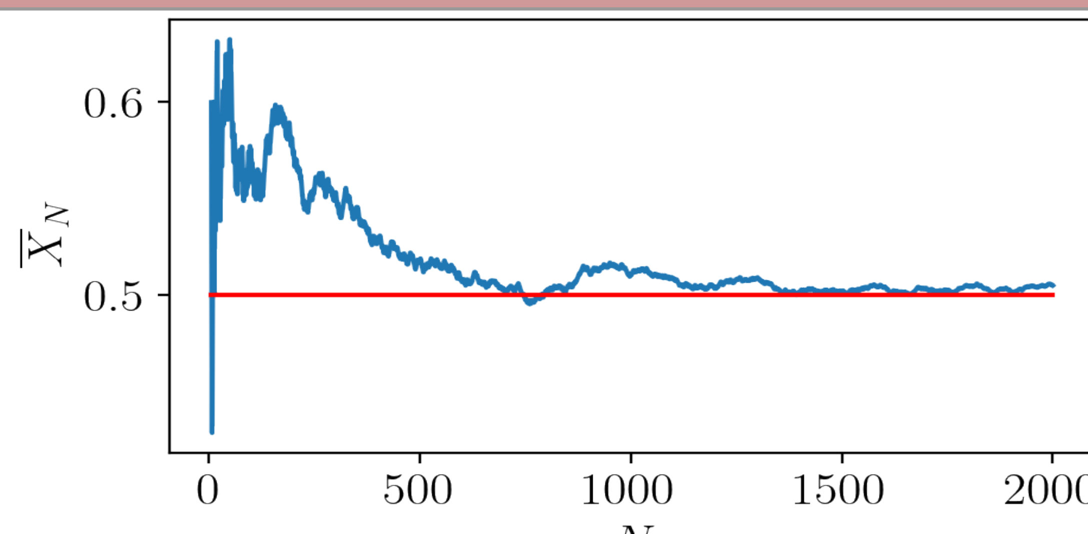
Classical Question : C

$\mathbb{E}[X]$



Strong Law of Large Nun
$\mathbb{E}[|X_1|] < \infty$ then

$$\overline{X}_N := \frac{1}{N} \sum_{i=1}^{N} X_i \xrightarrow[N \to \infty]{a.s.} \mathbb{E}[X]$$

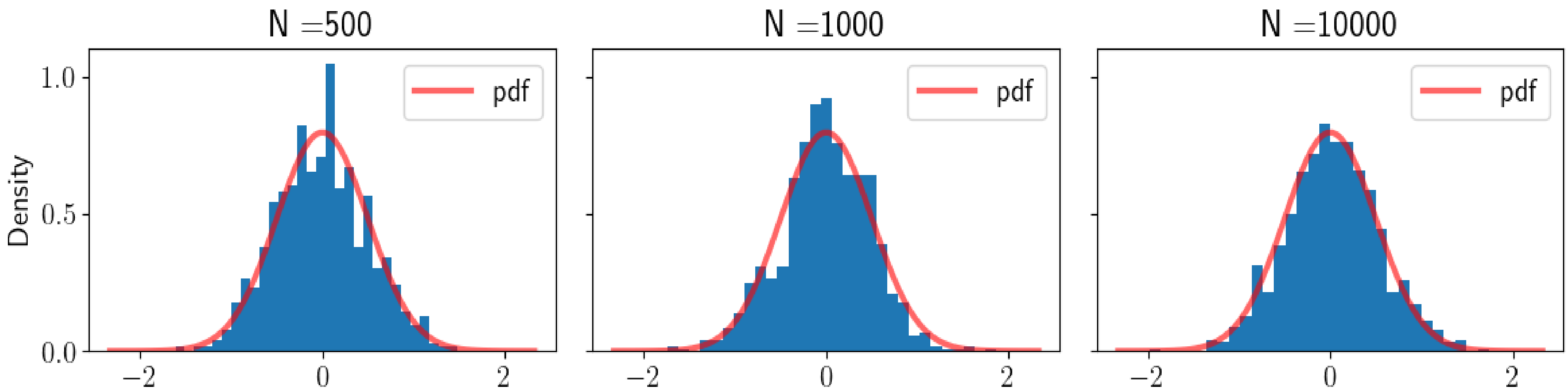


Classical Framework for Statisticians

$$\left(\sqrt{N}\left(\overline{X}_N^{(1)} - \mathbb{E}\left[X\right]\right), \sqrt{N}\left(\overline{X}_N^{(2)} - \mathbb{E}\left[X\right]\right) \ldots, \sqrt{N}\left(\overline{X}_N^{(M)} - \mathbb{E}\left[X\right]\right)\right)$$

Classical Framework for Statisticians

$$\left(\sqrt{N}\left(\overline{X}_N^{(1)} - \mathbb{E}\left[X\right]\right), \sqrt{N}\left(\overline{X}_N^{(2)} - \mathbb{E}\left[X\right]\right) \ldots, \sqrt{N}\left(\overline{X}_N^{(M)} - \mathbb{E}\left[X\right]\right)\right)$$

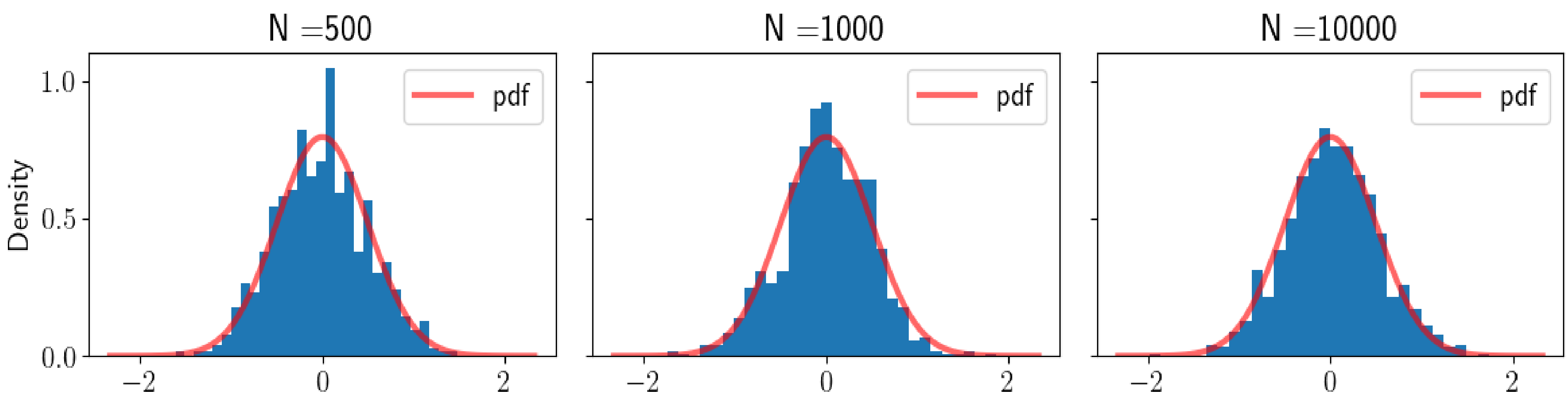


Central Limit Theorem : If $X_1, X_2, \ldots$ is an i.i.d sequence and $\mathbb{E}\left[|X_1|^2\right] < \infty$ then

$$\sqrt{N}\left(\overline{X}_N - \mathbb{E}\left[X\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \mathrm{Var}(X)\right)$$

Classical Framework for Statisticians

$$\left(\sqrt{N}\left(\overline{X}_N^{(1)} - \mathbb{E}\left[X\right]\right), \sqrt{N}\left(\overline{X}_N^{(2)} - \mathbb{E}\left[X\right]\right) \dots, \sqrt{N}\left(\overline{X}_N^{(M)} - \mathbb{E}\left[X\right]\right)\right)$$



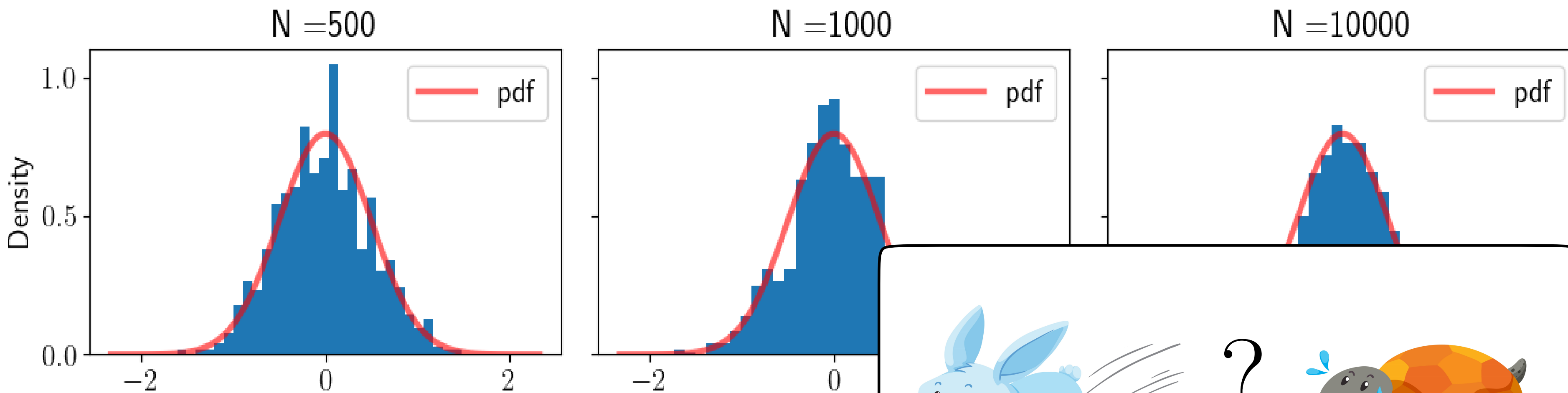
Central Limit Theorem : If $X_1, X_2, \dots$ is an i.i.d sequence and $\mathbb{E}\left[|X_1|^2\right] < \infty$ then

$$\sqrt{N}\left(\overline{X}_N - \mathbb{E}\left[X\right]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X))$$

✌️ $\overline{X}_N \approx \mathbb{E}\left[X\right] + \frac{1}{\sqrt{N}} Z + o\left(\frac{1}{\sqrt{N}}\right)$ ✌️

Classical Framework for Statisticians

$$\left(\sqrt{N}\left(\overline{X}_N^{(1)} - \mathbb{E}[X]\right), \sqrt{N}\left(\overline{X}_N^{(2)} - \mathbb{E}[X]\right) \dots, \sqrt{N}\left(\overline{X}_N^{(M)} - \mathbb{E}[X]\right)\right)$$



Central Limit Theorem : If $X_1, X_2, \dots$ is an ... en

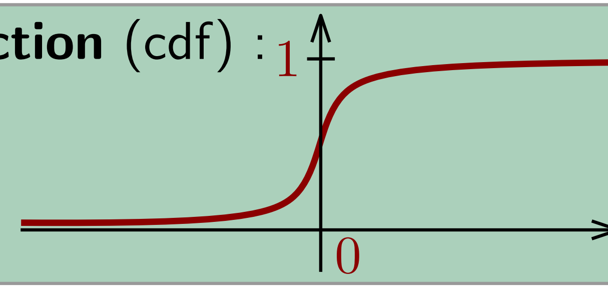
$$\sqrt{N}\left(\overline{X}_N - \mathbb{E}[X]\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathrm{Var}(X))$$

$$\overline{X}_N \approx \mathbb{E}[X] + \frac{1}{\sqrt{N}} Z + o\left(\frac{1}{\sqrt{N}}\right)$$

Classical Framework for Statisticians

Cumulative Distribution Function (cdf) :

$$F_P(x) := P(X \leq x)$$

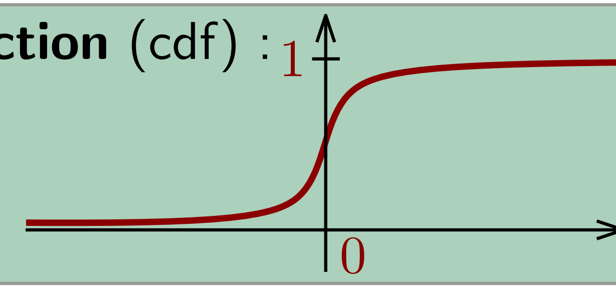


$$P \iff F_P$$

Classical Framework for Statisticians

Cumulative Distribution Function (cdf) :

$$F_P(x) := P(X \leq x)$$

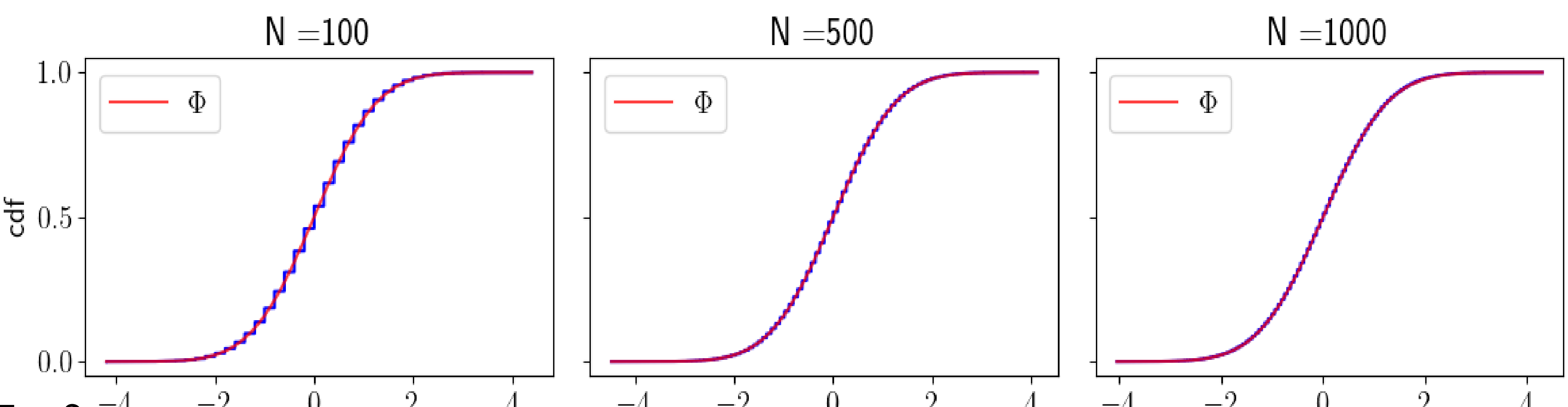


$$P \iff F_P$$

Berry-Essen Theorem :

F_N cdf of $\frac{\sqrt{N}}{\sigma(X)}\left(\overline{X}_N - \mathbb{E}\left[X\right]\right)$
Φ cdf of $\mathcal{N}(0,1)$, then

$$\|F_N - \Phi\|_\infty = O\left(\frac{1}{\sqrt{N}}\right)$$



Generalization

$\mathbb{R}^d$

Gaussian Vector : $\mathcal{N}_{\mathbb{R}^d}(\mu, \Sigma)$

$$Z \sim \mathcal{N}_{\mathbb{R}^d} \iff \forall a \in \mathbb{R}^d,\ a^T.Z \sim \mathcal{N}$$

$$\mu_i = \mathbb{E}\left[Z_i\right] \quad and \quad \Sigma_{i,j} = \mathrm{Cov}(Z_i, Z_j).$$

- Strong Law of Large Number : ✓
- Central Limit Theorem : ✓
- Cumulative Distribution Function : $\simeq$
- Berry-Essen Theorem : $\simeq$

Generalization

$\mathbb{R}^d$

Gaussian Vector : $\mathcal{N}_{\mathbb{R}^d}(\mu, \Sigma)$

$$Z \sim \mathcal{N}_{\mathbb{R}^d} \iff \forall a \in \mathbb{R}^d,\ a^T.Z \sim \mathcal{N}$$

$$\mu_i = \mathbb{E}\left[Z_i\right] \quad and \quad \Sigma_{i,j} = \mathrm{Cov}(Z_i, Z_j).$$

- Strong Law of Large Number : ✓
- Central Limit Theorem : ✓
- Cumulative Distribution Function : $\simeq$
- Berry-Essen Theorem : $\simeq$

$\mathcal{C}([0,T])$

Gaussian Process : $\mathcal{N}_{\mathcal{C}}(\mu, \kappa)$

$$(Z_t) \sim \mathcal{N}_{\mathcal{C}} \iff \forall d,\ \forall (t_1, \ldots, t_d),\ \begin{pmatrix} Z_{t_1} \\ \vdots \\ Z_{t_d} \end{pmatrix} \sim \mathcal{N}_{\mathbb{R}^d}$$

$$\mu_t = \mathbb{E}\left[Z_t\right] \quad and \quad \kappa(t,s) = \mathrm{Cov}(Z_t, Z_s).$$

- Strong Law of Large Number : ✓
- Central Limit Theorem : ✓
- Cumulative Distribution Function : ✗
- Berry-Essen Theorem : ✗

Generalization

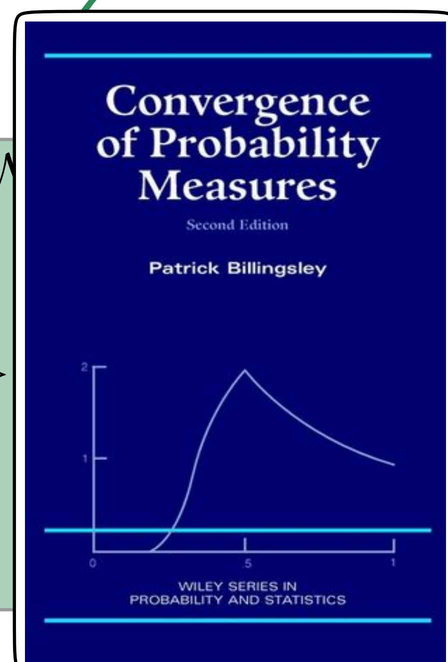
$\mathbb{R}^d$

Gaussian Vector : $\mathcal{N}_{\mathbb{R}^d}(\mu, \Sigma)$

$$Z \sim \mathcal{N}_{\mathbb{R}^d} \iff \forall a \in \mathbb{R}^d,\ a^T.Z \sim \mathcal{N}$$

$$\mu_i = \mathbb{E}\left[Z_i\right] \quad and \quad \Sigma_{i,j} = \mathrm{Cov}(Z_i, Z_j).$$

- Strong Law of Large Number : ✓
- Central Limit Theorem : ✓
- Cumulative Distribution Function : $\simeq$
- ssen Theorem : $\simeq$



$\mathcal{C}([0, T])$

Gaussian Process : $\mathcal{N}$

$$(Z_t) \sim \mathcal{N}_{\mathcal{C}} \iff \quad \begin{pmatrix} Z_{t_1} \\ \vdots \\ Z_{t_d} \end{pmatrix} \sim \mathcal{N}_{\mathbb{R}^d}$$

$$\mu_t = \mathbb{E}\left[Z_t\right] \quad \text{ov}(Z_t, Z_s).$$

Gilvenko-Cantelli Theorem

- ~~Strong Law of Large Number~~ : ✓
- ~~Central Limit Theorem~~ : ✓

Donsker Theorem

- Cumulative Distribution Function : ✗
- Berry-Essen Theorem : ✗

Gaussian Approximation

Ⓡ Assuming $\mathbb{E}[X] = 0$, $\sigma(X) = 1$, $\quad Y_N := \frac{1}{\sqrt{N}} \sum X_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$

Problem :

Can we : draw $X_1, \ldots, X_N \sim P$

draw $Z \sim \mathcal{N}(0,1)$

Highly Dependent

$$s.t. \qquad |Y_N - Z| \leq C \,.\, r_N \qquad w.h.p$$

What is the best rate?

Gaussian Approximation

Assuming $\mathbb{E}[X] = 0$, $\sigma(X) = 1$, $\quad Y_N := \frac{1}{\sqrt{N}} \sum X_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$

Problem :

Can we : draw $X_1, \ldots, X_N \sim P$
draw $Z \sim \mathcal{N}(0,1)$

Highly Dependent

$$s.t. \qquad |Y_N - Z| \leq C \,.\, r_N \qquad w.h.p$$

What is the best rate?

Reformulation :

Can we : draw $X_1, \ldots, X_N \sim P$
draw $Z_1, \ldots, Z_N \sim \mathcal{N}(0,1)$

$$s.t. \qquad \left|\sum X_i - \sum Z_i\right| \leq C \,.\, \tilde{r}_N \qquad w.h.p$$

Gaussian Approximation

Ⓡ Assuming $\mathbb{E}[X]=0$, $\sigma(X)=1$, $\quad Y_N := \frac{1}{\sqrt{N}}\sum X_i \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$

Problem :

Can we : draw $X_1,\ldots,X_N \quad \sim P$

draw $Z \quad \sim \mathcal{N}(0,1)$

Highly Dependent

$$s.t. \qquad |Y_N - Z| \leq C\ .\, r_N \qquad w.h.p$$

What is the best rate?

Reformulation :

Can we : draw $X_1,\ldots,X_N \quad \sim P$

draw $Z_1,\ldots,Z_N \quad \sim \mathcal{N}(0,1)$

$$s.t. \qquad \left|\sum X_i - \sum Z_i\right| \leq C\ .\ \tilde{r}_N \qquad w.h.p$$

Bartfai, P. (1966) Die Bestimmung der zu einem wiederkehrenden Prozess gehorenden Verteilungfunktion aus den mit Fehlern behafteten Daten einer einzigen Realisation, Studia Sci. Math. Hungar.

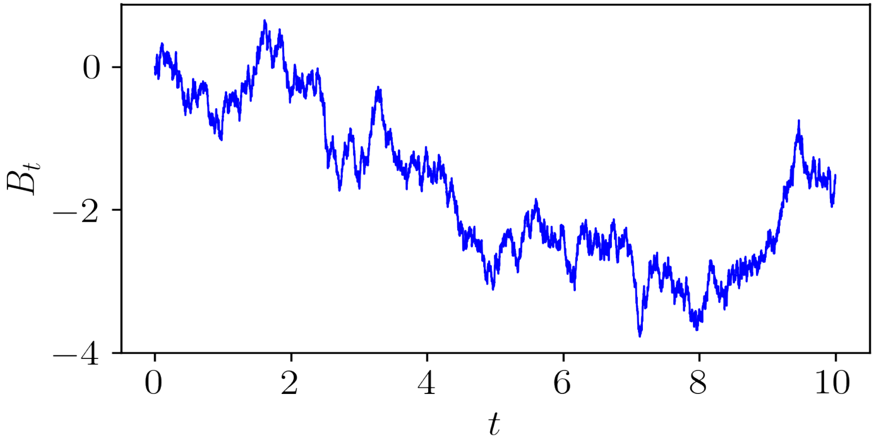
If $\mathbb{E}\left[e^{tX}\right] < \infty$, for $|t| \leq \eta$, then we can't do better than :

$$\tilde{r}_N = \log N.$$

Skorokhod Embedding

Brownian Motion

- $B(0) = 0$
- B is continuous
- B has independent increments
- $B(t) - B(s) \sim \mathcal{N}(0, t-s)$, $t \geq s$

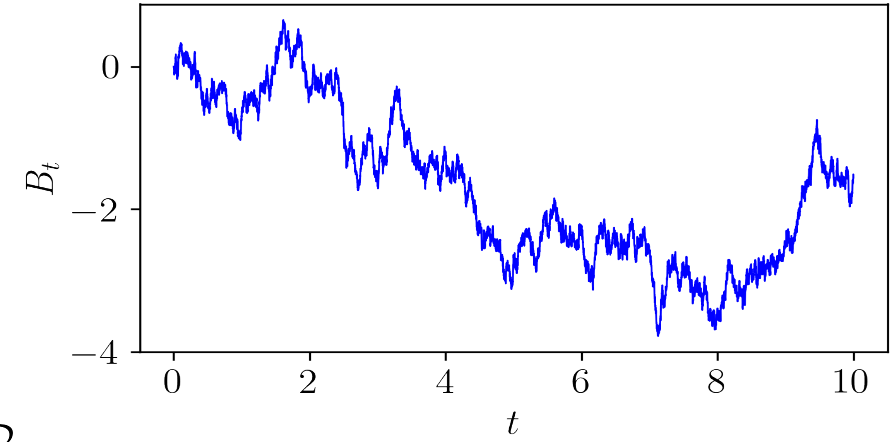


$$B(n) \overset{\mathcal{D}}{=} \mathcal{N}(0, n) \overset{\mathcal{D}}{=} Z_1 + \cdots + Z_n$$

Skorokhod Embedding

Brownian Motion

- $B(0) = 0$
- B is continuous
- B has independent increments
- $B(t) - B(s) \sim \mathcal{N}(0, t-s)$, $t \geq s$

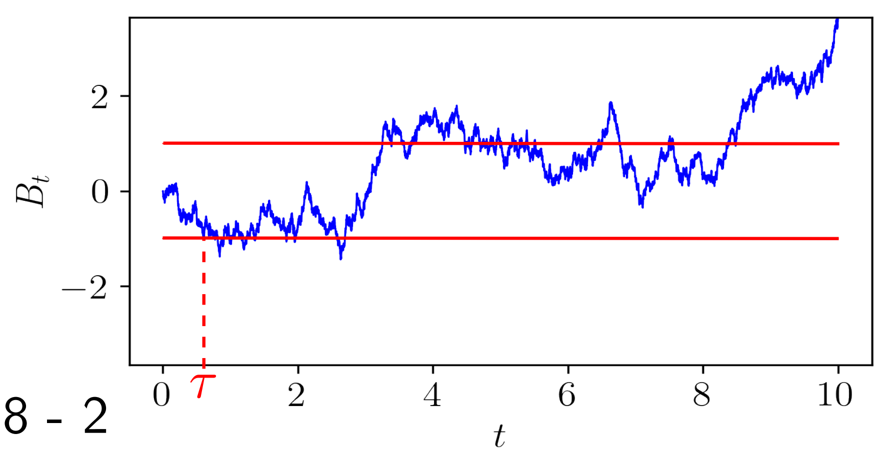


$$B(n) \stackrel{\mathcal{D}}{=} \mathcal{N}(0, n) \stackrel{\mathcal{D}}{=} Z_1 + \cdots + Z_n$$

Skorokhod Embedding

If $\mathbb{E}[X] = 0$ and $Var(X) = 1$, then there exists a Markov Time τ such that

$$B(\tau) \stackrel{\mathcal{D}}{=} X \quad \text{and} \quad \mathbb{E}[\tau] = 1$$



$$\tau = \inf\{t \geq 0, B(t) = -1 \text{ or } B(t) = 1\}$$

$$\mathbb{P}(B(\tau) = 1) = \mathbb{P}(B(\tau) = -1) = \frac{1}{2}$$

Skorokhod Embedding

Goal : $\quad |\sum X_i - \sum Z_i| \leq C \ . \ \tilde{r}_N \qquad w.h.p$

$\sum Z_i \stackrel{\mathcal{D}}{=} B(N)$

Skorokhod Embedding

Goal : $|\sum X_i - \sum Z_i| \leq C \, . \, \tilde{r}_N \qquad w.h.p$

$\overset{\mathcal{D}}{=} B(N)$

$\tau_1, \ldots, \tau_N \underset{i.i.d}{\overset{\mathcal{D}}{=}} \tau$

$X_1 := B(\tau_1)$

$X_2 := B(\tau_1 + \tau_2) - B(\tau_1) \quad (\perp X_1)$

$$X_1 + X_2 = B(\tau_1 + \tau_2)$$

$\vdots$

$X_1 + \cdots + X_N = B(\tau_1 + \cdots + \tau_N)$

Skorokhod Embedding

Goal : $\quad |\sum X_i - \sum Z_i| \leq C \, . \, \tilde{r}_N \qquad w.h.p$

$$\sum Z_i \overset{\mathcal{D}}{=} B(N)$$

$$\tau_1, \ldots, \tau_N \underset{i.i.d}{\overset{\mathcal{D}}{=}} \tau$$

$$X_1 := B(\tau_1)$$

$$X_2 := B(\tau_1 + \tau_2) - B(\tau_1) \quad (\perp X_1)$$

$$X_1 + X_2 = B(\tau_1 + \tau_2)$$

$$\vdots$$

$$X_1 + \cdots + X_N = B(\tau_1 + \cdots + \tau_N)$$

✌️ $\mathbb{E}[\tau] = 1$

$\tau_1 + \cdots + \tau_N \simeq N$

$B(\tau_1 + \cdots + \tau_N) \simeq B(N)$ ✌️

Skorokhod Embedding

Goal : $\quad |\sum X_i - \sum Z_i| \leq C \, . \, \tilde{r}_N \qquad w.h.p$

$$\sum Z_i \overset{\mathcal{D}}{=} B(N)$$

$$\tau_1, \ldots, \tau_N \underset{i.i.d}{\overset{\mathcal{D}}{=}} \tau$$

$$X_1 := B(\tau_1)$$

$$X_2 := B(\tau_1 + \tau_2) - B(\tau_1) \quad (\perp X_1)$$

$$X_1 + X_2 = B(\tau_1 + \tau_2)$$

$\vdots$

$$X_1 + \cdots + X_N = B(\tau_1 + \cdots + \tau_N)$$

Strassen, V.(1967) Almost sure behaviour of sums of independent random variables and martingales, Proc. Fifth Berkeley Sympos. Math. Statist. and Probability.

If additionally $\mathbb{E}\left[|X|^4\right] < \infty$,

$$|B(\tau_1 + \cdots + \tau_N) - B(N)| = O\left(N^{1/4} (\log N)^{1/2} (\log \log N)^{1/4}\right) \quad a.s.$$

Quantile Transformation

A elementary problem :

Draw $Z, Z' \sim \mathcal{N}(0,1)$, to minimize $|Z - Z'|$.

Draw $Z \sim \mathcal{N}(0,1)$. And take $Z' = Z$.

Quantile Transformation

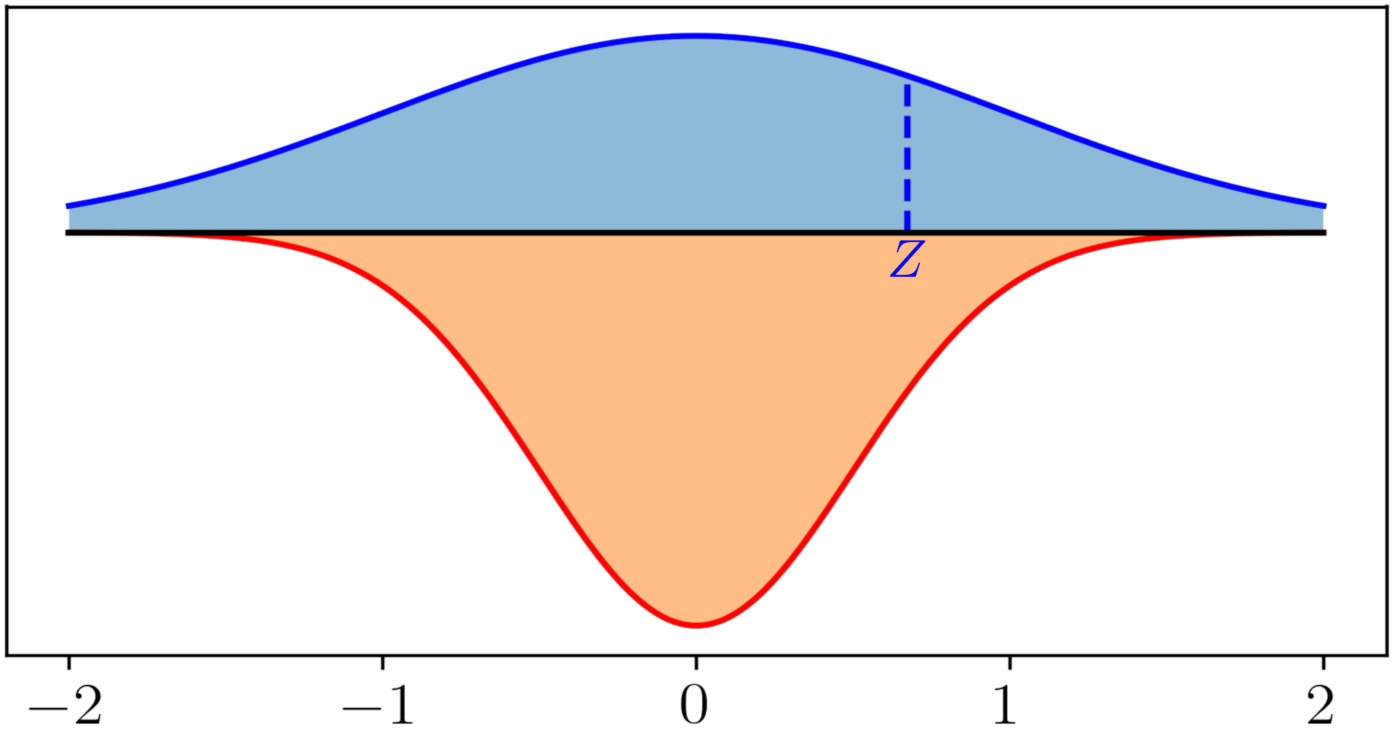
A elementary problem :

Draw $Z, Z' \sim \mathcal{N}(0,1)$, to minimize $|Z - Z'|$.

Draw $Z \sim \mathcal{N}(0,1)$. And take $Z' = Z$.

Another problem :

Draw $X \sim P$ and $Z \sim \mathcal{N}(0,1)$, to minimize $|X - Z|$.



Quantile Transformation

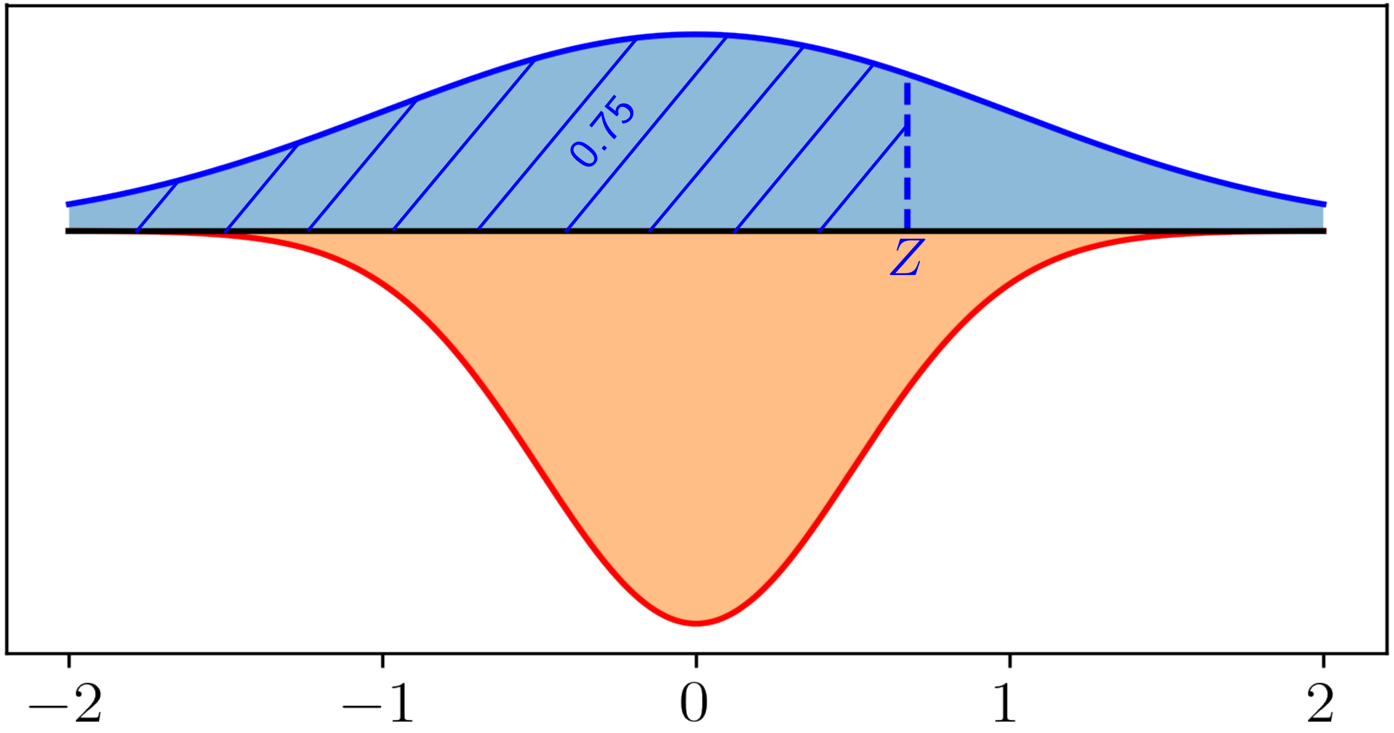
A elementary problem :

Draw $Z, Z' \sim \mathcal{N}(0,1)$, to minimize $|Z - Z'|$.

Draw $Z \sim \mathcal{N}(0,1)$. And take $Z' = Z$.

Another problem :

Draw $X \sim P$ and $Z \sim \mathcal{N}(0,1)$, to minimize $|X - Z|$.



Quantile Transformation

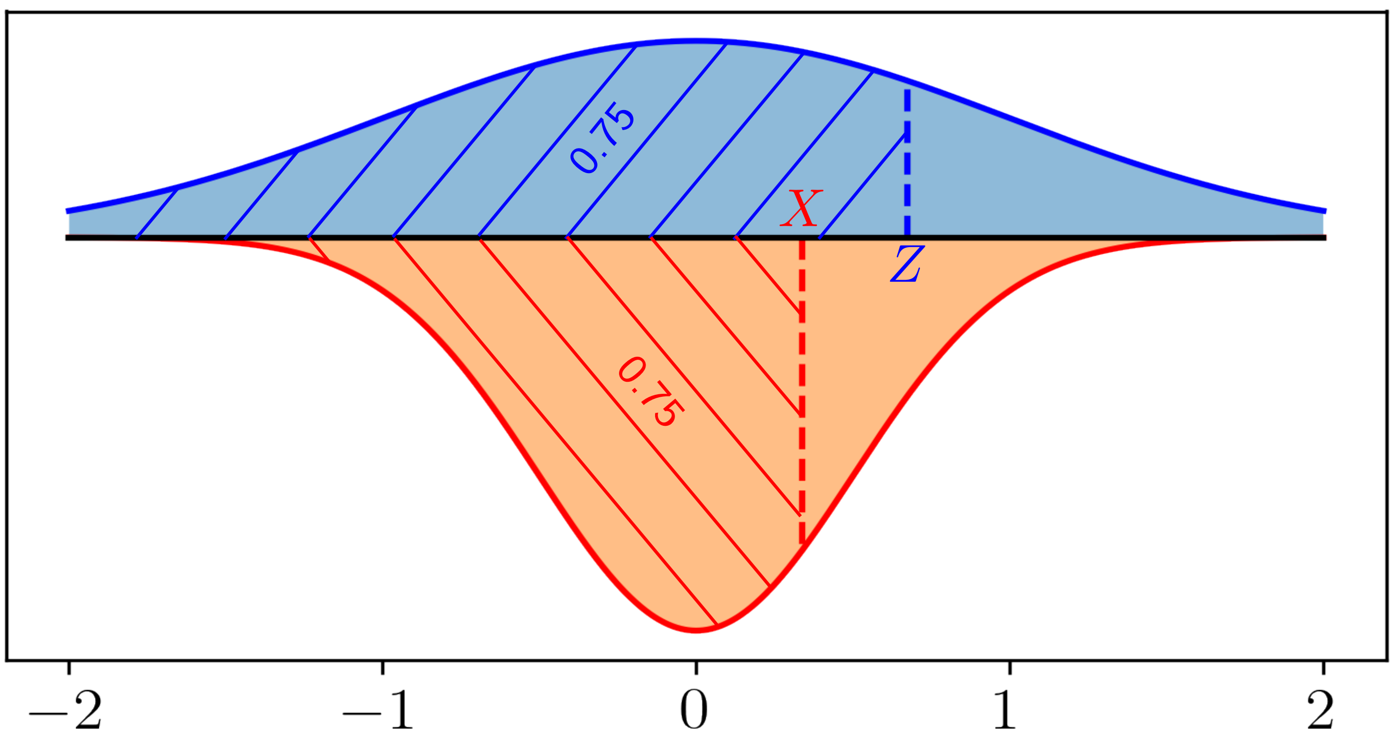
A elementary problem :

Draw $Z, Z' \sim \mathcal{N}(0,1)$, to minimize $|Z - Z'|$.

Draw $Z \sim \mathcal{N}(0,1)$. And take $Z' = Z$.

Another problem :

Draw $X \sim P$ and $Z \sim \mathcal{N}(0,1)$, to minimize $|X - Z|$.



Quantile Transformation

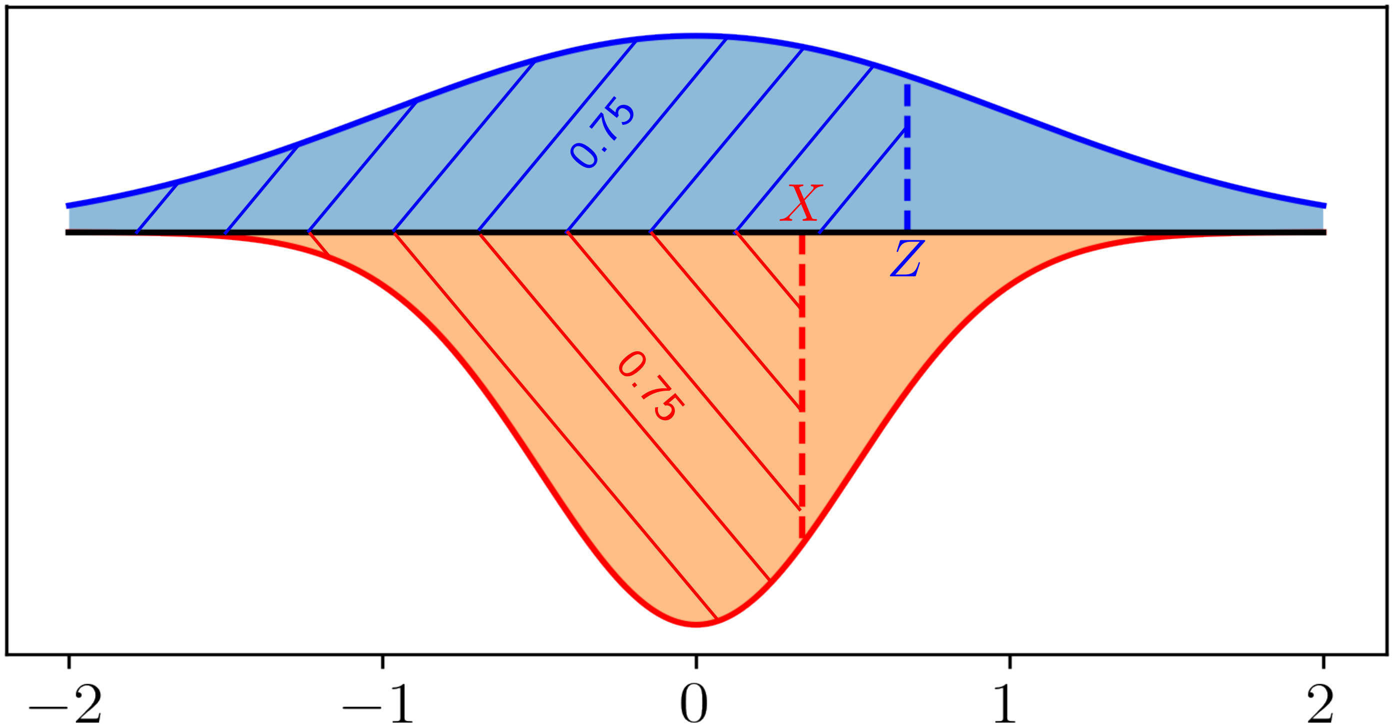
A elementary problem :

Draw $Z, Z' \sim \mathcal{N}(0,1)$, to minimize $|Z - Z'|$.

Draw $Z \sim \mathcal{N}(0,1)$. And take $Z' = Z$.

Another problem :

Draw $X \sim P$ and $Z \sim \mathcal{N}(0,1)$, to minimize $|X - Z|$.



$$F_P(X) = F_{\mathcal{N}(0,1)}(Z) \quad \Longrightarrow \quad X := (F_P)^{-1}(F_{\mathcal{N}(0,1)}(Z))$$

Application to the Gauss. Approx. Problem

Goal : Find an approximation of

$$S_N = X_1 + \cdots + X_N \quad \text{by} \quad T_N = Z_1 + \cdots + Z_N$$

Let's assume that $N = 2^n$

First idea :

- Draw the $Z_1, \ldots, Z_N$.
- Quantile Transformation : $S_N := (F_{S_N})^{-1} \left(F_{T_N} (Z_1 + \cdots + Z_N) \right)$

But who are the X_i's?

Application to the Gauss. Approx. Problem

Goal : Find an approximation of

$$S_N = X_1 + \cdots + X_N \quad \text{by} \quad T_N = Z_1 + \cdots + Z_N$$

Let's assume that $N = 2^n$

First idea :

- Draw the $Z_1, \ldots, Z_N$.
- Quantile Transformation : $S_N := (F_{S_N})^{-1}\left(F_{T_N}(Z_1 + \cdots + Z_N)\right)$

But who are the X_i's?

$$\underbrace{X_1 + \cdots + X_{2^{n-1}}}_{U_L} + \underbrace{X_{2^{n-1}+1} + \cdots + X_{2^n}}_{U_R} \qquad \underbrace{Z_1 + \cdots + Z_{2^{n-1}}}_{V_L} + \underbrace{Z_{2^{n-1}+1} + \cdots + Z_{2^n}}_{V_R}$$

$\tilde{U} = U_L - U_R$ $\qquad$ $\tilde{V} = V_L - V_R$

$$\tilde{U} := (F_{\tilde{U}})^{-1}\left(F_{\tilde{V}}(\tilde{V})\right)$$

Application to the Gauss. Approx. Problem

Goal : Find an approximation of

$$S_N = X_1 + \cdots + X_N \quad \text{by} \quad T_N = Z_1 + \cdots + Z_N$$

Let's assume that $N = 2^n$

First idea :

- Draw the $Z_1, \ldots, Z_N$.
- Quantile Transformation : $S_N := (F_{S_N})^{-1}(F_{T_N}(Z_1 + \cdots + Z_N))$

But who are the X_i's?

$$\underbrace{X_1 + \cdots + X_{2^{n-1}}}_{U_L} + \underbrace{X_{2^{n-1}+1} + \cdots + X_{2^n}}_{U_R} \qquad \underbrace{Z_1 + \cdots + Z_{2^{n-1}}}_{V_L} + \underbrace{Z_{2^{n-1}+1} + \cdots + Z_{2^n}}_{V_R}$$

$\tilde{U} = U_L - U_R$ $\qquad\qquad$ $\tilde{V} = V_L - V_R$

$$\cancel{\tilde{U} := (F_{\tilde{U}})^{-1}\left(F_{\tilde{V}}(\tilde{V})\right)} \qquad \tilde{U} := (F_{\tilde{U}}(\cdot|S_N))^{-1}\left(F_{\tilde{V}}(\tilde{V})\right)$$

Application to the Gauss. Approx. Problem

Goal : Find an approximation of

$$S_N = X_1 + \cdots + X_N \quad \text{by} \quad T_N = Z_1 + \cdots + Z_N$$

Let's assume that $N = 2^n$

First idea :

- Draw the $Z_1, \ldots, Z_N$.
- Quantile Transformation : $S_N := (F_{S_N})^{-1} \left(F_{T_N}(Z_1 + \cdots + Z_N)\right)$

But who are the X_i's?

$$\underbrace{X_1 + \cdots + X_{2^{n-1}}}_{U_L} + \underbrace{X_{2^{n-1}+1} + \cdots + X_{2^n}}_{U_R} \qquad \underbrace{Z_1 + \cdots + Z_{2^{n-1}}}_{V_L} + \underbrace{Z_{2^{n-1}+1} + \cdots + Z_{2^n}}_{V_R}$$

$$\tilde{U} = U_L - U_R \qquad \tilde{V} = V_L - V_R$$

$$\cancel{\tilde{U} := (F_{\tilde{U}})^{-1} \left(F_{\tilde{V}}(\tilde{V})\right)} \qquad \tilde{U} := \left(F_{\tilde{U}}(\cdot|S_N)\right)^{-1} \left(F_{\tilde{V}}(\tilde{V})\right)$$

$$U_L := \frac{1}{2}\left(S_N + \tilde{U}\right)$$

$$U_R := \frac{1}{2}\left(S_N - \tilde{U}\right)$$

Application to the Gauss. Approx. Problem

Recall $S_k = X_1 + \cdots + X_k$ and $T_k = Z_1 + \cdots + Z_k$

Komlós, J., Major, P., Tusnády, G., (1975)
An approximation of partial sums of independent rv's and the sample df. I. Z. Wahrsch. verw. Gebiete

If $\mathbb{E}\left[e^{tX}\right] < \infty$, for $|t| \leq \eta$,
then for all N we can find $Y_1, \ldots, Y_N$ and $X_1, \ldots X_N$, such that

$$\mathbb{P}\left(\sup_{1 \leq k \leq N} |S_k - T_k| \geq C.\log N + x\right) \leq K e^{-Lx}$$

Corollary :

$$\sup_{1 \leq k \leq N} |S_k - T_k| = O_P(\log N)$$

Generalization : Multidimensional

$\mathbb{R}^d$

Zaitsev, A. Y. (1998) Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. ESAIM: Probability and Statistics.

Suppose that $\exists \tau \geq 0$, s.t.

$$\phi(z) = \log \mathbb{E}\left[e^{\langle z, X\rangle}\right] < \infty, \qquad \tau|z| < 1$$

and $d_u d_v^2 \varphi(z) \leq \tau \|u\| \langle Cov(X)v, v\rangle$, then

$$\sup_{1\leq k\leq N} \|S_k - T_k\|_\infty = O_P\left(d^{21/4} \log d.\ \log N\right)$$

- Quantile Transformation :
  - Coordinate by coordinate, with conditional CDF.
- same Dyadic Scheme.

Generalization : Emp. Processes

$\mathcal{F}$

$(\mathbb{X}, \mathcal{A})$ a measurable space,
$\xi_1, \xi_2, \ldots \overset{i.i.d}{\sim} P$, $\mathbb{X}$-valued variables.
$\mathcal{F} \subset L_2(\mathbb{X}, dP)$ a family of functions.

$$Z_N(f) = \sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} f(\xi_i) - \mathbb{E}\left[f(\xi)\right]\right)$$

A Centered Gaussian process on $L_2(\mathbb{X}, dP)$:

$W : L_2(\mathbb{X}, dP) \mapsto \mathbb{R}$ centered random function, s.t.

$$\mathbb{E}\left[W(f)W(g)\right] = \mathbb{E}\left[f(\xi)g(\xi)\right] - \mathbb{E}\left[f(\xi)\right]\mathbb{E}\left[g(\xi)\right]$$

Koltchinskii, V. I., (1994), Komlós-Major-Tusnády approximation for the general empirical process and Haar expansions of classes of functions. Journal of Theoretical Probability

Problem :

Can we find $\xi_1, \ldots, \xi_N$ and W_N such that

$$\sup_{f\in\mathcal{F}} |Z_N(f) - W_N(f)| \leq C.r_N \qquad w.h.p$$

$\mathcal{F}$

Two major hypothesis :

Entropy :

$N_d(\mathcal{F}, \epsilon)$: minimal number of balls of radii ϵ neededeed to cover $\mathcal{F}$.
$H_d(\mathcal{F}, \epsilon) = \log N_d(\mathcal{F}, \epsilon)$.

How fast does it grow, when $\epsilon \to 0$?

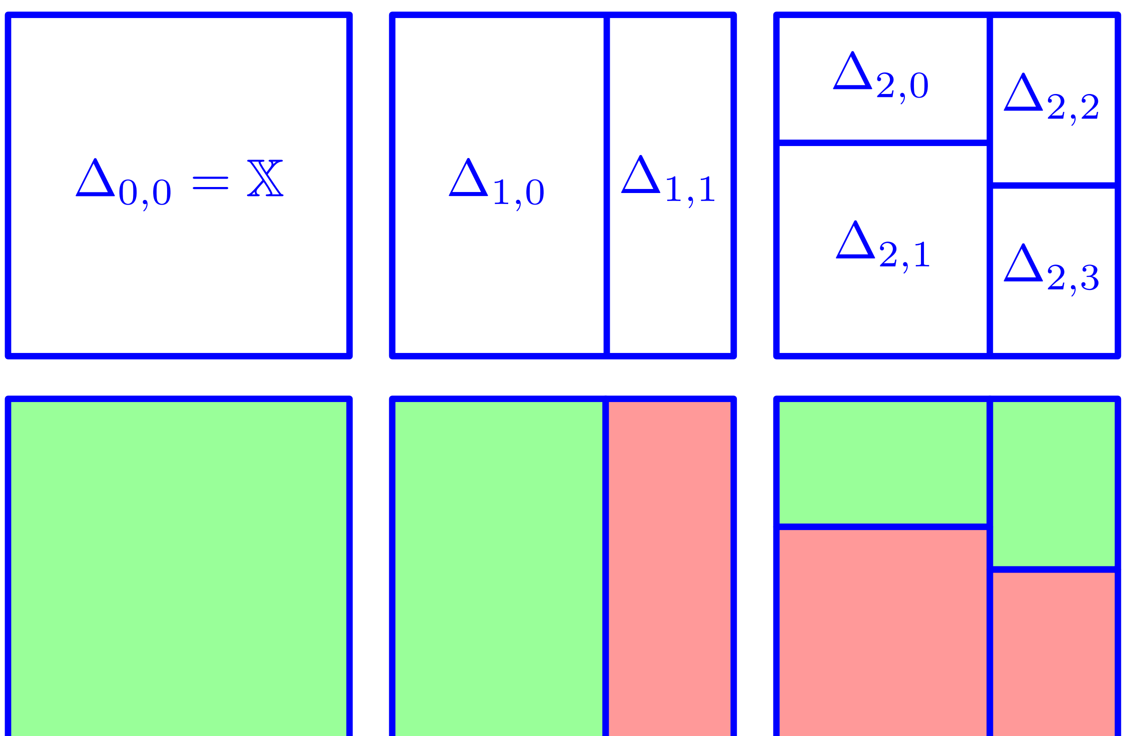
Two major hypothesis :

Entropy :

$N_d(\mathcal{F}, \epsilon)$: minimal number of balls of radii ϵ neededed to cover $\mathcal{F}$.
$H_d(\mathcal{F}, \epsilon) = \log N_d(\mathcal{F}, \epsilon)$.

How fast does it grow, when $\epsilon \to 0$?

Haar Expansion :



$\bullet\bullet\bullet$

$$P(\Delta_{k,l}) = 2^{-k}$$

$$h_{i,k} = 2^{i/2}(1_{\Delta_{i+1,2k}} - 1_{\Delta_{i+1,2k+1}})$$

h_0 $h_{0,0}$ $h_{1,0}$ $h_{1,1}$

$\mathcal{F}$

$$f_i = \langle f, h_0 \rangle h_0 + \sum_{j=0}^{i-1} \sum_{k=0}^{2^j - 1} \langle f, h_{j,k} \rangle h_{j,k}$$

Regularity of the functions in $\mathcal{F}$:

Can we find $\beta > 0$, such that :

$$\sup_{f \in \mathcal{F}} \|f - f_i\|^2_{L_2(\mathbb{X}, dP)} = O\left(2^{-\beta i}\right)$$

Current work

$\mathcal{C}([0,T])$

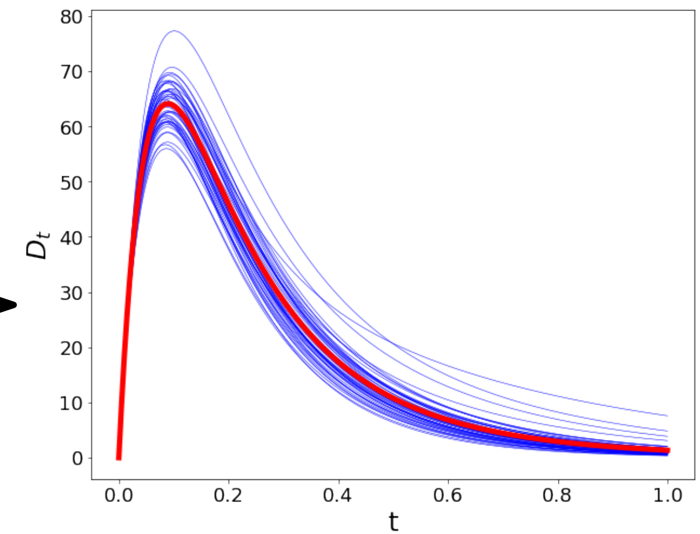
(G_0, G_1) (G_2, G_3)

(G_4, G_5) $(G_{\dots}, G_{\dots})$

(G_{n-1}, G_n)

Dataset

$D_t(G, G') = \|e^{-tL(G)} - e^{-tL(G')}\|_F$



$\forall (G, G')$,

$$\begin{array}{ccc} [0,T] & \to & \mathbb{R}^+ \\ t & \to & D_t(G,G') \end{array} \text{ is } k\text{-lipschitz.}$$

k independent of (G, G').

The family $\{D_t(\cdot,\cdot),\ t \in [0,T]\}$ is Donsker.

Confidence band around $\overline{D}$.

Two sample testing : $\max_t \left| \overline{D}_t^{(1)} - \overline{D}_t^{(2)} \right|$

Can we get a Gaussian Approximation with a good rate?

Good rates ?

$\mathcal{C}([0,T])$

Koltchinskii's Approach :

- $\mathbb{X} = \{$pairs of graphs of size $n\} \subset [0,1]^d$, where $d = \#$edges $= n(n-1)$.
- $\mathcal{F} = \{D_t, t \in [0,T]\}$

$N_d(\mathcal{F}, \epsilon) = O\left(\frac{1}{\epsilon}\right)$
$\beta = 2/d$

$$r_N = \frac{\log^{3/2} N}{N^{1/d}}$$

Good rates ?

$\mathcal{C}([0,T])$

Koltchinskii's Approach :

- $\mathbb{X} = \{$pairs of graphs of size $n\} \subset [0,1]^d$, where $d = \#$edges $= n(n-1)$.
- $\mathcal{F} = \{D_t, t \in [0,T]\}$

$N_d(\mathcal{F}, \epsilon) = O\left(\frac{1}{\epsilon}\right)$
$\beta = 2/d$

$$\boxed{r_N = \frac{\log^{3/2} N}{N^{1/d}}}$$

Finite Dimensional Approach :

- From random functions $f : [0,T] \to \mathbb{R}$, construct a random vector

$$\begin{pmatrix} f(t_1) \\ \vdots \\ f(t_k) \end{pmatrix}$$

$$\boxed{r_N = \frac{\log^{9/14} N}{N^{1/7}}}$$

Good rates ?

$\mathcal{C}([0,T])$

Koltchinskii's Approach :

- $\mathbb{X} = \{$pairs of graphs of size $n\} \subset [0,1]^d$, where $d = \#$edges $= n(n-1)$.
- $\mathcal{F} = \{D_t, t \in [0,T]\}$

$$N_d(\mathcal{F}, \epsilon) = O\left(\frac{1}{\epsilon}\right)$$
$$\beta = 2/d$$

$$\boxed{r_N = \frac{\log^{3/2} N}{N^{1/d}}}$$

Finite Dimensional Approach :

- From random functions $f : [0,T] \to \mathbb{R}$, construct a random vector

$$\begin{pmatrix} f(t_1) \end{pmatrix}$$

Berthet, P., Mason, D. M., (2006). Revisiting two strong approximation results of Dudley and Philipp, High dimensional probability.

$$\boxed{r_N = \frac{\log^{9/14} N}{N^{1/7}}}$$

Thank you for your attention!