12/1/22

Part 5 : Limiting Theorems

$X_1, X_2, \ldots, X_i, \ldots, X_n \overset{iid}{\sim} f(x)$

$E(X_i) = \mu$

$\text{var}(X_i) = \sigma^2$

sample avg. $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$

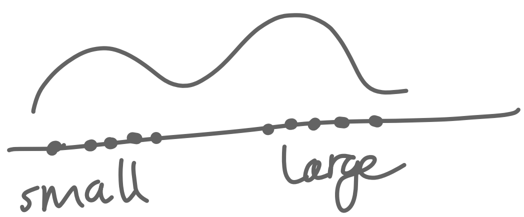
$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$

$\text{var}(\bar{X}) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^{n} X_i\right) \overset{\text{by independence}}{=} \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i)$

$= \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$ averaging $\Rightarrow$ decreases variances

ex.)

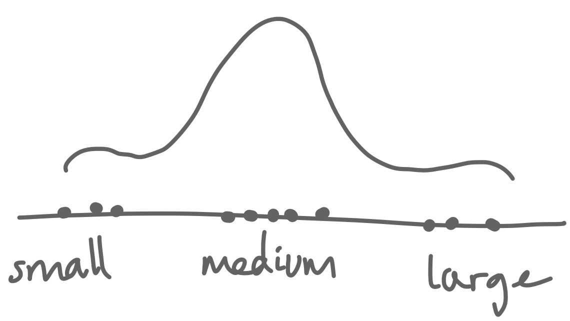
$n=1$



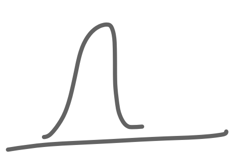
averaging causes:
smaller variance
smoother shape

$n=2$ $\bar{X} = \frac{X_1 + X_2}{2}$

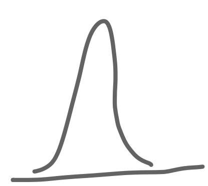
	X_2: small	X_2: large
X_1: small	small average	medium average
X_1: large	medium average	large average



$n=10$



$n=100$



$n \to \infty, \quad \frac{\sigma^2}{n} \to 0$

Law of Large Numbers

$$P(|\bar{x}-\mu| > \underbrace{\varepsilon}_{\text{fixed } \forall n, \text{ "large deviation"}}) \xrightarrow[\infty]{n} 0$$

convergence in probability

Central Limit Theorem

distribution

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{Var(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} N(0,1)$$

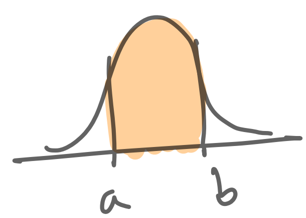
converges to 0 in probability

$$\sqrt{n}\,(\bar{x} - \mu) \xrightarrow{D} N(0, \sigma^2)$$

convergence in distribution

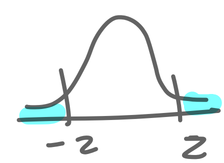
magnify

$P(Z \in (a,b)) \to$



$$= \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

$$P\Big(|\bar{x}-\mu| > \underbrace{\frac{z}{\sqrt{n}}\sigma}_{\text{"small deviation"}}\Big) = P\left(\left|\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right| > z\right)$$



understanding meaning of LLN

Assume $f(x) \sim Unif\,[0,1]$

$$P(\underbrace{|\bar{X}_n - \mu| > \varepsilon}_{B_{n,\varepsilon}}) \xrightarrow[\infty]{n} 0$$

avg of n #'s

$(x_1, \dots, x_n) \in \Omega_n = [0,1]^n$ n-dim cube

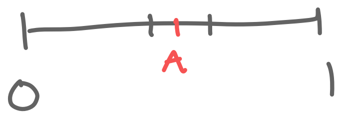
$B_{n,\varepsilon} \in \Omega_n \qquad |B_{n,\varepsilon}| \to 0$

$$B_{n,\varepsilon} = \{(x_1, x_2, \dots, x_n) : |\bar{x}_n - \tfrac{1}{2}| > \varepsilon\}$$

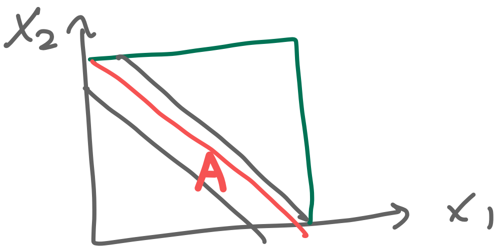
$A_{n,\varepsilon} = B_{n,\varepsilon}^{c}$

$|A_{n,\varepsilon}| \to 1$ weak LLN

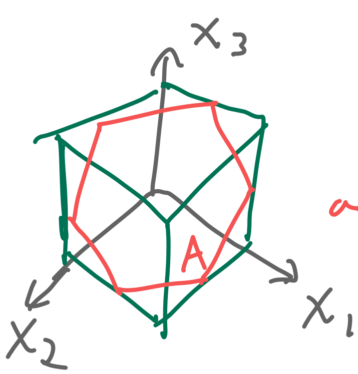
$n = 1$



$n = 2$



$n = 3$



$\boxed{\text{Volume of diagonal piece}}$

add thickness

$|\bar{x} - \mu| < \varepsilon$ LLN

$|\bar{x} - \mu| < \frac{\sigma}{\sqrt{n}} z$ CLT

Markov inequality

$Z \geq 0$

$$P(Z \geq t) \leq \frac{E(Z)}{t}$$

$$P(Z \geq t) = \int_t^\infty f(z)\,dz$$

$$E(Z) = \int_0^\infty z f(z)\,dz \geq \int_t^\infty z f(z)\,dz \geq \int_t^\infty t f(z)\,dz$$

$$= t \cdot \int_t^\infty f(z)\,dz = t\,P(Z \geq t)$$

Chebychev inequality

$$P(|\bar{x} - \mu| > \varepsilon) = P(\underbrace{(\bar{x} - \mu)^2}_{Z} > \varepsilon^2) \leq \frac{E((\bar{x} - \mu)^2)}{\varepsilon^2}$$

$$= \frac{\mathrm{var}(\bar{x})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \longrightarrow 0$$

Weak LLN

so $P(|\bar{x} - \mu| > \varepsilon) \to 0$

sharpen bound:

Chernoff trick

$E(X_i) = 0 \qquad X_i \leftarrow X_i - \mu$

$X_1 \perp X_2 \qquad f(x_1, x_2) = f_1(x_1)\, f_2(x_2)$

$$E\left(h_1(x_1)\,h_2(x_2)\right) = \iint h_1(x_1)\,h_2(x_2)\,f_1(x_1)f_2(x_2)\,dx_1 dx_2$$

$$= \int h_1(x_1)f_1(x_1)\,dx_1 \int h_2(x_2)f_2(x_2)\,dx_2$$

$$= E\left(h_1(x_1)\right) E\left(h_2(x_2)\right)$$

$$\operatorname{cov}\left(h_1(x_1), h_2(x_2)\right) = 0$$ because $x_1 \perp x_2$

want to bound with $> t$

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n) > t$$ where $t > 0$

$$\Rightarrow \quad x_1 + \dots + x_n > nt$$

$$\underbrace{e^{\lambda(x_1 + \dots + x_n)}}_{Z} > \underbrace{e^{n\lambda t}}_{\text{new } t}$$ where $\lambda > 0$

$$P(\bar{x} > t) = P\left(e^{\lambda(x_1 + \dots + x_n)} > e^{n\lambda t}\right)$$

$$\leq \frac{E\left(e^{\lambda(x_1 + \dots + x_n)}\right)}{e^{n\lambda t}}$$ by markov inequality

$$= \frac{E\left(e^{\lambda x_1} e^{\lambda x_2} \cdots e^{\lambda x_n}\right)}{e^{n\lambda t}}$$

$$= \frac{E\left(e^{\lambda x_i}\right)^n}{e^{\lambda t \cdot n}} = \left(\frac{M(\lambda)}{e^{\lambda t}}\right)^n$$

becomes function of λ

Moment Generating Function: $M(\lambda) = E(e^{\lambda x})$

$M(0) = 1$

$M'(\lambda) = E(e^{\lambda x} \cdot x)$ $\qquad M'(0) = E(x)$

$M''(\lambda) = E(e^{\lambda x} \cdot x^2)$ $\qquad M''(0) = E(x^2)$

$\downarrow$

moments: derivatives @ 0

$e^{-n(\lambda t - \log M(\lambda))}$

$\downarrow$ maximize λ, λ will be > 0

$D(t) \rightarrow$ rate, D_{KL}

$P(\bar{x} > t) \leq e^{-n D(t)}$

"large deviation"

Connect to Gaussian

$X \sim N(0, \sigma^2)$

integrate over Gaussian identity

$$M(\lambda) = E(e^{\lambda x}) = \int e^{\lambda x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x^2 - 2\sigma^2 \lambda x)} dx$$

$$= \underbrace{\int \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}[(x^2 - \sigma^2\lambda)^2 - (\sigma^2\lambda)^2]}}_{N(\sigma^2\lambda, \sigma^2)} dx$$

$$= e^{\frac{1}{2}\sigma^2\lambda^2} \longrightarrow M(\lambda) \text{ of Gaussian}$$

Assum sub-Gaussian

$$X \sim f(x)$$

$$M(\lambda) \leq e^{\frac{1}{2}\sigma^2\lambda^2}$$

$$P(\bar{x} > t) \leq \left(\frac{M(\lambda)}{e^{\lambda t}}\right)^n \leq e^{n\left(\frac{1}{2}\sigma^2\lambda^2 - \lambda t\right)}$$

$$= e^{n \cdot \frac{1}{2}\sigma^2\left(x^2 - 2\frac{\lambda}{\sigma^2}t\right)}$$

$$= e^{n\frac{1}{2}\sigma^2\left(\left(\lambda - \frac{t}{\sigma^2}\right)^2 - \left(\frac{t}{\sigma^2}\right)^2\right)}$$

optimize = 0

$$\underbrace{P(\bar{x} > t) < e^{-\frac{nt^2}{2\sigma^2}}}_{\text{concentration inequality / Hoeffling / Bernstein}} \rightarrow \text{gaussian tail}$$

main idea: establish sub-gauss mgf $\rightarrow$ make bound

one more step to prove CLT. (small deviation)

$e^{\frac{1}{2}\sigma^2\lambda^2}$: MGF of $N(0, \sigma^2)$

Central Limit Theorem:

$E(X_i) = 0 \qquad Var(X_i) = 1$

$\bar{X} \cdot E(\bar{X}) = 0 \qquad Var(\bar{X}) = \frac{1}{n} \qquad \bar{X} \xrightarrow{P} 0$ $\leftarrow$ converges in probability

$E(\sqrt{n}\,\bar{X}) = 0 \qquad Var(\sqrt{n}\,\bar{X}) = 1$

$$\sqrt{n}\,\bar{x} \xrightarrow{D} N(0,1)$$

$$\begin{tikzcd}
\sqrt{n}\,\bar{x} \arrow[r, "D"] \arrow[d, "\mathrm{MGF}"'] & N(0,1) \arrow[d, "\mathrm{MGF}"] \\
M_{\sqrt{n}\bar{x}}(\lambda) \arrow[r] & e^{\frac{\lambda^2}{2}}
\end{tikzcd}$$

- convergence in distribution

$$E\left(e^{\lambda\sqrt{n}\,\bar{x}}\right) = E\left(e^{\lambda\sqrt{n}\,\frac{x_1+\dots+x_n}{n}}\right)$$

$$= E\left(e^{\frac{\lambda}{\sqrt{n}}(x_1+\dots+x_n)}\right) = \prod_{i=1}^{n} E\left(e^{\frac{\lambda}{\sqrt{n}}x_i}\right) \longrightarrow M\left(\frac{\lambda}{\sqrt{n}}\right)$$

recall $M(\lambda) = E(e^{\lambda x})$

$$= M\left(\frac{\lambda}{\sqrt{n}}\right)^n = \Big(\underset{\downarrow\, 1}{M(0)} + \underset{\downarrow\, 0}{M'(0)}\frac{\lambda}{\sqrt{n}} + \frac{1}{2}\underset{\downarrow\, 1}{M''(0)}\frac{\lambda^2}{n}$$

$$= \left(1 + \frac{1}{2}\frac{\lambda^2}{n}\right)^n$$

$$= e^{\frac{1}{2}\frac{\lambda^2}{n}n} = e^{\frac{\lambda^2}{2}}$$

* moment does not exist for heavy tailed functions

↳ use characteristic function instead

$$M(\lambda) = E(e^{\lambda x}) = \int e^{\lambda x} f(x)\,dx \rightarrow \text{Laplace transform}$$

Characteristic Function

$$\phi(w) = E\left(e^{iwx}\right)$$

$$= E\left(\cos wx + i\sin wx\right)$$

$$= \int e^{iwx} f(x)\,dx \longrightarrow \text{Fourier transform}$$

heat eqn

Strong Law of Large Numbers

$\forall \varepsilon, \exists N, \forall n \geq N, |\bar{x} - \mu| < \varepsilon$

↳ almost sure convergence : $P(\underbrace{\bar{X}_n \to \mu}_{A}) = 1$

$f(x) \sim \text{Unif}[0, 1]$

$A \subset \Omega = [0, 1]^\infty$ inf. dim cube

vol of $A = 1$

$\boxed{\text{Volume of diagonal piece}}$

$P(B = A^c) = 0$

$A_{n,\varepsilon} = \{(x_1, \ldots, x_n) : |\bar{x}_n - \mu| < \varepsilon\}$

$A_{n,\varepsilon} \subset [0,1]^n$

According to Weak LLN: $P(A_{n,\varepsilon}) \to 1$

$$A = \bigcap_{\substack{\varepsilon = \frac{1}{k} \\ \forall k}} \bigcup_{N=1}^{\infty} \bigcap_{n=1}^{\infty} A_{n,\varepsilon}$$

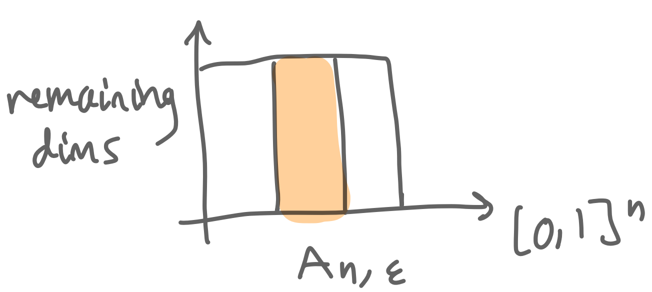
$A_{n,\varepsilon} \subset [0, 1]^n$ $\qquad A_{n,\varepsilon} = \{\underline{(x_1 \cdots x_n)} : \quad \}$

same volume

$A_{n,\varepsilon} \subset [0, 1]^\infty$ $\qquad A_{n,\varepsilon} = \{(\underline{x_1, \ldots, x_n}, \underbrace{x_{n+1}, \ldots}_{\text{remaining dims}}) : \}$

Lebsgue measure :

length, area, volume ...



basic shapes → measure

+ countable $\cup$ $\cap$ c

+ infinite additivity

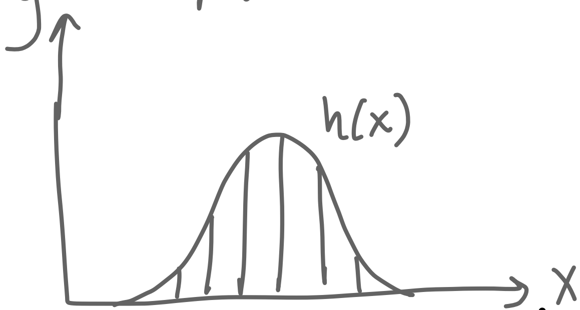
Wiener measure

{paths}

$\Longrightarrow$ σ algebra = {measurable sets $\subset \Omega$}

Lebsgue integral

Riemann



x: Can be complicated

h( ) : can be irregular

physical

Lebsgue

measurable function

$$\sum \mu(h(x) > y) \Delta y$$

Lebsgue measure

$\Delta y \to 0$

y : real #. simple

More generally defined

Convergence

$$f_i \longrightarrow f \Rightarrow \int f_i \to \int f$$

Final: . learning experience