Today's topic

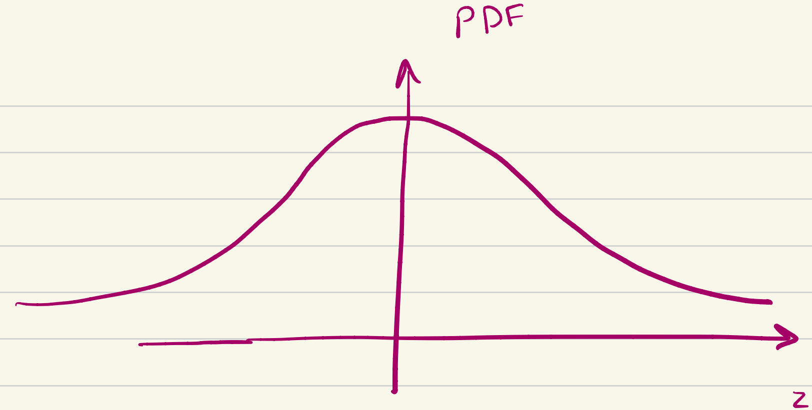
Gaussian distributions

CLT

Berry Esseen Theorem

Sub-Gaussians (Next lecture)

$$\phi(z) := \Pr[Z = dz] = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$



$$+\ Z_1 + Z_2 = Z$$

$Z_1 \to N(0,1)$, $Z_2 \to N(0,1)$, $Z \to N(0,2)$

CLT

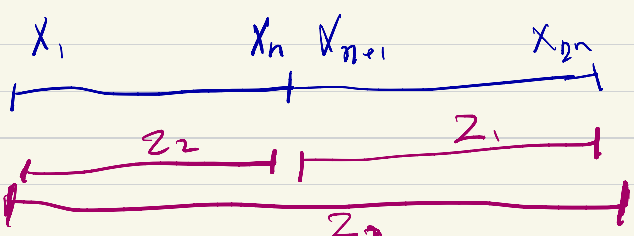
$E[X]$ and $Var[X] < \infty$

$X_1, \ldots, X_n$

$$\bar{X}_n = \frac{1}{n} \sum X_i$$

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \to Z$$

$N(0,1)$



$$\frac{1}{2n}\sum_{i=1}^{2n} X_i = \frac{1}{2}\left(\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i}_{Z_1} + \underbrace{\frac{1}{n}\sum_{i=n+1}^{2n} X_i}_{Z_1}\right)$$

$$+\ a Z_1 \quad + \quad b Z_2 \quad = \quad Z$$

$Z_1 \swarrow N(\mu_1, \sigma_1^2)$

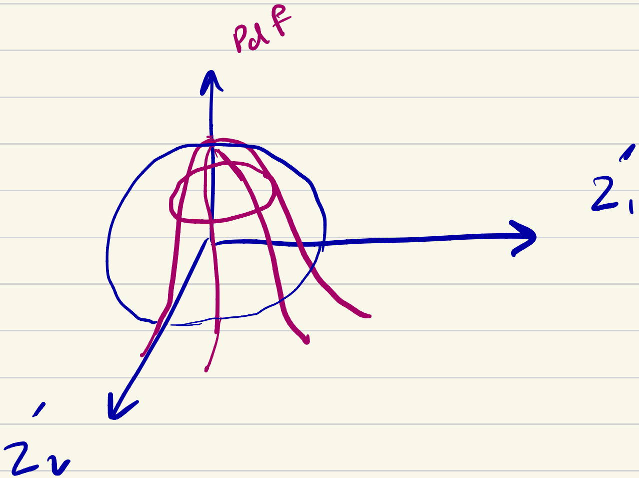
$Z_2 \swarrow N(\mu_2, \sigma_2^2)$

Z: $N(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2)$

$$Z_1 = \mu_1 + \sigma_1 Z_1' \qquad Z_1' \sim N(0,1)$$

$$Z_2 = \mu_2 + \sigma_2 Z_2' \qquad Z_2' \sim N(0,1)$$

$$a Z_1 + b Z_2 = (a\mu_1 + b\mu_2) + a\sigma_1 Z_1' + b\sigma_2 Z_2'$$



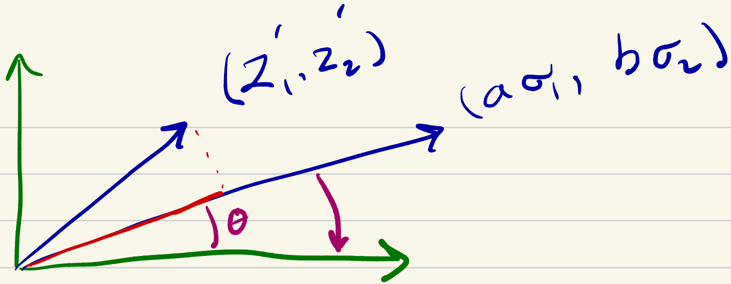
$$\mathrm{PDF}(z_1', z_2')$$

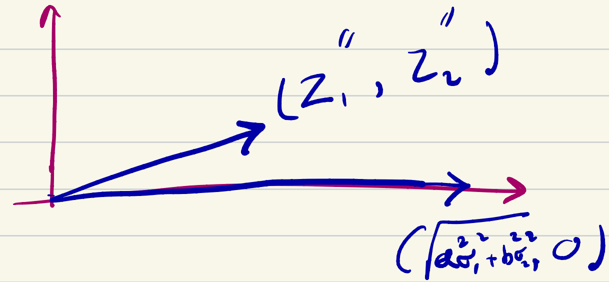
$$= \Pr[Z_1' = dz_1' \text{ and } Z_2' = dz_2']$$

$$= \frac{1}{2\pi} e^{-\frac{z_1'^2 + z_2'^2}{2}}$$

$$a\sigma_1 \cdot Z_1' + b\sigma_2 Z_2' = \langle (a\sigma_1, b\sigma_2), (z_1', z_2') \rangle$$

$*$





$$\Pr[Z_1' = z_1 \text{ and } Z_2' = z_2 \xrightarrow{\text{rotation } \theta} z_1'', z_2'']$$

$$= \Pr[Z_1'' = z_1'', \quad Z_2'' = z_2'']$$

$$* = \left\langle \left(\sqrt{a^2\sigma_1^2+b^2\sigma_2^2}, 0\right), \left(Z_1'', Z_2''\right) \right\rangle = \sqrt{a^2\sigma_1^2+b^2\sigma_2^2}\, Z_1''$$

$$\Rightarrow \quad aZ_1 + bZ_2 \sim N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

Berry Esseen Theorm

$$E[X_i] = 0 \quad , \quad Var[X_i] = \sigma^2$$

$$S_n = X_1 + X_2 + \cdots + X_n$$

$$E[S] = 0 \qquad Var[S] = 1$$

CLT

$E[X]$ and $Var[X] < \infty$

$$X_1, \ldots, X_n$$

$$\overline{X}_n = \frac{1}{n} \sum X_i$$

$$\underbrace{\frac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma}}_{Y_n} \to \underset{N(0,1)}{Z}$$

$$Y \to \sqrt{n}\,\frac{Y - E[Y]}{\sigma}$$

$$\lim_{n\to\infty} \sup_t \left| \Pr[Y_n \leq t] - \phi(t) \right| \to 0$$

B.E Thm:

CDF of $N(0,1)$

$$\forall t: \left| \Pr[S_n \le t] - \Phi(t) \right| \le C \cdot \sum_{i=1}^{n} E[X_i^3]$$

$\simeq 0.5$

≥ 0.01

---

version 2

$X_1 \ldots X_n$

$E[X_i] = 0 \quad E[X_i^2] = \sigma^2$

$E[|X_1|^3] = \rho < \infty$

$$S_n = \frac{X_1 + \ldots X_n}{n}$$

$$\left| \Pr\left[\frac{S_n \sqrt{n}}{\sigma} \le t \right] - \Phi(t) \right| \le \frac{C\,\rho}{\sigma^3 \sqrt{n}}$$

Example $X_1, \ldots, X_n \sim \mathrm{Ber}(1/2)$

$$X_1 = \begin{cases} +1 & \text{w.p. } 0.5 \\ -1 & \text{w.p. } 0.5 \end{cases}$$

$\mathrm{Var}[\mathrm{Ber}(1/2)] = p(1-p)$

$= \underbrace{\frac{1}{4}}_{\sigma_2}$

$\rho = 1$

$$S_n = \frac{\#(+1) - \#(-1)}{n}$$

$$\Pr\left[\frac{\#(+1)}{n} \geq \frac{1}{2} + \varepsilon\right] =$$

$$\Pr\left[\#(+1) \geq \frac{n}{2} + \varepsilon \cdot n\right] =$$

$$\Pr\left[\#(+1) \geq \frac{\#(1) + \#(-1)}{2} + \varepsilon n\right] =$$

$$\Pr\left[\frac{\#(+1) - \#(-1)}{2} \geq \varepsilon n\right]$$

$$= \Pr\left[\frac{S_n \cdot n}{2} \geq \varepsilon n \right] = \Pr\left[S_n \geq 2\varepsilon \right]$$

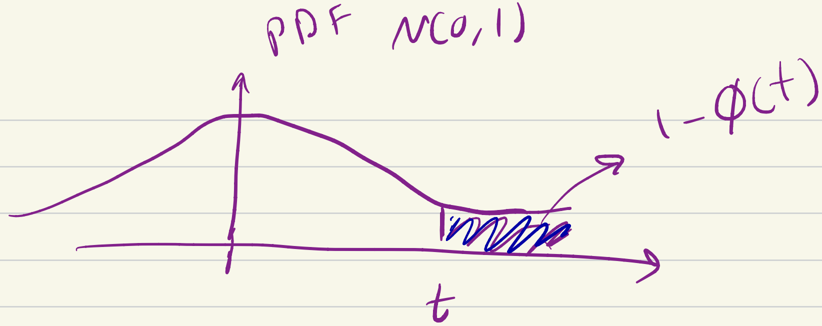
$$= \underbrace{\Pr\left[\frac{S_n \sqrt{n}}{\sigma} \geq \underbrace{\frac{2\varepsilon\sqrt{n}}{\sigma}}_{t} \right]}_{A}$$

$\Pr[Z \leq t]$

B-E:

$$\left| A - \underbrace{\left(1 - \phi_{\mathrm{CDF}}(t)\right)}_{} \right| \leq \frac{C \cdot E[|X_i|^3]}{\sigma^3 \sqrt{n}} \leq \frac{0.5 \cdot 1}{\frac{1}{8}\sqrt{n}}$$

$$\left| A - 0.05 \right| \leq 0.000001 \qquad \approx \Theta\left(\frac{1}{\sqrt{n}}\right)$$



$$t \cdot \square = \int_{z=t}^{\infty} t \cdot \phi(z)\, dz$$

$$\le \int_{z=t}^{\infty} z\, \phi(z)\, dz = -\phi(z)\Big|_{t}^{\infty}$$

$$= \underbrace{-\phi(\infty)}_{0} + \phi(t) = \phi(t)$$

$$\frac{d\,\phi(z)}{dz} = \frac{d}{dz}\left(\frac{1}{\sqrt{2\pi}} e^{-z^2/2}\right)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \cdot (-z)$$

$$= -\phi(z) \cdot z$$

$$\text{area} = \Pr[Z \geq t] \leq \frac{\phi(t)}{t} \approx \frac{e^{-t^2}}{t}$$

$$\left(\frac{1}{t} - \frac{1}{t^3}\right) \cdot \phi(t) < \Pr[Z \geq t] \leq \frac{\phi(t)}{t}$$

$$\Pr[Z \geq t] \leq e^{-t^2}$$