11/30/2021

CS70 LECTURE NOTES 25 NORMAL DISTRIBUTION

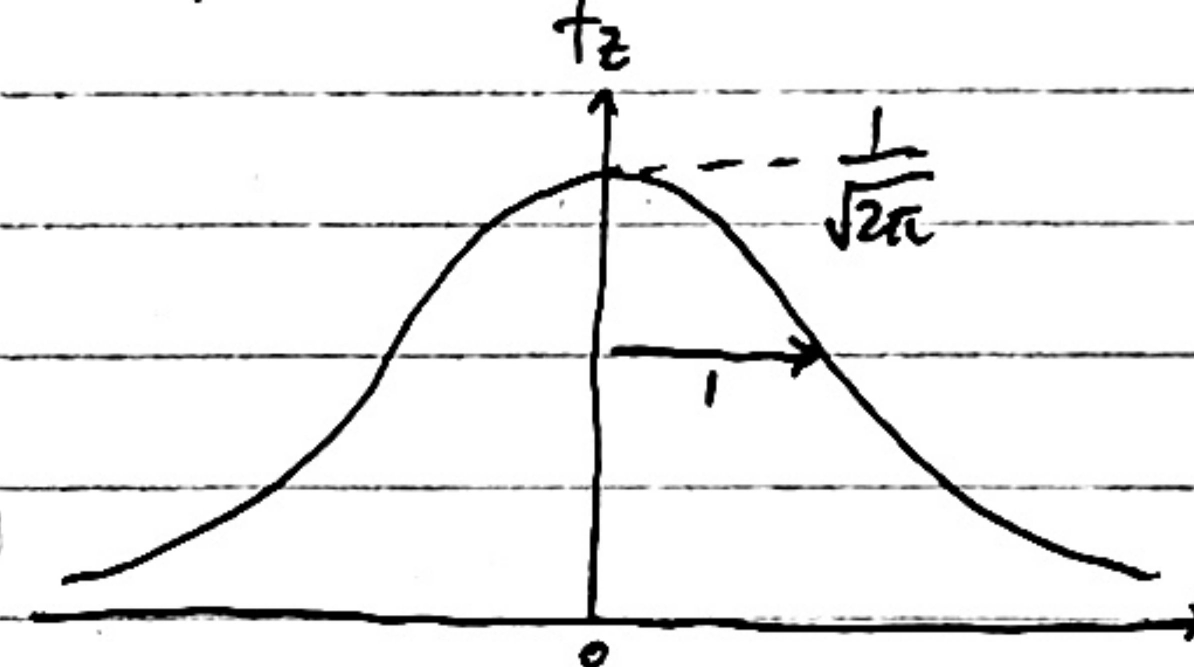
$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Z\left(\frac{x-\mu}{\sigma}\right)$$

Using chain rule $u(x) = \frac{x-\mu}{\sigma}$, $\frac{du}{dx} = \frac{1}{\sigma}$

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

If $\sigma < 0$, then $f_X(x) = \frac{-1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$

If $\sigma \neq 0$, $f_X(x) = \frac{1}{|\sigma|} f_Z\left(\frac{x-\mu}{\sigma}\right)$. $\boxed{X = \sigma Z + \mu}$

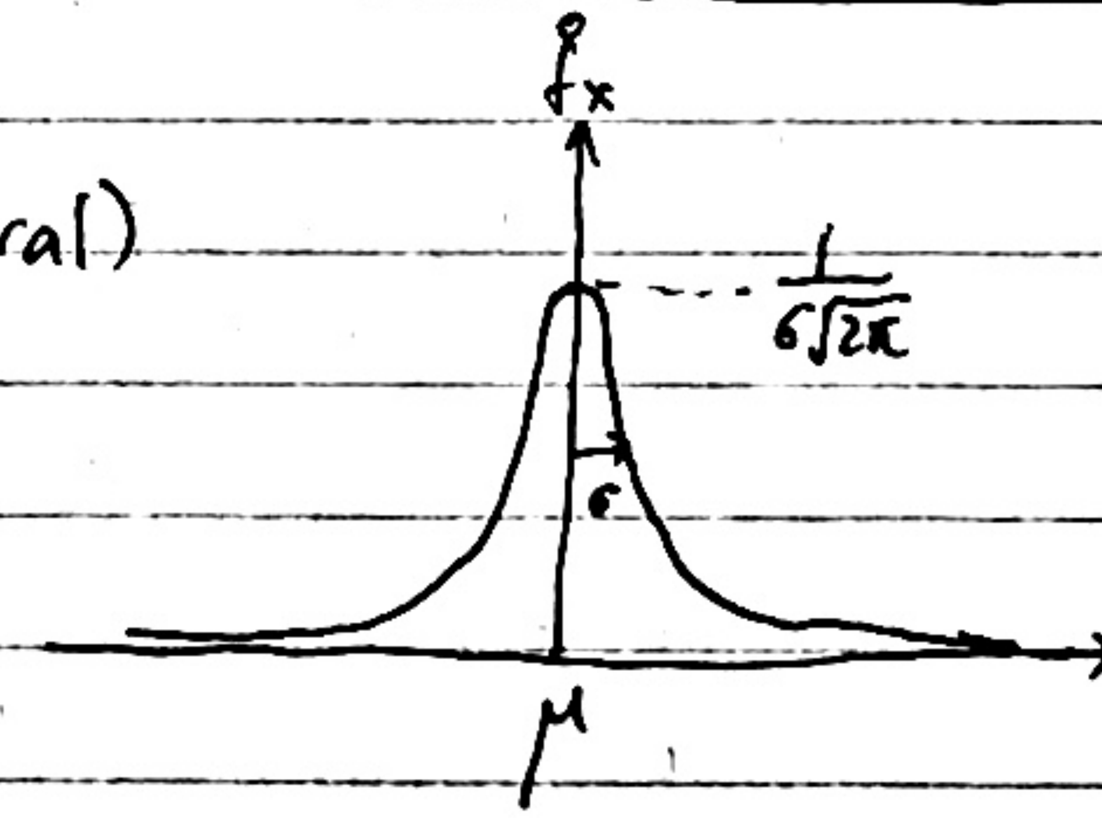


Example: $Z \sim N(0,1)$ i.e. $\mu = 0$, $\sigma^2 = 1$

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

substitute $z = \frac{x-\mu}{\sigma}$

Then $X = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$

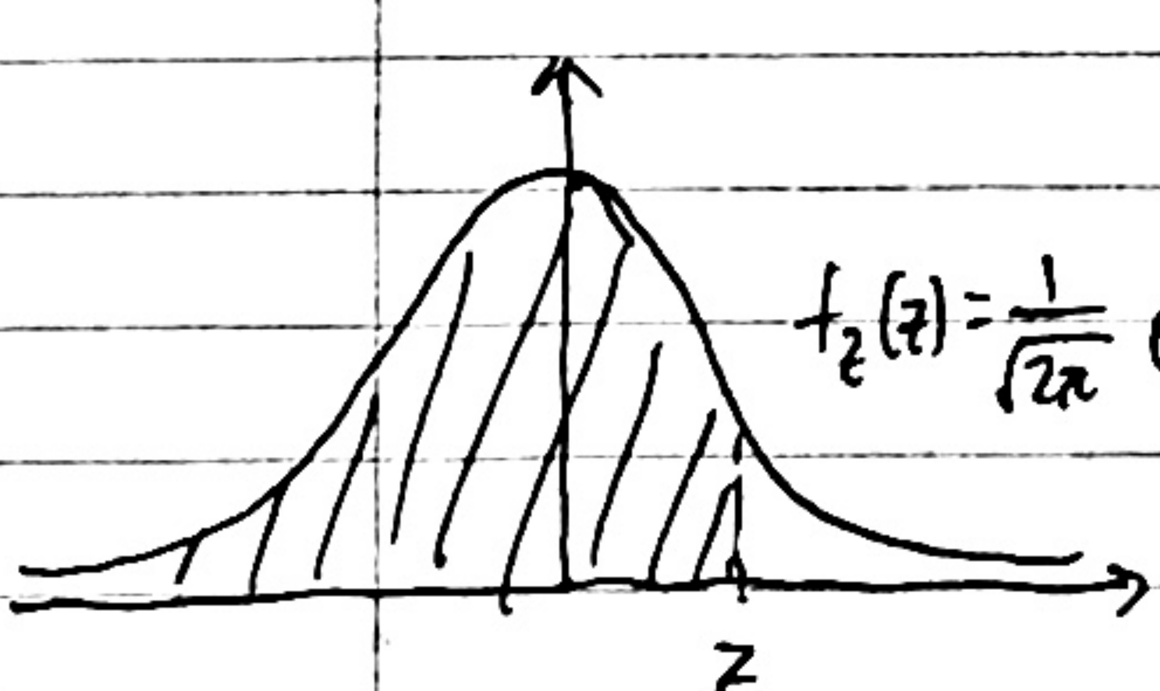


$E[X] = \mu$ $\quad \int_{-\infty}^{\infty} f_X(x)\,dx = 1$ (proof by double integral)

$\sigma_X^2 = \sigma^2$

$$F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{z^2}{2}} dz$$

No closed form expression



For nonstandard normal distributions
1. Standardize it
2. Refer to table or computation methods.

look up in the table

$$F_X(x) = P[X \le x] = P[X - \mu \le x - \mu] = P\Big[\underbrace{\frac{X-\mu}{\sigma}}_{Z} \le \frac{x-\mu}{\sigma}\Big] = P\left[Z \le \frac{x-\mu}{\sigma}\right]$$

Define $\Phi(\alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-\frac{z^2}{2}} dz = P[Z \le \alpha]$

$$P[Z > \alpha] = 1 - P[Z \le \alpha] = 1 - \Phi(\alpha)$$

$\therefore P[Z > 1] = 1 - \Phi(1) = 1 - 0.8413 = 0.1587$

$P[Z \le -\alpha] = P[Z > \alpha]$ $\quad P[-\alpha \le Z \le \alpha] = P[|Z| \le \alpha] = \Phi(\alpha) - \Phi(-\alpha)$

$= 2\Phi(\alpha) - 1$

$\alpha = 1 \quad \mathbb{P}[|Z| \leq 1] = \mathbb{P}[|X-\mu| \leq \sigma] = 0.6827 \quad (2 \cdot \underbrace{0.8413}_{\phi(1)} - 1)$

$\alpha = 2 \quad \mathbb{P}[|Z| \leq 2] = \mathbb{P}[|X-\mu| \leq 2\sigma] = 0.9545 \quad (2 \cdot \underbrace{0.9772}_{\phi(2)} - 1)$

$\alpha = 3 \quad \mathbb{P}[|Z| \leq 3] = \mathbb{P}[|X-\mu| \leq 3\sigma] = 0.9973 \quad (2 \cdot \underbrace{0.9987}_{\phi(3)} - 1)$

Central Limit Theorem

Let $X_1, \ldots, X_n$ be identical independent distributed random variables and $\mathbb{E}[X_i] = \mu \ \forall i$ and $\sigma^2 = \sigma_{X_i}^2 \ \forall i$

Assume $|\mu| < \infty$, $\sigma^2 < \infty$ (i.e. mean and variance finite)

Define sample mean $M_n = \frac{X_1 + \cdots + X_n}{n}$. Then $\mathbb{E}[M_n] = \mu$.
$\sigma_{M_n}^2 = \frac{\sigma^2}{n}$

Define $Z_n = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\Rightarrow \mathbb{E}[Z_n] = 0$ and $\sigma_{Z_n}^2 = 1$ (Does not matter what n is)

CLT says $\lim_{n\to\infty} F_{Z_n}(z) = \lim_{n\to\infty} \mathbb{P}[Z_n \leq z] = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$

(pointwise convergence)

i.e. As n increases, the cumulative probability function converges to the normal CDF function
No assumptions on distribution of X_is (could be discrete whatsoever).

Pollster Problem Revisited

μ: $X_i = 1$
$1-\mu$: $X_i = 0$

$M_n = \frac{X_1 + \cdots + X_n}{n}$ is an estimate of μ.

How many people do we poll s.t. $\mathbb{P}[\underbrace{\mu}_{\text{true mean}} - \varepsilon \leq M_n \leq \mu + \underbrace{\varepsilon}_{\text{accuracy}}] \geq \underbrace{0.95}_{\text{confidence level}}$

Set $\varepsilon = 0.03$

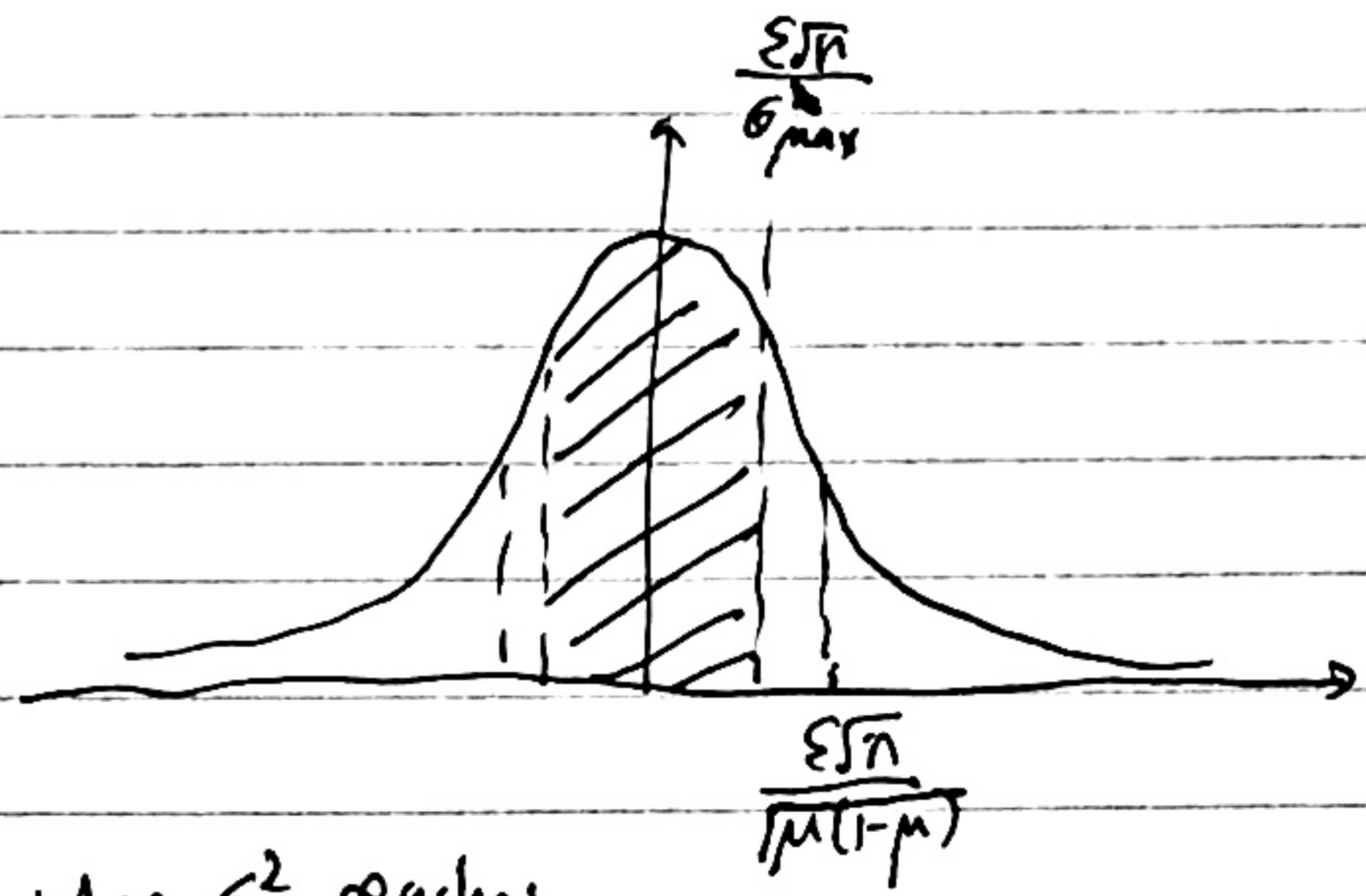
$\mathbb{P}[\mu - \varepsilon \leq M_n \leq \mu + \varepsilon] = \mathbb{P}[-\varepsilon \leq M_n - \mu \leq \varepsilon]$
$= \mathbb{P}[|M_n - \mu| \leq \varepsilon] \geq 0.95$

Standardize. Let $Z_n = \frac{M_n - \mu}{\sigma_{M_n}} = \frac{M_n - \mu}{\frac{\sigma}{\sqrt{n}}}$

$\mathbb{P}[|M_n - \mu| \leq \varepsilon] \geq 0.95 \Rightarrow \mathbb{P}\left[\frac{|M_n - \mu|}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\varepsilon}{\frac{\sigma}{\sqrt{n}}}\right] \geq 0.95$

$$\mathbb{P}\left[\frac{|M_n-\mu|}{\frac{\sigma}{\sqrt{n}}} \le \frac{\varepsilon\sqrt{n}}{\sigma}\right] \ge 0.95$$

For a Bernoulli, $\sigma^2 = \mu(1-\mu)$



So $$\mathbb{P}\left[\frac{|M_n-\mu|}{\frac{\sigma}{\sqrt{n}}} \le \frac{\varepsilon\sqrt{n}}{\sqrt{\mu(1-\mu)}}\right] \ge 0.95$$

Consider the worst case scenario. Consider when σ^2 reaches its maximum. $\sigma_{max}^2 = \frac{1}{4}$

So $$\frac{\varepsilon\sqrt{n}}{\sqrt{\mu(1-\mu)}} \longrightarrow \frac{0.03\sqrt{n}}{\frac{1}{2}}$$

$\Rightarrow 2\Phi\left(\frac{0.03\sqrt{n}}{\frac{1}{2}}\right) - 1 \ge 0.95 \Rightarrow$ can get a lower bound on n (Inverse look up on the probability table)