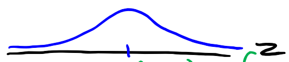
CS 70 ## 4.23 LECTURE 26

GAUSSIAN RVs

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



even; $E(Z) = 0$ $\int f_Z(z)\,dz = 1$

$E(Z) = \int \underbrace{z}_{\text{odd}} \underbrace{f_Z(z)\,dz}_{\text{even}} = 0$ $E(Z^2) = \int \underbrace{z^2}_{\text{even}} \underbrace{f_Z(z)\,dz}_{\text{even}} = \sigma_Z^2 = 1 \longrightarrow \sigma_Z = 1$

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

$F_Z(z) = \int_{-\infty}^{z} f_Z(z)\,dz$ $P_r(Z \le \alpha) = \Phi(\alpha)$ $P_r(Z > \alpha) = 1 - \Phi(\alpha)$

$P(-\alpha \le Z \le \alpha) = 2\Phi(\alpha) - 1$

$P(|Z| \le 1) = 2\Phi(1) - 1 = 2 \cdot 0.8413 - 1 = 0.6827$ one standard deviation

$P(|Z| \le 2) = 2\Phi(2) - 1 = 2 \cdot 0.9772 - 1 = 0.9545$ two standard deviations

$P(|Z| \le 3) = 2\Phi(3) - 1 = 2 \cdot 0.9987 - 1 = 0.9974$ three standard deviations

STANDARDIZING A RV

X, $E(X)$, σ_X^2 define intermediate RV $Y = X - E(X)$ $E(Y) = 0$

variance hasn't changed (scalar shift) $\sigma_Y^2 = \sigma_X^2$

$\sigma_Y = \sigma_X$, which we can use to get a unit-variance RV. $Z = Y/\sigma_Y$

$\sigma_Z^2 = \left(\frac{1}{\sigma_Y}\right)^2 \sigma_Y^2 = 1$ now, Z has $E(Z) = 0$ and $\sigma_Z^2 = 1$, regardless of X!

$Z = (X - E(X))/\sigma_X$ will always convert $X \sim N(\mu, \sigma^2)$ to $Z \sim N(0,1)$

$X = \sigma Z + \mu$ this operation is reversible $E(X) = \sigma E(Z) + E(\mu) = \mu$

$\sigma_X^2 = \sigma^2 \,\mathrm{var}(Z) = \sigma^2$

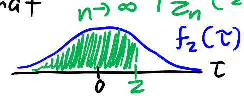
CENTRAL LIMIT THEOREM

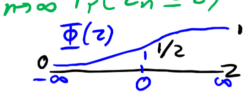
Suppose $X_1, X_2, \ldots, X_n$ are IID with $\mu = E(X)$ and $\sigma^2 = \sigma_X^2$.

Sample mean $M_n = \frac{X_1 + X_2 + \ldots + X_n}{n}$ where $E(M_n) = \mu$ and $\sigma_{M_n}^2 = \sigma^2/n$.

Let $Z_n = \frac{M_n - \mu}{\sigma}$. The CLT says that $\lim_{n\to\infty} F_{Z_n}(z) = \lim_{n\to\infty} P_r(Z_n \le z) = \Phi(z)$

But what does $\Phi(z)$ look like?





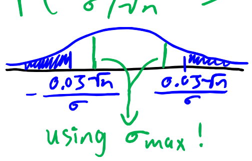
POLLSTER PROBLEM REVISITED

μ: $X_i = 1$; $1-\mu$: $X_i = 0$ — i^{th} voter believes in legalization. Estimate $\mu \rightarrow$ how many people should be polled so $P_r(\mu - \epsilon \le M_n \le \mu + \epsilon) \ge 0.95$?

$P_r(-\epsilon \le M_n - \mu \le \epsilon) \rightarrow P_r(|M_n - \mu| \le \epsilon) \ge 0.95 \longrightarrow P(|M_n - \mu| > \epsilon) \le 0.05$

Let $\epsilon = 0.03$. Standardize M_n: $Z_n = \frac{M_n - E(M_n)}{\sigma_{M_n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(M_n - \mu)}{\sigma}$

$P\left(\frac{|M_n - \mu|}{\sigma/\sqrt{n}} > \frac{0.03}{\sigma/\sqrt{n}}\right) \le 0.05 \longrightarrow P\left(|Z_n| > \frac{0.03\sqrt{n}}{\sigma}\right) \le 0.05$ CLT!



$\sigma^2 = \mu(1-\mu) \rightarrow \sigma_{max}^2 = 0.25 \rightarrow \sigma_{max} = 1/2$

$P(|Z_n| > \underbrace{0.06\sqrt{n}}_{1.96}) = 2(1 - \Phi(0.06\sqrt{n})) \le 0.05$

$1.96 \longrightarrow 1.96 \le 0.06\sqrt{n} \rightarrow \boxed{n \ge 1068}$