Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 11
Lecture - 40
Large Sample Confidence Intervals

ye hum confidence interval ke baare mein dekh rahe the aur ye paribhasha humne pichhle samay ant mein dekhi thi, jise main phir se dohrata hoon: hamare paas ek random sample hai normal distribution se mean mu aur variance sigma² ke saath, jahan sigma² humein exactly pata hai; kisi bhi alpha ke liye hum 100·(1–alpha)% ka confidence interval for mu nikalna chahte hain; agar alpha = 0.05, to hum 95% confidence interval dekh rahe hain; method wahi hai: sample mean x-bar ke around interval x-bar ± Z(α/2)*(σ/√n); is interval ki length depend nahi karti x-bar par, balki depend karti hai alpha, sigma, aur n par; pichhle example mein hamara n=31, sigma=2, x-bar=80; alpha=0.05 (95%) ke liye interval tha 79.296 se 80.704, length = 1.408; ab alpha=0.1 (90%) ke liye Z(0.05)=1.645, to interval banta hai 79.4090 se 80.5910, length = 1.182; ab alpha=0.01 (99%) ke liye Z(0.005)=2.58, to interval banta hai 79.0732 se 80.9268, length = 1.8536; ab hum do sawal puchte hain: agar 99% confidence interval humein aur zyada confidence deta hai, to phir 95% ya 90% interval kyon use karein? answer: kyunki jitna zyada confidence lenge, utni hi interval ki length badh jayegi 99% mein interval sabse lamba, 95% mein medium, 90% mein sabse chhota; matlab agar humein high precision (chhota interval) chahiye, to hum lower confidence choose karte hain, agar humein high confidence chahiye, to broad interval accept karna padta hai ye trade-off statistics ka fundamental concept hai.1.18, 95% confidence interval ki length se kam hai jo 1.4 thi, aur jaise-jaise hum confidence badhata ja rahe hain waise-waise interval ki length bhi badh rahi hai 99% confidence interval ki length 1.85 ho rahi hai, to yeh jo length hai wo precision ko darshati hai; hamara confidence to badh raha hai lekin precision ghat raha hai; agar humein 100% confidence chahiye to vo tab hoga jab hum kahenge ki humara interval poori real line hai usmein mu zaroor aayega but precision zero ho jayegi, completely useless; humein chhota-chhota interval chahiye to confidence kam hota jayega; yeh ek give-and-take relationship hai precision aur confidence ke beech; agar hum chaahein to hum confidence ko fixed rakh kar interval width ko limit kar sakte hain; example: agar response time of editing command is normally distributed with sigma = 25 ms and mu unknown, and we need a 95% confidence interval for mu with maximum width 10, to length formula 2·z(α/2)·σ/√n use karte hain; σ=25, α=0.05 → z(0.025)=1.96, to 2·1.96·25 / √n ≤ 10; solve karne par √n = 9.8, so n = 96.04, but sample size integer hota hai so minimum n = 97; is tarah desired confidence aur desired precision dono achieve kiye ja sakte hain by choosing proper n; upar wale examples mein humne assume kiya tha ki sigma known hai, lekin real world mein sigma unknown hota hai; isliye large-sample confidence interval for mean mu use hota hai; population mean mu aur population variance sigma² ke saath sample size large ho to x-bar approximately normal hota hai by central limit theorem; to z = (x-bar – mu) / (sigma/√n) approximately standard normal; but jab sigma unknown hai to usko replace karte hain s se (sample standard deviation), so large-sample 100·(1–α)% confidence interval for mu banta hai:

$$P\left(-\tau_{\frac{\alpha}{2}} < \left(\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}\right) < \tau_{\frac{\alpha}{2}}\right) = 1-\alpha$$

yahan humne distribution ke shape ko assume nahi kiya, sirf mean aur variance ka exist hona zaroori hai; isi tarah hum $sigma^2$ ke liye bhi confidence interval nikal sakte hain using estimator s^2 aur uske distribution ke basis par, lekin wo hum is course mein detail mein nahi kar rahe sirf itna jaanna kaafi hai ki kisi bhi parameter ke liye confidence interval estimator ke distribution ya approximate distribution par depend karta hai.