Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 9
Lecture - 32
Distribution of sample Mean

wo likh rahe hain expected value kya hai, bilkul wahi aata hai 0.3; expected value of x bar = 0.3, jo humare original Bernoulli population ka bhi mean tha; ab variance of x bar bhi calculate kar lete hain variance ke hisaab se pehle expected value of x bar² nikalte hain: x² times probability lenge, to milega: $0^2 \cdot 0.49 + (1/2)^2 \cdot 0.42 + 1^2 \cdot 0.09 = 0 + 1/4 \cdot 0.42 + 0.09 = 0.105 + 0.09 = 0.195$; phir variance of x bar = E[x bar²] − (E[x bar])² = $0.195 - (0.3)^2 = 0.195 - 0.09 = 0.105$; ab compare karo population variance = 0.21 tha, aur sample mean ka variance (jab n=2) aa gaya 0.21/2 = 0.105 bilkul same; ye result proof karta hai jo proposition me likha tha:

$$E(\bar{X}) = \mu$$

$$V(\bar{X}) = \frac{\sigma^2}{N}$$

Aur approach perfectly match karta hai; isliye x bar ek unbiased estimator hai population mean ka, aur jaise jaise sample size n badhta hai, variance of sample mean reduce hota hai (sigma² / n), matlab uncertainty kam hoti hai aur accuracy badhti hai population mean ko estimate karne me; isi liye sample mean statistics aur data-science me sabse zyada upyog hone wala statistic hai.ye expected value se bilkul milti hai jo hamare proposition me likha tha agar population ka mean mu hai to expected value of x bar bhi mu rahegi; variance kya hoga x bar ka sigma square / n jo original variance sigma square hai usko n se bhaag denge, yahan hamara n = 2; chalo dekhte hain kya ye example me sahi hai, to pehle variance of x bar nikalne ke liye expected value of x bar² nikalte hain: sum over $x^2 \cdot p(x\ bar)$; to hoga $0^2 \cdot 0.49 + (1/2)^2 \cdot 0.42 + 1^2 \cdot 0.09 = 0 + (1/4)*0.42 + 0.09 = 0.105 + 0.09 = 0.195$; ab variance of x bar = $0.195 - (0.3)^2 = 0.195 - 0.09 = 0.105$; ye 0.105 bilkul 0.21 / 2 ke barabar hai, kyunki population ka variance 0.21 tha; fir expected value of T bhi nikal sakte hain T = n·x bar isliye expectation = 2·0.3 = 0.6; variance of T = $n^2 \cdot Var(x\ bar) = 4 \cdot 0.105 = 0.42$; n=2 ke liye PMF hum directly bana sakte hain, lekin kyunki T bas 2·x bar hai, isliye expectation aur variance directly calculate ho jate hain; ab agar aap n=3 try karo, x bar = (x1 + x2 + x3)/3, to uska expectation fir bhi 0.3 hi milega aur variance hoga sigma²/3 = 0.21/3 = 0.07, jo 0.105 se chhota hai; fir ham histogram plots dekhte hain n=1 me Bernoulli 0.3 ka distribution, 0 ke upar bar height 0.7 aur 1 ke upar bar height 0.3; n=2 ke liye bars 0.49, 0.42, 0.09; n=3 ke liye values 0, 1/3, 2/3, 1 par bars dikhenge; n=5 aur n=10 ke liye bars aur smooth ho jate hain ye sampling distribution approach ko visually confirm karta hai; jaise jaise n badhta hai sampling distribution of x bar narrow hota hai, spread kam hota hai, estimation zyada accurate hota hai; theory me proposition me sirf population ka mean mu aur variance sigma^2 ka assumption tha distribution kaisa hai ye nahi maana gaya; lekin jab specific case ho normal population ka (N(mu, sigma^2)), to hum x bar aur T ka exact distribution also identify kar sakte hain, aur ye hum next steps me theorem ke form me dekhenge; x1, x2, … xn random sample from normal population ho to unke basis par x bar ke distribution ke baare me hum aur zyada strong result likh sakte hain.i equal to 1 se n tak sum xi divided by n hum continue karenge notation x bar ka hi use karke, taaki notation me complication na aaye, lekin x bar n pe dependent hota hai; tab theorem ke hisaab se x bar ka distribution bhi normal hota hai; matlab x bar ~ Normal(mu, sigma^2/n); pichhle proposition me humne sirf mean aur variance assume kiye the, lekin jab population ka distribution normal ho ye extra assumption dete hi conclusion much stronger ho jata hai; phir sample total T ka distribution bhi normal hota hai n·mu mean ke saath aur n·sigma^2 variance ke saath; ab isko ek real-life example me use karte hain: maan lo loan application form bharne ke liye lagne wala samay normally distributed hai mean 10 minutes aur standard deviation sigma = 2 minutes; pehle din 5 log form bhare unke samay x1 se x5 doosre din 6 log form bhare unke samay y1 se y6; ye sab random variables independent aur identically distributed hain normal(10, 4); ab sawal: probability sample average time 11 minutes sena ho? yani P(x bar ≤ 11 and y bar ≤ 11); x bar aur y bar independent hain isliye total probability = P(x bar ≤ 11) · P(y bar ≤ 11); now x bar ~ Normal(10, 4/5), sigma_xbar = 2/√5; normalize karte hain: P(x bar ≤ 11) = P((x bar−10)/(2/√5) ≤ (11−10)/(2/√5)) = P(Z ≤ 1.12) = 0.8686; similarly y bar ~ Normal(10, 4/6), sigma_ybar = 2/√6; normalize: P(y bar ≤ 11) = P((y bar−10)/(2/√6) ≤ (11−10)/(2/√6)) = P(Z ≤ 1.22) = 0.8888; total probability = 0.8686 × 0.8888 = 0.7720 (rounded to 4 decimal places); isliye lagbhag 77.2% chance hai ki dono din ke sample averages 11 minutes se kam honge; aur ye sab possible hua kyunki x bar aur y bar ka exact distribution normal hume pata tha, na ki sirf mean aur variance.