Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 11
Lecture - 39
Invariance Principle; Confidence Intervals

Hum dekh rahe the maximum likelihood estimators (MLE) jise hum short me MLE bolte hain uchchatam sambhavyata aakalan; estimator (aalkak) hota hai random variable aur estimate (aakalan) hota hai uski observed numeric value; isme hum likelihood function f(x1, x2, …, xn) ko parameters ke relation me maximize karte hain; F ko maximize karna equivalent hota hai log F ko maximize karne ke, aur log-likelihood ke derivatives lena jyada aasaan hota hai; jab humne normal distribution ke liye

$$\hat{\mu} = \bar{x}$$

$$\hat{\sigma}^2 = (\frac{1}{n})\sum_{i=1}^{n}(x_i - \bar{x})^2$$

nikala tha tab dekha ke σ^2-hat unbiased nahi hota lekin phir sawal aata hai: fir bhi MLE kyun use karein jab unbiasedness ek desirable property hai? Kyunki MLE ki kuch apni bahut mazboot properties hoti hain; unme se ek important property hai invariance principle; iska arth ye: agar $\theta_1, \theta_2, ..., \theta_m$ ke MLE respectively $\hat{\theta}_1, \hat{\theta}_2 \dots \hat{\theta}_m$ nikale ja chuke hain, aur $h(\theta_1, \theta_2, ..., \theta_m)$ koi function hai, to h ka MLE directly $h(\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m)$ hoga; h ke liye alag se likelihood likhne, derivative lene aur solve karne ki zaroorat nahi; chhota example: normal distribution me σ^2-hat mila = $(1/n)\sum(x_i - \bar{x})^2$; hume standard deviation σ ka MLE chahiye; kyunki $\sigma = \sqrt{\sigma^2}$, to direct: σ-hat = $\sqrt{(\sigma^2\text{-hat})}$; hume fir se derivative lene ki koi zaroorat nahi; dusra example: binomial(10, p) sample general joint likelihood

$f(x_1, \dots, x_n; p) = \prod$ (10 choose xi) p^Σxi (1−p)^{Σ(10−xi)},

$$f(x_1, x_{2\dots;} p) = \prod_{i=1}^{n}\binom{10}{x_i}p^{x_i}(1-p)^{(10-x_i)} = \prod_{i=1}^{n}\binom{10}{x_i}p^{\sum x_i}(1-p)^{\sum(10-x_i)}$$

$$\ln(f) = \ln(\prod_{i=1}^{n}\binom{10}{x_i}) + (\sum x_i)\ln(p) + (\sum(10 - x_i))\ln(1-p)$$

log-likelihood log f = const + (Σxi)log p + (Σ(10−xi))log(1−p); derivative se solve karne par p-hat = Σxi / (10 n) = $\bar{x}/10$ (MLE), aur agar h(p) = p^3 ya p/(1−p) ya e^p to MLE respectively h(p-hat) hi hoga; yahi hai invariance principle MLE transformations ke liye stable rehta hai, jo is method ko powerful banata hai.log of p aur ye hai log of 1–p; isko agar derivative lenge aur 0 se equate karenge, to derivative

d(ln f)/dp me pehla term 0 hoga, doosre term me aayega $(\Sigma xi)/p$ aur teesre term me aayega $+ \Sigma(10–xi) * (1/(1–p)) * (–1)$ to ye ban jayega $– \Sigma(10–xi)/(1–p)$; isko 0 ke sath equate karte hain to milta hai $(\Sigma xi)/p = (\Sigma(10–xi))/(1–p)$, reciprocals lene ke baad milta hai $p/(\Sigma xi) = (1–p)/(\Sigma(10–xi))$, aur cross-multiply karke milega $p(\Sigma(10–xi)) = (1–p)(\Sigma xi)$; expand karne par: $p(10n – \Sigma xi) = \Sigma xi – p(\Sigma xi)$ to $–p(\Sigma xi)$ dono side cancel ho jayega aur bachega $10np = \Sigma xi$, aur isse directly p-hat = $(\Sigma xi)/(10n)$ milta hai, jo $\bar{x} / 10$ ke barabar hai, yani MLE of p = sample mean/10; agar hume binomial 10p ka variance ka MLE chahiye jo hota hai 10p(1–p) to by invariance principle, sigma²-hat = 10 * p-hat * (1–p-hat) = $10*(\bar{x}/10)*(1–\bar{x}/10)$; ab ek bahut important caution example– uniform (0, θ) population: waiting times at bus stop uniform distribution parameter θ (unknown), sample values x1…xn observed; likelihood function f = $\prod(1/\theta) = 1/\theta^n$ only if $\theta \geq$ maximum sample value, otherwise zero; as function of θ, f differentiable nahi hai kyunki threshold boundary par discontinuity hoti hai calculus fail ho jata hai; isliye yahan MLE direct reason se milta hai: maximize f tab hota hai jab θ bilkul smallest possible ho jo sample max se chhota na ho, isliye θ-hat = max(x1, x2, ..., xn); yahi maximum likelihood estimate; aur corresponding estimator capital X(n), jo sample ka maximum hota hai; is example se pata chalta hai ke MLE hamesha calculus se nahi milta; agar likelihood differentiable nahi ho to logical reasoning se maximize karte hain; MLE ek strong method hai estimators dhoondhne ke liye, invariance principle jese strong properties sath me deti hai; ab hum agle topic: confidence intervals par badhte hain jaise example: population normal with unknown mean μ, known sigma=2, sample size n=31, sample mean = 80; hum μ ke liye 95% confidence interval nikalna seekhenge.kafi karanon ke kaaran hai jisse sample mean is population mean ka bahut accha estimate hai ye unbiased hai, method of movement se bhi humein yehi milta hai, MLE bhi x-bar hi hota hai mu ka, to ye bahut accha estimator hai; lekin sirf ye number 80 dene se zyada information nahin milti, agar uske saath standard error bhi report karein to behtar hota; isliye hum confidence interval ke baare mein dekh rahe hain; humein pata hai ki is situation mein natural estimator mu ka x-bar hai; x-bar ka EXACT distribution bhi humein pata hai x-bar normally distributed hota hai, mean mu aur variance sigma²/n ke saath, yahan sigma² = 4/31; phir agar hum x-bar ko standardize karein, $(\text{x-bar} – \text{mu})/(2/\sqrt{31})$, to ye standard normal Z hota hai; standard normal curve symmetric hota hai zero ke around; hum jaante hain ki −1.96 se 1.96 ke beech Z ke aane ki probability 0.95 hai; isi se milta hai: probability$[−1.96 < (\text{x-bar} – \text{mu})/(2/\sqrt{31}) < 1.96] = 0.95$; isse algebra se mu ko beech mein rakh kar milta hai: probability$[\text{x-bar} – 1.96*(\text{sigma}/\sqrt{n}) < \text{mu} < \text{x-bar} + 1.96*(\text{sigma}/\sqrt{n})] = 0.95$; yahan mu random nahi hai—ye ek fixed unknown constant hai, randomness x-bar mein hai, jo sample se aaya; jab hum actual sample values daalte hain jaise x-bar = 80, sigma = 2, n = 31 tab humko fixed numeric interval milta hai: 79.296 se 80.704; ab is statement ko "probability mu is in [79.296, 80.704] = 0.95" nahi bol sakte, kyunki yahan koi randomness nahi hai; mu bhi fixed, interval bhi fixed; sahi interpretation ye hai: agar hum sampling baar-baar repeat karein, to aise 95% intervals mu ko cover karenge; ye jo humne procedure use kiya, uska confidence 95% hai; isliye iska naam hai

95% confidence interval; general definition: agar X1..Xn normal population se ho jahan sigma known ho, to mu ka (1–alpha) confidence interval hota hai: x-bar ± $Z(\alpha/2)$ * sigma/$\sqrt{n}$; aur interval ki length hoti hai: 2 * $Z(\alpha/2)$ * sigma/$\sqrt{n}$; yahin par hum aaj rok rahe hain agle session mein confidence interval ko aur detail mein dekhenge.