Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 11
Lecture - 38
Maximum Likelihood Estimation

abhi hum dekhenge method of maximum likelihood, jise hum bolte hain uchchatam sambhavyata aakalan; agar humare paas population ka koi parameter theta unknown hai, to hum uske baare me jaan'ne ke liye sample lete hain x1, x2, …, xn, aur unke observed values hoti hain small x1, x2, …, xn; phir hum ek statistic choose karte hain jo estimator hota hai ye random variable hota hai jo sample values ka function hota hai, aur jab hum actual sample ke numbers usme dal dete hain to hume estimate milta hai; ab maximum likelihood method kya karta hai: binaomial wale example me population binomial(10, p) hai, jisme n=10 fix hai, only p unknown hai; hum sample ke roop me iid do observations lete hain x1, x2;
Joint PMF

$$p(x_1, x_2) = p_{x_1}(x_1)p_{x_2}(x_2)$$

$$= \binom{10}{x_1} p^{x_1}(1-p)^{(10-x_1)} \binom{10}{x_2} p^{x_2}(1-p)^{(10-\ _2)}$$

inka joint PMF hota hai marginal ka product: p(x1)p(x2) = (10 choose x1)p^{x1}(1−p)^{10−x1} × (10 choose x2)p^{x2}(1−p)^{10−x2}; ab suppose observe value hai x1=4, x2=7; to joint PMF (jo ab likelihood function ban jayega) hota hai: L(p) = (10 choose 4)(10 choose 7)p^{11}(1−p)^9; isme sirf p hi variable hai, baaki sab constants; maximum likelihood principle kehta hai: wo p choose karo jo likelihood function ko maximum kare; intuition: probability ki jagah hum ulta soch rahe hain sample fix hai, p varying; question: "kis p ke liye ye sample sabse zyada sambhavy (likely) hota?"; ab is concept ko general bana kar likhte hain: let x1…xn have joint PMF/PDF f(x1,x2,…,xn ; theta), aur theta unknown parameter hai; tab L(theta) = f(x1,x2,…,xn ; theta) ko likelihood function kehte hain; phir maximum likelihood estimator theta-hat wo value hoti hai jo L(theta) ko maximize karti hai; maximum likelihood estimation ka basic concept.jaise humne bahut saare examples me dekha binomial distribution depend karta hai parameter p par, normal distribution depend karta hai parameters mu aur sigma par jab hum koi distribution study kar rahe hote hain, to uska joint PMF ya joint PDF kuch unknown parameters par depend karta hai, jinhe hum denote karte hain theta1, theta2, …, thetaM se; method of moments me bhi humne general approach dekha, waise hi maximum likelihood me bhi hum ek general formal definition de rahe hain: humare paas joint PMF/PDF f(x1,x2,…,xn; theta1,…,thetaM) hai, jisme theta1…thetaM unknown parameters hain; jab hum sample values observe karte hain small x1…xn, tab function f as a function of thetas hi likelihood function hota hai; maximum likelihood estimates (MLE) theta1_hat…thetaM_hat wo values hoti hain jo likelihood function ko maximize karti hain; corresponding estimators (random-variable versions) bhi same notation se denote kiye jate hain theta1_hat…thetaM_hat aur ye hota hai function sample values ka; ab wapas chalte hain hamare binomial example par: likelihood function L(p) = (10 choose 4)(10 choose 7) p^{11}(1−p)^9; hume p ki wo value chahiye jo is L(p) ko maximum kare; directly maximize karna mushkil hota hai isliye log-likelihood lete hain: lnL =

ln(10C4)+ln(10C7)+11 ln p + 9 ln(1−p); constant terms 0 ho jati hain derivative me, isliye derivative of lnL w.r.t p hota hai 11/p + 9*(−1/(1−p)) = 11/p − 9/(1−p); isko 0 ke barabar rakhte hain: 11/p = 9/(1−p) ⇒ cross multiply: 9p = 11(1−p) = 11−11p ⇒ 20p = 11 ⇒ p = 11/20; ye hamara MLE aata hai; isliye hum conclude karte hain: maximum likelihood estimate of p based on sample (4,7) is 11/20 ≈ 0.55; isi tarah general cases me bhi hum likelihood maximize karke theta ke estimates nikalte hain.P hat = 11/20; to phir ek baar main dahurata hoon—humne likelihood function liya, uska log likelihood function liya, fir derivative liya aur usko 0 ke barabar rakhkar equation solve kari, jisse p hat = 11/20 mila; ye extreme point maximizer bhi ho sakta hai ya minimizer bhi, isliye second derivative check karna zaroori hota hai; humne d^2/dp^2 log f calculate kiya: $-11/p^2 - 9/(1-p)^2$ jo clearly < 0 rehta hai har p ∈ (0,1) ke liye, matlab ye point maximum hai, isliye MLE of p = 11/20; aur dekha jaa sakta hai ke ye (4 + 7)/(10 + 10) ≈ 0.55 = x̄ se coincide bhi karta hai; doosra example exponential distribution ka exponential(lambda); likelihood hai
Joint pdf

$$f(x_1, x_{2\ldots\ldots}x_n; \lambda) = \prod_{i=1}^{n} f_{x_i}(x_i; \lambda)$$

$$= \prod_{i=1}^{n} (\lambda e^{-\lambda x_i})$$

$$lnf(x_1, x_{2\ldots\ldots}x_n; \lambda) = n \, ln \, \lambda - \lambda \sum_{i=1}^{n} x_i$$

log likelihood: n ln λ − λ∑xi; derivative wrt λ: n/λ − ∑xi = 0 ⇒ λ hat = n / ∑xi = 1 / x̄; second derivative $-n/\lambda^2 < 0$ ⇒ max; noting that E[1/x̄] ≠ 1/E[x̄], to MLE unbiased nahi hota; teesra example normal distribution $N(\mu, \sigma^2)$: joint PDF product ke form me; log likelihood $-(n/2)\ln(2\pi) - (n/2)\ln\sigma^2 - (1/(2\sigma^2))\sum(x_i-\mu)^2$; derivative wrt μ: ∑(xi−μ)=0 ⇒ μ hat = x̄; derivative wrt σ^2: $-(n/(2\sigma^2)) + (1/(2\sigma^4))\sum(x_i-\mu)^2 = 0$ ⇒ σ^2 hat = $(1/n)\sum(x_i-\bar{x})^2$; yahan dhyan dene wali baat: yeh sample variance ka biased version hai kyunki denominator n hai, n−1 nahi; MLE $sigma^2$ ke liye $s^2 = 1/n \sum(x_i-\bar{x})^2$ deta hai; to summary: normal case me MLE of μ is x̄ and MLE of σ^2 is $(1/n)\sum(x_i-\bar{x})^2$; maximum likelihood method systematically likelihood maximize karke estimators produce karta hai jo aksar achhi statistical properties rakhte hain, especially large samples me.