Introduction to Probability & Statistics

Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi

Week - 10

Lecture - 36

Standard Error of Estimators

ham dekh rahe hain inference jiska matlab hai hume population ke unknown parameters jaise mu, sigma square, p, lambda ya generic theta ke baare me jaankari prapt karni hai; iske liye hum random sample x1, x2, …, xn lete hain aur ek statistic choose karte hain jo estimator kehlata hai yeh estimator ek random variable hota hai kyunki yeh sample values ka function hota hai; last time humne dekha tha ki estimator agar unbiased ho to accha hota hai, aur unbiased hone ke saath–saath jiska variance kam ho use hum aur zyada prefer karte hain; lekin estimator to random hota hai, jabki actual data milne par jo number hum compute karte hain usko kehte hain estimate; yani estimator = akalak (random variable), estimate = akalan (actual number jo sample values se nikla); jab hum theta ko estimate karte hain to use hum theta hat se denote karte hain, par dono contexts me kabhi random, kabhi computed value symbol same hota hai aur context se clear ho jata hai; ab agar hum ek statistician ki tarah estimate report karte hain, jaise ki "estimate = 5.3", to sirf number dena kaafi nahi hota hume yeh bhi batana hota hai ki estimate theta ke kitne paas hai, yani estimate ki precision kya hai; isi liye hum define karte hain precision: ki estimator theta hat values θ ke around kitni close hain isko hum measure karte hain standard error se; formally: standard error of theta hat hai uska standard deviation, jisko hum likhte hain $\sigma(\hat{\theta})$; isme dikkat yeh hoti hai ki variance ya sd nikalte waqt unknown parameter θ ki value involve hoti hai, jaise uniform$(0,\theta)$ wale example me humne dekha tha variance formula me θ^2 aata tha; jab standard error nikalte hain to uski value hoti hai $\theta / \sqrt{n(n+2)}$; par yahan θ unknown hai, to hum real–world me standard error directly compute nahi kar sakte; isliye hum uska estimate banate hain estimate of standard error ko likhte hain $\hat{\sigma}(\hat{\theta})$, jisme unknown θ ko hum replace kar dete hain uske estimate se; yahi statistical inference ka kaam hai estimate + precision dono saath report karna. hai yeh kaise nikaalte hain estimate of the standard error, sigma-hat theta-hat, is gotten by replacing theta by its estimate theta-hat; to previous uniform-0-theta example me sigma(theta-hat) tha theta / $\sqrt{n(n+2)}$, to sigma-hat(theta-hat) me hum theta ko replace kar denge theta-hat se, yani sigma-hat(theta-hat) = $(n+1)/n \cdot$ x(bracket n) / $\sqrt{n(n+2)}$; isi tarah jab hum kisi point estimate ko report karte hain to hum estimate ke saath uska standard error bhi batate hain, aur agar standard error directly calculate nahi ho sakta to estimated standard error batate hain; example ke liye Bernoulli population ke sample me hamara estimator x-bar = p-hat hota hai, jisme variance(p-hat) = $p(1-p)/n$, to standard error = $\sqrt{p(1-p)/n}$, lekin kyunki p unknown hota hai hum estimated standard error use karte hain jisme p ko replace kar dete hain p-hat se; dono earlier samples me n = 100 tha sample-1 me p-hat = 0.43 to sigma-hat(p-hat) = $\sqrt{0.43\times0.57/100}$ = 0.0495; sample-2 me p-hat = 0.49 to sigma-hat(p-hat) = $\sqrt{0.49\times0.51/100}$ = 0.0499; to pehle sample ke liye hum report karenge estimate "0.43 ± 0.0495" aur dusre ke liye "0.49 ± 0.0499", jisse pata ki dono estimates mein precision similar hai aur dono approximately population parameter p ke kareeb hain. to hum aise report kar sakte hain ki estimate of p, sample size 100 par, 0.43 hai aur uska estimated standard error 0.0495 hai, aur sample-2 ke liye hum report karenge ki hamara point estimate 0.49 hai with standard error 0.0499; ab example lete hain random sample from normal population jiska mean mu aur variance sigma-square hai pehle case me sigma = 2 known hai to sirf ek unknown parameter mu bachta hai, jiska best estimator x-bar hi hai kyunki x-bar unbiased bhi hai aur MVUE bhi; is case me variance(x-bar) = 4/n to standard error = 2/√n aur agar n=25 ho aur sample ka x-bar =12.3 ho to mu-hat=12.3 aur standard error=2/√25=0.4 yahan koi estimation ki zarurat nahi; lekin agar sigma unknown ho to sigma(theta-hat)=sigma/√n hota hai aur usko hum estimate karte hain sigma-hat/√n se jahan sigma-hat =√s-square; agar sample variance s^2=4.4 mila to sigma-hat=√4.4 aur sigma-hat(mu-hat)=√4.4/√25=0.4195; ab sawal ye hai ki standard error report karna zaroori kyun hai jab n large hota hai to CLT ke tehat estimator theta-hat approx normal hota hai mean theta aur variance $sigma^2$(theta-hat) ke saath, isliye (theta-hat − theta)/sigma(theta-hat) approx standard normal hota hai, aur hume pata hai standard normal ka 95% area −1.96 aur +1.96 ke beech hota hai, isliye P(−1.96 ≤ (theta-hat − theta)/sigma(theta-hat) ≤ 1.96)=0.95, jo algebraic rearrangement ke baad ban jata hai P(theta-hat −1.96·sigma(theta-hat) ≤ theta ≤ theta-hat +1.96·sigma(theta-hat))=0.95; iska matlab hai ki yeh interval ek random interval hai jisme left endpoint aur right endpoint random hain, aur 95% chance hai ki theta ki true value is interval ke andar hogi; agar standard error chhota hai to interval chhota hoga aur precision zyada agar standard error bada hai to interval bada aur precision kam; ab ek non-standard pdf example lete hain: x1, x2, …, xn random sample from population with pdf f(x;theta)=1/[2(1+x·theta)] for x in (−1,1), aur 0 otherwise, jahan theta ek unknown parameter hai jisko hum f(x;theta) notation se darsha rahe hain, aur isi theta ke baare me hume inference karna hai.actually humne abhi jo kiya wahi karenge agar population ka standard distribution hota to mean variance already known hota, par yahan non-standard distribution diya hai to pehle hum uska mean dhoondte hain; expected value of xi nikalne ke liye hum integral lenge −1 se +1 tak half × (x + x^2·theta) dx, integral solve karne par value aati hai theta/3; iska matlab xi ka mean theta/3 hai, aur isi se hume milta hai expected value of $\bar{x}$ = theta/3 aur expected value of 3·$\bar{x}$ = theta, isliye theta-hat = 3·$\bar{x}$ ek unbiased estimator ho gaya; ab standard error ke liye variance of theta-hat nikalenge: variance(theta-hat) = variance(3·$\bar{x}$) = 9·variance($\bar{x}$), jahan variance($\bar{x}$) = $sigma^2$ / n; ab hume $sigma^2$ chahiye, to expected value of xi^2 integrate karte hain: −1 se 1 tak half × (x^2 + x^3·theta) dx; yahan x^3·theta ka symmetric integral cancel ho jata hai aur result milta hai 1/3; phir variance(xi) = E[xi^2] − (E[xi])^2 = 1/3 − (theta/3)^2 = 3/9 − $theta^2$/9 = (3 − $theta^2$)/9; isliye variance(theta-hat) = 9 × (3 − $theta^2$)/9n = (3 − $theta^2$)/n aur standard error = sqrt((3 − $theta^2$)/n); par theta unknown hai to hum estimated standard error nikalte hain jahan theta ki jagah theta-hat rakha jata hai; ab sample values lete hain: 0.2, −0.3, 0.01, 0.23, −0.75; inka sum −0.61 aata hai aur $\bar{x}$ = −0.61/5 = −0.122, isliye theta-hat = 3 × (−0.122) = −0.366; ab estimated standard error = sqrt(3 − (theta-hat)^2)/sqrt(n) = sqrt(3 − (−0.366)^2)/sqrt(5) ≈ 0.1997.