Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 9
Lecture - 31
Sampling Distributions

pichhle time humne dekh liya tha ek definition jisko main abhi dohraane wala hoon statistic fir ek baar main dohraata hoon: ye statistic spelling mistake nahi hai, ye "s" ant me nahi hai jaise subject statistics me hota hai; statistic ko main alag shabdon me likh raha hoon: statistic is any function whose values can be calculated from sample values, yaani statistic is a function of observations; humne examples dekhe the sample mean x-bar, sample total T, sample variance S^2, sample standard deviation S; ye sabhi random variables hote hain; ab ek example dekhte hain: coin toss jo pichhle baar dekha tha par is baar hum ek toss kar rahe hain, to x_1 aur x_2 dono Bernoulli(0.3) random variables hain, independent; to x-bar = $(x_1 + x_2)/2$ bhi ek random variable hai; hum uska distribution nikal sakte hain: x_1 aur x_2 values le sakte hain: 0-0, 0-1, 1-0, 1-1, aur x-bar accordingly hoga: 0, 1/2, 1/2, aur 1; kyunki independence hai, joint PMF $P(x_1,x_2) = P(x_1)\cdot P(x_2)$; to values milti hain: P(0,0)=0.49, P(0,1)=P(1,0)=0.21, P(1,1)=0.09; ab x-bar ke PMF me x-bar=0 tabhi aayega jab x_1=0 aur x_2=0, so P(x-bar=0)=0.49; x-bar=1 tabhi jab x_1=1,x_2=1 so P(x-bar=1)=0.09; x-bar=1/2 dono cases me x_1=1,x_2=0 or x_1=0,x_2=1 to P(x-bar=1/2)=0.21+0.21=0.42; to humne x-bar ka PMF calculate kar liya aur dekha ki sample mean bhi ek random variable hai jiska distribution hum sample data aur independence ke saath derive kar sakte hain. calculate kiya hai humne X-bar ka PMF calculate kiya hai; therefore PMF of X-bar (jisko main darshaaoonga $P_{\bar{X}}(x)$ se) X-bar teen alag-alag values le sakta hai: x = 0, x = 1/2, x = 1, aur unki probabilities hain x=0 par 0.49, x=1/2 par 0.42, x=1 par 0.09, aur baaki kisi bhi x ke liye probability 0; to is tarah humne sample mean ka distribution find kiya jab n=2 hai aur X1, X2 Bernoulli(0.3) hain; isi tarah agar hamare paas 3 independent Bernoulli(0.3) random variables ho to X-bar = (X1+X2+X3)/3 ka PMF bhi isi tarah nikaal sakte hain pehle joint PMF likhenge independence use karke, phir possible value-combinations se X-bar ke outcomes nikaalenge; matlab sample statistic ka distribution dhoondhna kabhi-kabhi zaroori hota hai; ab hum population ka concept samajhte hain statistics me: population ka matlab yahan "desh ki nahi balki wo saare objects jin par study ki ja rahi hai jaise sari tubelights ek factory ki, ya sari marks ek class ke bachchon ke; unme se ham sample lete hain sample of size n aur sampling do tarah se hoti hai: with replacement jahan chosen element dobara bhi choose ho sakta hai, aur without replacement jahan chosen member side me rakh diya jata hai aur phir dobara nahi aata; aur phir definition: X1,X2,…,Xn ko hum random sample kehte hain agar (1) observations independent ho aur (2) unka distribution ek jaisa ho, yaani identical; isko alag notation me likhte hain: independent and identically distributed IID; independent hona condition-1 se aata hai aur identically distributed hona condition-2 se; concept hum future me repeatedly use karenge statistics ke analysis me. vo likh rahe hain aur point number two hai ki sabhi ka distribution same hai to ve identically distributed hain, to random sample agar bolenge to iska matlab iid hota hai;

dono ka meaning exactly same independent aur identically distributed; iid ko hum Hindi me bol sakte hain svaroopi aur swatantra bantit; population kya hai? wo sari units jinse study ho rahi hai; unka ek distribution hota hai, aur hum us population se sample lete hain; agar sampling exactly iid conditions ko satisfy kare lekin practical world me population infinite nahi hota, aur sampling generally without replacement hoti hai; phir bhi, agar sample size n bahut chhota hai population size N ke mukable (numerically $n/N < 0.05$), to iid assumption ek bahut accha approximation hota hai; hence random sample ≈ iid; ab ek example: maan lijiye bank transaction time exponential(λ) distribution follow karta hai; aur maan lijiye X1 aur X2 is distribution se ek random sample hain, yaani X1 aur X2 independent aur identical exponential(λ) random variables; practical interpretation: bank me do customers randomly choose kiye gaye unke transactions ke waqt X1 aur X2 measured hue; total waiting time = T = X1 + X2 ye ek statistic hai, sample values ka function; sample size yahan n = 2 hai; X1 aur X2 continuous random variables hain, to T bhi continuous hoga; PDF of T nikalne ke liye first step: joint PDF of (X1,X2) kyunki sample iid hai, joint PDF = product of marginal PDFs; fx1,x2 = fx1(x1) * fx2(x2); aur exponential distribution ka marginal pdf hota hai: λ e^(−λx1) × λ e^(−λx2), jab x1 ≥ 0 aur x2 ≥ 0, warna joint pdf = 0; ab T ka distribution nikalne ke liye hum CDF ka use karenge; t > 0 ke liye

$$P(T \le t) = P(X_1 + X_2 \le t)$$

$$= \int_0^t \int_0^{t-x} (\lambda^2 e^{-\lambda x_1}\, e^{-\lambda x_1}) dx_2 dx_1$$

$$= \int_0^t \lambda e^{-\lambda x_1} [1 - e^{-\lambda(t-x_1)}$$

$$P(T \le t) = \left(1 - e^{-\lambda t}\right) - \lambda t e^{-\lambda t}$$

F_T(t) = P(T ≤ t) = P(X1 + X2 ≤ t); isko hum double integration se calculate karenge over the region x1≥0, x2≥0, x1+x2 ≤ t; is tarah se hum T ka CDF paayenge aur differentiate karke T ka PDF nikal lenge; isi tarah se hum bade n ke liye sample total aur sample mean ke distributions bhi study karenge ye statistics ke core tools hain.joint PDF jo hai vo isko main is tarah se calculate kar raha hoon: probability calculate karni hai is ghatna ki, to continuous random variables ke liye do integrals lagenge joint PDF ko integrate karenge us region par jahan condition x1 + x2 ≤ t satisfy hoti hai; isliye integration limits honge x1 from 0 to t, aur x2 from 0 to t − x1, aur joint PDF hai λ² e^(−λx1) e^(−λx2); kyunki x2 wala integral ke liye e^(−λx1) constant hai, to usko bahar nikal sakte hain aur andar bachega integral 0 se t−x1 tak λ e^(−λx2) dx2 jo exponential CDF deta hai: 1 − e^(−λ(t−x1)); phir total integration milega: 0 to t tak λe^(−λx1) dx1 − 0 to t tak λe^(−λt) dx1, jo simplify karke banega: 1 − e^(−λt) − λt e^(−λt); ye hai CDF of T; ab PDF chahiye to differentiate karenge: derivative of 1 is 0; derivative of e^(−λt) is −λe^(−λt); aur derivative of λt e^(−λt) chain rule se milega λ² t e^(−λt); final result: f_T(t) = λ²

$t e^{-\lambda t}$ for $t > 0$, aur $f_T(t) = 0$ for $t < 0$; is tarah se humne sample total ka distribution nikaal liya iska PDF mil gaya; aur ye jo distributions of statistics hum nikaalte hain jaise pichle example me $\bar{x}$ (sample mean) ke liye, aur is example me sample total ke liye ise kehte hain sampling distribution.