Introduction to Probability & Statistics
Prof. Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 5
Lecture - 16
Binomial Distribution

Aaj hum kuch standard discrete distributions ke baare me pata karenge; saadhaaran asatatmantam hum pehle hi dekh chuke hain; pehla udaharan hum dekhenge jo pehle se padh chuke hain: ek random variable jo sirf do values deta hai 0 aur 1; is random variable ke maan 1 lene ki probability ko hum P se darshate hain to probability X=0 ho jati hai 1–P; yeh ek discrete distribution hai kyunki P + (1–P) = 1 ho jata hai; P ko probability of success bola jata hai, ya jaise coin toss me heads ko success bola jaye to P uski probability hoti hai; aur P hamesha 0 aur 1 ke beech hota hai, extreme cases me P=0 aur P=1 bhi include kiye jaate hain; yeh jo random variable hai usko hum Bernoulli random variable ya Bernoulli distribution with parameter P bolte hain; jaise maine kaha, P ki values 0 se 1 ke beech hoti hain; iske kuch generalization hum aage dekhenge; sabse simple udaharan hai coin toss: ek toss me do parinaam head ya tail aur hum labels assign kar dete hain 1 aur 0 ke roop me; ishi ko hum Bernoulli random variable kehte hain; ab isko aage badhate hain: agar hum coin ko do baar uchhaalein, har uchhal me 2 parinaam aur dono independent trial hote hain, to total parinaam honge 2 × 2 = 4 jo multiplication rule ke hisaab se aata hai; 4 outcomes honge: HH, HT, TH, TT; ab is par agar hum ek random variable X define karein, ki X = number of heads in two independent tosses of the coin, to HH me X=2, HT me X=1, TH me X=1, aur TT me X=0; to X ki values hoti hain 0,1,2. Jo maine abhi bola, yeh value leta hai 0 agar outcome TT ho, yani dono toss me tail aaye to heads ki sankhya 0 hoti hai; X=1 kab hota hai Jab outcome HT ya TH ho, aur X=2 tab hota hai jab outcome HH ho; yani X jo humne define kiya hai number of heads in two tosses uske teen possible maan hain: 0, 1, aur 2; agar probability of H = p aur probability of T = 1–p ho, to hum probability ke niyam ke hisaab se is random variable X ka PMF nikaal sakte hain; P(0) hoga probability ki X=0 yani TT ho, jo independence ke kaaran $(1–p)\times(1–p) = (1–p)^2$ hota hai;

$$p(x) \geq 0 \quad \sum_{x=0}^{2} p(x) = (1-p)^2 + 2p(1-p) + p^2$$

isi tarah P(1) nikaalte hain: X=1 jab HT ya TH ho sakta hai, HT aur TH exclusive events hain, to unki probabilities add hongi; probability of H in first toss = p aur T in second = (1–p) to p(1–p); TH ke liye T first toss aur H second toss, jo (1–p)p hoga; dono milake P(1)= 2p(1–p); aur P(2) hoga probability of HH = $p\times p = p^2$; to PMF hai: $P(X=0) = (1–p)^2$, $P(X=1)=2p(1–p)$, $P(X=2)=p^2$; kisi bhi dusre real X ke liye P(X)=0 hoga; har term non-negative hai, kyunki $(1–p)^2 \geq 0$, $p^2 \geq 0$, aur $2p(1–p)$ bhi ≥ 0 kyunki p aur (1–p) dono 0 se 1 ke beech hain; ab agar in teeno probabilities ko jodenge to milta hai $(1–p)^2 + 2p(1–p) + p^2$; ye bilkul a+b whole square ke expansion jaisa hai: $a^2+2ab+b^2$, jahan a=(1–p) aur b=p; to $P(0)+P(1)+P(2)= (1–p+p)^2 = 1^2 = 1$; to is tarah confirm hota hai ki ye PMF valid probability mass function hai. Toh ye jo binomial expansion isme aata hai uske kaaran is tarah ke jo yaadrit chikchar hote hain woh ek family me aa jaate hain, unko bolte hain binomial random variables ya binomial distribution; ye kaise aata hai isko main iski satik paribhasha me thodi der me dunga, lekin pehle dekhte hain ki binomial random variable ka udaharan kaise milta hai humne abhi dekha ek sikke ko do baar uchhalna; lekin yahan sikka important nahi hai, concept ye hai ki ek trial me parinaam sirf do hi ho sakte hain jaise coin toss me H ya T ya assembly line par component inspection me "defective" ya "not defective" do hi outcomes ek ko hum kahenge success aur ek ko kahenge failure; "success" ya "failure" ke naam ka meaning important nahi hai, yeh sirf labels hain; har trial me success aane ki probability fix hoti hai, pehle trial me ho ya paanchve trial me, woh badalti nahi usko hum P se darshate hain; P ka maan 0 se 1 ke beech hota hai; agar koi experiment in 4 conditions ko satisfy kare to use hum bolte hain binomial experiment; aur agar binomial experiment ho to random variable X define kiya jaata hai as number of successes in n trials; yeh supremely important model hai aur bahut jagah use hota hai; jo example humne pehle dekha do tosses, heads ki sankhya, jahan P heads ki probability hai woh isi general model ka special case tha, wahan n=2 tha; general binomial random variable ke do parameters hote hain n aur p jahan n = number of trials aur p = probability of success in each trial; notation: X follows binomial distribution with parameters (n, p); yahan n ek dhan-purnaank hota hai (1,2,3,…), aur p ka maan 0 se 1 ke beech; extreme cases p=0 aur p=1 se variable random nahi rehta, balki ek constant random variable ban jaata hai agar p=1 to har trial success hoga, agar p=0 to har trial failure hoga; remark: agar X ~ binomial(1, p) ho to ye exactly Bernoulli(p) distribution ke barabar hota hai, jaisa humne earlier dekha tha; ab general binomial random variable X ~ binomial(n, p) ko thoda detail me samajhte hain:

$$p(0) = p(X = 0) = p(FF..F) = (1 - p)^n$$

$$p(n) = p(X = n) = p(SS \ldots..S) = (p)^n$$

X kaun se values le sakta hai? agar n trials hain to X ke possible maan honge: 0 (agar sab failures aaye), 1 (agar ek success aaye), 2,3,…,n (agar sab successes aaye); yani X ki values 0 se n tak total n+1 values; ab PMF nikaalte hain: P(X=0) = probability ki sab failures aaye = (1–p)^n; P(X=n) = probability ki sab successes aaye = p^n; P(X=1) = probability ki exactly ek success aaye; iske liye ek trial success aur baaki sab failures general counting ke concept se multiple ways possible hote hain, jaisa hum aage derive karenge; poora PMF hum likhenge but key idea ye hai ki yeh binomial distribution hai jisme essential structure hota hai: repeated independent Bernoulli trials, fixed success probability p, aur X = number of successes; yeh hi binomial random variable ki aadhaarbhut paribhasha hai. Usme hume chahiye ki iske parinaamon ko dekhen, to ek parinaam ho sakta hai pehle trial me S aaya aur baaki sab trials me F aaya, is parinaam ki probability hogi p (success ke liye) aur baaki n–1 trials ke liye (1–p), aur kyunki sab trials independent hain to is parinaam ki total probability hogi p(1–p)^(n–1); agar X=1 hota hai to matlab sirf ek hi success hua hai, aur ye success kisi bhi trial me aa sakta hai — pehle, dusre, teesre… kisi bhi ek trial me S aayega aur baaki sab me F; har aise parinaam ki probability same hogi p(1–p)^(n–1), lekin ye parinaam kitne alag–alag tareeqon se ho sakta hai? Hume n slots me se 1 slot choose karna hai jahan S aayega, aur baaki sab F honge; ye n choose 1 = n tareeqon se ho sakta hai; isliye P(X=1) = nC1 · p(1–p)^(n–1); isi ko aage badhate hue P(X=2)= nC2 · p^2 · (1–p)^(n–2), kyunki hume n slots me se 2 slots choose karne hain jahan S aaye aur baaki me F; general case me P(X=x) = nCx · p^x · (1–p)^(n–x), jab x = 0,1,2,…,n ho, aur baaki sab x values ke liye P(X)=0; is tarah is binomial random variable ke PMF ko define kiya jaata hai; aur agar in saari probabilities ka sum karein from x=0 to n: Σ [nCx · p^x · (1–p)^(n–x)] to ye binomial expansion ke hisaab se (p + (1–p))^n = 1^n = 1 hota hai, isliye PMF valid hai; agar hum n=1 rakhein to ye PMF bilkul Bernoulli(p) ke barabar ho jaata hai; ab expected value E[X] nikaalte hain: E[X] = Σ x · P(X=x) from x=0 to n; x=0 term zero hoti hai, baaki terms ke liye formula apply karke simplify karne par E[X] = n · p milta hai; yaani binomial random variable ka apekshit maan hota hai np; variance ke liye Var(X)=E[X^2] – (E[X])^2; aur ye bhi derive kiya ja sakta hai ki Var(X)= np(1–p); derivation is tarah hota hai: note karo X(X–1)=X^2–X; to E[X(X–1)] = E[X^2] – E[X]; aur E[X(X–1)] nikaalne par milta hai n(n–1)p^2; isko rearrange karne par E[X^2]= n(n–1)p^2 + np; iske baad Var(X)=E[X^2] – (E[X])^2 = [n(n–1)p^2 + np] – (np)^2 = np(1–p); is tarah hum conclude karte hain ki binomial(n,p) random variable ke liye expected value = np aur variance = np(1–p).

$$E(X) = \sum_{x} xP(x) = \sum_{x=0}^{n} x \binom{n}{x} p^{x}(1-p)^{n-x}$$

$$= 0 + \sum_{x=1}^{n} x \frac{n!}{x!\,(n-x)!} p^{x}(1-p)^{n-x}$$

$$= \sum_{x=1}^{n} x \frac{n(n-1)!}{x(x-1)!\,((n-1)-(x-1))!} p^{x}(1-p)^{(n-1)-(x-1)}$$

$$= np \sum_{y=0}^{n-1} \binom{n-1}{y} p^{y}(1-p)^{(n-1)-y}$$