Introduction to Probability & Statistics
Prof. Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 5
Lecture - 19
Hypergeometric and Poisson Distribution

ki jo sample chuna gaya hai usme se kitne successes hain, to yeh bilkul hypergeometric situation hai;

$$p(x) = P(X = x) = e^{-\mu} * \mu^x / x!$$

$$e = 2.71$$

$$E(X) = \mu, V(X) = \mu$$

ismein X ek hypergeometric random variable hai jiske parameters ke roop me capital N = 20, capital M = 12 (inkyjet printers), aur sample size small n = 5; humein probability nikaalni hai ki X = 0, 1, 2, 3, 4, ya 5 hone par kya hoga; jaise X = 2 ke case me event hai 5 me se exactly 2 inkjet aur 3 laser printers number of ways of choosing 2 inkjet out of 12 = 12C2; selecting 3 lasers out of 8 = 8C3; dono selection ek saath karne ke tarike = (12C2)(8C3); denominator = total number of ways of choosing any 5 out of 20 = 20C5; selection bilkul random hai to har sample equally likely hai; isliye P(X=2) = (12C2 * 8C3) / (20C5); general case me P(X=x) = (M choose x)((N–M) choose (n–x)) / (N choose n), jahan x ki valid values 0 se min(n, M) tak hoti hain, lekin condition n–x ≤ N–M bhi satisfy honi chahiye; expected value hoti hai E[X] = n*(M/N) aur variance hota hai Var(X) = n*(M/N)(1–M/N)((N–n)/(N–1)); ye extra factor ((N–n)/(N–1)) jab 1 ke kareeb hota hai jab population size N sample n se bahut bada hota hai, tab hypergeometric approx. binomial ban jata hai. jo hai binomial prachal n aur p ke saath, aur maan lijiye ki n tends to infinity yani number of trials bahut bada ho raha hai aur p bahut chhota ho ja raha hai (p → 0), lekin iss tarah se ki unka product n·p kisi ek positive real number mu ki taraf converge ho raha ho (n·p → mu > 0); agar hum binomial random variable X ki probability dekhen ke X = x, to hume binomial PMF milta hai n choose x, p^x, (1–p)^(n–x); in teen cheezon ke saath agar n → ∞ aur p → 0 aur n·p → mu hota hai, to mathematically dekha jata hai ki yeh binomial probability approach karti hai Poisson probability

$$\binom{n}{x} p^x (1-p)^{n-x} \to \frac{e^{-\mu}\mu^x}{x!}$$

yani jab binomial trials me n kaafi bada ho aur p kaafi chhota ho, lekin n·p constant ho, tab Poisson PMF directly use karke probability calculate ki ja sakti hai; p ka chhota hona batata hai ki success event rare hai yani S-event rare hai isliye Poisson distribution rare events ke liye ek behtareen model hota hai; agar n bada ho aur p chhota ho to practical rule of thumb hai: agar binomial experiment me $p \le 0.1$ ho aur $n \ge 50$ ho to Poisson approximation of PMF used ki ja sakti hai; ab ek example dekhte hain: ek publisher har page ko carefully proofread karta hai taaki errors kam se kam rahein, lekin errors kabhi kabhi reh jaate hain; probability that randomly chosen page contains at least one error = 0.005, jo rare hai; aur yeh page-to-page independent mana hai; to hum binomial situation me hain total pages n = 400, aur p = 0.005; ab random variable X lete hain = number of pages with at least one error; hume P(X = 1) nikalna hai; theoretically X ek binomial random variable hai B(400, 0.005), to P(X=1) = 400C1 * (0.005)^1 * (0.995)^399; isko actual calculator se nikaalenge to milta hai approx 0.270669; lekin Poisson approximation se X approx Poisson(mu= n·p = 2), to P(X=1) = e^(–2) * 2^1 / 1! = 0.270671; dono answers lagbhag identical; lekin agar hume P(X=200) nikalna hota, to 400C200 evaluate karna practically impossible hota simple calculator par, jabki Poisson approximation se bahut aasani se result mil sakta hai; dono points clear hote hain: Poisson distribution kaafi important hai, aur chaahe directly koi physical random experiment na bhi ho, Poisson ko binomial distribution ki limiting form ke roop me samjha ja sakta hai; aur practical calculations me iska use bada hi upyogi hota hai.