Introduction to Probability & Statistics
Prof. Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute (Delhi)
Week - 4
Lecture - 13
Cumulative Distribution Function

Humne discrete random variables asatat ya aduruchchik char ka abhyas shuru kiya tha. Aaj ek aur example se shuruaat karte hain. Random variable X ka PMF tabular form mein diya gaya hai: X ki values hain –1 (probability 0.1), 0 (probability 0.25), 1 (probability 0.35), aur 2 (probability 0.3). In probabilities ka sum 1 hota hai, isliye yeh ek valid PMF hai aur iska matlab hai ki random variable X sirf yeh chaar values hi leta hai. Kisi event ki probability PMF se nikalna aasaan hai—jaise $P(X = 0) = 0.25$, $P(X = 1) = 0.35$. Agar hum $P(X = 1.3)$ poochein, to 1.3 PMF table me nahi diya gaya, isliye $P(1.3) = 0$. Probability $X \leq 1$ jaise events ko PMF use karke values jodkar nikalte hain:

$$P(X \leq 1) = P(-1) + P(0) + P(1) = 0.1 + 0.25 + 0.35 = 0.7.$$

Isi tarah $P(X \leq 1.5)$ bhi 0.7 hi hoga kyunki 1.5 X ki kisi value ke beech ka point hai. Agar $X < -1$ ho, to event impossible hai, to $P(X \leq x) = 0$. Agar $-1 < x < 0$ ho, to sirf –1 count hoga, to $P = 0.1$.

Agar $0 < x < 1$ ho, to $P(X \leq x) = P(-1) + P(0) = 0.35$. Agar $1 < x < 2$ ho, to $P = 0.7$. Agar $x \geq 2$ ho to $P = 1$. Is tarah humne alag-alag ranges me cumulative probabilities calculate ki. Isi ko ek function ke roop me define kiya jata hai jise CDF—cumulative distribution function—kehte hain, aur isse $F(x)$ likhte hain. Definition: $F(x) = P(X \leq x) = \Sigma\, p(y)$ sabhi y ke liye jinke $y \leq x$. CDF har real x ke liye define hota hai. Upar wale example ka CDF ek increasing step function banta hai—$x < -1$ par $F(x) = 0$; $-1 \leq x < 0$ par 0.1; $0 \leq x < 1$ par 0.35; $1 \leq x < 2$ par 0.7; aur $x \geq 2$ par 1.

Graph me iski shape seediyon jaisi lagti ha har step rising hota hai aur right-endpoint open hota hai. Yeh property har discrete random variable ke CDF ki hoti hai. Is example se yeh bhi samajh me aata hai ki hume sample space ya actual experiment jaanna zaroori nahi hai; sirf PMF de diya jaye to hum X se sambandhit sabhi events ki probabilities nikal sakte hain. Ab hum aage ek aur example lenge jisme black, green aur red balls di hui hongi aur usme probability properties ka istemaal karke PMF derive karenge.

Jab hum calculations karenge tab humein difference pata chal jayega. Samajh lijiye main ek random variable Y define karta hoon jo batata hai ki humne kitne balls nikale jab tak red ball nahi mil jaati experiment tab khatam hota hai jab red ball aa jaata hai. Y ki alag-alag values ho sakti hain: $Y = 1$ tab hoga jab pehle hi draw me red ball mil jaaye. $Y = 2$ tab ho sakta hai jab pehla ball red na ho (ya to black ya green ho) aur doosra ball red ho. Isi tarah $Y = 3, 4, 5$ sab ho sakte hain. Y = 6 bhi ho sakta hai kyunki total 7 balls me 2 red hain aur 5 non-red. Agar pehle 5 balls me koi bhi red na aaye, to 6th ball zaroor red hogi. Lekin Y = 7 nahi ho sakta kyunki agar 6 draws tak red nahi aaya, iska matlab 5 non-red pehle hi nikal gaye aur bachhi hui 2 balls dono red hongi, to 6th draw me hi red aana hi aana hai. Ab PMF nikalte hain: P1 = probability(Y = 1) matlab pehle hi draw me red aaye; total 7 balls me 2 red, to P1 = 2/7. P2 = probability(Y = 2) matlab pehle draw me red na aaye aur doosre me red aaye. Pehle draw me red na aane ka chance 5/7, phir remaining 6 balls me 2 red bachi hoti hain, to second draw me red aane ka chance 2/6. To P2 = (5/7) × (2/6) = 10/42. Isi tarah P3 ke liye: pehle do draws me red nahi aata (5/7 × 4/6), phir teesre me red aata (2/5), to P3 = 8/42. Isi pattern se P4, P5, P6 calculte hote hain: P4 = 6/42, P5 = 4/42, P6 = 2/42. In sabka sum 42/42 = 1 hota hai, jo confirm karta hai ki Y sirf inhi chhe values ko leta hai. Ab isi Y ka CDF nikalne par number line 7 intervals me banti hai—(–∞,1), [1,2), [2,3), [3,4), [4,5), [5,6), [6,∞). Har interval me F(x) cumulative sum hota hai: F(x<1)=0; F(1≤x<2)=P1; F(2≤x<3)=P1+P2; F(3≤x<4)=P1+P2+P3; F(4≤x<5)=P1+P2+P3+P4; F(5≤x<6)=P1+P2+P3+P4+P5; F(x≥6)=1. Isi tarah humne coin-toss wale example me geometric PMF bhi dekha jiska PMF hai P(X=x) = (1–p)^(x–1) × p. Iska CDF nikalne par sum geometric series ban jaata hai aur result milta hai F(x) = 1 – (1–p)^x for x = 1,2,3,… aur F(x)=0 jab x<1, aur F(x)=F(floor(x)) jab x non-integer ho. Is tarah geometric random variable ka CDF bhi derive hota hai.

$$F(x) = \sum_{y \le x} P(y) = \sum_{y \le x} (1-p)^{y-1} p$$

$$= p \sum_{y=1}^{x} (1-p)^{y-1}$$

$$= p \frac{1-(1-p)^{(x-1)+1}}{1-(1-p)}$$

$$F(x) = 1 - (1-P)^x$$