Introduction to Probability & Statistics

Prof. Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi

Week - 6

Lecture - 21

CDF Of A Continuous Distribution

ki koi bhi yadrūchik char ke liye, ab pehle humne jab discrete random variable padhe the tab uska sanchayi bantanfalan dekha tha, jise humne define kiya tha probability $X \leq x$ ke roop me yaani ki random variable X ke woh saare values jo small x se kam ya barabar hote hain, un sabki probability ka sum; isi tarah hum cumulative distribution function CDF yani sanchayi bantanfalan ko kehte hain; aur yeh concept X discrete ho ya continuous, dono me lagu hota hai farq sirf is baat me hota hai ki probability calculate kaise karte hain; continuous random variable ke saath joda hota hai PDF probability density function (praikta ghanatva phalan), aur agar X continuous random variable hai jiska PDF small f(x) hai, to uska CDF capital F(x) define hota hai capital

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(y)\, dy,$$

for all real x; pichhle lecture me humne PDF ke kaafi examples dekhe the, ab unke corresponding CDF nikalenge; pehla example — uniform distribution: jahan $f(x) = 1/(b–a)$ for $a \leq x \leq b$, aur $f(x)=0$ otherwise; agar is f(x) ka graph dekhen to yeh ek constant horizontal line hoti hai between x=a and x=b, height = 1/(b–a), aur a ke pehle aur b ke baad ye zero hota hai; ab CDF capital F(x) nikalein

$$F(x) = \int_{-\infty}^{x} f(y)dy$$

$$x < a \; F(x) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{x} 0dy = 0$$

$$a \leq x < b \; F(x) = \int_{-\infty}^{x} f(y) = \int_{-\infty}^{a} f(y)dy + \int_{a}^{x} f(y)dy$$

$$= 0 + \int_{a}^{x} \frac{1}{b-a} dy = \left(\frac{y}{b-a}\right) \; (y = a) \rightarrow x = \frac{x-a}{b-a}$$

$$b < x \; F(X) = \int_{-\infty}^{x} f(y)dy = \int_{-\infty}^{a} f(y)dy + \int_{a}^{b} f(y)dy + \int_{b}^{x} f(y)dy$$

isi tarah agar x aur y jaise random variables lagein jo continuous hain to CDF har jagah smooth hogi, monotonic non-decreasing hogi, aur kabhi decrease nahi karegi; aur khas baat — continuous random variable ke liye $P(X = c) = 0$ hota hai kisi bhi c ke liye, isliye $X < c$, $X \leq c$, $X = c$ — sab ke probability expressions me koi farq nahi padta, sab ka result same hi hota hai jab hum integrals se probability nikalte hain. Alag–alag antaralon me, to yeh jo integral $-\infty$ se x tak hai, isko hume alag–alag intervals me todna padta hai: pehle $-\infty$ se 0 tak alag se calculate karenge (jahan $f(y)=0$ hai), aur 0 se x tak alag se (jahan $f(y) \neq 0$ hai). Pehla to hamesha 0 hi aata hai, yeh hum bahut baar kar chuke hain. Jab uniform distribution ka example liya tha jahan $f(x) = 1/(b-a)$ for $a \leq x \leq b$ tha, wahan CDF F(x) nikalne ke liye humne teen cases liye: $x < a$, $a \leq x \leq b$, aur $x \geq b$. Jab x a aur b ke beech tha, tab $F(x) = \int$a to x $(1/(b-a))$ dy aata hai, jo hota hai $(x-a)/(b-a)$. Jab $x \geq b$, to $F(x) = \int$a to b $(1/(b-a))$ dy $= (b-a)/(b-a) = 1$. Isse humne final CDF likha: $F(x)=0$ agar $x<a$; $(x-a)/(b-a)$ agar $a\leq x<b$; aur 1 agar $x\geq b$. Iska graph banane par a se pehle $F(x)=0$, a se b tak ek straight line (linear increase) 0 se 1 tak, aur b ke baad $F(x)=1$ milta hai. Aap check kar sakte hain ki $x=a$ par $F(x)=0$ aur $x=b$ par $F(x)=1$ — yani F continuous function hai. Doosri baat, F(x) non-decreasing (increasing) hota hai.

Ab doosra example jo pichhle class me dekha tha: PDF $f(y) = -y$ for $-1<y<0$, $f(y)=y$ for $0<y<1$, aur 0 otherwise. Yani real line ko 4 intervals me baanta gaya: $(-\infty,-1)$, $(-1,0)$, $(0,1)$, $(1,\infty)$; inme f alag–alag define hai: pehle aur chauthe interval me 0, beech ke dono me $-y$ aur y. CDF F(x) nikalne ke liye phir se 4 cases lene honge: $x<-1$, $-1\leq x<0$, $0\leq x<1$, $x\geq 1$.

$$x < -1: F(x) = \int_{-\infty}^{x} f(y)\, dy = 0$$

kyunki $f(y)=0$ hai.

$$-1 \leq x < 0:\ F(x) = \int_{-\infty}^{-1} f(y)\, dy + \int_{-1}^{x} f(y)\, dy = 0 + \int_{-1}^{x} f(y)\, dy$$

Iska integral hota hai $-y^2/2$, limits -1 se x; value aati hai $(1/2 - x^2/2) = 1/2 - x^2/2$.

$$0 \le x < 1:\ F(x) = \int_{-\infty}^{-1} 0\, dy + \int_{-1}^{x} 0(-y)\, dy = 0 + \int_{0}^{x} y\, dy$$

Pehla 0, doosra integral 1/2 deta hai, teesra integral $x^2/2$; to $F(x)=1/2 + x^2/2$. Jab $x \ge 1$: $F(x)=1$ (check kar sakte hain: –1 se 0 tak area 1/2, 0 se 1 tak 1/2, total 1).

To final CDF: $F(x)=0$ if $x<-1$; $F(x)=1/2 - x^2/2$ if $-1 \le x<0$; $F(x)=1/2 + x^2/2$ if $0 \le x<1$; $F(x)=1$ if $x \ge 1$. Yeh CDF non-decreasing hai, aur jaisa continuous case me hota hai, inequality me $\le$ ya $<$ likhne se probability par farq nahi padta. CDF continuous random variables ke liye bahut useful hota hai, shayad discrete se bhi zyada, kyunki agar CDF pata hai to probabilities aasani se nikal sakte hain. General result: agar X continuous random variable hai with PDF f(x) aur CDF F(x), to:

$$P(X > a) = 1 - F(a)$$

$$P(a \le X \le b) = F(b) - F(a)$$

Yeh kyun sahi hai? PDF ke graph se intuitive explanation milta hai: poora area under curve 1 hota hai; F(a) left side ka area deta hai ($-\infty$ se a tak), to right side (a se ∞) ka area $1-F(a)$ hota hai. Aur [a,b] ke beech ka area $F(b) - F(a)$. Ab ek bridge load ka example: maan lijiye bridge par traffic aur trucks ki wajah se jo load aata hai, uska magnitude X naam ke random variable se darshate hain, kisi unit (jaise Newton) me; X ka PDF diya hai $f(x) = 1/2 + x$ for $0 \le x \le 1$, aur 0 otherwise. Isse turant pata chal jata hai ki X ke values sirf [0,1] me aayenge (kyunki baahar f=0 hai). CDF ke liye phir se 3 intervals: $x<0$, $0 \le x<1$, $x \ge 1$. For $x<0$: $F(x)=0$; for $x \ge 1$: $F(x)=1$. $0 \le x<1$ ke liye:

$$F(x) = \int_{-\infty}^{0} 0\ dy + \int_{0}^{x} (1/2 + y)\, dy = \int_{0}^{x} (1/2 + y)\, dy$$

$$= [y/2 + y^2/2]_0^x = x/2 + x^2/2$$

Ab $P(X > 1/2)$ nikalni ho to direct formula use kar sakte hain:

$P(X > 1/2)=1 - F(1/2)$. $F(1/2) = (1/2)/2 + (1/2)^2/2 = 1/4 + 1/8 = 3/8 = 0.375$;

to probability hogi $1 - 0.375=0.625$.

Isi tarah $P(1/4 \le X \le 3/4) = F(3/4) - F(1/4)$. $F(3/4)= (3/4)/2 + (3/4)^2/2 = 3/8 + 9/32 = (12+9)/32=21/32$;

$F(1/4)= (1/4)/2 + (1/4)^2/2 = 1/8 + 1/32 = 5/32$; difference $= 21/32 - 5/32 = 16/32 = 1/2 = 0.5$.

Ab ek important observation: agar hume small f pata ho to integrals se capital F nikal sakte hain; agar capital F pata ho aur F kisi point x par differentiable ho, to uska derivative waha small f(x) ke barabar hota hai. Formal roop: for every x jahan CDF differentiable hai, $F'(x) = f(x)$. Yani hum F se f aur f se F, dono tarah ja sakte hain. Agla concept jo sab random variables ke liye (discrete + continuous) par lagu hota hai lekin zyada tar use continuous ke saath hota hai — woh hai percentiles. Koi bhi number p lo jahan $0 < p < 1$; $100 \cdot p$ 0 aur 100 ke beech hoga (jaise p=0.9 to 100p=90, jo 90th percentile hai). General me 100p-th percentile ek aisa number hota hai distribution ka, jise hum $\eta(p)$ se darshate hain (eta of p), jo X ke range me aata hai aur is equation ko satisfy karta hai: $F(\eta(p)) = p$. Yani CDF ka value p ho jaye jis point par, wahi us distribution ka 100p-th percentile hota hai. To 0 se eta-p tak small f(x) dx—agar hume capital F pata hai ya small f pata hai, dono ka use karke ham eta-p ki value nikaal sakte hain. Is concept ko samajhne ke liye phir ek baar graph se dekhte hain: yeh small f ka graph hamesha x-axis ke upar hota hai aur poora area 1 hota hai; eta-p wahi number hota hai lagbhag jiske left side ke area ka maan small p hota hai. Yani eta-p aisa number hai jiske liye F(eta-p)=p; agar p=0.9 ho to eta-0.9 he 90th percentile hoga. Ab uniform distribution ka pehla example revisit karein: agar X uniform(a,b) ho to $f(x)=1/(b-a)$ for $a \le x \le b$ aur 0 otherwise; CDF $F(x)=0$ if $x<a$, $(x-a)/(b-a)$ if $a \le x < b$, aur 1 if $x \ge b$. Capital F differentiable hota hai sab x par except x=a aur x=b par jahan graph me sharp corner hota hai; derivative check karein: $x<a$ par $F'(x)=0=$ small f(x); $x>b$ par bhi derivative 0=small f(x); aur beech me $F'(x)=1/(b-a)=$small f(x). Yani capital F se small f derivative lekar nikal sakte hain. Uniform distribution ke percentile me agar 90th percentile (eta 0.9) chahiye to F(eta)=0.9; to eta-a / (b–a)=0.9; yani eta=0.9b+0.1a; general me $eta(p)=p \cdot b + (1-p) \cdot a$; aur agar p=0.5 ho to eta-0.5 = (a+b)/2—interval ka midpoint—jise hum median bhi kehte hain.

Ab doosra example jahan capital $F(x)=0$ if $x<0$, $F(x)=x/2 + x^2/2$ if $0 \le x < 1$, aur $F(x)=1$ if $x \ge 1$; ab agar eta-p nikalna hai, to kyunki p 0 aur 1 ke beech hoga to eta-p bhi 0 aur 1 ke beech hi hoga; isliye F(eta-p)= eta/2 + eta²/2 = p; simplify karne par milta hai $eta^2 + eta - 2p = 0$, quadratic equation; to root milta hai $(-1 \pm \sqrt{(1+8p)})/2$; ab jo root 0–1 ke beech hoga wahi percentile hoga: eta-p=$-1/2 + (\sqrt{(1+8p)})/2$. Pehle uniform me equation linear thi, yahan quadratic aa rahi hai—percentile ke liye CDF se equation banakar solve karna hota hai. Aur jaisa ki kaha, p=0.5 par 100p=50; 50th percentile ko hi median kehte hain; median ko μ~ (mu-tilde) se denote kiya jata hai; aur median woh number hai jiske liye $F(\mu\sim)=0.5$; median ke left ka area 0.5, right ka area bhi 0.5; yani distribution ko do equal halves me baant deta hai—isi ko hum kehte hain median ya madhyika.