Introduction to Probability & Statistics
Prof Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 8
Lecture - 28
Covariance

maan lijiye ki humare paas do random variables X aur Y hai jo sanyukt roop se bantit hain agar dono discrete hon to hum joint PMF P(X,Y) use karte hain jisse marginal PMFs PX aur PY nikaale jaa sakte hain, aur agar dono continuous hon to joint PDF f(x,y) se marginal PDFs mil jaati hain;

agar h koi function hai X aur Y ka, to expected value E[h(X,Y)] kaise nikaalte hain ye hum dekh chuke hain ab kuch vishisht functions ke saath hum moments ko samjhenge, jaise h(X,Y)=X ho to E[h(X,Y)] = E[X], jise discrete case me hum sum x·P(X,Y) over all x,y se ya simple sum x·PX(x) se nikaal sakte hain (kyunki $PX(x)=\Sigma_\gamma P(X,Y)$); continuous me yahi integral form me hota hai:

Discrete:

$$E(X) = \sum_x \sum_y xp(x,y) = \sum_x xp_x(x)$$

$$E(Y) = \sum_x \sum_y yp(x,y) = \sum_x yp_y(y)$$

Contineous:

$$E(X) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xf(x,y)dydx = \int_{-\infty}^{\infty} xf_x(x)dx$$

$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} yf(x,y)dydx = \int_{-\infty}^{\infty} yf_y(y)dy$$

isi tarah E[Y] nikalta hai aur $E[X^2]$, $E[Y^2]$ se variance bhi ab ek naya moment aata hai jab do variables saath-mein ho jisko bolte hain covariance ya sah-prasaran, jo measure karta hai ki X aur Y ek-dusre ke saath kitna related motion rakhte hain; iski paribhasha hai

Cov(X,Y)=E[(X−μX)(Y−μY)],

jahan μX aur μY respective means hain; discrete case me

$Cov(X,Y)=\Sigma_x\Sigma_\gamma$ (x−μX)(y−μY)P(X,Y),

aur continuous case me

$Cov(X,Y)=\iint(x-\mu X)(y-\mu Y)f(x,y)dxdy$;

ek remark $Cov(X,X)=E[(X-\mu X)^2]$ jo ki variance hi hota hai, yani variance ek special case hai covariance ka, aur covariance saamanyata ye batata hai ki do random variables apne apne mean se upar-neeche simultaneously jaate hain ya nahi. unka jis tarah se aapas me phaila hua hai woh covariance ek tarah se darshata hai, aur jaise variance ke liye humare paas shortcut formula tha, waise hi covariance ke liye bhi shortcut formula hota hai: $Cov(X,Y)=E[XY]-\mu X\cdot\mu Y$

yani $E[XY]-E[X]\cdot E[Y]$, aur ye zyadatar cases me calculation ko bahut aasaan bana deta hai; variance hamesha non-negative hota hai, kyunki uski definition hi squared deviation par based hoti hai, lekin covariance negative bhi ho sakta hai aur positive bhi agar X aur Y ke beech majboot sakaratmak sambandh ho (jab X zyada hota hai to Y bhi zyada hota hai), tab covariance ≥ 0 hota hai; agar unke beech nakaratmak sambandh ho (jab X zyada hota hai to Y kam hota hai), tab covariance ≤ 0 hota hai; isko crispy tarah se samajhne ke liye visualization le sakte hain jahan hum X-axis par X values, Y-axis par Y values plot karte hain—agar data diagonal upar ki taraf jata dikhe to positive relationship, agar diagonal neeche ki taraf jata dikhe to negative relationship, aur agar koi clear pattern na ho to covariance lagbhag zero ke paas aayega; iske baad hum discrete aur continuous dono cases me covariance ka calculation practice karte hain—pehle discrete me, jahan PMF P(X,Y) diya hota hai aur hum

$Cov(X,Y)=\Sigma_x\Sigma_y\,(x-\mu X)(y-\mu Y)P(X,Y)$

$Cov(X,Y)=\iint(x-\mu X)(y-\mu Y)f(x,y)dxdy$ se;

to covariance asal me measure karta hai ki X aur Y apne respective mean se deviation kis direction me lete hain—agar dono ek saath upar ya ek saath neeche jate ho to positive covariance, agar ek upar aur doosra neeche jaaye to negative covariance, aur agar relationship random ho to covariance kareeb-kareeb zero.2 alag-alag maan leta hai—X ho sakta hai 100 ya 250, aur Y ki values ho sakti hain 0, 100, ya 200; aur jaise humne pehle kiya tha, table me PXY ya sanyukt praayikta man-phalan (joint PMF) ki values di gayi hain: x=100 par y=0,100,200 ke corresponding probabilities 0.2, 0.1, 0.2; aur x=250 par y=0,100,200 ke corresponding 0.05, 0.15, 0.3; inse hum dono marginal PMFs nikaal sakte hain—PX(100)=0.5, PX(250)=0.5, aur PY(0)=0.25, PY(100)=0.25, PY(200)=0.5; ab shortcut formula ke liye hume chahiye E[X], E[Y], aur E[XY]; $E[X]=100\cdot0.5 + 250\cdot0.5 = 175$; $E[Y]=0\cdot0.25 + 100\cdot0.25 + 200\cdot0.5 = 125$; E[XY] joint PMF se nikaalte hain: $100\cdot0\cdot0.2 + 100\cdot100\cdot0.1 + 100\cdot200\cdot0.2 + 250\cdot0\cdot0.05 + 250\cdot100\cdot0.15 + 250\cdot200\cdot0.3 = 23{,}750$; ab covariance ke shortcut formula par aate hain: $Cov(X,Y)=E[XY] - E[X]\cdot E[Y] = 23{,}750 - 175\cdot125 = 1{,}875$; matlab first example me X aur Y me positive covariance hai; fir continuous example me joint PDF $f(x,y)=24xy$ hai region $0\leq x\leq1$, $0\leq y\leq1$ aur $x+y\leq1$ ke andar; marginal PDFs milte hain $fX(x)=12x(1-x)^2$ aur $fY(y)=12y(1-y)^2$; expected values $E[X]=E[Y]=2/5$; $E[XY]=\iint xy\cdot f(x,y)dxdy = 2/15$; aur covariance $= E[XY] - E[X]E[Y] = (2/15) - (2/5)(2/5) = -2/75 \approx -0.0267$, jo negative covariance batata hai; yaad rakhein variance of X kabhi negative nahi ho sakta lekin covariance negative bhi ho sakta hai, positive bhi ho sakta hai, aur zero bhi; iske baad important property aati hai $Cov(aX+b, cY+d)=ac \cdot Cov(X,Y)$, jo batati hai ki covariance measurements aur units ke scale par depend karta hai, aur isi wajah se alag-alag examples ke covariance values ko direct compare nahi karna chahiye.