Introduction to Probability & Statistics
Prof. Abhay Gopal Bhatt
Department of Statistics
Indian Statistical Institute Delhi
Week - 7
Lecture - 24
Normal Distribution

Aaj hum normal distribution ko padhenge jisko prasaamanya bantan kehte hain, aur probability aur statistics me ye sabse mahatvapurn distribution maana jata hai; vastavik zindagi ke kaafi saare data normal distribution se achhe se model ho jate hain, exact nahi lekin kaafi accurate approximation milta hai, aur kyun milta hai iska jawab hum aage lectures me ek theorem ke saath dekhenge; normal distribution ko define karne ke liye hum PDF — probability density function use karte hain:

$$f(x;\mu,\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right) e^{-(x-\mu)^2/2\sigma^2}$$

jahan mu koi bhi real number ho sakta hai aur sigma shunya se bada hota hai; isme pi ki approximate value 3.14159 aur e ki value lagbhag 2.71828 hoti hai; normal PDF har x ke liye positive hota hai aur symmetric hota hai mu ke around, yani f(x−mu) = f(mu−x); is distribution ka mean (expected value) mu hota hai aur variance sigma square; graph me mu ke left aur right area 0.5–0.5 hota hai, isliye mu median bhi hota hai; agar sigma kam ho to distribution zyada peaked aur concentrated hota hai mu ke paas, aur sigma bada ho to data spread out hota hai; probability a ≤ X ≤ b ke liye PDF integrate karna padta hai, jo normal case me exact solve nahi kiya ja sakta, isliye numerical approximation use hota hai; iske liye hum ek special case define karte hain

$$P(a \leq X \leq b) = \int_a^b \left(\frac{1}{\sqrt{2\pi}\sigma}\right) e^{-(x-\mu)^2/2\sigma^2} dx$$

Standard Normal Distribution jisme mu = 0 aur sigma = 1 hota hai, aur isko Z se denote karte hain: Z ~ Normal(0,1) jiska PDF hota hai fZ(z)=1/sqrt(2pi)*e^{-z^2/2}, aur isi curve ko standard normal curve kaha jata hai. Toh zyadaatar hum Z se isko darshate hain, usi tarah uske CDF ko sancheyi maan phalan ya cumulative distribution function ko bhi ek vishisht symbol se darshate hain usko hum capital Phi se darshate hain; hamesha koi bhi random variable ke CDF ko hum capital F se darshate hain lekin ye ek kaafi special distribution hai, kaafi special random variable hai, isko zyadaatar books me capital Phi se darshaya jata hai, Phi of Z; iska definition woh hi hai:

probability (Z ≤ z), aur isko calculate karne ka tarika bhi wahi hai integrate karenge minus infinity se lekar z tak integrate karenge PDF;

$$\emptyset(z) = p(Z \leq \varepsilon) = \int_{-\infty}^{\varepsilon} (\frac{1}{\sqrt{2\pi}}) e^{-x^2/2} dx$$

z toh fixed hota hai, limit of integration isliye integration variable ko x se darshaya hai; lekin problem wahi hai jo pehle bataya: is integral ki exact value hum bilkul calculate nahi kar sakte, koi technique nahi jisse exact solution nikale; lekin bahut saare z ke liye approx value nikalne ke liye tables aur computer algorithms available hain, aur isi ke liye hum standard normal probability tables ka upyog karte hain; ye jo Phi function hai uski values numerical approximation se nikaali gayi hain aur unko table form me tabulate kiya gaya hai; ye table internet par, ya kisi bhi statistics ki kitab me mil jayega; is standard normal table ko hum read karna seekhte hain row headings me left side par Z values di hoti hain: −3.4, −3.3, etc. aur ye 0.0 tak jata hai, aur next page par 0.0 se 3.4 tak; column headings hoti hain 0.00 se 0.09 tak; row Z ke first decimal tak value deti hai aur column second decimal point ko represent karta hai; jaise phi(−3.30) ke liye row −3.3 aur column 0.00 entry hoti hai 0.0005; isi tarah phi(−3.35) = 0.0004 row −3.3 aur column 0.05; table ke upar diya diagram ye dikhata hai ki ye shaded area hota hai curve ke neeche z ke left side ka area i.e., CDF; example values: probability(Z ≤ 1.36), probability(Z ≥ 1.36), probability(Z ≤ −1.36), probability(−0.29 ≤ Z ≤ 1.36); pehle probability(Z ≤ 1.36) = phi(1.36); agar Z ke decimal 2 se zyada hain toh usko round off karke 2 decimals tak likhna padta hai; standard normal table ke second page par positive Z values hoti hain; uske row me 1.3 aur column me 0.06 dekh kar hum phi(1.36) nikaal sakte hain; negative page par negative Z left side ke liye aur positive page par positive Z right side ke liye hota hai, lekin concept wahi rehta hai curve ke niche z ke left region ka area hi cumulative probability hota hai. 869 jo bilkul pichhle answer se milta hai aur woh milna zaroori bhi hai kyunki hume pata hai ki normal density standard normal curve mu ke around symmetric hota hai, yahan pe mu = 0 hai; chhota diagram imagine karein: ye hai 1.36, ye hai −1.36; second example me humne kya kiya: 1.36 ke daaye jo area hai curve ke niche, woh table use karke nikala 0.869; teesre example me −1.36 ke baaye area nikala; symmetry ke kaaran standard normal curve symmetric hai zero ke around right side ka red area aur left side ka red area same hota hai, aur table values se bhi confirm hota hai; chauthe question me probability −0.29 ≤ Z ≤ 1.36 kyunki Z continuous random variable hai yeh probability hogi phi(1.36) minus phi(−0.29); phi(1.36) hum pehle dekh chuke hain, phi(−0.29) ke liye row −0.2 aur column 0.09 dekhenge table se value milegi 0.3859; subtraction karne par milta hai 0.5272; is tarah standard normal probabilities nikaali jaati hain; ab standard normal ke percentiles dekhte hain agar P 0 aur 1 ke beech koi number ho to 100p-th percentile (eta_p) wahi number hoga jiske liye phi(eta_p) = p; isme hum reverse process karte hain: pehle Z le kar phi value nikaalte the, ab phi value dekar Z dhoondhte hain; maan lijiye p = 0.67, to hume 67th percentile chahiye table me aisi value dhoondhenge jo 0.6700 ho aur woh 0.44 par milti hai yani phi(0.44) = 0.6700; symmetry se phi(−0.44) = 1 − 0.6700 = 0.3300 to 33rd percentile hai

−0.44; 50th percentile obvious hai: 0; 99th percentile ke liye phi(z)=0.99, table me 2.32 par 0.9898, aur 2.33 par 0.9901; approximate value milegi 2.33; symmetry se 1st percentile hoga −2.33; isi tarah 95th percentile ke liye phi(z)=0.9500 table me 1.64 par 0.9495 aur 1.65 par 0.9505 midpoint se approximately 1.645; symmetry se 5th percentile hoga −1.645; isse juda hua Z-alpha notation bahut use hota hai statistics me critical points of standard normal distribution; Z-alpha woh point hota hai jiske right side ka area alpha hota hai, left ka area 1 − alpha; isliye Z-alpha = 100·(1−alpha)-th percentile; agar alpha = 0.05 to right side area 0.05 yani left side area 0.95 — to Z0.05 = 1.645; ab tak jo table use ki usme mu = 0 aur sigma = 1 tha lekin agar mu aur sigma kuch aur ho? Har mu aur sigma ke liye table banana padega? Nahi kyunki agla theorem useful hai: agar X normal distributed hai mu aur sigma square ke saath to hum transformation Z = (X − mu)/sigma kar dete hain yani subtract mean and divide by standard deviation to ye Z standard normal distributed hota hai yani N(0,1); is transformation ko normalization ya standardization kehte hain; is theorem ko use karke hum koi bhi normal distribution ko standard normal me convert karke table se values padh sakte hain ye hi powerful technique statistics me baar-baar use hoti hai. Agar hamare paas ek non-standard random variable X hai jiske prachhal mu aur sigma hain, aur hume probability dekhni hai ke X a aur b ke beech ho, yani probability(a ≤ X ≤ b), to hum X ko normalize karte hain subtract the mean, divide by standard deviation aur ye hi transformation agar hum a aur b ke saath bhi karen to inequality same rehti hai;

$$P(a \le X \le b) = P(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma})$$

$$P(a \le X \le b) = \emptyset\left(\frac{b-\mu}{\sigma}\right) - \emptyset(\frac{a-\mu}{\sigma})$$

ye dono events equivalent hote hain isliye unki probability bhi same hoti hai, aur ab kyunki Z standard normal hai, to probability calculate karne ke liye hum standard normal table use kar sakte hain: phi((b − mu)/sigma) minus phi((a − mu)/sigma); isi formula ke zariye koi bhi normal distribution ko table se solve kiya ja sakta hai; ab kuch examples: maan lijiye X ~ Normal(1, 4), to sigma = 2; agar hume probability(X ≤ 3) chahiye to ye hoga probability((X − 1)/2 ≤ 1), yani probability(Z ≤ 1), jiska table se value hota hai 0.8413; doosra example: probability(−1 ≤ X ≤ −1/2) ke liye hum karenge probability((−1 − 1)/2 ≤ Z ≤ (−1/2 − 1)/2) yani probability(−1 ≤ Z ≤ −0.75); table me phi(−0.75)=0.2266 aur phi(−1)=0.1587, to final result 0.2266 − 0.1587 = 0.0679; isi tarah percentiles bhi non-standard distribution ke liye nikaale ja sakte hain: agar X ~ Normal(mu, sigma^2) aur Z = (X − mu)/sigma to Z standard normal hota hai; agar p 0 aur 1 ke beech koi value ho, aur eta_p standard normal ka 100p-th percentile hai jiske liye phi(eta_p)=p, to percentile ko X ke terms me likhte hue milta hai: (X − mu)/sigma ≤ eta_p ⇒ X ≤ sigma·eta_p + mu; isse hume final important formula milta hai: normal distribution ke liye X ka 100p-th percentile = mean + (standard deviation × standard normal ka 100p-th percentile), aur is formula ka use hum baar-baar karenge.