EXAMPLE 4: E, V, σ ∞

find median, expected value, variance, & SD of

$f(x) = 1/x^2$, $(1, \infty)$

$E(X) = \int_1^\infty x \frac{1}{x^2}\,dx = \int_1^\infty \frac{1}{x}\,dx$ (divergent)

so integral does not exist.

↳ this integral has a median, but no mean, variance, or SD bc the integral D.N.E

THEOREM

IF a + b are real constants, then to find the new random variable:

$E(aX + bY) = a\,E(X) + b\,E(Y)$

$V(aX + bY) = a^2\,V(X) + b^2\,V(Y)$

$SD(aX + bY) = \sqrt{a^2 V(X) + b^2 V(Y)}$

assuming x + y are independent

4.2 FAMILIES OF CONTINUOUS DISTRIBUTIONS

WK 6-2

SPECIAL DISTRIBUTIONS

1 EXPONENTIAL DISTRIBUTION:

a continuous random variable has a exponential distribution if its PDF has the form of

GENERAL:

$f(x) = \lambda e^{-\lambda x}, x > 0$

can compute manually

$F(x) = 1 - e^{-\lambda x}$

$E(X) = 1/\lambda$, $V(X) = 1/\lambda^2$

· USUALLY USED TO MODEL TIME BETWEEN RARE EVENTS / LIFE TIME (don't rec life time, there are better ways to model)

λ = frequency - # of rare events per time unit

$1/\lambda$ = time

EXAMPLE: there are 6 hurricanes per year.

so frequency between 2 hurricanes is 1/6 of a year = 6 month

MEMORYLESS PROPERTY:

$P(X > t_1 + t_2 \mid X > t_1) = P(X > t_2)$

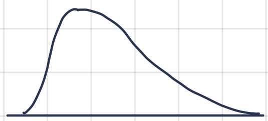
GEOMETRIC DISTRIBUTION ALSO HAS THE MEMORYLESS PROPERTY

IMAGINE X = TIME, GIVEN X IS > t_1, THE PROBABILITY OF X > $t_2 + t_1$ IS THE SAME AS X > t_2

this makes it not good for life time events

2 GAMMA DISTRIBUTION

WE ARE MORE INTERESTED IN THIS



$T = T_1 + T_2 + \dots + T_\alpha$

↳ T_i are times between rare events, they all follow exponential distribution

+ add them all up + you get gamma!

PDF → $f(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}$, $t > 0$, $\lambda > 0$, α is a positive int

$E(T) = \frac{\alpha}{\lambda}$

$V(T) = \frac{\alpha}{\lambda^2}$ ← can get from exponential

THE GENERAL GAMMA FUNCTION:

$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t}\,dt$

$\Gamma(\alpha + 1) = \alpha\,\Gamma(\alpha), \alpha > 0$

$\Gamma(1) = 1$

$\Gamma(n) = (n-1)!$

GAMMA FUNCTION

(THIS IS THE ACTUAL ONE, VERY COMPLICATED)

WE USE GAMMA TO MODEL LIFE TIME, IT CANNOT BE CALCULATED BY HAND

SINCE T is continuous n X is discrete

$P(T > t) = P(T \geq t) = P(X < \alpha)$

$P(T \leq t) = P(T < t) = P(X \geq \alpha)$

GAMMA - POISSON FORMULA →

T = distribution of time of α rare events

λ = how many rare events per time unit

Event $\{T > t\}$ means that the α-th event occurs after the moment t,

therefore, fewer than α events occur before time t

→ $P(T > t) = P(X < \alpha)$

→ $P(T \leq t) = P(X \geq \alpha)$ (complement)

$\{T > t\} = \{X < \alpha\}$, where X is a r.v. rep. the # of rare events, therefore, X has a poisson distribution w/ param. λt

EXAMPLE (GAMMA-POISSON):

ON THE AVERAGE, IT TAKES 25 SECONDS TO DOWNLOAD A FILE FROM THE INTERNET. IF IT TAKES AN EXPONENTIAL AMOUNT OF TIME TO DOWNLOAD ONE FILE, THEN WHAT IS THE PROBABILITY THAT IT WILL TAKE MORE THAN 70 SECONDS TO DOWNLOAD 3 INDEPENDENT FILES?

avg time = 25 seconds

exponential distr. for time

↳ multiple exponential = gamma

to compute:

λ = frequency, $1/25$ (avg time), $1/\text{avg time}$

t = time to download 3 files → $T_1 + T_2 + T_3$, $T_i \sim$ exponential ($\lambda = 1/25$)

α = 3

$T \sim$ Gamma ($\lambda = 1/25$, $\alpha = 3$)

$P(T > 70) = P(X < 3)$, $X \sim$ Poisson ($\lambda t = 70/25$) ($14/5$)

$P(X \leq 2)$ = poissoncdf($14/5$, 2) = .4695

HOW ABOUT LESS THAN 80 SECONDS?

$P(T < 80) = P(X \geq 3)$

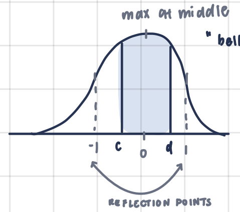
$X \sim$ Poisson ($80/25$, 2) (λt)

$= 1 - P(X \leq 2)$ still use cdf

$= 1 -$ Poissoncdf($80/25$, 2)

$=$ 0.6201

3 NORMAL DISTRIBUTION



HAS TWO PARAMETERS: μ + σ

$X \sim$ Normal (μ, σ) if X has pdf

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

↳ but usually we simplify + take out constants

→ NOT POSSIBLE BY HAND, THERE ARE ∞ NORMAL DIST.'S, SO THEY USUALLY ONLY CALCULATE THE STANDARD NORMAL DISTRIBUTION

↳ $f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ ⇝ so! standard normal table is aka z table

↳ with other normals, we convert non-standard to a standard normal (what textbook does)

· but we can do non-standard using technology :)

PROPERTIES OF STANDARD NORMAL

1) bell shaped curve w/ max at $x = 0$

2) symmetrical w/ vertical line through 0

3) x-axis is a horizontal asymptote, but curve n horizontal axis never meet

4) point of reflection at -1 + 1

5) half the area is left of 0, other half lies to the right

6) about 68% of the area lies between -1 + 1
   95% of the area lies between -2 + 2
   99.7% of the area lies between -3 + 3
   EMPIRICAL RULES

CALCULATOR:

P(lower < X < upper) = Normalcdf(lower, upper, mean, SD)

P(X < ?) = area, ? = invNorm(area to the left, mean, SD) compute cut off point

↓ CAN CHOOSE AREA SIDE ON CALC ACTUALLY

will answer two types of questions.

1) probability between 2 numbers: upper n lower

2) cut off point: area one

EXAMPLE (NORMAL DISTRIBUTION) → UPPER + LOWER

ADULT MEN'S HEIGHTS ARE NORMALLY DISTRIBUTED WITH μ = 70 inches and σ = 2.5 inches. WHAT IS THE PROBABILITY THAT A RANDOMLY SELECTED MAN WILL HAVE A HEIGHT LESS THAN 66 INCHES?

X = height $\sim$ Normal(μ = 70, σ = 2.5)

$P(X < 66)$ = normalcdf(-1000, 66, 70, 2.5)

should be -∞ but just use a large - #

= 0.0548

BETWEEN 67 + 71 INCHES?

$P(67 < X < 71)$ = normalcdf(67, 71, 70, 2.5)

= 0.5404

GREATER THAN 76 INCHES?

Normalcdf(76, 1000, 70, 2.5)

= 0.0082

WHAT IS THE VARIABLE?

HEIGHT

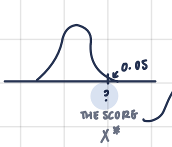
WHAT DISTRIBUTION DOES IT FOLLOW?

NORMAL, USE NORMAL CDF

CUT OFF POINT EXAMPLE

FOR A CERTAIN POPULATION OF HS STUDENTS, THE SAT-M SCORES ARE NORMALLY DISTRIBUTED W/ μ = 500 + σ = 100. A CERTAIN ENG. UNI. WILL ONLY ACCEPT STUDENTS W/ SCORES IN THE TOP 5%. WHAT'S THE MINIMUM SAT-M SCORE FOR THIS PROGRAM?

X = score $\sim$ Normal (μ = 500, σ = 100)



*USING INVERSE! DONT USE .05,

$P(X < ?)$ = .95 ← area on left

USE .95 ← AREA TO LEFT.

X^* = invNorm(.05, 500, 100, RIGHT)

or invNorm(.95, 500, 100, LEFT)

= 664.4854