Ch.5: Continuous Random Variables

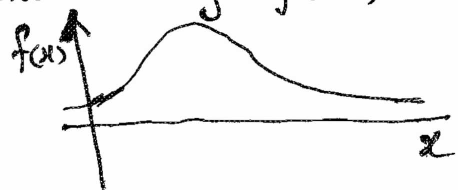
We have already discussed the discrete random variables ---> Binomial Distribution

↓

Continuous R.V. ---> Normal Distribution

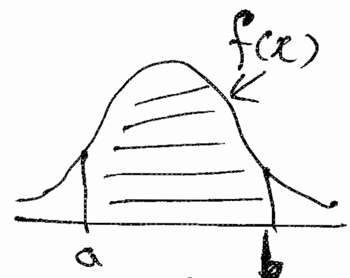
- Continuous R.V. can assume any numerical value on an interval
  - The time needed to finish a certain task
  - Temperatures
  - Weights

- The graphical form of the probabilities of a continuous r.v. X is a <u>smooth</u> <u>curve</u>, denoted by $f(x)$



$f(x)$ is called either

- Probability density function (pdf)
- or frequency function
- Probability distribution

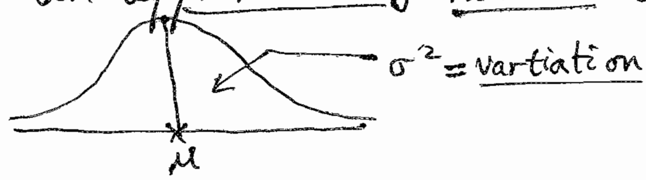


- $P(a < X < b)$ = Area under $f(x)$ between a and b
- $P(X = a) = 0$ [No Probability at any single point]
- $P(a \leq X \leq b) = P(a < X < b)$

  [No Prob.'s at end points]

Normal Distribution

- It is well-known distribution, called bell-shaped distribution.
- Many data follow an approximately normal distribution



- The probability density function (pdf) of $X \sim N(\mu, \sigma^2)$ is

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

where μ = mean of the normal r.v. X

σ = standard deviation

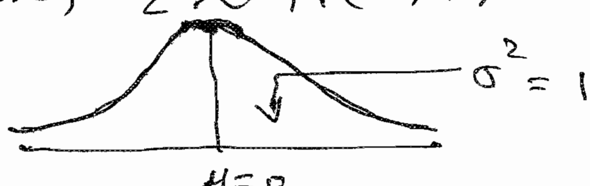
π = 3.1416...

e = 2.71828...

- The pdf of $X \sim N(\mu, \sigma^2)$ depends on 2 parameters
  - μ = mean of X
  - σ^2 = variance of X
- If $\mu = 0$ and $\sigma^2 = 1$, we will have standard normal distribution

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, \quad -\infty < z < \infty$$

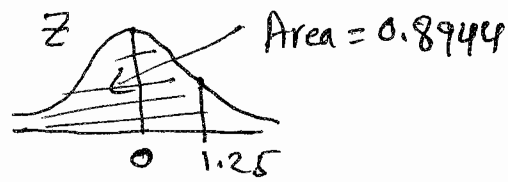
In this case, $Z \sim N(0, 1)$



Standard Normal Distribution

Normal Table

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0										
0.1										
0.2										
0.3										
⋮										
1.0	0.8413									
1.1										
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
⋮										
3.4										

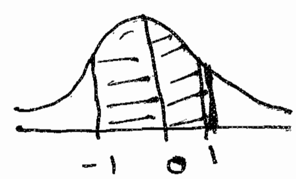


- $P(Z \leq 1.25) = T(1.25) = 0.8944$

- The Normal distribution is symmetric

$$P(Z > 1.25) = P(Z < -1.25)$$

$$P(-1 < Z < 0) = P(0 < Z < 1.00)$$



- $P(-1.25 < Z < 1.25) = T(1.25) - T(-1.25)$

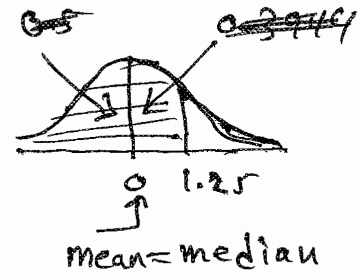
$$= 0.8944 - 0.1056$$
$$= \boxed{0.7888}$$

- $P(1 < Z < 1.25) = T(1.25) - T(1)$

$$= 0.8944 - 0.8413$$
$$= \boxed{0.0531}$$

- $P(Z > 1.25) = 1 - T(1.25)$
$= 1 - 0.8944$
$= \boxed{0.1056}$

- ~~$P(Z < 1.25) = 0.5 + 0.3944$~~
~~$= 0.8944$~~



Example 1: The grades of STAT.101 are normally distributed with mean 70 and standard deviation 10.

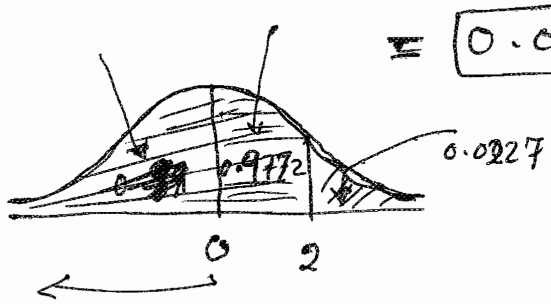
What is the probability that a student can get more than 90?

Solution: Need $P(X > 90)$

(1) write $\mu = 70$, $\sigma = 10$
(2) Change $X \longrightarrow Z$

$x = 90 \longrightarrow z = \frac{90 - 70}{10} = \boxed{2}$

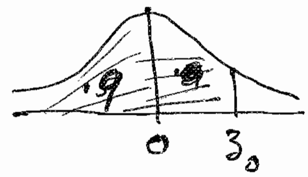
(3) Find $P(Z > 2) = 1.0 - T(2)$
$= [1.0 - 0.9772]$
~~$1 - 0.9772$~~
$= \boxed{0.0228}$



$\Rightarrow P(X > 90) = P(Z > 2) = \boxed{0.0228}$

Percentiles of $N(0,1)$

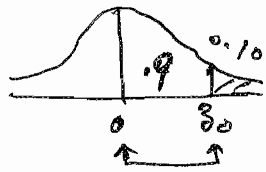
- If $P(Z \leq z_0) = 0.90$, find z_0. Note $z_0 = 90^{th}$ percentile



From normal table $P(Z \leq 1.28)$

$$= 0.8997$$
$$\cong 0.90$$
$$\Rightarrow z_0 = 1.28$$

- If $P(Z > z_0) = 0.10$, find z_0



$$\Rightarrow z_0 = 1.28$$

Back to Example 1: Find the grade x_0 ~~that~~ such that 90% of grades fall below x_0 and 10% fall above x_0. That is,

$$P(X \leq x_0) = 0.90$$

$$P\left(Z \leq \frac{x_0 - 70}{10}\right) = 0.90$$

$$\Rightarrow \frac{x_0 - 70}{10} = 90\text{th percentile of } N(0,1) = 1.28$$

$$\frac{x_0 - 70}{10} = 1.28$$

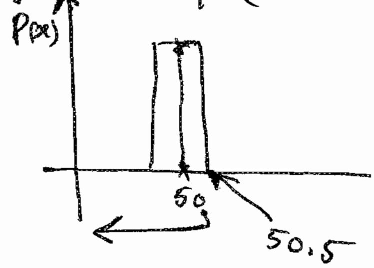
$$\Rightarrow x_0 = 70 + (1.28)(10)$$
$$= 70 + 12.8$$
$$= \boxed{82.8}$$

Approximating a Binomial Distribution by Normal Distribution

- Binomial Probabilities can be obtained for $n = 1, 2, \dots, 10, 15, 20, 25$. For this, we need approximation method for computing the binomial Prob's.
- For large n, a normal distribution can be used to provide an approximation to binomial distribution.
- Note the binomial distribution is discrete dist.
  normal distribution is continuous dist.

  They are different families

- If $X \sim B(100, 0.4)$, find (i) $P(X \le 50)$ (ii) $P(X = 50)$



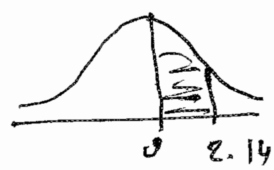
(i) $P_B(X \le 50) \simeq P_N(X \le 50.5)$

[Continuity correction]

$\mu = np = 100 \times 0.4 = \boxed{40}$

$\sigma = \sqrt{100 \times .4 \times .6} = \sqrt{24} = \boxed{4.9}$

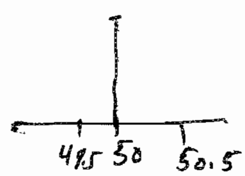
$= P_N\left(Z \le \frac{50.5 - 40}{4.9}\right)$



$= P_N(Z \le 2.14)$

$= \cancel{0.4838}\ T(2.14) = \boxed{0.9838}$

$= \cancel{0.5 + 0.4838 = 0.9838}$



(ii) $P_B(X = 50) = P_N(\underbrace{49.5}_{x_1} \le X \le \underbrace{50.5}_{x_2})$

$x_1 \longrightarrow z_1 = \frac{49.5 - 40}{4.9} = \boxed{1.94}$

$x_2 \longrightarrow z_2 = \frac{50.5 - 40}{4.9} = \boxed{2.14}$

$P(1.94 < Z < 2.14) = T(2.14) - T(1.94) = 0.9838 - .9738$

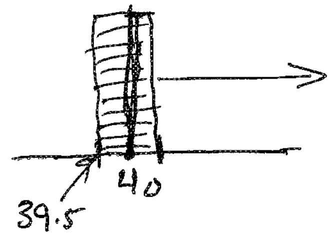
$= \boxed{0.0100}$

Example 2 : It is known that 60% of students are smoking. A sample of size 50 students is taken. Find the probability that at least 40 students are smoking.

- X: # of smokers $\sim B(50, 0.6)$

$$P_B(X \geq 40) = \underbrace{P(40) + P(41) + \cdots + P(50)}_{\text{You can't use Binomial Table}}$$

- We use normal approx.



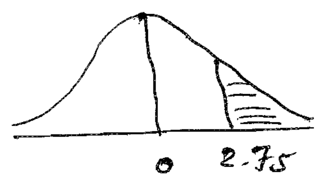
$$\approx P_N(X \geq 39.5)$$

$$\mu = 50 \times 0.6 = \boxed{30}$$

$$\sigma = \sqrt{50 \times 0.6 \times 0.4} = \sqrt{12} = \boxed{3.46}$$

$$39.5 \longrightarrow z = \frac{39.5 - 30}{3.46} = 2.75$$

$$P(Z \geq 2.75) = 1.0 - T(2.75)$$



$$= 1.0 - 0.9970$$

$$= \boxed{0.003}$$