Chapter 6 Discreet Probability Distributions

((chpt 7: Continuous Prob. Dist.))

In chpt 6 and 7 we will introduce the popular distributions of data that we see in stats.

6.1 Random Variables

In algebra we use x to represent a variable

Ex Solve for x: $3x + 2 = 4(x-1)$

In statistics our variables are a numerical outcom of an probability experiment.

Ex Let X = the number of heads that appear after tossing 5 coins on the table

X is a random Variable

Ex Exaples of random variables

- Numbers from the roll of a die
- Height of a randomly choosen male
- Number of siblings in a randomly choosen family
- Values of electricity bills in your neighborhood

Def: Discreet Random Vars represent data that comes in chunks, or quanta

ex: the number of eggs gathered in the morning

Continuous Random Var represents data that allows further refinement in a value

ex: gallons of milk from the cow

Def: Probability distribution for a discreet random var specifies the probability that a variable's value has of being selected

EX Sum of two dice

Red and Blue dice

- sample space

$S =$

	1	2	3	4	5	6
1	11	21	31	41	51	61
2	12	22	32	42	52	62
3	13	23	33	43	53	63
4	14	24	34	44	54	64
5	15	25	35	45	55	65
6	16	26	36	46	56	66

2 3 4 5

9 10 11 12

- Prob. Distribution Tally

X Sum	Outcomes	Freq.
2	11	1/36 = 0.028
3	12, 21	2/36 =
4	13, 22, 31	3/36 =
5	14, 23, 32, 41	4/36 = 0.111
6	15, 24, 33, 42, 57	5/36 =
7	16, 25, 34, 43, 52, 61	6/36 =
8	26, 35, 44, 53 62	5/36 =
9	36, 45, 54, 63	4/36 =
10	46, 55, 64	3/36 =
11	56, 65	2/36 =
12	66	1/36 =

- Prob. Histogram

- Probability Distribution for the sum of two dice (table)

X	2	3	4	5	6	7	8	9	10	11	12
P(X)	0.028	0.056	0.083	0.111	0.138	0.166	0.138	0.111	0.083	0.054	0.028

→ sum = 1.00

Properties

$$\boxed{\sum P(x) = 1.0}$$

Ex sum of two dice

$0.028 + 0.056 + 0.083 + 0.111 + 0.138 + 0.166$
$+ 0.138 + 0.111 + 0.083 + 0.054 + 0.028 \quad = 1.0$

- area under the probability histogram is 1

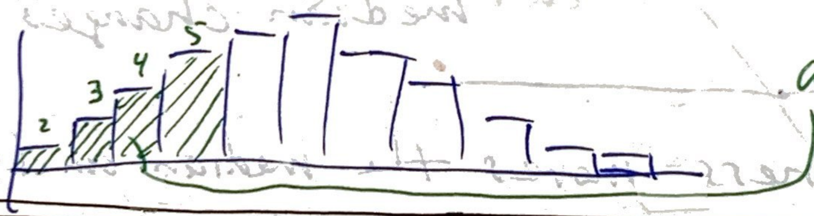
Ex
- $P(\text{sum}=5) = 0.111$ from the X vs. $P(X)$ table
- $P(\text{sum} \leq 5) = P(2 \text{ or } 3 \text{ or } 4 \text{ or } 5)$

$= P(2) + P(3) + P(4) + P(5)$

$= 0.028 + 0.056 + 0.08 + 0.111 \quad = 0.389$

$\approx \boxed{0.39}$

So 39% chance of rolling a sum of 5 or less



area = 0.39

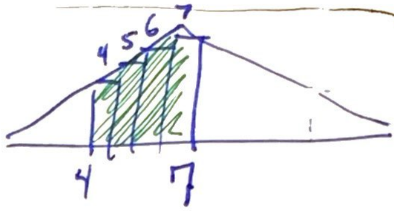
Ex
- $P(\text{sum} > 5)$

$= 1 - P(\text{sum} \leq 5)$

$= 1 - 0.39 = \boxed{0.61}$

0.61

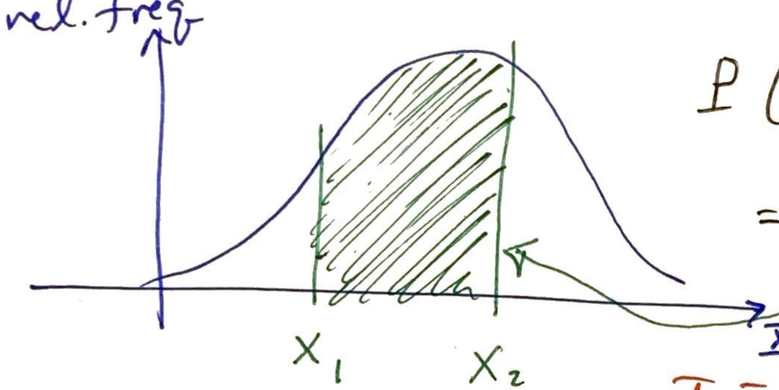
EX $P(4 \leq \text{sum} \leq 7)$



$= P(4) + P(5) + P(6) + P(7)$

$= 0.083 + 0.111 + 0.138 + 0.166 = \boxed{0.498}$

⊛ The probability of a Random Variable having a certain range of values is the area under the prob. distribution between those ranges



$P(x_1 \leq X \leq x_2)$

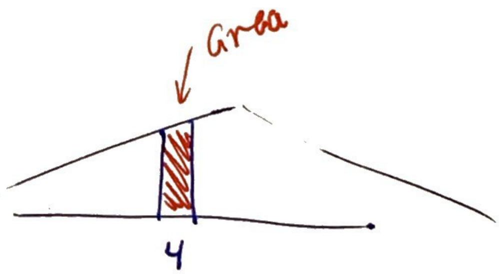
$=$ area between values x_1 & x_2

TOTAL area = 1.00

⊛ Again, the probability of a random variable tells us how frequently we can expect that outcome X

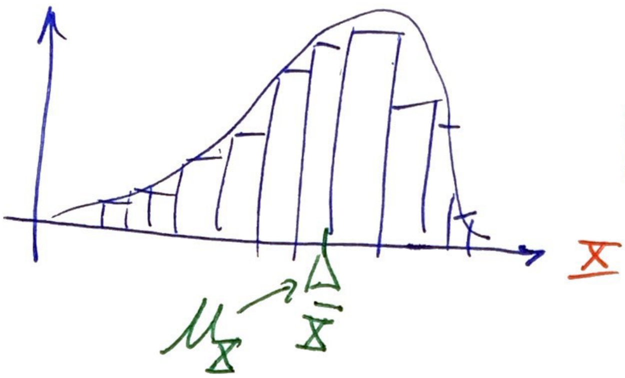
EX For discreet data, toss two dice

$P(\text{sum} = 4) = 0.083$



((For continuous data we do not address like P(Bessy has 4 gal. of milk today) only ranges 3.8–4.2 g))

⊛ mean value of a Random Variable



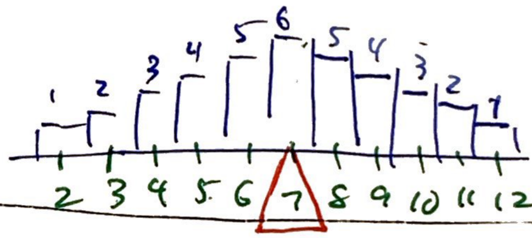
$$\boxed{\mu_X = \sum_{1}^{N} X \cdot P(X)}$$

Ex Two dice Problem : Find the mean

$$\boxed{\mu_X} = 2\cdot(0.028) + 3(0.055) + 4(0.083) + 5(0.111)$$
$$+ 6(0.138) + 7(0.166) + 8(0.138) + 9(0.111) + 10(0.083)$$
$$+ 11(0.055) + 12(0.028)$$

2nd Σx

so $\boxed{\mu_X} = \underline{\underline{7.00}}$



Ex A Computer Monitor is rarely perfect as there are over a million pixels on the screen.

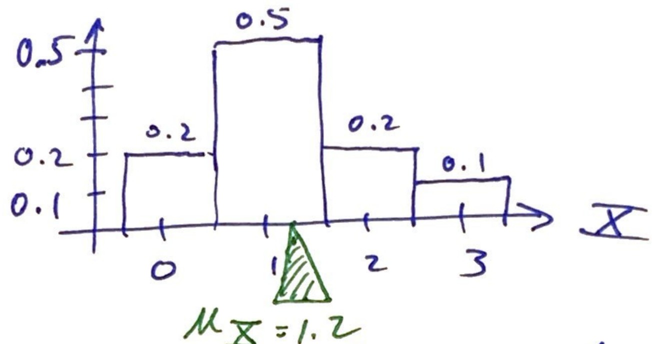
The following sampling process yeilded the

- prob. table

X	0	1	2	3
$P(X)$	0.2	0.5	0.2	0.1

← # of defective pixels out of factory

- Prob. Histogram



- mean of X = mean defective pixels

$$\mu_X = \sum X \cdot P(X)$$

$$= 0(0.2) + 1(0.5) + 2(0.2) + 3(0.1)$$

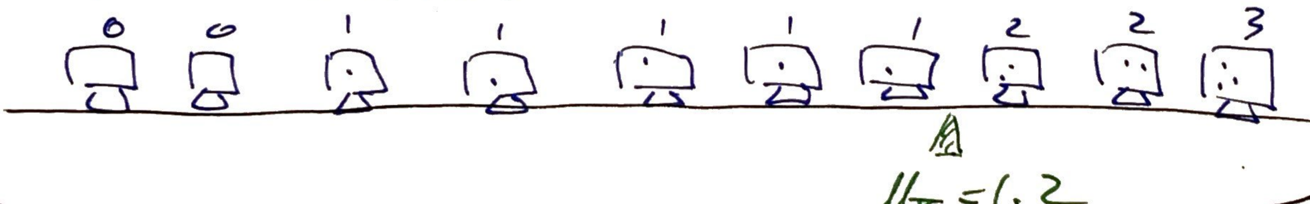
$$= \boxed{1.2 \text{ def. Pixels}}$$

Law of Large numbers

* This is also called the Expected Value, E_X

Consider a sample of 10 monitors

Order these monitors



* The **Law of Large Numbers** tells us that the larger the sample the closer our actual number are to the expected (mathematical) numbers

* The mean of a random variable, μ_X, is the mean of the population's variable

$$\mu_X = \sum X \cdot P(X)$$

$$\mu = \frac{\sum f \cdot x}{N}$$

Same

Ex let X be the number of boys in a family of 5 children. Theoretically, given a 50:50 chance of a boy being born, we produce the following probability table:

Num of Boys → X	0	1	2	3	4	5
$P(X)$	0.031	0.156	0.313	0.313	0.156	0.031

- Test: is $\sum P(X) \stackrel{?}{=} 1$ yes 1.0 exactly (no rounding errors)
- Q: what is the expected number of boys born in a family w/ 5 children?

$$\mu_X = \sum X \cdot P(X)$$
$$= 0\cdot(0.031) + 1\cdot(0.156) + 2\cdot(0.313) + 3\cdot(0.313)$$
$$+ 4\cdot(0.156) + 5\cdot(0.031)$$
$$= \boxed{2.50}$$

$\boxed{E(X) = 2.50}$, not unexpected since we have 5 children and a 50:50 chance of a boy.

Variance of a Random Variable

Recall variance (std. dev.) shows how spread out data is

Def: The variance of a random var is

$$\boxed{\sigma_X^2 \quad \sum [X - \mu_X]^2 \cdot P(X)}$$

where $\mu_X = \sum X \cdot P(X)$

an often easier formula is

$$\boxed{\sigma_X^2 = \sum X^2 \cdot P(X) - (\mu_X)^2}$$

Def: the standard deviation of a random var. is

$$\boxed{\sigma_X = \sqrt{\sigma_X^2}}$$

Ex Defective Pixels Problem revisited

X	0	1	2	3
$P(X)$	0.2	0.5	0.2	0.1

previous example $\mu_X = 1.2$

use a table to compute σ_X

X	$X - \mu_X$	$(X-\mu_X)^2$		$P(x)$		$(X-\mu_X)^2 \cdot P(x)$
0	$0 - 1.2 = -1.2$	1.44	*	0.2	=	0.288
1	$1 - 1.2 = -0.2$	0.04	*	0.5	=	0.020
2	$2 - 2.2 = 0.8$	0.64	*	0.2	=	0.128
3	$3 - 2.2 = 1.8$	3.24	*	0.1	=	0.324 +
						0.760

So $\sigma_X^2 = 0.760$ and the std. dev $= \sqrt{0.760} = 0.872$

Ex **Roulette Table**

- 38 pockets: 1 – 36 w/ 18 red, 18 black, plus also a "0" & "00" green
- We bet \$1 on red
- If red comes up you get the dollar back + one more dollar.
- If black or green come up you win "−\$1" (you loose the dollar).

X	-1	$+1$
$P(X)$	$\frac{18+2}{38}$	$\frac{18}{38}$

$=$

-1	$+1$
$20/38$	$18/38$

$=$

-1	$+1$
0.53	0.47

So then the expected winnings are

$$E(X) = \mu_X = X_1 \cdot P(X_1) + X_2 \cdot P(X_2)$$

$$= (-1)\left(\frac{20}{38}\right) + (+1)\left(\frac{18}{38}\right)$$

$$= -\frac{2}{38} = \boxed{-0.0526}$$

House always wins!!

- We see an expected loss. That is every time we play we expect a 5.26¢ loss

EX Mineral Economists estimate a mining venture has a

- 0.4 probability of \$30 million loss
- 0.5 probability of \$20 million profit
- 0.1 prob. of \$40 million profit.

• What is the expected earnings every time they start mining?

X	$-30M$	$+20M$	$+40M$
$P(X)$	0.4	0.5	0.1

So ~~the~~ expected profit is

$$E(X) = X_1 P(X_1) + X_2 P(X_2) + X_3 P(X_3)$$

$$E(X) = (-30M)(0.4) + (20M)(0.5) + (40M)(0.1)$$

$E(X) = \boxed{2M}$ profitable, otherwise they would not start