$\boxed{\text{MATH 115} \\ \text{LECTURE 16}}$: Random Variables (~ Chap 5)

We want a more convenient (and mathematical) way of describing events, and that also allows to deal with events that are not equally probable.

For example : Consider a pair of fair dice. We tossed them.

We know that the sample space is

$$S = \{ (1,1), (1,2), \ldots, (1,6),\\ (2,1), \ldots \quad , (2,6),\\ \vdots\\ (6,1), \ldots \quad , (6,6) \}.$$

Let's say we are interested in the following information:

- Sum of numbers that appear when the dice are tossed.
- Maximum of the two numbers that appear.

Later we will want to know things such as "probability that the maximum of the two numbers is 6", etc.

So we define two "random variables" :

$X \to$ sum of numbers,

$Y \to$ max. of numbers.

Although we call them "variables", these are functions:

$X((1,1)) = 2, \ X((2,3)) = 5, \ldots$

$Y((1,1)) = 1, \ X((1,6)) = 6, \ldots$

- Def: A <u>random variable</u> X is a function from the sample space to the real numbers. $X: S \to \mathbb{R}$.

- <u>Examples</u>:

1) Flip a coin 5 times. Let X be the number of heads. Then,

Sample space: $S = \{(H,H,H,H,H), (H,H,H,H,T), \ldots\}$ (2^5 elements)

$X((H,T,T,H,H)) = 3, \ldots$

→ Range or image of X: $R(X) = \{0,1,2,3,4,5\}$

2) Measure the height of people and record the result.

X = height in feet, Y = height in inches.

↳ We can do the usual operations with random variables:

$Y = 12X$ means $Y(\text{person } i) = 12\, X(\text{person } i)$.

1.1 Discrete random variables

Consider again the example of two dice and the sum of the numbers. There we could ask questions such as

$P(X=3)$,

$P(X \geq 2)$,

$P(2 \leq X \leq 4)$,

$\vdots$

These can be computed if we know $P(X=2)$, $P(X=3)$, $P(X=4)$.

• **Def.** Let X be a <u>discrete random variable</u>, that is, its range is discrete $R(X) = \{x_1, x_2, \dots, x_n\}$ (n might be infinite).

Then, X induces a function f as follows

$$f(x_k) = P(X = x_k)$$

If 1) $f(x_k) \geq 0$ for all k,
2) $\sum_k f(x_k) = 1$

the f is called the (probability) <u>distribution of X</u>.

• <u>Remark</u>: For discrete random variable the distribution is usually given as a table,

x	x_1	x_2	$\dots$
$f(x)$	$f(x_1)$	$f(x_2)$	$\dots$

• <u>Exercise</u>: Consider the example of two fair dice, and the random variables X = sum of the two numbers, Y = max. of the two numbers.

Find the distributions of X and Y.

<u>Sol</u>:

<u>X</u>: X can take 11 values ($R(X) = \{2, 3, \dots, 12\}$), and we know that $P(X=2) = \frac{1}{36}$, $P(X=3) = \frac{2}{36}$, ...

Summarizing,

x	2	3	4	5	6	7	8	9	10	11	12
$f(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

<u>Y</u>: $R(Y) = \{(1,2,3,4,5,6)\}$

$P(Y=1) = f(1) = 1$, $P(Y=2) = f(2) = \frac{3}{36}$, ..., $P(Y=6) = \frac{11}{36}$

y	1	2	3	4	5	6
$f(y)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

- **Exercise**: Suppose a coin is tossed three times, and suppose the coin is weighted so heads happens $2/3$ of the time.

  Find:
  1) Sample space
  2) Range of X, where X
  3) Distribution of X.

Sol:

1) $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

2) $R(X) = \{0, 1, 2, 3\}$

3) $P(X=0)$? $P(X=1)$? ...

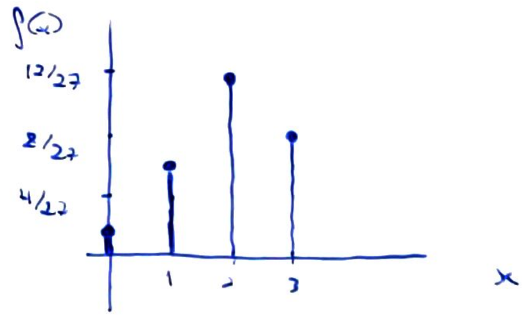
$$P(X=0) = P(TTT) = \left(\frac{1}{3}\right)^3 = \frac{1}{27}$$

$$P(X=1) = P(HTT \cup THT \cup TTH) = P(HTT) + P(THT) + P(TTH) =$$

$$= \frac{2}{3}\frac{1}{3}\frac{1}{3} + \frac{1}{3}\frac{2}{3}\frac{1}{3} + \frac{1}{3}\frac{1}{3}\frac{2}{3} = 3\,\frac{1}{3^2}\,\frac{2}{3} = \frac{6}{27}$$

Analogously for $P(X=2)$, $P(X=3)$, so in the end

x	0	1	2	3
$f(x)$	$\frac{1}{27}$	$\frac{6}{27}$	$\frac{12}{27}$	$\frac{8}{27}$



- General Binomials Trials

Consider an event which can only succeed or fail. The probability of success is p, and thus for failure is $q = 1 - p$:

$P(\text{"success"}) = p$, $P(\text{"failure"}) = q = 1 - p$.

The experiment is: repeat the event, call "trial", n times, and record the number of successes.

Define the random variable X = number of successes.

We want to find the distribution of the so-called Binomial random variable with parameters n and p (here X).

Remark: We know how to do this from previous exercises.

$P(X = k)$ with $k \in \{0, \dots, n\}$?

As usual, we have n trials $\rightarrow$ $\overbrace{\_\_ \;\; \_\_ \;\; \_\_ \;\; \dots \;\; \_\_}^{n \text{ times}}$

Consider those with k successes $\rightarrow$ $\underbrace{\_\_ \;\; \_\_ \;\; \_\_ \;\; \dots \;\; \_\_}_{k \text{ successes}}$

probability of each one of these trials $p^k q^{n-k}$

$\rightarrow$ Number of times this happens: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Therefore $\rightarrow$

$P(X=k) = \binom{n}{k} p^k q^{n-k}$ → So this describes the distribution of X.

$f(k) = \binom{n}{k} p^k q^{n-k}$.

Remark: Is this well-defined? → $f(k) \geq 0$ for all k?

→ $\sum_{k=0}^{n} f(k) = 1$? (Binomial Theorem)

- Examples.

1) A baseball player with a 0.3 batting average comes to bat 4 times in a game. Find.

a) Probability that he gets 2 hits.

b) Probability he gets at least 3 hits.

2) Approx. 20% of a mall's customers are over 65 years old. A new store decides to give a gift certificate to the first 4 people over 65 to enter the mall (to encourage seniors to become customers).

What is the probability that the 4th person over 65 is the 10th customer to go in?

- <u>Expectation and variance</u> of a discrete random variable.

- <u>Def</u>: The expected value $E(X)$ of a discrete random variable with probability distribution $f(x)$ is

$$\boxed{E(X) = \sum_{k} x_k f(x_k)}$$

Notice that this is a weighted average, where more likely outcomes have a bigger weight. Indeed, for a <u>uniform distribution</u>, that is, if all events are equally probably, then

$$P(X = x_k) = f(x_k) = \frac{1}{n} \qquad (R(X) = \{x_1, \dots, x_n\}).$$

$$E(X) = \sum_{k=1}^{n} x_k f(x_k) = \frac{1}{n} \sum_{k=1}^{n} x_k = \frac{x_1 + x_2 + \dots + x_n}{n}$$ | <u>Average</u>

- <u>Exercise</u>: A fair coin is tossed until a heads or five tails occurs. Find the expected number of tosses of the coin.

<u>Sol</u>:

$$S = \{H, TH, TTH, TTTH, TTTTH, TTTTT\}$$

$$X(H) = 1,\ X(TH) = 2,\ X(TTH) = 3, \dots \to R(X) = \{1, 2, 3, 4, 5\}.$$

$$P(X=1) = \frac{1}{2},\ P(X=2) = \frac{1}{4},\ P(X=3) = \frac{1}{8},\ P(X=4) = \frac{1}{16},$$

$$P(X=5) = 2 \cdot \frac{1}{32} = \frac{1}{16}$$

· Therefore,

$$E(X) = 1\cdot\frac{1}{2} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{8} + 4\cdot\frac{1}{16} + 5\cdot\frac{1}{16} = \frac{8+8+6+4+5}{16} = \frac{31}{16}.$$

- <u>Exercise</u>: Let X be a binomial random variable with parameters n and p. Find $E(X)$.

<u>Sol</u>:

$$E(X) = \sum_{k=1}^{n} k\, f(k) = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k}$$

Now we notice the following trick:

$\frac{\partial}{\partial p}(p^k) = k\, p^{k-1} \Rightarrow k\,p^k = p\,\frac{\partial}{\partial p}(p^k)$, thus we can write

$$E(X) = \sum_{k=1}^{n} p \binom{n}{k} \frac{\partial}{\partial p}(p^k)\, q^{n-k} = p\,\frac{\partial}{\partial p}\left(\sum_{k=1}^{n}\binom{n}{k} p^k q^{n-k}\right) =$$

$$= p\,\frac{\partial}{\partial p}(p+q)^n = pn\,(p+q)^{n-1} = pn.$$

(This corresponds to our intuition: if the event has a probability p of success, and we repeat the trial n times, the expected value of successes is np).