Class 8

Thursday, September 16, 2021 9:21 PM

CHAPTER 3

3.1.1 Random Variables

MOTIVATION:

See Section 2.1.2 : **Bernoulli Trials and Binomial Distribution:**
Example 2.9
Suppose that I have a coin for which $P(H) = p$ and $P(T) = 1-p$. I toss the coin 5 times.

- What is the probability that the outcome is *THHHH*? → by independence → $(1-p)^1 p^4$
- What is the probability that the outcome is *HTHHH*?
- What is the probability that the outcome is *HHTHH*?
- **What is the probability that I will observe exactly four heads and one tail?**
- **What is the probability that I will observe exactly three heads and two tails?**
- **If I toss the coin *n* times, what is the probability that I observe exactly *k* heads and *n*−*k* tails?**

Bernoulli Trials:

outcomes

↳ Tossing a coin → 2 types − {H, T}
− n trials each with same 2 possible outcomes < S success → 1 ; F Failure → 0 (numeric coding to)
trials are independent, (randomness)
− $P(\text{Success}) = P(S) = p$ is constant across trials

note: $S^c = F$!!

→ We $\boxed{\text{count}}$ the # heads in n trials

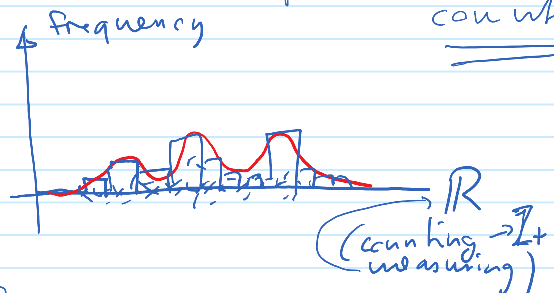
variable : X

X is a "random" variable → outcomes have been mapped to number counts !

Random Experiment

- Cards
- Coins
- lottery
- Disease occurrence

mapping →



- Disease occurrence

$\to \mathbb{R}$ (counting $\to \mathbb{Z}_+$, measuring)

$X =$ # heads in n trials.
BTW: # tails $= n - X$ } only look at X

range of X: (possible values of X):

$r_X = \{0, 1, 2, \ldots, n\}$

$$P(\underset{\substack{k \text{ heads} \\ n-k \text{ tails}}}{X = k}) = \binom{n}{k} p^k (1-p)^{n-k}; \quad k = 0, 1, \ldots, n; \quad 0 < p < 1$$

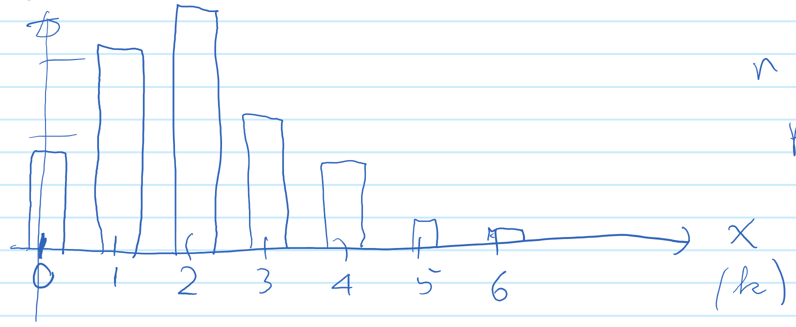
called the Binomial random variable < Bin(n, p)

→ all prob's are called the $\boxed{\text{Binomial}(n, p) \text{ distribution}}$

This is a discrete distribution because the range is countable.

Graph all probabilities: (eg bar graph → pin plot)

$P(X = k) = p(k)$ → PMF (prob. mass fct.)



$n = 6$

$p = .3$

Note $\sum_{k=0}^{6} p(k) = 1$

In general, to analyze random experiments, we usually focus on some numerical aspects of the experiment.

For example, in a soccer game we may be interested in the number of goals, shots, shots on goal, corners kicks, fouls, etc.

If we consider an entire soccer match as a random experiment, then each of these numerical results gives some information about the outcome of the random experiment.

These are examples of *random variables*. ***In a nutshell, a random variable is a real-valued variable whose value is determined by an underlying random experiment.***

These are examples of *random variables*. ***In a nutshell, a random variable is a real-valued variable whose value is determined by an underlying random experiment.***

$\to \mathbb{R}$ (or a subset) continuous RV's

$\to \mathbb{N}, \mathbb{Z}$ (or subset) discrete RV's

Random Variables:
A random variable X is a function from the sample space to the real numbers.

$$X : S \to \mathbb{R}$$

Example (again)
I toss a coin five times. This is a random experiment and the sample space can be written as *S*={*TTTTT*,*TTTTH*,...,*HHHHH*}.
We can define a random variable *X* whose value is the number of observed heads. The value of *X* will be one of 0,1,2,3,4or 5 depending on the outcome of the random experiment.

Example 3.2

$\Pr(\text{head}) = p \quad \Pr(\text{tail}) = 1-p$

Find the range for each of the following random variables.

- I toss a coin 100 times. Let be the number of heads I observe. $r = \{0, 1, 2, \ldots, 100\}$
- I toss a coin until the first head appears. Let *Y* be the total number of coin tosses.

Some possible outcomes: eg: TTTH, TTTTH, H, etc $\to \infty$

"Stopping rule" range of Y: $\{1, 2, 3, 4, \ldots\}$

$P(Y=1) = P(H) = p \qquad 0 < p < 1$

$P(Y=2) = P(T,H) = (1-p) \cdot p$

$P(Y=3) = P(T,T,H) = (1-p)^2 p$

The geometric(p) distribution

$$P(Y=m) = (1-p)^{(m-1)} \cdot p$$

$0 < p < 1$

$m = \{1, 2, 3, 4, \ldots\}$

Do these probs add to 1? (They should!)

Show $\sum_{m=1}^{\infty} (1-p)^{(m-1)} \cdot p \overset{?}{=} 1$

$= p\left((1-p)^0 + (1-p)^1 + (1-p)^2 + \ldots\right) = p\left(\underbrace{\sum_{k=0}^{\infty} (1-p)^k}_{\frac{1}{1-(\ldots)}}\right)$

$0 < (1-p) < 1 \qquad p$

$= p \cdot \frac{1}{p} = 1$ ✓✓

$0 < (1-p) < 1 \quad p$ ✓✓ $(= p \cdot \frac{1}{p} = 1)$ $\frac{1}{1-(1-p)}$

- The random variable T is defined as the time (in hours) from now until the next earthquake occurs in a certain city.
  - **Solution** $T \in \mathbb{R}_+$ (continuous) or $T \geq 0$

3.1.2 Discrete Random Variables

> X is a discrete random variable, if its range is countable.

3.1.3 Probability Mass Function (PMF)

$$R_X = \{x_1, x_2, x_3, \ldots\}.$$

$$A = \{s \in S | X(s) = x_k\}.$$; mapping

$P(A) = P(X = x_k)$ prob in original S → prob's that define the random variable.

> **Definition 3.1**
> Let X be a discrete random variable with range $R_X = \{x_1, x_2, x_3, \ldots\}$ (finite or countably infinite). The function
> $$P_X(x_k) = P(X = x_k), \text{ for } k = 1, 2, 3, \ldots,$$
> is called the *probability mass function (PMF)* of X.

Example 3.3:
I toss a fair coin twice, and let X be defined as the number of heads I observe. Find the range of X, R_X, as well as its probability mass function P_X.

✓

For notation purposes:

Extending the range to all nonnegative integers (or all real numbers):

$$P_X(x) = \begin{cases} P(X = x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$ then we can ... mathematical

$$P_X(x) = \begin{cases} P(X=x) & \text{if } x \text{ is in } R_X \\ 0 & \text{otherwise} \end{cases}$$

then we can apply mathematical operation the the entire real or integer numbers

Plotting the PMF as either bargraph or pinplot:

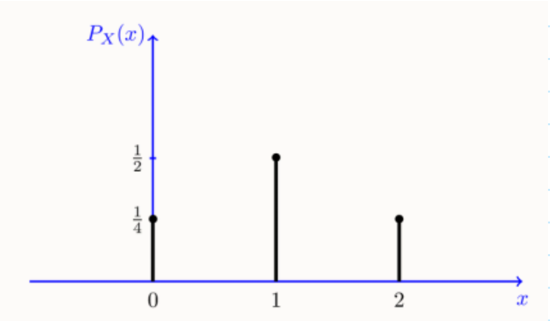


Fig.3.1 - PMF for random Variable X in Example 3.3.

Example 3.4:

I have an unfair coin for which $P(H)=p$, where $0<p<1$. I toss the coin repeatedly until I observe a heads for the first time. Let Y be the total number of coin tosses. Find the distribution of Y.

Solution:

done

Note that by definition the PMF is a probability measure, so ***it satisfies all properties of a probability measure***. In particular, we have

Properties of PMF:

- $0 \le P_X(x) \le 1$ for all x;
- $\sum_{x \in R_X} P_X(x) = 1$;
- for any set $A \subset R_X, P(X \in A) = \sum_{x \in A} P_X(x)$.

3.1.4 Independent Random Variables

Studying more than one variable at a time:

Often two or more random variables may be dependent (or may be "correlated")

Examples:

When random variables are independent, it is easier to calculate probabilities of "joint" events:

eg: P(A and B happening), where A is an event of variable 1 and B is an event of variable 2.

The concept of independent random variables is very similar to independent events. Remember, two events A and B are independent if we have $P(A,B) = P(A)P(B)$ (remember comma means *and*, i.e., $P(A,B) = P(A \text{ and } B) = P(A \cap B)$). Similarly, we have the following definition for independent discrete random variables.

Definition 3.2
Consider two discrete random variables X and Y. We say that X and Y are independent if

$$P\Big(X = x, Y = y\Big) = P(X = x)P(Y = y), \qquad \text{for all } x, y.$$

In general, if two random variables are independent, then you can write

$$P\Big(X \in A, Y \in B\Big) = P(X \in A)P(Y \in B), \qquad \text{for all sets } A \text{ and } B.$$

Definition 3.3
Consider n discrete random variables $X_1, X_2, X_3, \ldots, X_n$. We say that $X_1, X_2, X_3, \ldots, X_n$ are independent if

$$P\Big(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\Big)$$
$$= P(X_1 = x_1)P(X_2 = x_2)\ldots P(X_n = x_n), \qquad \text{for all } x_1, x_2, \ldots, x_n.$$

3.1.6 Solved Problems: Discrete Random Variables

Do Solved Problems 1, 2 & 8.