Essence of Probability

Theme:

Explain all the videos from bayesian viewpoint (without informing the audience about this):
With frequentists, probability only has meaning in the context of repeating a measurement. As we measure something, we'll see slight variations due to variances in the equipment we use to collect data. As we measure something a large number of times, the frequency of the given value indicates the probability of measuring that value. They rely on many blind trials of an experiment before making statements about an estimate of a variable.

With the Bayesian approach, we extend the idea of probability to cover the aspect of certainty about statements. The probability gives us a statement of our knowledge of what the measurement result will be. For Bayesians, our own knowledge about an event is fundamentally related to probability. They deal with "beliefs" (/"distributions") about the variable and update their beliefs about the variable as new information comes in.

List of videos:

Note: Will explain the basics of these topics and then ask them this question within the first two minutes of the video. To increase their confidence and make them feel that they know this intuitively.

1. Probability rules: In this video I would be discussing rules of sum/product, probabilistic principle of inclusion-exclusion, rule of complement along with an example of each of these and applications of probability.
2. Measure Theory: Discuss probability spaces, distributions, random variables, integration and its properties, expected value, product measures.
3. Conditional Probability: Question: What is the probability of getting 2 aces when selecting 2 cards at random from a deck of cards (52)? Answer: For the first card, the probability would be 4/52, for the second card it would be **3/51**. The desired probability is (4/52) x (3/51). The probability of getting the second ace would be updated by the information that the first card was an ace.
4. Discrete Random Variable: Question: If a random variable X is defined as the smallest of two cards in the experiment of selecting two cards (one by one) from a deck of cards, then the random variable X assigns value of min(C1,C2) to the outcome (C1,C2) of the chance experiment. A random variable X gets its value only after the underlying experiment has been performed. Before the experiment is performed, we can describe the set of possible values of X. Answer: Here Bayesian way of thinking about it would be that the smallest card would have the highest likelihood of getting assigned to X as whenever it's selected it would be assigned to X.
5. Binomial Distribution (Discrete Distributions): Question: What is the probability of selecting an ace from a deck of cards 2, 3, 4 times in a row an experiment which has 10 independent trials (selecting an ace = "success", any other card = "failure")? Answer: Here without doing any calculation you can intuitively say that the probability of getting it 2 times > probability of getting it 3 times > probability of getting it 4 times.

6. Continuous Random Variables: Question: Let X be a uniform, continuous random variable on [0,100] what is the probability that X $\geq$80? Answer: Here without doing any calculation you can answer this question intuitively to be 20%.
7. Normal Distribution (Continuous Distributions): Question: Which of the following situations can be described well by a normal distribution? (a) the sat scores of students at a college (b) number of days it takes for a person to pay a bill (c) rolling a fair die (d) time between two flashes of lightning A: (a)
8. Expectation (Expected Value), Variance, Covariance:
   Question1: What would be the expected value of rolling a fair, six-sided die? A: ⅙(1) + ⅙(2) + ⅙(3) + ⅙(4) + ⅙(5) + ⅙(6) = 3.5
   Question2: After defining variance show them two graphs one with points that clustered near their expected value and another with values scattered and ask them about which one has low variance.
   Question3: After explaining covariance show them two graphs one in which two variables are linearly related and another in which they are not.
9. Bayes Rule: Question: How likely it is to rain if you see dark clouds in the sky?
   P(H | D) = P(D | H) x P(H) / P(D) Note: H = hypothesis D = data
   Task is to figure out probability of rain given dark clouds,
   P(H) = probability of rain & P(D) = probability of dark clouds
10. Markov Chains: Question: A certain frog lives in a pond with two lily pads, east and west. A long time ago, he found two coins at the bottom of the pond and brought one up to each lily pad. Every morning, the frog decides whether to jump by tossing the current lily pad's coin. If the coin lands heads up, the frog jumps to the other lily pad. If the coin lands tails up, he remains where he is. Here this example can be interpreted using "bayesian" reasoning without knowing any more information. Where east and west are the state spaces and the probability of each individual coin toss are the probability transition functions.

Before answering each question mentioned above would place a (pause and ponder Πcreature) on the screen. Only in case of videos 9 and 10 would directly start with a question without any explanation.
Also, while answering these question I would not be mentioning bayesian.
In the last video of the series I would define what bayesianism and frequentism is.

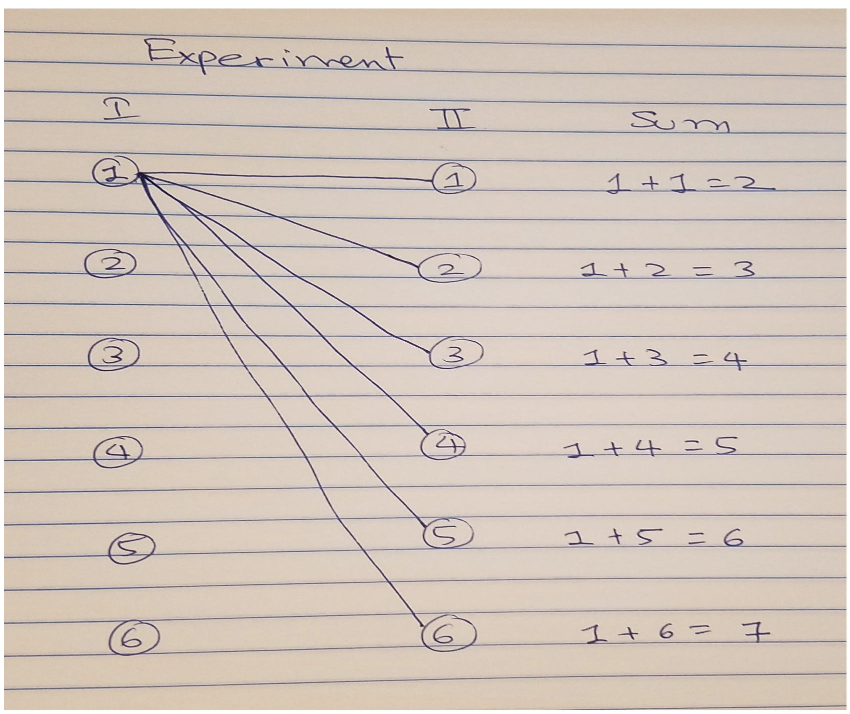
Video 4: Discrete Random Variables

A **random variable** assigns a single numeric value to each basic outcome in the sample space. A **discrete random variable** can take on a list of possible values ("finitely many" or "countably infinitely many" values).
Question: If a random variable X is defined as the smallest of two cards in the experiment of selecting two cards (one by one) from a deck of cards, then the random variable X assigns a value of min(C1,C2) to the outcome (C1,C2) of the chance experiment. A random variable X gets its value only after the underlying experiment has been performed. Before the experiment is performed, can you describe the set of possible values of X.
Answer: Here "bayesian"(would be replaced by "one") way of thinking about it would be that the smallest card would have the highest likelihood of getting assigned to X as whenever it's selected it would be assigned to X.

If a random variable X is defined as the sum of two values in the experiment of rolling two fair dice, then the random variable X assigns a value of sum(V1,V2) to the outcome of this experiment. Before the experiment is performed, we can describe the set of possible values of X.

Caption: Along with each explanation (one figure and two tables below) I would include two dice and show different values that the would take and how it would change the values of P(X) and F(x).

The number of possible combinations is 36. Here in the example below, we have all the possible combinations of the result of experiment 1 taking value 1 and the experiment 2 taking values 1, 2, 3, 4, 5, 6 respectively.



Caption: By showing this one iteration in which each of these arrows would appear one after another to show what value random variable X would take in each case.

As you can see there are 6 possible pairs when the outcome of experiment 1 is 1, viz. (1,1), (1,2), (1,3), (1,4), (1,5), (1,6). In the same way, there are 6 possible pairs each for 2, 3, 4, 5, 6. Which makes the total number of pairs to be 36. You can calculate this by multiplying the number of possible values each experiment takes. Suppose if in another scenario our two experiments are a coin toss and dice roll then the number of possible combinations would be 6x2=12 (as discussed in the video 1 (product rule)).
Getting back to the problem of two fair dice roll, how many values can our random variable X take.
(Pause and Ponder Π $creature$)
The number of possible values our random variable X can take is from 2 to 12 (total of 11 values). As 1 is the smallest and 6 the largest possible outcome for each dice roll, if we take the case where both the dice roll results in 1 we get the value of X as 2. If we get maximum value 6 for both the outcomes we get the value of X to be 12.
Now if we want to assign numerical probability values to each of these values then for sum value i.e. random variable X taking value 2 is 1/36. As there is only one combination (1,1) which will generate the sum of 1. Number 36 is the total number of combinations that are possible. For X to take value 3 the combinations are (1,2), (2,1), i.e. P(X=3) = 2/36. For X to take value 4 the combinations are (1,3), (2,2), (3,1), i.e. P(X=4) = 3/36. Here in the table below you can see that the number of combinations that would get us the value 7 is the maximum P(X=7) = 6/36 or (17%), list all the combinations that would get us the result 7; (1,6), (2,4), (3,3), (4,2), (5,2), (6,1).

x	P(X=x)	Occurrence
2	0.03	1
3	0.06	2
4	0.08	3
5	0.11	4
6	0.14	5
7	0.17	6
8	0.14	5
9	0.11	4
10	0.08	3
11	0.06	2
12	0.03	1

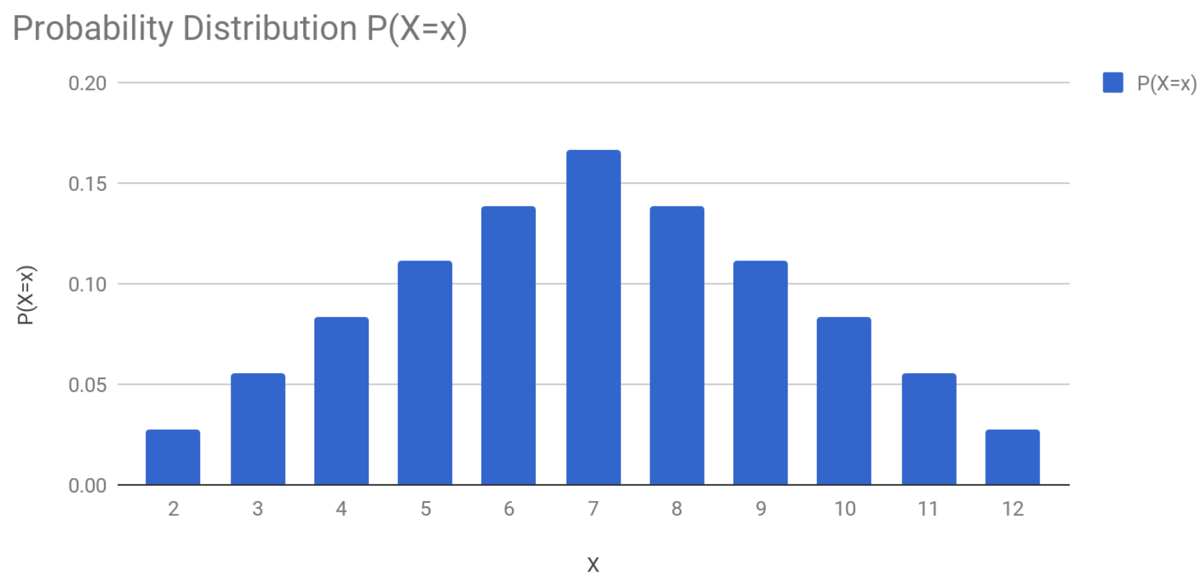
Caption: Would populate this table row by row along with the explanation on the right side of the screen, along with a graphic showing the dice values that would get us these results.
The **cumulative distribution function** of a random variable X is a function that F(x) that, when evaluated at point x, gives the probability that the random variable will take on a value less than or equal to x; i.e. P(X$\leq$ x). For example, a random variable representing two dice rolls would have probability distribution
F(2) = P(X$\leq$ 2) = 1/36
F(3) = P(X$\leq$ 3) = (1+2)/36 = 3/36
F(4) = P(X$\leq$ 4) = (1+2+3)/36 = 6/36

I.e. we go on adding the occurrences of the probability of the value of X being 4 or less than 4 (in this case 2, 3). There is another way of getting to these values as well, if we add the values of probabilities till that point then also we would get the same values.
$F(2) = P(X=2) = 0.03$
$F(3) = P(X=2) + P(X=3) = 0.03+0.06 = 0.09$
$F(4) = P(X=2) + P(X=3) + P(X=4) = 0.03+0.06+0.08 = 0.17$

$$F(x) = P(X \leq x) = \sum_{xi \leq x} P(X = x_i)$$

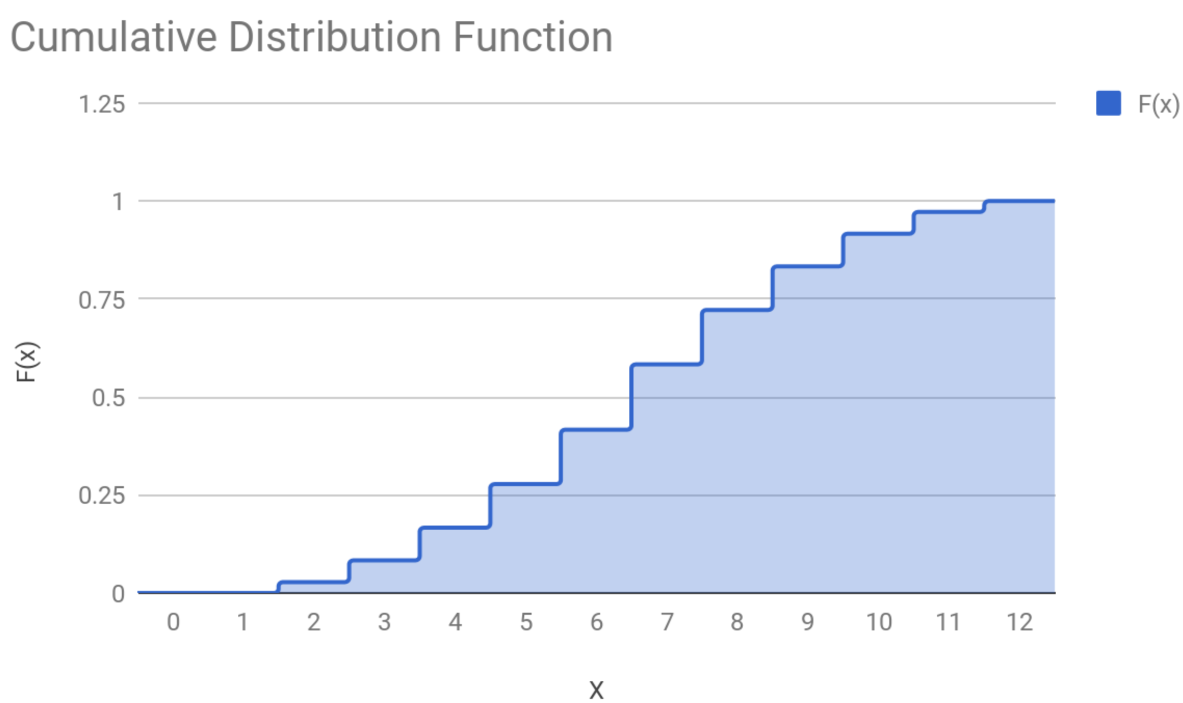
x	P(X=x)	F(x)
2	0.03	0.03
3	0.06	0.09
4	0.08	0.17
5	0.11	0.28
6	0.14	0.42
7	0.17	0.58
8	0.14	0.72
9	0.11	0.83
10	0.08	0.92
11	0.06	0.97
12	0.03	1.00

Caption: The column F(x) would be added one by one row-wise along with each of the calculation for each of those rows on the right.
Notice the value of F(12) is 1 as the set values that are less than or equal to 12 is the entire set of all the values that X can take from 2 to 12. Thus the value of F(12) is 36/36 which is equal to 1.
Probability distribution:
If we plot these values of P(X=x) on a graph then we get the graph below. The X-axis shows the values that the random variable X would take and the Y-axis shows the probability of how likely each value is going to show up. From these, we can see that the likelihood of X taking value 7 is the highest, that of 6 and 8 is the same and that of 1 and 12 is the lowest.



Caption: Along with these bars each one would appear from left to right with the combinations that generated them and the specific value i.e. for P(X=2) = 1/36.

As we discussed cumulative probability distribution while generating the table containing the values for F(x), now if we plot values of x on the X-axis and F(x) on the Y-axis in a graph we get the graph below.



Caption: This graph would be populated left to right with values of F(X) being shown for each iteration on the right.

We can use this above graph to get the probability of getting values less theΠ, as it would lie between 3 and 4 the value at that point indicates the probability of getting values less than Π.

I would like to end this video with a question about the topic that we would be discussing next week:
Q: What is the probability of selecting an ace from a deck of cards 2, 3, 4 times in a row an experiment which has 10 independent trials (selecting an ace = "success", any other card = "failure")?

Note: In all the videos we can ask questions about the card game in the beginning (can make it little challenging), but while discussing we can discuss simple examples (coin toss, dice roll) and make them more challenging as the video progresses. The idea behind this is to challenge the viewer (to get them feel more involved rather than passive consumption) and then start with the simple things and build up on top of that. This would get the required emotional response and might make them more interested in the video.

I would like to mention this at some point (in the beginning) of the series, the difference between probability and odds:
The probability of an event E is defined as:

P(E) = (Chances of E) / **(Total Chances)**

For example drawing an ace card (4) out of a deck of cards (52):

P(Ace) = 4/**52** = 0.077

The odds of an event E is defined as:

O(E) = (Chances of E) / **(Chances Against E)**

For the above card example, the odds of drawing an ace becomes:

O(E) = 4/**(52-4)** = 0.083

References:

1. Bayes Rule Example
2. Probability Theory Jaynes
3. Understanding Probability Henk Tijms
4. Probability: Theory and Examples Rick Durrett
5. brilliant.org probability
6. Deep Learning book
7. Deep Learning book 2
8. Miscellaneous
9. Markov Chains and Mixing Times
10. Markov Chains
11. Brilliant.org cdf
12. Discrete Random Variables - An Intuitive Introduction to Probability (Coursera)