86-631/42-631 Neural Data Analysis
Lecture 3: Review of probability I

1) Sets
    a. Definition of intersection, union, disjoint, etc.
2) Axioms of probability
    a. Three properties
    b. Total law of probability
3) Definition of conditional probability
4) Definition of independence
5) Random variables
    a. Bernoulli / Binomial introduction

We've mentioned that neurons are noisy, and we've gone into some of the sources of that noise. Therefore, to understand how neurons relate to the real world – how their firing corresponds to computation or representation, we need to take this noise into account. The most useful mathematical tool for understanding signal representation within noise is probability theory. We will therefore take some time to review some of the basics of probability theory that will be useful for this class. The basics of probability are relatively simple. However, things can get confusing in real life, when one tries to apply these techniques to the dirty world of experiments. *Most problems come from not being absolutely clear when defining the probability space you are working with.* I will give examples as we progress.

Sets

The calculus of probability is defined on event sets. Events are descriptions about the outcome of some experiment: I rolled a dice and got a '6', or I recorded from a neuron under some set of conditions and saw 3 APs within 200ms of stimulus onset. Both of these are statements about what happened.

We of course need to consider the set of all possible things that could happen. This set is the union of all possible events and is called the ***sample space***; it is typically denoted S.

> Example: coin flip. S={H,T}.
> Example: dice roll. S={1,2,3,4,5,6}.
> Example: neuron firing spikes in a 100ms window. S={non-negative integers}.

The sample space S is composed of events, and so if an event w is in S we would write $w \in S$. (And if it's not in S, we write $w \notin S$.)

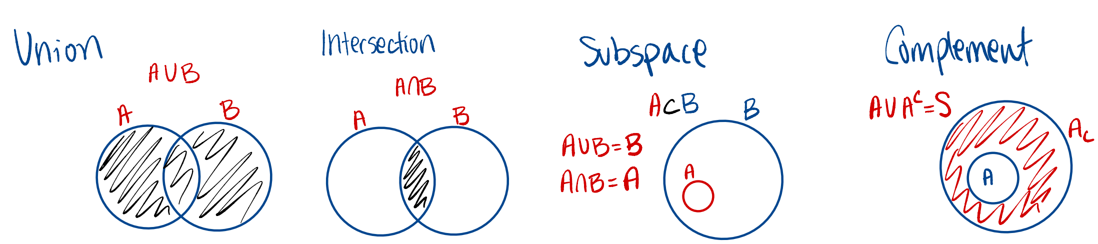
The ***union*** of two events A and B is denoted $A \cup B$, and consists of all outcomes that are either in A or in B or in both A and B. Example: Roll of a dice. Let A be the event that the result is ≤ 3, and B be the event that the result is odd. Then $A \cup B$ = {1,2,3,5}.

The ***intersection*** of two events A and B is denoted $A \cap B$ and consists of the events that are ONLY in both A and B. For the example above, $A \cap B$={1,3}. We say two events are ***mutually exclusive***, or ***disjoint***, if they have empty intersection. Example, if A is the event that the roll is even, and B is the event that the roll is odd, A and B are disjoint.

If the outcomes in A are entirely contained within B, we say that A is a ***subset*** of B, and denote this as $A \subseteq B$. When this is the case, $P(A \cap B)=P(A)$.

Finally, we denote by A^C the ***complement*** of A, it consists of all events that are not in A. Note that $S=A \cup A^c$.

Give Venn diagram examples of A, A^C, B, $A \cup B$, $A \cap B$, $A \subseteq B$, and S.



Axioms of probability:

1. For all events A, $P(A) \geq 0$.
2. $P(S)=1$
3. If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \ldots \cup A_n)=P(A_1)+P(A_2)+\ldots+P(A_n)$.

If two events are mutually exclusive is
the sum of their probabilities

Five properties of probability:

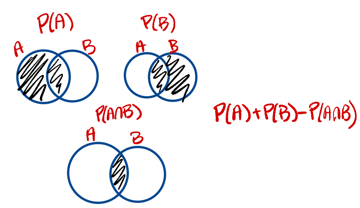
For any events A and B, we have

(i) $P(A) \leq 1$. Any events prob can't be greater than 1 (P(S))

(ii) $P(A^c)=1-P(A)$, where A^c is the complement of A.
Example: A=rolling a 6. $P(A^c)=5/6$.

(iii) If A and B are mutually exclusive, $P(A \cap B)=0$.
Example. A=rolling an even number. B=rolling an odd number. $P(A \cap B)$?

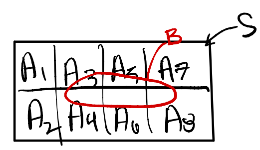
(iv) $P(A \cup B)=P(A)+P(B)-P(A \cap B)$.
Example. A=rolling an odd number. B=rolling a number ≤ 3. $P(A)=1/2$. $P(B)=1/2$.
$A \cap B = \{1,3\}$; $P(A \cap B)=1/3$. $P(A \cup B)=1/2+1/2-1/3=2/3$.
Also, $A \cup B = \{1,2,3,5\}$, so $P(A \cup B)=4/6=2/3$.



(v) Law of total probability
$P(B)=P(B \cap A)+P(B \cap A^c)$.
Example: draw Venn diagram example.

More rigorously, suppose $A_1, \ldots, A_n$ represent n mutually exclusive, exhaustive events. (Exhaustive means that the union of all these events spans the entire sample space). Divide S up into squares to demonstrate.

$$S=\bigcup_{i=1}^{n} A_i$$

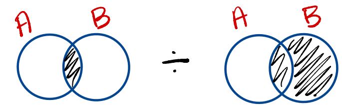


For this case, $p(B)=\sum_{i=1}^{n} p(B \cap A_i)$

$P(B)=\sum_{i=1}^{n} P(B \cap A_i)$

Add up area of
overlap for all tiles

Conditional Probability Assume P(B) > 0 → it happens

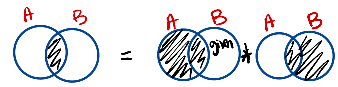


Assume that P(B)>0. The *conditional probability* of A given B is defined as

$$P(A|B) = \frac{P(A \overset{\text{and}}{\cap} B)}{P(B)}$$

Illustrative example with Venn diagrams – renormalizing the sample space to B.

Note that the definition of conditional probability also implies that $P(A \cap B) = P(A|B)P(B)$. This can sometimes be a convenient way of measuring the joint probability.



It is important to remember that conditional probability distributions obey all the same laws as regular probability distributions.

Conditioning and the law of total probability.

Let's go back to our law of total probability. We had $P(B)=P(B \cap A)+P(B \cap A^c)$. We can now define this in terms of conditional probabilities as well:

sum up each conditional probabilities

marginal probability → $P(B)=P(B|A)P(A)+P(B|A^c)P(A^c)$.

Similarly, if $A_1, \ldots, A_n$ represent n mutually exclusive, exhaustive events, then:

$$p(B) = \sum_{i=1}^{n} p(B \cap A_i) = \sum_{i=1}^{n} p(B|A_i)p(A_i)$$

Tuning curves (neuron in V1):
→ present gratings
A_0 = horizontal A_{90} = vertical
→ show bunch of stimuli
What is prob that neuron gives 100 spikes given I show each grating?

Now, it can be very easy to get confused with conditional probabilities. *You have to be crystal clear about what the events are!*

2 coin Problem:

- Flip 2 coins
- At least one came up heads
- Prob that other one came up heads?

A = event that both come up heads

B = event that at least one came up heads

$P(A) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$

$P(B) = \frac{3}{4}$ (tails, tails) {HH, TT, HT, TH}

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1}{\cancel{4}} \cdot \frac{\cancel{4}}{3} = \boxed{\frac{1}{3}}$$

Example, the three prisoners problem:

Three prisoners, A, B and C, are in separate cells and sentenced to death. The governor has selected one of them at random to be pardoned. The warden knows which one is pardoned, but is not allowed to tell. Prisoner A begs the warden to let him know the identity of one of the others who are going to be executed. "If B is to be pardoned, give me C's name. If C is to be pardoned, give me B's name. And if I'm to be pardoned, flip a coin to decide whether to name B or C."

The warden tells A that B is to be executed. Prisoner A is pleased because he believes that his probability of surviving has gone up from 1/3 to 1/2, as it is now between him and C. Prisoner A secretly tells C the news, who is also pleased, because he reasons that A still has a chance of 1/3 to be the pardoned one, but his chance has gone up to 2/3. What is the correct answer?

Below is the joint probability of all the events:

Being pardoned	Warden: "B will be executed"	Warden: "C will be executed"	Sum
A	1/6	1/6	1/3
B	0	1/3	1/3
C	1/3	0	1/3
Sum	1/2	1/2	1

So, the question amounts to: What is the probability that A will be pardoned given that the Warden says "B will be executed"? What is the probability that C will be pardoned given that the Warden says "B will be executed"? Let *b* be the event that the Warden says "B will be executed".

$P(A|b) = P(A,b) / P(b) = (1/6) / (1/2) = 1/3$

$P(C|b) = P(C,b) / P(b) = (1/3) / (1/3) = 2/3$

C is correct.

Independence

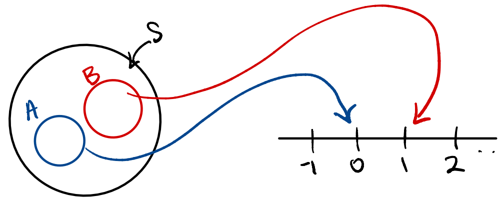
Two events are *independent* if and only if $P(A\cap B)=P(A)P(B)$. $A \perp B$

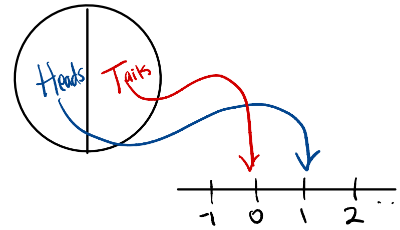
This definition agrees entirely with our intuition: if two events are independent, conditioning on one event shouldn't affect the probability of the other event. This turns out to be true. From the definition of conditional probability, $P(A|B)=P(A\cap B)/P(B)$. If A and B are independent:

$P(A|B)=P(A)P(B)/P(B)=P(A)$. This statement is symmetrical, $P(B|A)=P(B)$ if A and B are independent.

The definition of independence extends to multiple events:

If A_1, A_2, …, A_n are n independent events, then $P(A_1\cap A_2\cap\ldots\cap A_n)=P(A_1)P(A_2)\ldots P(A_n)$. Actually to be mutually independent, they have to obey the stricter rule that *every* subset of the n events must obey the multiplication rule.





Random Variables

Now, up to now we've only been talking about probability defined on events and sample spaces. This is rather limiting, however. For example, say we want to get a handle on the long term average or the spread in the different values you would see if you ran the experiment a bunch of times. How do we do that? What's the average of {H} and {T}?

To do these sorts of things, we need to introduce the concept of a *random variable.* A random variable takes *events* and maps them onto the real number line. So, for example, it could take tails {T} and map it to the number 0, and it could take heads {H} and map it to the number 1. Similarly, you could have another random variable that takes the outcomes of the roll of a dice {•}, etc. We can make a random variable that maps {•} to 1, {xx} to 2, etc.

Now because the events are random and can be described as having an associated probability, the outcomes of the random variable can also be described as having an associated probability. Specifically, for a random variable X which maps events E into real values x (draw picture), the probability that X will be x is given by the probability of the union of events E that all map into x:

$$p(X = x) = p\left(E = X^{-1}(x)\right)$$

All events of x that map into X

As an example, we could have a random variable that maps {I} , {II}, and {III} into 0, and {IV}, {V}, and {VI} into 2. The probability of getting a 0 would be ½, and the probability of getting a 2 would also be ½.

Note that to be a random variable, X must assign to every possible event a value. This means that:

$\in$ = set membership (is an element of)

$$\sum_{x \in X} p(x) = 1$$

Super Important!!!

That is, the random variable must map all possible outcomes into some set of real values. The *domain* of X is S, the *range* is written as $x \in X$. Summing the probabilities over the entire range of X gives 1, because you are summing the probability over the entire sample space S.

Example: binomial distribution

As an example, let's say that we measure whether or not a spike fired on multiple repeats of a particular stimulus. Say in any given trial, the neuron fires a spike with probability p. For a single trial, we have two possible events: we measure a spike or we do not: {spike, no spike}. Now, the random variable that maps these events to the real line might be X(no spike)=0 and X(spike)=1. This is an example of a Bernoulli distributed random variable with parameter q: you get 1 with probability q, and you get 0 with probability…

one trial

Anyone?Ans: (1-q). P(spike) = q
P(nospike) = 1-q

k=0 {no spike, no spike} = (1-q)(1-q)
$(1-q)^2$

Okay, let's repeat the test. Now there are a total of four possible events:

k=2 {spike, spike} = (q)(q)
q^2

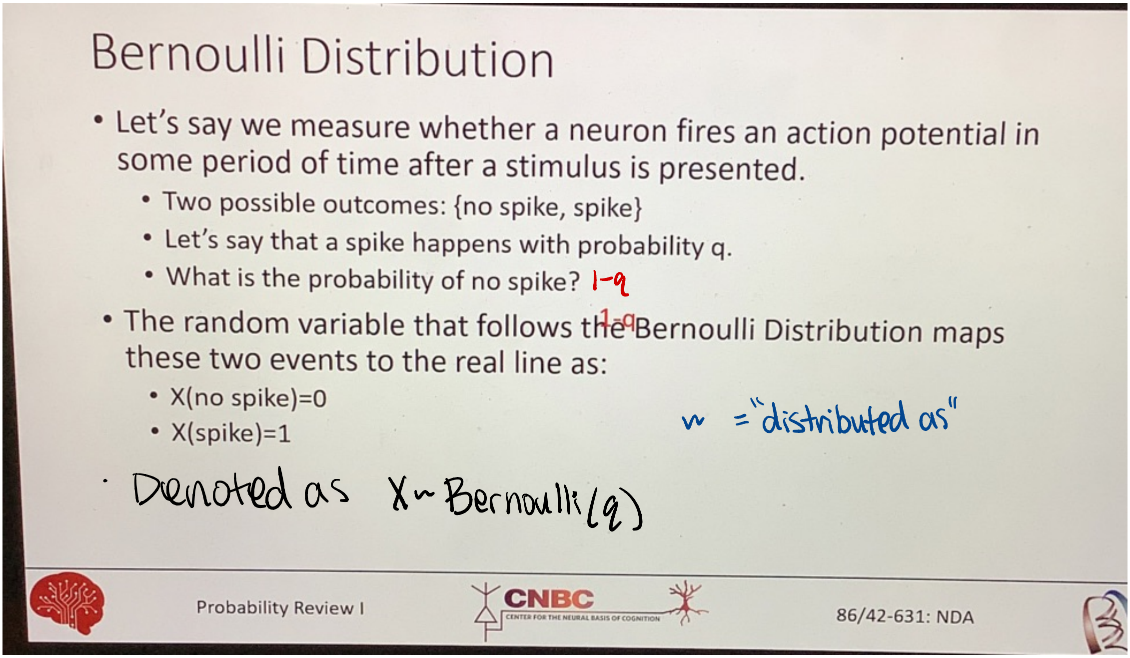
{no spike, no spike; no spike, spike; spike, no spike; spike, spike}.

mutually exclusive so have to add up

Sum up all… does it equal 1? It should.

k=1 {spike, nospike} {nospike, spike} = q(1-q)

Bernoulli Distribution:



n = no spike
s = spike

4 trials:

$k=0$ {n, n, n, n}

$\boxed{(1-q)^4}$

$k=1$ {n,n,n,s; n,n,s,n; n,s,n,n; s,n,n,n}

$k=2$ {n,n,s,s; s,s,n,n; s,n,s,n; n,s,n,s; n,s,s,n; s,n,n,s}

6

If we assume these two tests are independent, then we know the probability of measuring the event {spike,spike} is just q^2. Similarly, the probability of the event {no spike, no spike} is $(1-q)^2$, the probability of the event {spike, no spike} is q(1-q), and the probability of the event {no spike, spike} is (1-q)q.

Now what if we only care about the total number of spikes we observe during the experiment? We could create a random variable, K, that just equals the number of spikes observed over all trials. So, K({no spike, no spike})=0; K({no spike, spike})=1; K({spike, no spike})=1, and K({spike, spike}) would =2. This random variable is said to follow the *binomial distribution:*

$$p(K=0)=(1-q)^2; \quad p(K=1)=q(1-q) + (1-q)q = 2q(1-q), \; p(K=2)=q^2.$$

Since you measured two trials, each independent with probability q, the distribution is said to be binomial(2,q). Let's say you performed n trials, each of which could yield a spike with probability q, independent of all the other trials. We would say that the number of spikes, k, you would measure across all trials would be distributed as a binomial with parameters n and q. This is formally written as:

k=# of spikes

$$K \sim \text{binomial}(n,q).$$

For 3 trials, for example, you could get:

no spikes — k=0: {no spike, no spike, no spike}: $(1-q)^3$.

1 spike — k=1: {ns,ns,s; ns,s,ns; s,ns,ns}: $3q(1-q)^2$

2 spikes — k=2: {ns,s,s; s,ns,s; s,s,ns}: $3(1-q)q^2$

3 spikes — k=3: {s,s,s}: q^3

In general, it can be proven that

$$K \sim bin(3,0) : \{n,n,n\} = (1-q)(1-q)(1-q) = (1-q)^3$$

$$K \sim bin(3,1) : \{n,n,s; n,s,n; s,n,n\} = (1-q)(1-q)q + (1-q)q(1-q) + q(1-q)(1-q) = 3q(1-q)^2$$

$$K \sim bin(3,2) : \{n,s,s; s,s,n; s,n,s\} = (1-q)qq + qq(1-q) + q(1-q)q = 3q^2(1-q)^3$$

$$K \sim bin(3,3) : \{s,s,s\} = qqq = q^3$$

$$P(K=k) = \binom{n}{k} q^k (1-q)^{n-k} \qquad \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

4 Trials:

$$K=0: \{n,n,n,n\} = (1-q)(1-q)(1-q)(1-q) = \boxed{(1-q)^4}$$

$$P(K=0) = \binom{4}{0} q^0 (1-q)^{4-0} = \frac{(4\cdot3\cdot2\cdot1)}{1(4-0)!} \cdot 1 \cdot (1-q)^4 = 1 \cdot 1 \cdot (1-q)^4 = \boxed{(1-q)^4}$$

$$K=1: \{n,n,n,s; n,n,s,n; n,s,n,n; s,n,n,n\} = (1-q)(1-q)(1-q)q + (1-q)(1-q)q(1-q) + (1-q)q(1-q)(1-q) + q(1-q)(1-q)(1-q) = \boxed{4q(1-q)^3}$$

$$P(K=1) = \binom{4}{1} q^1 (1-q)^{4-1} = \frac{4\cdot\cancel{3\cdot2\cdot1}}{1(\cancel{3\cdot2\cdot1})} \cdot q \cdot (1-q)^3 = \boxed{4q(1-q)^3}$$

$$K=2: \{n,n,s,s; s,s,n,n; s,n,s,n; n,s,n,s; n,s,s,n; s,n,n,s\} = (1-q)(1-q)qq + \ldots + \ldots + \ldots + \ldots + \ldots = \boxed{6q^2(1-q)^2}$$

$$P(K=2) = \binom{4}{2} q^2 (1-q)^{4-2} = \frac{4\cdot3\cdot2\cdot1}{(2\cdot1)(2\cdot1)} q^2 (1-q)^2 = \frac{\cancel{4}\cdot3\cdot2\cdot1}{\cancel{4}} q^2 (1-q)^2 = \boxed{6q^2(1-q)^2}$$

$$K=3: \{n,s,s,s; s,n,s,s; s,s,n,s; s,s,s,n\} = (1-q)qqq + \ldots + \ldots + \ldots = \boxed{4q^3(1-q)}$$

$$P(K=3) = \binom{4}{3} q^3 (1-q)^{4-3} = \frac{4\cdot\cancel{3\cdot2\cdot1}}{(\cancel{3\cdot2\cdot1})(1)} \cdot q^3 (1-q)^1 = \boxed{4q^3(1-q)}$$

$$K=4: \{s,s,s,s\} = qqqq = \boxed{q^4}$$

$$P(K=4) = \binom{4}{4} q^4 (1-q)^{4-4} = \frac{\cancel{4\cdot3\cdot2\cdot1}}{(\cancel{4\cdot3\cdot2\cdot1})(1)} q^4 \cdot 1 = \boxed{q^4}$$

100 trials:

$K=50$: $P(K=50) = \binom{100}{50} q^{50} (1-q)^{100-50} = \frac{100!}{50!\,(100-50)!} q^{50} (1-q)^{50}$

$= \boxed{\frac{100!}{50!\,50!} q^{50} (1-q)^{50}}$

$K=2$: $P(K=2) = \binom{100}{2} q^{2} (1-q)^{98}$

$= \frac{100!}{2!\,(98!)} q^{2} (1-q)^{98} = \frac{100 \cdot 99}{2 \cdot 1} = \boxed{4950 q^{2} (1-q)^{98}}$