Lecture 2:

Probability and counting – 2

Counting rules – continued

Counting rules – continued: binomials

We've seen that there are:

- n^k ways to pick k objects from n objects **with** replacement
- $n! = n \cdot (n-1)\ldots 2 \cdot 1$ ways to **arrange** (permute) n objects = pick n objects from n objects **without** replacement

Question: how many ways are there to pick k **objects from** n objects **without** replacement?

Answer: **Binomial coefficient**, $\binom{n}{k}$ (reads "n choose k"), which equals $\binom{n}{k} = \dfrac{n!}{k!\,(n-k)!}$

Counting rules – continued: binomials

Question: how many ways are there to pick k **objects from** n objects **without** replacement?

Answer: **Binomial coefficient,** $\binom{n}{k}$ (reads "n choose k"), which

equals $\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$ – **notice** that just =

$$\frac{\text{\# of perm. of } n \text{ obj.}}{(\text{\# of perm. of } k\ \textbf{chosen} \text{ obj.})(\text{\# of perm. of } (n-k)\ \textbf{not chosen} \text{ obj.})}$$

Counting rules – continued: binomials

The **Binomial theorem** holds: $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} =$

$$= \underbrace{\binom{n}{0}}_{=1} a^0 b^n + \underbrace{\binom{n}{1}}_{=n} a^1 b^{n-1} + \underbrace{\binom{n}{2}}_{=\frac{n(n-1)}{2}} a^2 b^{n-2} + \ldots + \underbrace{\binom{n}{n}}_{=1} a^n b^0$$

To **prove** it, see that $(a+b)^n = \underbrace{(a+b)(a+b)\ldots(a+b)}_{n \text{ brackets}}$ and for

each term we're choosing k a-s from n brackets, thus choosing b

from the rest $(n-k)$ brackets – there are $\binom{n}{k}$ ways to do so.

Counting rules – continued: Bose-Einstein

***Question*:** Now, that, but **with** replacement – there are n objects to choose from, and we're making k choices, and the order of choices doesn't matter? (Would be n^k if order did matter).

*Btw, this problem was considered by physicists in the 1920-s, and let to discovery of what's called **Bose-Einstein** statistic.*

Btw, this is equivalent to finding the # of solutions $(x_1, \ldots, x_n)$ *to* $x_1 + x_2 + \ldots + x_n = k$ *with* x_i*-s being non-negative integers.*

Counting rules – continued: Bose-Einstein

***Question*:** Now, that, but **with** replacement – there are n object to choose from, and we're making k choices, and the order of choices doesn't matter? (Would be n^k if order did matter).

*Btw, this problem was considered by physicists in the 1920-s, and let to discovery of what's called **Bose-Einstein** statistic.*

Let's encode such choices with what's called **"stars and bars"**:
$| \star | \star \star | \star \star \star | \star |$ = a certain placement of 7 stars in 4 boxes.
If there are n boxes, there are $n+1$ bars. 2 outer bars are "fixed", so there are $(n-1)+k$ symbols written between them – of them we have to choose k places – to put $\star$-s there. So, $\binom{n+k-1}{k}$

Non-naive definition of probability

Non-naive definition of probability

Definition 1.6.1: (General definition of probability). A **probability space** consists of a **sample space** S and a **probability function** P which takes an event $A \subset S$ as input and returns $P(A)$, a real number between 0 and 1, as output. P satisfies these axioms:

1. $P(\emptyset) = 0, \quad P(S) = 1$

2. If $A_1, A_2, \ldots$ are disjoint (= mutually exclusive, $A_i \cap A_j = \emptyset,\ i \neq j$) events, then $$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P(A_j)$$

Non-naive definition of probability

Definition 1.6.1: (General definition of probability). A **probability space** consists of a **sample space** S and a **probability function** P.

In the naive formulation, we had pebbles (elementary events) of same mass, of total mass 1, events were (possibly overlapping) piles of pebbles.

With this definition, we can have pebbles of different masses. We can have **countably infinite** number of pebbles, as long as masses add up to 1.

We can even have **uncountable** sample spaces - regions on the real line $\mathbb{R}$, or in the plane $\mathbb{R}^2$, etc.

Non-naive definition of probability

Definition 1.6.1: (General definition of probability). A **probability space** consists of a **sample space** S and a **probability function** P.

The **frequentist** view is: probability = long-run frequency over a large number of repetitions of an experiment

The **Bayesian** view is: probability = degree of belief about the event in question

These two perspectives are complementary and both will be helpful. In both these views, the above axioms lead to the same properties of probability.

Non-naive definition of probability

Theorem 1.6.2: (Properties of probability). Probability has the following properties, for any events A and B

1. $P(A^c) = 1 - P(A)$

2. If $A \subset B$, then $P(A) \leq P(B)$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Non-naive definition of probability

Theorem 1.6.2: (Properties of probability). Probability has the following properties, for any events A and B

1. $P(A^c) = 1 - P(A)$

2. If $A \subset B$, then $P(A) \leq P(B)$

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof: 1. Since A and A^c are disjoint and their union is S, the 2nd axiom gives: $P(S) = P(A \cup A^c) = P(A) + P(A^c)$

By 1st axiom, $P(S) = 1$, so $P(A) + P(A^c) = 1$

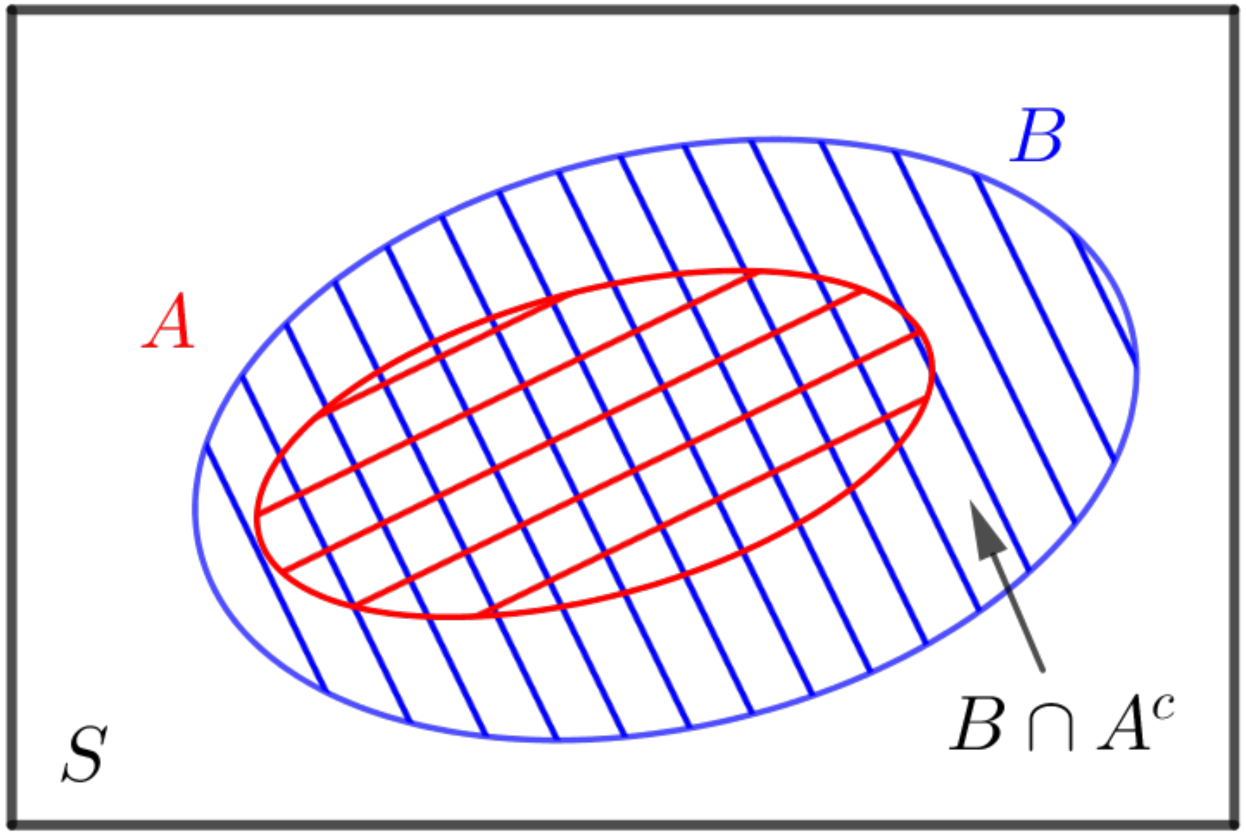
Non-naive definition of probability

Theorem 1.6.2: …

2. If $A \subset B$, then $P(A) \leq P(B)$

Proof:

If $A \subset B$, then $B = A \cup (B \cap A^c)$



Since A and $B \cap A^c$ are disjoint, by 2nd axiom

$$P(B) = P(A \cup (B \cap A^c)) =$$
$$= P(A) + P(B \cap A^c)$$

P is ≥ 0, so $P(B \cap A^c) \geq 0$, so $P(B) \geq P(A)$

Non-naive definition of probability

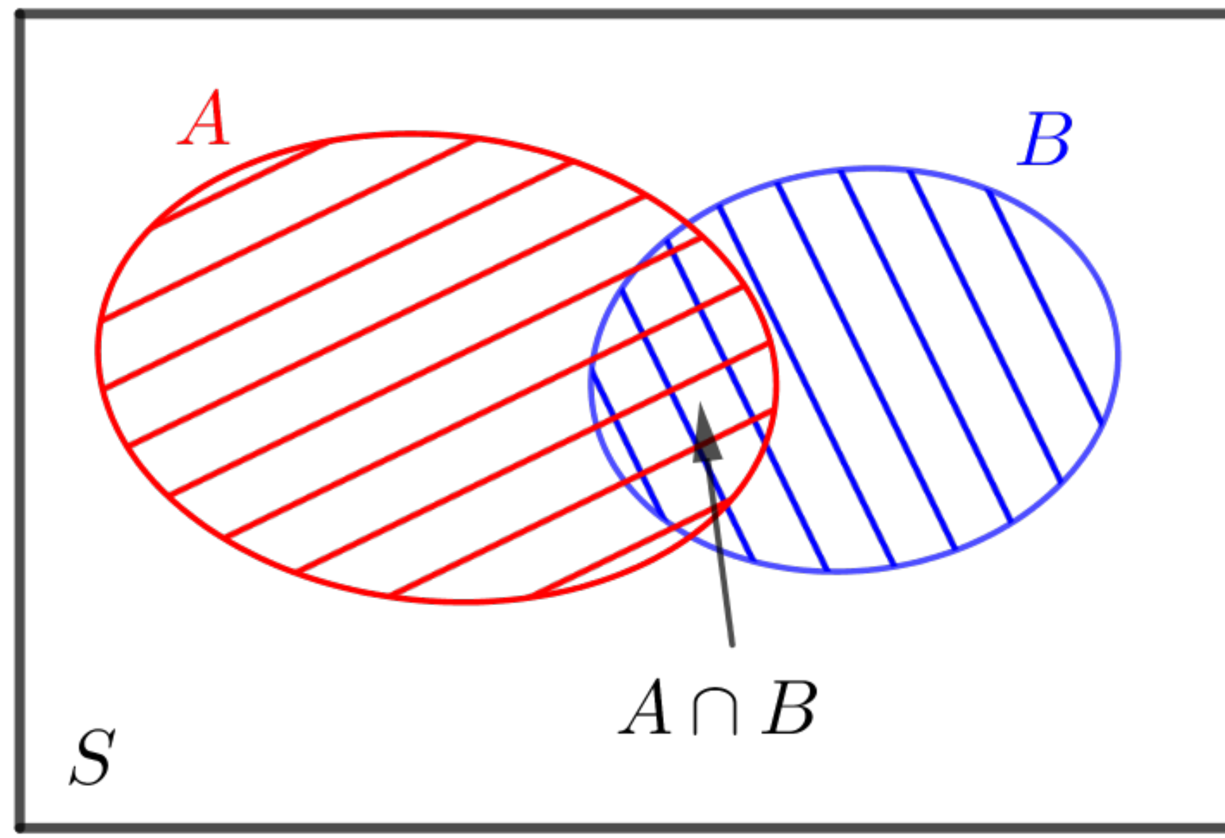
Theorem 1.6.2: …

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Proof:

$A \cup B = A \cup (B \cap A^c)$, so

$P(A \cup B) = P(A) + P(B \cap A^c)$



Since $B \cap A$ and $B \cap A^c$ are disjoint, by 2nd axiom,

$P(B \cap A) + P(B \cap A^c) = P(B)$

So $P(B \cap A^c) = P(B) - P(A \cap B)$, so

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

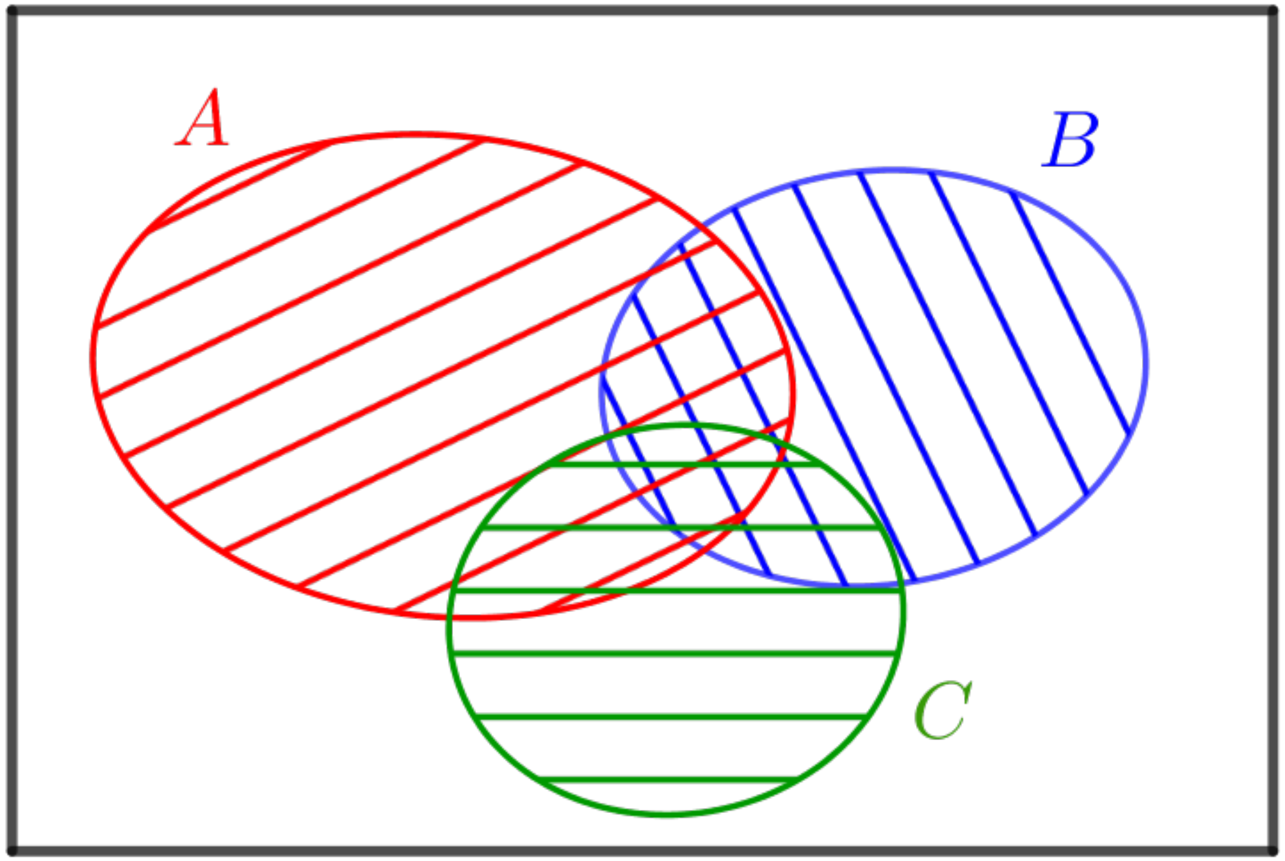
Non-naive definition of probability

Theorem 1.6.2: …

3. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

This is a certain case of **inclusion-exclusion**.

Consider 3 events $\rightarrow$



Intuition says:
(pairwise intersections are counted twice, triple – 3 times)

$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$
$$-P(A \cap B) - P(A \cap C) - P(B \cap C)$$
$$+P(A \cap B \cap C)$$

Non-naive definition of probability

Theorem 1.6.3 (Inclusion-exclusion). For any events $A_1, \ldots, A_n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j) +$$

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n)$$

(This formula is good to use if there's some symmetry of the events A_i. If there is no symmetry, one should try other tools first)

Non-naive definition of probability

Theorem 1.6.3 (Inclusion-exclusion). For any events $A_1, \ldots, A_n$

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i} P(A_i) - \sum_{i<j} P(A_i \cap A_j) +$$
$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \ldots + (-1)^{n+1} P(A_1 \cap \ldots \cap A_n)$$

Example: (de Montmort's matching problem). Consider a well-shuffled deck of n cards, labeled 1 to n. You flip them one by one, saying the numbers 1 through n as doing so. You **win** if, at some point, the number you say is the same as on the card currently flipped over. **Question:** What's the probability of winning?

Non-naive definition of probability

Example: (de Montmort's matching problem). Consider a well-shuffled deck of n cards, labeled 1 to n. You flip them one by one, saying the numbers 1 through n as doing so. You **win** if, at some point, the number you say is the same as on the card currently flipped over. **Question:** What's the probability of winning?

Solution: Let A_i = event that i-th card has number i written on it. We're interested in $P(A_1 \cup \ldots \cup A_n)$. (***Why?***)

First, $P(A_i) = \frac{1}{n}$. Second, $P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$ (***Why?***)

Non-naive definition of probability

First, $P(A_i) = \frac{1}{n}$. Second, $P(A_i \cap A_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$

Next, $P(A_i \cap A_j \cap A_k) = \frac{1}{n(n-1)(n-2)}$ and so on.

In the inclusion-exclusion formula, there will be n terms involving 1 event, $\binom{n}{2}$ terms involving 2 events, $\binom{n}{3}$ for 3 events, etc, so

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \frac{n}{n} - \frac{\binom{n}{2}}{n(n-1)} + \frac{\binom{n}{3}}{n(n-1)(n-2)} - \ldots + (-1)^{n+1}\frac{1}{n!}$$

$= 1 - \frac{1}{2!} + \frac{1}{3!} - \ldots + (-1)^{n+1}\frac{1}{n!}$ – that resembles the Taylor series for $e^{-1} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots$, only it has an extra 1 and signs are in another order, so (for large n) $P(\text{win}) = 1 - 1/e \approx 0.63$