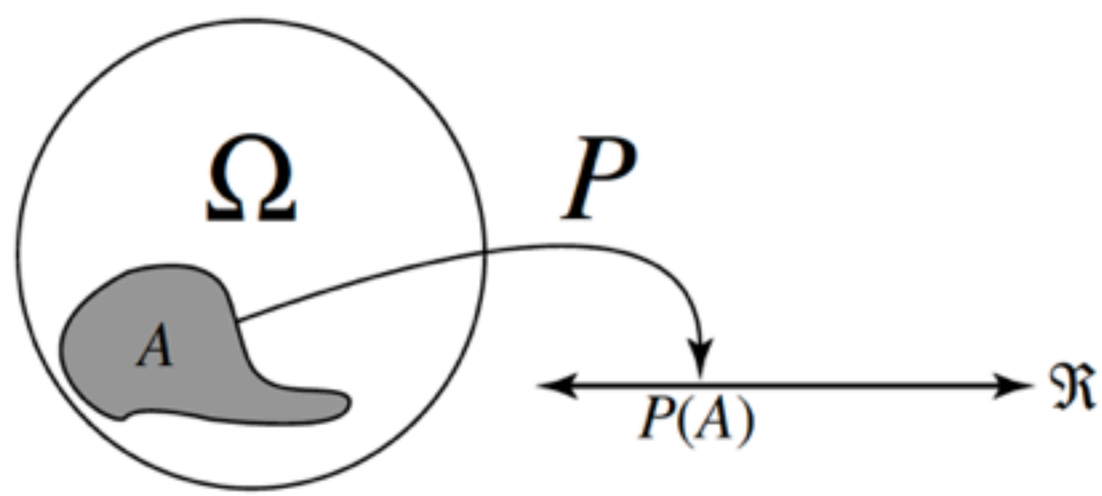
February 8th
Probability for Linguists*

*with slides grabbed from Mark Paskin's Short Course on Graphical Models

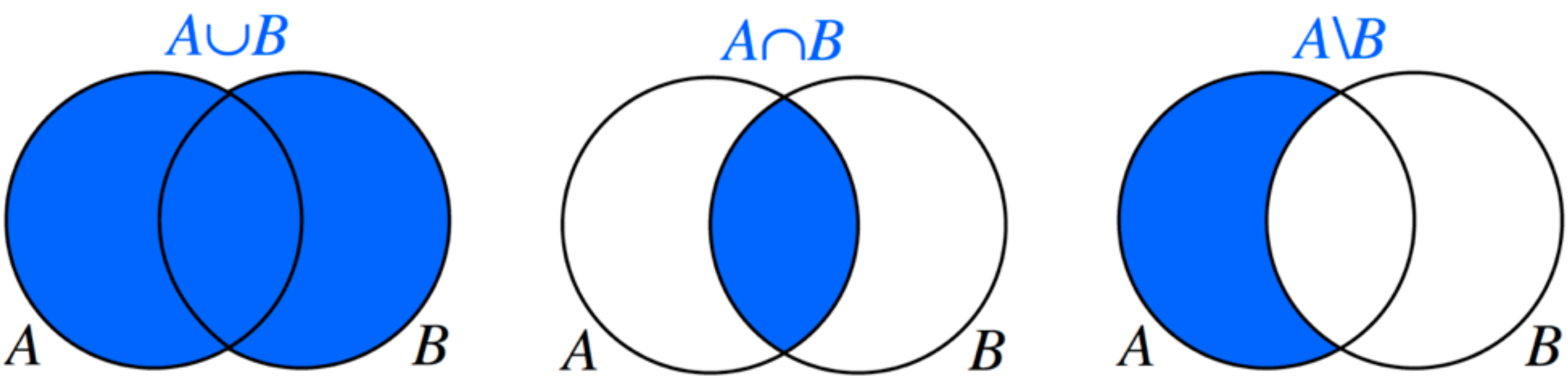
Probability spaces

- A *probability space* represents our uncertainty regarding an *experiment*.
- It has two parts:
  1. the *sample space* Ω, which is a set of *outcomes*; and
  2. the *probability measure* P, which is a real function of the subsets of Ω.



- A set of outcomes $A \subseteq \Omega$ is called an *event*. $P(A)$ represents how likely it is that the experiment's *actual* outcome will be a member of A.

The only prerequisite: Set Theory



For simplicity, we will work (mostly) with finite sets. The extension to countably infinite sets is not difficult. The extension to uncountably infinite sets requires Measure Theory.

Complex events

Let E be the number of raindrops, out of a four-drop spritz, that land on the Righthand stone

$A \subset \Omega$	$\mathbb{N}$
LLLL	0
RLLL,LRLL,LLRL,LLLR	1
LLRR,LRRL,RRLL,RLLR,RLRL,LRLR	2
LRRR,RRRL,RLRR,RRLR	3
RRRR	4

Complex events

Let E be the number of raindrops, out of a four-drop spritz, that land on the Righthand stone

$A \subset \Omega$	$\mathbb{N}$
LLLL	0
RLLL,LRLL,LLRL,LLLR	1
LLRR,LRRL,RRLL,RLLR,RLRL,LRLR	2
LRRR,RRRL,RLRR,RRLR class A_3	3
RRRR	4

Complex events

Let E be the number of raindrops, out of a four-drop spritz, that land on the Righthand stone

$A \subset \Omega$		$\mathbb{N}$
LLLL		0
RLLL,LRLL,LLRL,LLLR	class A_1	1
LLRR,LRRL,RRLL,RLLR,RLRL,LRLR		2
LRRR,RRRL,RLRR,RRLR	class A_3	3
RRRR		4

Complex events

Let E be the number of raindrops, out of a four-drop spritz, that land on the Righthand stone

$A \subset \Omega$		$\mathbb{N}$
LLLL		0
RLLL,LRLL,LLRL,LLLR	class A_1	1
LLRR,LRRL,RRLL,RLLR,RLRL,LRLR		2
LRRR,RRRL,RLRR,RRLR	class A_3	3
RRRR	class A_4	4

First die	Second die											
die	1	2	3	4	5	6						
6	7	8	9	10	11	12						
5	6	7	8	9	10	11						
4	5	6	7	8	9	10						
3	4	5	6	7	8	9						
2	3	4	5	6	7	8						
1	2	3	4	5	6	7						
x		2	3	4	5	6	7	8	9	10	11	12
$p(X = x)$		$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Figure 2.2 A random variable X for the sum of two dice. Entries in the body of the table show the value of X given the underlying basic outcomes, while the bottom two rows show the pmf $p(x)$.

You are my density

Probability mass function, probability density function, or just "probability distribution" means the probability $P(X = x_k)$ that r.v. X takes on the value x_k. Sometimes written $f(x)$.

$$
\begin{aligned}
f(0) = \quad & P(E = 0) & = \left(\tfrac{1}{2}\right)^4 \text{ or } \tfrac{1}{16} \text{ or } 0.0625 \\
f(1) = \quad & P(E = 1) & = \tfrac{4}{16} \text{ or } 0.25 \\
f(2) = \quad & P(E = 2) & = \tfrac{6}{16} \text{ or } 0.375 \\
f(3) = \quad & P(E = 3) & = \tfrac{4}{16} \text{ or } 0.25 \\
f(4) = \quad & P(E = 4) & = \tfrac{1}{16} \text{ or } 0.0625
\end{aligned}
$$

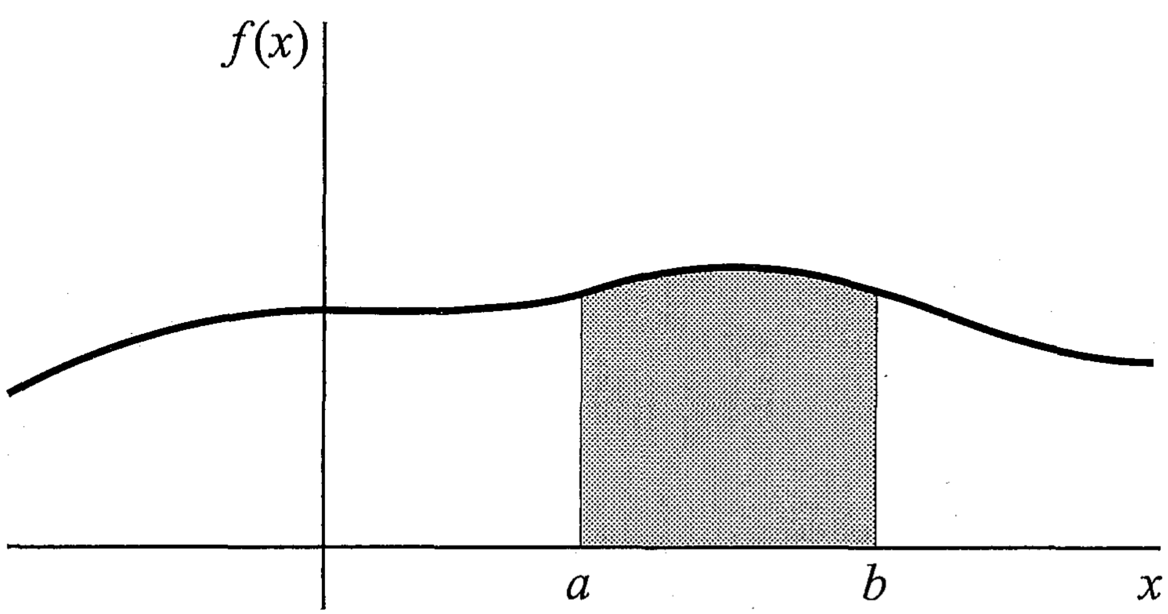
You have the same success prob, every time

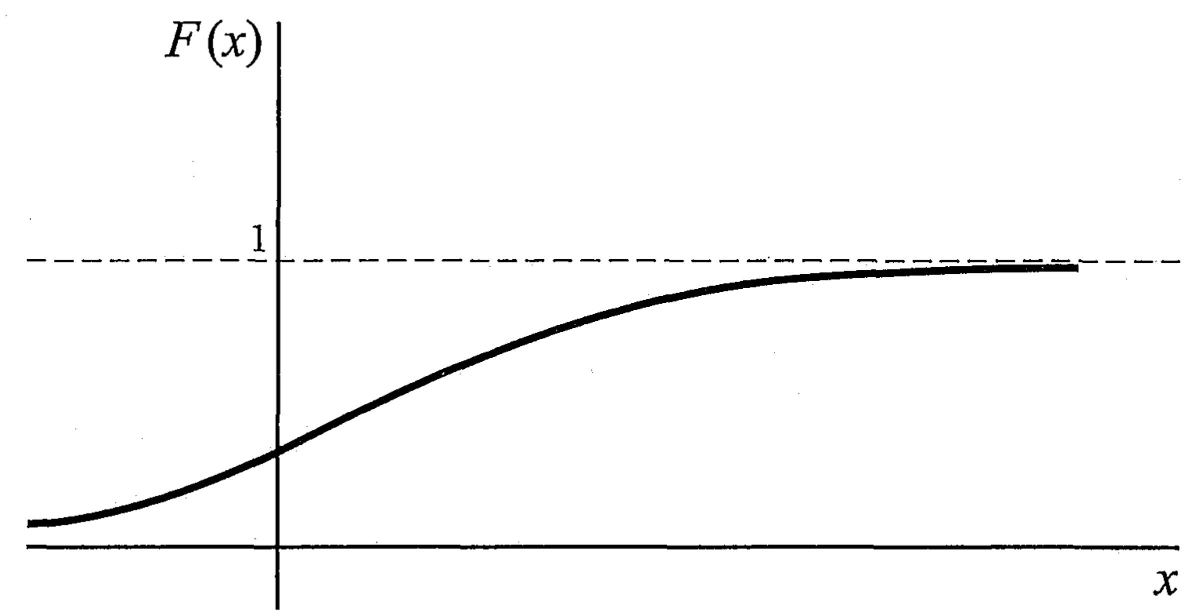
```
binomialprobability <- function(n,p,k){
  choose(n,k)*p^k*(1-p)^(n-k)
}
```

You are my CDF

The cumulative distribution function $F(x)$ for a r.v. X is the probability that X takes on a value of x or less.

$$F(x) = P(X \le x)$$

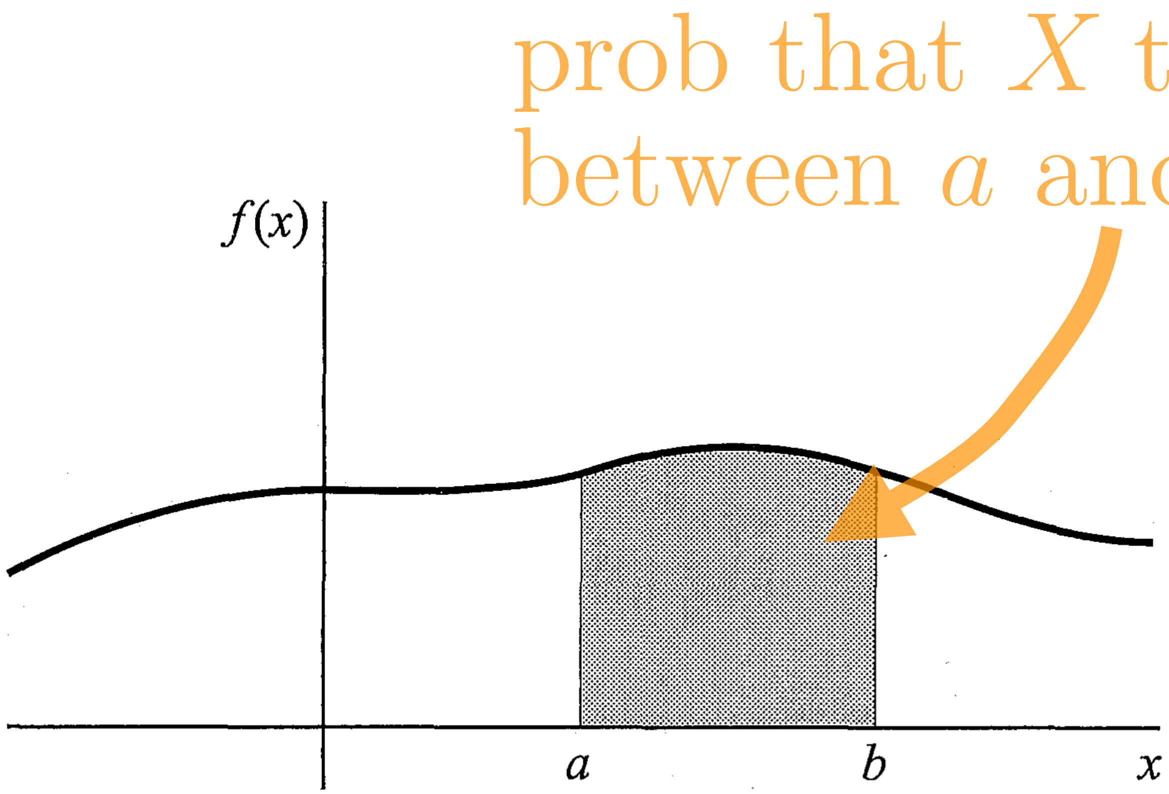


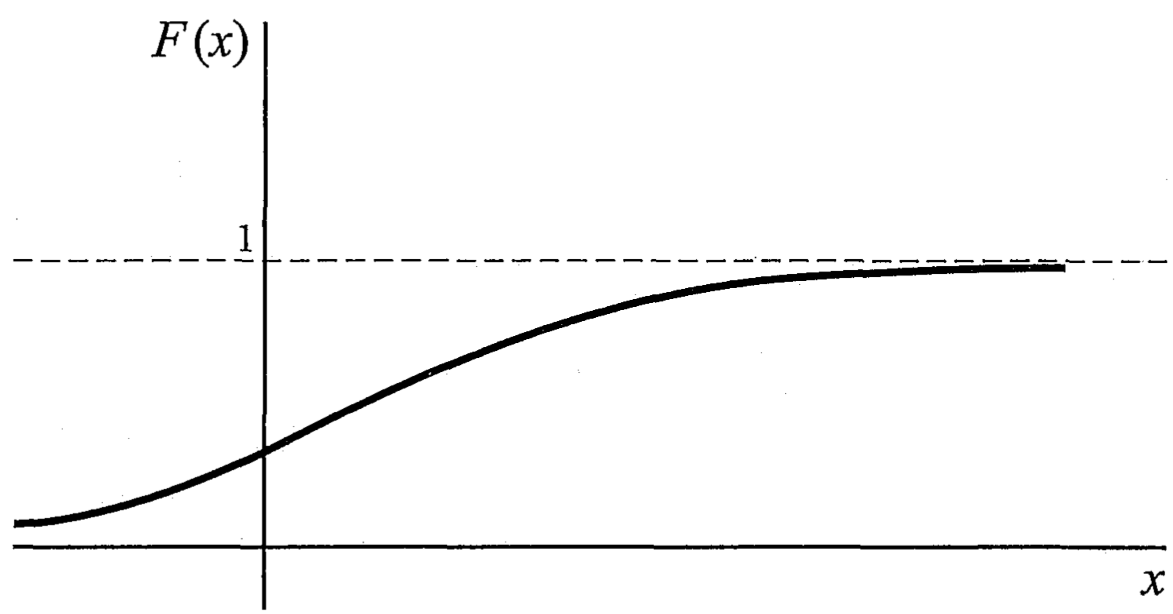


You are my CDF

The cumulative distribution function $F(x)$ for a r.v. X is the probability that X takes on a value of x or less.

$$F(x) = P(X \leq x)$$





An example probability space

- If our experiment is to deploy a smoke detector and see if it works, then there could be four outcomes:

  $$\Omega = \{(\textit{fire}, \textit{smoke}), (\textit{no fire}, \textit{smoke}), (\textit{fire}, \textit{no smoke}), (\textit{no fire}, \textit{no smoke})\}$$

  Note that these outcomes are *mutually exclusive*.

- And we may choose:
  - $P(\{(\textit{fire}, \textit{smoke}), (\textit{no fire}, \textit{smoke})\}) = 0.005$
  - $P(\{(\textit{fire}, \textit{smoke}), (\textit{fire}, \textit{no smoke})\}) = 0.003$
  - $\ldots$

- Our choice of P has to obey three simple rules...

Conditional probability picture

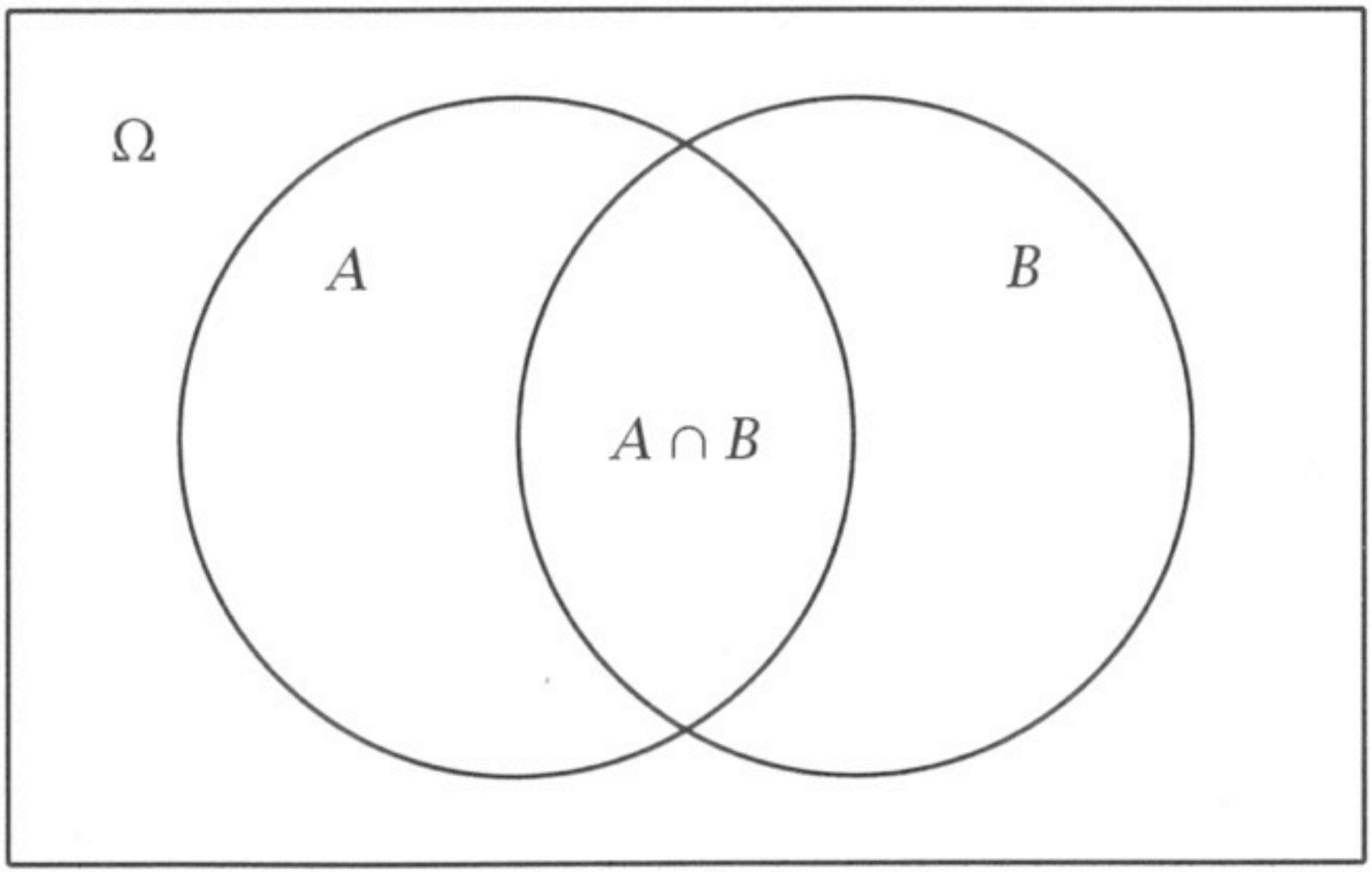


Figure 2.1 A diagram illustrating the calculation of conditional probability $P(A|B)$. Once we know that the outcome is in B, the probability of A becomes $P(A \cap B)/P(B)$.

Independence example

1.43. The probabilities that a husband and wife will be alive 20 years from now are given by 0.8 and 0.9, respectively. Find the probability that in 20 years (*a*) both, (*b*) neither, (*c*) at least one, will be alive.

Let H, W be the events that the husband and wife, respectively, will be alive in 20 years. Then $P(H) = 0.8$, $P(W) = 0.9$. We suppose that H and W are independent events, which may or may not be reasonable.

(*a*) $P(\text{both will be alive}) = P(H \cap W) = P(H)P(W) = (0.8)(0.9) = 0.72.$

(*b*) $P(\text{neither will be alive}) = P(H' \cap W') = P(H')P(W') = (0.2)(0.1) = 0.02.$

(*c*) $P(\text{at least one will be alive}) = 1 - P(\text{neither will be alive}) = 1 - 0.02 = 0.98.$

Independent?

Table 5.2. *Numbers of monolingual or bilingual adults in two hypothetical populations cross-tabulated by sex*

Population A

	Male	Female	Total
Bilingual	2 080	1 920	4 000
Monolingual	3 120	2 880	6 000
	5 200	4 800	10 000

=0.6*0.52

Population B

	Male	Female	Total
Bilingual	2 500	1 500	4 000
Monolingual	2 700	3 300	6 000
	5 200	4 800	10 000

≠0.6*0.4