Event, \( A \)

The complement of the Event \( A \), \( A^c \)

State space: The possible events.

The Probability of an event \( \Pr(A) \) satisfies the following

(i) \( \Pr(A) \geq 0 \),  (ii) \( \Pr(A) \leq 1 \),  (iii) Probabilities sum to one

Computing probabilities for multiple events

It may often be easier to compute the probability for the complement of a complex event, and then use

\[ \Pr(A^c) = 1 - \Pr(A) \]

---

**Example 1.** Consider the toss of a coin, which is heads with probability \( p \). What is the probability of seeing one or more heads in 100 flips? Well, one could compute this as

\[
\sum_{i=1}^{100} \Pr(k \text{ heads}) = \Pr(\text{one head}) + \Pr(\text{two heads}) + \cdots + \Pr(\text{all heads})
\]

or much more simply as

\[ 1 - \Pr(\text{no heads in 100 flips}) = 1 - (1 - p)^{100} \]

---

Probability of two (or more) independent events: The **And rule**. If the events \( A \) and \( B \) are independent, then

\[ \Pr(A \text{ and } B) = \Pr(A) \cdot \Pr(B) \]

In other words, for joint independent events, multiply probabilities,

\[ \Pr(A_1 \text{ and } A_2 \text{ and } \cdots \text{ and } A_n) = \prod_{i=1}^{n} \Pr(A_i) \]
Note that we usually use the shorthand of writing \( \Pr(A \text{ and } B) = \Pr(A, B) \), the so called joint probability of events \( A \) and \( B \).

If \( A \) and \( B \) are not independent, then known that (say) \( B \) has occurred provides information on the probability that \( A \) occurs. This conditional probability of event \( A \) given that event \( B \) is seen (or of \( A \) given \( B \)), is written \( \Pr(A \mid B) \). Likewise, the probability of Event \( A \) regardless of whether or not event \( B \) occurred, \( \Pr(A) \) is called the marginal probability of \( A \). The joint, marginal, and conditional probabilities are related as follows:

\[
\Pr(x, y) = \Pr(x \mid y) \cdot \Pr(y)
\]

Hence, we can compute the conditional probability from

\[
\Pr(x \mid y) = \frac{\Pr(x, y)}{\Pr(y)}
\]

**Example 2.** Suppose the genotypes \( AA \) and \( Aa \) give a yellow flower, while \( aa \) is pink. Suppose the frequency of \( A \) is 0.75 in a population and that Hardy-Weinberg holds. What is the probability that a yellow flower is \( AA \)?

Freq(\( AA \)) = 0.75\(^2\) = 0.5625 (9/16), while

Freq(\( Aa \)) = \( 2 \times (3/4)(1/4) = 0.375 \) (3/8). Thus,

\[
\Pr(\text{Yellow}) = \Pr(\text{AA}) + \Pr(\text{Aa}) = 0.9375 (15/16),
\]

and

\[
\Pr(\text{AA} \mid \text{yellow}) = \frac{\Pr(\text{yellow and } AA)}{\Pr(\text{yellow})} = \frac{\Pr(\text{AA})}{\Pr(\text{yellow})} = \frac{9/16}{15/16} = 0.6.
\]

**Bayes’ theorem.** Suppose there are \( n \) possible outcomes \( (b_1, b_2, \cdots, b_n) \) of a random variable that we cannot observe. Given the observed outcome of a correlated variable \( A \), what is the probability of \( b_j \)? From the definition of a conditional probability, \( \Pr(b_j \mid A) = \Pr(b_j, A) / \Pr(A) \). We can decompose this further, by noting that \( \Pr(b_j, A) = \Pr(b_j) \Pr(A \mid b_j) \) and \( \Pr(A) = \sum_i^n \Pr(b_i) \Pr(A \mid b_i) \). Putting these together gives Bayes’ theorem,

\[
\Pr(b_j \mid A) = \frac{\Pr(b_j) \Pr(A \mid b_j)}{\sum_{i=1}^n \Pr(b_i) \Pr(A \mid b_i)}
\]
Example 3. Consider a genetic condition that causes all offspring to be female, and suppose the frequency of this disorder in the population is 0.01. Suppose you observe a family with seven girls (and no boys). What is the chance that they have the disorder? Here $b_1 = \text{disorder family}$, $b_2 = \text{normal family}$.

\[
\Pr(7 \text{ girls} \mid \text{disorder}) = 1, \quad \Pr(7 \text{ girls} \mid \text{normal}) = (1/2)^7
\]

\[
\Pr(7 \text{ girls}) = \Pr(7 \text{ girls} \mid \text{dis}) \cdot \Pr(\text{dis}) + \Pr(7 \text{ girls} \mid \text{nor}) \cdot \Pr(\text{nor})
\]

\[
= 1 \cdot 0.01 + (1/2)^7 \cdot 0.99 = 0.0178
\]

Putting these together,

\[
\Pr(\text{dis} \mid 7 \text{ girls}) = \frac{\Pr(\text{dis}) \cdot \Pr(7 \text{ girls} \mid \text{dis})}{\Pr(7 \text{ girls})} = \frac{0.01 \cdot 1}{0.0178} = 0.564
\]

The Or Rule: the probability of a set of disjoint events. If the events $A$ and $B$ are disjoint (non-overlapping), then

\[
\Pr(A \text{ or } B) = \Pr(A) + \Pr(B)
\]

If $A$ and $B$ overlap, then

\[
\Pr(A \text{ or } B) = \Pr(A) + \Pr(B) - \Pr(A, B)
\]

Probability distributions: Discrete

For discrete events, $\Pr(X = x)$, where $X$ is a random variable and $x$ a particular outcome (for example, $X$ is a roll of the dice, and $X = 5$ is the outcome of a roll of five).

Common discrete distributions

1. Bernoulli. Parameter $p$. Outcomes are $X = 0$ or $X = 1$ (success or failure), where $p = \text{pop. success}$

\[
P(x) = \begin{cases} 
0 & \text{with probability } 1 - p \\
1 & \text{with probability } p 
\end{cases}
\]

Mean $\mu = p$, variance $\sigma^2 = p(1 - p)$

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \text{ for } 0, 1, 2, \cdots, n$$

Mean $\mu = np$, variance $\sigma^2 = np(1-p)$


$$P(x) = p(1-p)^{x-1}, \text{ for } 1, 2, \cdots, \infty$$

Mean $\mu = (1-p)/p$, variance $\sigma^2 = (1-p)/p^2$

4. Negative binomial. Parameters $r, p$. Total number of failure before $r$ successes occur

$$P(x) = \frac{(r+x-1)!}{(r-1)! x!} p^r (1-p)^x, \text{ for } 0, 1, 2, \cdots, \infty$$

Mean $\mu = r(1-p)/p$, variance $\sigma^2 = r(1-p)/p^2$

5. Poisson. Parameter $\lambda$. Number of events occurring when the expected number of events in $\lambda$.

$$P(x) = \frac{e^{-\lambda} \lambda^x}{x!}, \text{ for } 0, 1, 2, \cdots, \infty$$

Mean $\mu = \sigma^2 = \lambda$

---

Example 4. The poisson correction for sequence comparison. Suppose were examining the rate of evolution of a protein sequence. Two sequences are aligned from two species of interest, and we we note that the sequences match at 40 of 100 sites. Assuming each mutation is different, what is the estimated per site rate of substitution? Note that two sights that differ could each have fixed different mutations, so simply using $60/100 = 0.60$ per site underestimates the rate. If $\lambda$ is the true rate, the probability that a site unremains unchanged is $e^{-\lambda}$, which we can estimate from the frequency of sites with no mutations. Hence

$$e^{-\lambda} = 0.4, \text{ or } \lambda = -\ln(0.4) = 0.916$$