

## Experiments and Events:

**Def:** An experiment is a process whose outcome is not known in advance with certainty.

**Sample Space:** Collection of *all* possible outcomes of an experiment.  $S$  or  $\Omega$ . Each outcome is an element of the sample space  $s \in S$ .

## Operations:

**Union:**  $x \in S : A \cup B = \{x \in A \text{ or } x \in B\}$

$$A \cup B = B \cup A$$

$$A \cup A = A$$

$$A \cup \emptyset = A$$

$$A \cup S = S$$

$$A \subset B \Rightarrow A \cup B = B$$

$$A_1, A_2, \dots, A_n \Rightarrow A_1 \cup A_2 \cup \dots \cup A_n = \bigcup_{i=1}^{i=n} A_i$$

$$\bigcup_{i=1}^{\infty} A_i \rightarrow \bigcup_{i \in I} A_i$$

$$(A \cup B) \cup C = A \cup (B \cup C) = A \cup B \cup C$$

**Intersections:**  $A \cap B = \{x \in A \text{ and } x \in B\} = AB$

$$A \cap B = B \cap A$$

$$A \cap A = A$$

$$A \cap \emptyset = \emptyset$$

$$A \cap S = A$$

$$A \subset B \Rightarrow A \cap B$$

$$\bigcap_{i \in I} A_i = \bigcap_{i=1}^{\infty} A_i$$

$$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$$

$$(A \cap B) \cap C = A \cap (B \cap C) = A \cap B \cap C$$

**Complements:**  $A^c = \{x \in S : x \notin A\}$

$$(A^c)^c = A$$

$$\emptyset^c = S$$

$$S^c = \emptyset$$

$$A \cup A^c = S$$

$$A \cap A^c = \emptyset$$

## Disjoint Events:

$A$  and  $B$  are *disjoint* or *mutually exclusive* if  $A$  and  $B$  have no outcomes in common. This happens only if  $A \cap B = \emptyset$ .

**Def:** A collection  $A_1, \dots, A_n$  is a collection of disjoint events if and only if  $A_i \cap A_j = \emptyset, \forall i, j, i \neq j$

$$\left( \bigcup_{i \in I} A_i \right)^c = \bigcap_{i \in I} A_i^c$$

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

$$x \in (A \cap B)^c \Rightarrow x \notin A \text{ and } x \notin B \Rightarrow x \in A^c \text{ and } x \in B^c \Rightarrow x \in A^c \cap B^c$$

## Probabilities:

Functions over  $S$  that measure the likelihood of events.

$$\forall A : Pr(A) \geq 0$$

$$Pr(S) = 1$$

For every *infinite sequence* of *disjoint* events

$$A_1, A_2, \dots (A_i \in S) : Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} Pr(A_i) Pr(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = Pr(A_1) + Pr(A_2) + \dots + Pr(A_n) + \dots$$

$$Pr(\emptyset) = 0$$

$$Pr\left(\bigcup_{i=1}^n A_i\right) = Pr\left(\bigcup_{i=1}^n A_i + \bigcup_{i=n+1}^{\infty} \emptyset\right) = \sum_{i=1}^n Pr(A_i)$$

$$Pr(A^c) = 1 - Pr(A)$$

$$A \subset B \Rightarrow Pr(A) \leq Pr(B)$$

$$\forall A : 0 \leq Pr(A) \leq 1$$

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

## Finite Sample Spaces:

$$S := \{s_1, s_2, \dots, s_n\}$$

To obtain a probability distribution over  $S$  we need to specify  $Pr(s_i) = P_i, \forall i = 1, 2, \dots, n$ , such that  $\sum_{i=1}^n P_i = 1$

**Def:** A sample space  $S$  with  $n$  outcomes  $s_1, \dots, s_n$  is a *simple sample space* if the probability assigned to each outcome is  $\frac{1}{n}$ . If  $A$  contains  $m$  outcomes then  $Pr(A) = \frac{m}{n}$ .

## Counting Methods:

**Multiplication Rule:** Suppose an experiment has  $k$  parts ( $k \geq 2$ ) such that the  $i^{th}$  part of the experiment has  $n_i$  possible outcomes,  $i = 1, \dots, k$ , and that *all possible outcomes can occur regardless of which outcomes have occurred in other parts*. The sample space  $S$  will contain vectors of the form  $(u_1, u_2, \dots, u_k)$ .  $u_i$  is one of the  $n_i$  possible outcomes of part  $i$ . The total number of vectors is  $n_1 \cdot n_2 \cdot \dots \cdot n_k$ .

**Permutations:** Given an array of  $n$  elements the first position can be filled with  $n$  different elements, the second with  $n - 1$ , and so on.  $n \cdot (n - 1) \cdot (n - 2) \cdot \dots \cdot 1 = n!$

$$P_{n,k} = \frac{n!}{(n-k)!}$$

$$P_{n,n} = n!$$

**Combinations:** In general we can "combine"  $n$  elements taking  $k$  at a time in  $C_{n,k} = \frac{P_{n,k}}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$ .

**Multinomial Coefficients:** Consider splitting  $n$  elements into  $k$  ( $k \geq 2$ ) groups in a way such that group  $j$  gets

$n_j$  elements and  $\sum_{j=1}^k n_j = n$ . The  $n_1$  elements in the first

group can be selected in  $\binom{n}{n_1}$ , the second in  $\binom{n-n_1}{n_2}$ , the third in  $\binom{n-n_1-n_2}{n_3}$  and so on until we complete the  $k$  groups. Then:  $\binom{n}{n_1} \cdot \binom{n-n_1}{n_2} \cdot \binom{n-n_1-n_2}{n_3} \cdot \dots \cdot \binom{n_k}{n_k} = \binom{n}{n_1, n_2, \dots, n_k}$

**Probability of union:** If  $A_1, A_2, \dots, A_n$  are *disjoint events* then  $Pr(A_1 \cup A_2 \cup \dots \cup A_n) = Pr\left(\bigcup_{i=1}^n A_i\right) =$

$$Pr(A_1) + Pr(A_2) + \dots + Pr(A_n) = \sum_{i=1}^n Pr(A_i)$$

If the events are not disjoint:  
 $A_1, A_2 : Pr(A_1 \cup A_2) = Pr(A_1) + Pr(A_2) - Pr(A_1 \cap A_2)$   
 $A_1, A_2, A_3 : Pr(A_1) + Pr(A_2) + Pr(A_3) - Pr(A_1 \cap A_2) - Pr(A_1 \cap A_3) - Pr(A_2 \cap A_3) + Pr(A_1 \cap A_2 \cap A_3)$

## Conditional Probability:

If  $A, B$  are events such that  $Pr(A) > 0$  and  $Pr(B) > 0$  then  $Pr(B|A) = \frac{Pr(A \cap B)}{Pr(A)}$  and  $Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}$  Furthermore:  
 $Pr(A \cap B) = Pr(B|A) \cdot Pr(A)$  and  
 $Pr(A \cap B) = Pr(A|B) \cdot Pr(B)$ . In general:  
 $Pr(A_1 \cap A_2 \cap \dots \cap A_n) = Pr(A_1) \cdot Pr(A_2|A_1) \cdot \dots \cdot Pr(A_n|A_1 \cap A_2 \cap \dots \cap A_{n-1})$

**Independence:**  $A, B$  are independent events if  $Pr(A|B) = Pr(A)$  and  $Pr(B|A) = Pr(B)$ . Then, if  $A, B$  are independent:  $Pr(A \cap B) = Pr(A|B) \cdot Pr(B) = Pr(A) \cdot Pr(B)$  and  $Pr(A \cap B) = Pr(B|A) \cdot Pr(A) = Pr(B) \cdot Pr(A)$ . In general if  $A_1, A_2, \dots, A_n$  are independent,  $Pr(A_1 \cap A_2 \cap \dots \cap A_n) = Pr(A_1) \cdot Pr(A_2) \cdot \dots \cdot Pr(A_n)$ . Note that if  $A \cap B = \emptyset$  then the two events are *not independent*. note that if  $A, B$  are independent then  $A, B^c$  are also independent.

**Conditionally Independent:**  $A_1, \dots, A_k$  are *conditionally independent* given  $B$  if, for every subset  $A_{i_1}, \dots, A_{i_m} : Pr(A_{i_1} \cap \dots \cap A_{i_m}|B) = Pr(A_{i_1}|B) \cdot \dots \cdot Pr(A_{i_m}|B)$ .

**Partitions:** Let  $B_1, \dots, B_k$  be such that  $B_i \cap B_j = \emptyset \forall i \neq j$  and  $\bigcup_{i=1}^k B_i = S$ . Then these events form a partition of  $S$ .

$$A = A \cap S = A \cap \left( \bigcup_{i=1}^k B_i \right) = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_k).$$

Then:

$$Pr(A) = Pr(A \cap S) = Pr\left(A \cap \left(\bigcup_{i=1}^k B_i\right)\right) = Pr(A \cap B_1) + Pr(A \cap B_2) + \dots + Pr(A \cap B_k) = Pr(A|B_1) \cdot Pr(B_1) + Pr(A|B_2) \cdot Pr(B_2) + \dots + Pr(A|B_k) \cdot Pr(B_k) = \sum_{i=1}^k Pr(A|B_i) \cdot Pr(B_i).$$

So, if  $B_1, \dots, B_k$  are a partition of  $S$ :

$$Pr(A) = \sum_{i=1}^k Pr(A|B_i) \cdot Pr(B_i)$$

**Bayes' Theorem:**  $B_1, \dots, B_k$  := a partition of  $S$  such that  $Pr(B_j) > 0, j = 1, \dots, k$ . Assume you have  $A$  such that  $Pr(A) > 0$ . Then:

$$Pr(B_i|A) = \frac{Pr(A|B_i) \cdot Pr(B_i)}{Pr(A)} = \frac{Pr(A|B_i) \cdot Pr(B_i)}{\sum_{j=1}^k Pr(A|B_j) \cdot Pr(B_j)}$$

## Random Variables:

**Def:** A real-valued function on  $S$  is a random variable. A random variable  $X$  is a function that assigns a real number  $X(s) = x$  to each possible outcome  $s \in S$ :  $X : S \Rightarrow \mathcal{D}$ .  $x :=$  a realization of the random variable,  $x \in \mathcal{D}$ .

**Notation:** We will be computing  $Pr(X \in E)$  for  $E \subset \mathcal{D} = Pr(s \in S : X(s) \in E)$

**Discrete Probability Distributions:** A r.v.  $X$  has a discrete distribution if it takes a countable number of values. The probability function of a discrete r.v. is  $f_X(x) = Pr(X = x)$ . Properties:

$$0 \leq f_X(x) \leq 1$$

$$\forall x \notin \mathcal{D} : f_X(x) = 0$$

$$\sum_{x \in \mathcal{D}} f_X(x) = 1$$

$$Pr(X \in A) = \sum_{x \in A} f_X(x)$$

**Uniform Distribution:**  $X = x, x \in \{1, 2, \dots, k\}$  with all values  $x$  equally likely. The p.f. is  $f_X(x) = Pr(X = x) =$

$$\begin{cases} \frac{1}{k} & x = 1, 2, \dots, k \\ 0 & o.w. \end{cases}$$

**Bernoulli Distribution:** An event  $A$  happens with probability  $p$   
 $X =$

$$\begin{cases} 1 & \text{if } A \text{ happens} \\ 0 & \text{if } A^c \text{ happens} \end{cases}$$

$$f_X(x) =$$

$$\begin{cases} (1-p) & x = 0 \\ p & x = 1 \\ 0 & o.w. \end{cases}$$

**Binomial:**  $n$  Bernoulli trials repeated independently with probability of success  $p$ .  $X :=$  number of success in  $n$  trials.  
 $x \in \{0, 1, \dots, n\}$ .  $f_X(x) = Pr(X = x) =$

$$\begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0, \dots, n \\ 0 & o.w. \end{cases}$$

$$X \sim Bin(n, p)$$

**Hypergeometric:** A box with  $A$  red balls and  $B$  blue balls.  $n$  balls are drawn *without replacement*.  $X :=$  number of red balls.  $X \leq \min(n, A)$ .  $\max(n - B, 0) \leq X \leq \min(n, A)$ .  
 $f_X(x) = Pr(X = x) =$

$$\begin{cases} \frac{\binom{A}{x} \cdot \binom{B}{n-x}}{\binom{A+B}{n}} & \text{for } \max(n - B, 0) \leq x \leq \min(n, A) \\ 0 & o.w. \end{cases}$$

**Negative-Binomial:** We repeat Bernoulli trials until  $r$  successes are observed.  $X :=$  number of failures  $= \{0, 1, \dots\}$ .  
 $p :=$  probability of success.

$$Pr(X = x) = Pr(x \text{ failures before } r \text{ successes}) = Pr(x \text{ failures and } r - 1 \text{ successes in } x + r - 1 \text{ trials}) \cdot Pr(\text{one success in last trial}) = \left[ \binom{x+r-1}{x} (1-p)^x p^{r-1} \right] \cdot p = \binom{x+r-1}{x} (1-p)^x p^r. f_X(x) =$$

$$\begin{cases} \binom{x+r-1}{x} (1-p)^x p^r & x = 0, 1, 2, \dots \\ 0 & o.w. \end{cases}$$

**Geometric:** Negative binomial with

$$r = 1. f_X(x) = \begin{cases} (1-p)^x p & x = 0, 1, \dots \\ 0 & o.w. \end{cases}$$

**Poisson:** Counts occurrences of an event.  $X$  is a Poisson r.v. with parameter  $\lambda$  (intensity) if the p.f. is

$$f_X(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & x = 0, 1, 2, \dots \\ 0 & o.w. \end{cases} \text{ with } \lambda > 0.$$

**Continuous Random Variables:** A r.v.  $X$  has a continuous distribution if there is a non-negative  $f$  such that  $Pr(a \leq X \leq b) = \int_a^b f(x) dx$ .  $f$  is the probability density function p.d.f.

**Cumulative Distribution Function:** (c.d.f.) For any r.v.  $X$  the c.d.f. is given by  $F(x) = Pr(X \leq x)$ . Properties:

$$\forall x : 0 \leq F(x) \leq 1$$

$F(x)$  is non-decreasing, i.e. if

$$x_1 < x_2 \Rightarrow \{X \leq x_1\} \subset \{X \leq x_2\} \text{ and so}$$

$$Pr(X \leq x_1) \leq Pr(X \leq x_2) \Rightarrow F(x_1) \leq F(x_2)$$

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ and } \lim_{x \rightarrow \infty} F(x) = 1$$

For a continuous r.v.:

$$F(x) = Pr(X \leq x) = \int_{-\infty}^x f(t) dt$$

$$F'(x) = f(x)$$

$$Pr(a < X \leq b) = Pr(a \leq X \leq b) = Pr(a \leq X < b) = Pr(a < X < b)$$

In general,  $X \sim Unif[a, b] \Rightarrow f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & o.w. \end{cases}$ . The

$$\text{c.d.f.: } F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$

**Quantile Function:**  $X$  continuous r.v.  $F^{-1}(p)$  is the quantile function of  $X$  for  $0 \leq p \leq 1$ .  $F^{-1}(p) = x \Rightarrow p = F(x)$ .

**Joint Continuous Distributions:** Joint p.d.f. given by  $f_{X,Y}(x, y) = Pr((X, Y) \in A) = \iint_A f(x, y) dx dy$ .

**Other Stuff:**

$$\sum_{i=1}^n i = \frac{n(n+1)}{2}, \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

$$\sum_{i=0}^n c^i = \frac{c^{n+1} - 1}{c - 1}, c \neq 1; \sum_{i=0}^{\infty} c^i = \frac{1}{1 - c}$$