Definitions

Sample space (S): set of all possible outcomes

- Aka *sure event*

Sample point: outcome in sample space, $p \in S$
Event: subset of sample space, $E \subseteq S$

- No elements: *null event*, $\emptyset$

Event operations & relationships

* Applicable to n events
Union: $A \cup B = \{x : x \in A \vee x \in B\}$
Intersection: $A \cap B = \{x : x \in A \wedge x \in B\}$
Complement: $A' = \{x : x \in S \wedge x \notin A\}$
Mutually exclusive: $A \cap B = \emptyset$
Contained: $A \subset B$
Equivalent: $A \subset B \wedge B \subset A \Rightarrow A = B$
Others:

- $A \cap A' = \emptyset$
- $A \cap \emptyset = \emptyset$
- $A \cup A' = S$
- $(A')' = A$
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- $A \cup B = A \cup (B \cap A')$
- $A = (A \cap B) \cup (A \cap B')$
- $(A_1 \cup A_2 \cup \ldots \cup A_n)' = A_1' \cap A_2' \cap \ldots \cap A_n'$
- $(A_1 \cap A_2 \cap \ldots \cap A_n)' = A_1' \cup A_2' \cup \ldots \cup A_n'$

Counting methods

Multiplication principle: r different experiments performed sequentially, producing $n_1 \times n_2 \times \ldots \times n_r$ outcomes
Addition principle: r different procedures performed sequentially, producing $n_1 + n_2 + \ldots + n_r$ ways (non-overlapping) to perform an experiment
Permutation: selection and arrangement of r objects out of n where order matters (i.e. $\{a, b\} \neq \{b, a\}$)

$$P_r^n = nPr = \frac{n!}{(n-r)!} = n(n-1)(n-2)\ldots(n-(r-1))$$

Combination: selection of r objects out of n where order does not matter

$$C_r^n = nCr = \binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!(n-r)!}$$

Probability

$P(\cdot)$ is a function on the collection fo events of the sample space S satisfying:

- Axiom 1. For any event A, $0 \leq P(A) \leq 1$
- Axiom 2. For the sample space, $P(S) = 1$
- Axiom 3. For any two mutually exclusive events A and B, $A \cap B = \emptyset$ and $P(A \cup B) = P(A) + P(B)$

Properties:

- $P(\emptyset) = 0$
- If $A_1, A_2, \ldots, A_n$ are mutually exclusive events, then $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$
- $P(A') = 1 - P(A)$
- $P(A) = P(A \cap B) + P(A \cap B')$
- **Inclusion-exclusion principle:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $A \subset B \Rightarrow P(A) \leq P(B)$

Finite sample space with equally likely outcomes: $S = \{a_1, a_2, \ldots, a_k\}$ and all outcomes are equally likely, so any event occurring is where $A \subset S$

$$P(A) = \frac{|A|}{|S|}$$

Probability of repeated event: if the outcome is always the same, then $P(K) = P(A)^n$

Conditional probability

For any two events A and B with $P(A) > 0$, the conditional probability of B given that A has occurred is

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

Since A has occurred, A becomes the reduced sample space
Multiplication rule: $P(A \cap B) = P(A)P(B|A)$ if $P(A) \neq 0$
Inverse probability formula: $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$

Independence

Events are independent ($A \perp B$) iff $P(A \cap B) = P(A)P(B)$

$$P(A) \neq 0 \Rightarrow P(A|B) = P(A)$$

$$P(B) \neq 0 \Rightarrow P(B|A) = P(B)$$

Intuition: A and B if the knowledge of A does not change the probability of B
Independence vs Mutually Exclusive:

- $P(A) > 0 \wedge P(B) > 0, A \perp B \Rightarrow$ not mutually exclusive
- $P(A) > 0 \wedge P(B) > 0, A, B$ not mutually exclusive $\Rightarrow A \not\perp B$
- S and $\emptyset$ are independent of any other event
- $A \perp B \Rightarrow A \perp B', A' \perp B, A' \perp B'$

Total probability

Partition: if $A_1, A_2, \ldots, A_n$ are mutually exclusive events and $\bigcup_{i=1}^n A_i = S$, then $A_1, A_2, \ldots, A_n$ is a partition of S (i.e. how to split the sample space up into parts)
Law of total probability: given a partition of S, for any even B,

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

Applied to single event A and B:

$$P(B) = P(A)P(B|A) + P(A')P(B|A')$$

Bayes' theorem

Give a partition of S, then for any event B and $k = 1, 2, \ldots, n$,

$$P(A_k|B) = \frac{P(A_k)P(B|A_k)}{\sum_{i=1}^n P(A_i)P(B|A_k)}$$

Applied to single event A and B:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A')P(B|A')}$$

Random variables

A function X which assigns a real number to every $s \in S$ (mapping of values from sample space to some value representing a property of that value in the sample space)

$$X : S \mapsto \mathbb{R}$$

Range space: set of real numbers produced by random variable X

$$R_X = \{x | x = X(s), s \in S\}$$

Notations:

- $\{X = x\} = \{s \in S : X(s) = x\} \subset S$
- $\{X \in A\} = \{s \in S : X(s) \in A\} \subset S$
- $P(X = x) = P(\{s \in S : X(s) = x\})$
- $P(X \in A) = P(\{s \in S : X(s) \in A\})$

Describing random variables: (1) range of inputs to outputs, (2) constructing a table/formula

Probability distribution

Probability distribution: $(x_i, f(x_i))$ where $f(x)$ is the probability function

Discrete random variables

Number of values in R_X is finite or countable
Probability mass function: for a discrete random variable X, the probability (mass) function is:

$$f(x) = \begin{cases} P(X = x), x \in R_X \\ 0, x \notin R_X \end{cases}$$

Properties: $f(x)$ must satisfy the following

- $f(x_i) \geq 0 \forall x_i \in R_X$ (all fractional and ≤ 1)
- $f(x) = 0 \forall x \notin R_X$
- $\sum_{i=1}^{\infty} f(x_i) = \sum_{x_i \in R_X} f(x_i) = 1$

Extension: for any set $B \subset \mathbb{R}$,

$$P(X \in B) = \sum_{x_i \in B \cap R_X} f(x_i)$$

Continuous random variables

R_X is an interval or collection of intervals
Probability density function: quantifies probability that X is in a certain range
Properties: $f(x)$ must satisfy the following

- $f(x) \geq 0 \forall x \in R_X$
- $f(x) = 0 \forall x \notin R_X$
- $\int_{-\infty}^{\infty} f(x)dx = \int_{R_X} f(x)dx = 1$

Extension 1: given that $a \leq b$,

$$\begin{aligned} &P(a \leq X \leq b) \\ &= P(a \leq X < b) \\ &= P(a < X \leq b) \\ &= P(a < X < b) \\ &= \int_a^b f(x)dx \end{aligned}$$

Extension 2:

$$P(X = x) = 0$$

Cumulative distribution function

Probability distribution over a range (both discrete and continuous)

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

Properties:

- $F(x)$ is always non-decreasing
- Ranges of $F(x)$ and $f(x)$ satisfy
  - $0 \leq F(x) \leq 1$
  - for discrete distributions, $0 \leq f(x) \leq 1$
  - for continuous distributions, $f(x) \geq 0$ but *not necessary* that $f(x) \leq 1$

Discrete random variables

$$\begin{aligned} F(x) &= \sum_{t \in R_X; t \leq x} f(t) \\ &= \sum_{t \in R_X; t \leq x} P(X = t) \end{aligned}$$

CDF is a step function and can be represented as such (note that probability is cumulated to reach 1):

$$F(x) = \begin{cases} 0, x < 0 \\ 1/4, 0 \leq x < 1 \\ 3/4, 1 \leq x < 2 \\ 1, 2 \leq x \end{cases}$$

For any two numbers $a < b$,

$$P(a \leq X \leq b) = P(X \leq b) - P(X < a) = F(b) - F(a-)$$

$F(a-)$ is the largest value in R_X that is smaller than a

Continuous random variable

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

$$f(x) = \frac{dF(x)}{dx}$$

For any two numbers $a < b$,

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a)$$

Note that if there are multiple functions per interval and a, b run across multiple intervals, separately integrate each interval with the functions for each interval

Expectation

Expectation, also known as mean, of random variable is the average value of X after repeating the experiment many times. This value may not be a possible value of X.
Discrete random variable:

$$\mu_X = E(X) = \sum_{x_i \in R_X} x_i f(x_i)$$

Continuous random variable:

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_{x \in R_X} x f(x)dx$$

Properties:

- $E(aX + b) = aE(X) + b$
- $E(X + Y) = E(X) + E(Y)$
- let $g(\cdot)$ be an arbitrary function,

$$E(g(X)) = \sum_{x \in R_X} g(x)f(x)$$

or

$$E(g(X)) = \int_{R_X} g(x)f(x)dx$$

Variance

Calculates the deviation of X from its mean (expectation)

$$\sigma_X^2 = V(X) = E(X - \mu_X)^2 = E(X^2) - E(X)^2$$

Applicable regardless of discrete/continuous random variable.

$$V(X) = \sum_{x \in R_X} (x - \mu_X)^2 f(x)$$

or

$$V(X) = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx$$

Properties:

- $V(X) \geq 0$ if $P(X = E(X)) = 1$ where X is a constant
- $V(aX + b) = a^2V(X)$
- standard deviation of X: $\sigma_X = \sqrt{V(X)}$

Joint distributions

(X, Y) is a two-dimensional random vector/random variable
Range space: $R_{X,Y} = \{(x, y) | x = X(s), y = Y(s), s \in S\}$ (effectively looking at all pairs of (x, y); generalizable to n dimensions)
Discrete two-dimensional random variable: number of possible values of $(X(s), Y(s))$ is finite or countable (both X and Y are discrete)
Continuous two-dimensional random variable: number of possible values of $(X(s), Y(s))$ can assume any value in some region of the Euclidean space $\mathbb{R}^2$ (both X and Y are continuous)

Joint probability function

Discrete joint probability function

$$f_{X,Y}(x, y) = P(X = x, Y = y)$$

Properties:

- $f_{X,Y}(x, y) \geq 0 \forall (x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0 \forall (x, y) \notin R_{X,Y}$
- $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$
- above is the same as $\sum \sum_{(x,y) \in R_{X,Y}} f(x, y) = 1$

Let $A \subset R_{X,Y}$,

$$P((X, Y) \in A) = \sum \sum_{(x,y) \in A} f_{X,Y}(x, y)$$

Continuous joint probability function

$$\begin{aligned} &P((X, Y) \in D) \\ &= P(a \leq X \leq b, c \leq Y \leq d) \\ &= \int_a^b \int_c^d f_{X,Y}(x, y)dydx \end{aligned}$$

*The order of integration does not matter
Properties:

- $f_{X,Y}(x, y) \geq 0 \forall (x, y) \in R_{X,Y}$
- $f_{X,Y}(x, y) = 0 \forall (x, y) \notin R_{X,Y}$
- $\int \int_{(x,y) \in R_{X,Y}} f_{X,Y}(x, y)dxdy = 1$

*Focus of this module is ranges where x and y do not depend on each other (not the same as independence)

Marginal probability distribution

Isolating X or Y from a joint probability distribution (projection of joint distribution to univariate distribution). To find X, use Y, and vice versa

$$P(X=x)=f_X(x)=\sum_y f_{X,Y}(x,y)$$

or

$$f_X(x)=\int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

Marginal probability distributions are probability functions.

Conditional distribution

Distribution of Y given that the random variable X is observed to take the value x
Conditional probability function of Y given that $X=x$:

$$f_{Y|X}(y|x)=\frac{f_{X,Y}(x,y)}{f_X(x)}$$

when given values, it finds $P(Y|X=x)$
Only defined for x such that $f_X(x)>0$ (same for y)
$f_{Y|X}(y|x)$ **is not a probability function of** x: the requirements of probability functions do not need to hold
Applications: you can also use summation for discrete joint random variables

$$P(Y\le y|X=x)=\int_{-\infty}^{y} f_{Y|X}(y|x)dy$$

$$E(Y|X=x)=\int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy$$

Reading discrete joint probability tables

x/y	y_1	y_2	y_3	$f_X(x)$
x_1	a	b	c	a + b + c
x_2	d	e	f	d + e + f
x_3	g	h	i	g + h + i
$f_Y(y)$	a + d + g	b + e + h	c + f + i	1

$E(Y|X=x)$ (same steps for $E(X|Y=y)$) (using $x=x_1$):

1. Sum of probability given $X=x_1=f_X(x_1)$, $a+b+c=K$
2. Divide each value in $X=x_1$ by K, a/K, b/K, c/K
3. Multiply each by the corresponding Y value, $\frac{ay_1}{K}$, $\frac{by_2}{K}$, $\frac{cy_3}{K}$
4. Sum the values: $E(Y|X=x_1)=\dfrac{ay_1+by_2+cy_3}{K}$

$E(X)$ (same steps for $E(Y)$): $x_1\cdot(a+b+c)+x_2\cdot(d+e+f)+x_3\cdot(g+h+i)$
Simplified $E(X)$ (same steps for $E(Y)$): $x_1\cdot f_X(x_1)+x_2\cdot f_X(x_2)+x_3\cdot f_X(x_3)$

Independent random variables

X does not decide Y and vice versa
X and Y are independent iff for any x and y (all pairs),

$$f_{X,Y}(x,y)=f_X(x)f_Y(y)$$

*Must manually check all combinations
Product feature: necessary condition for independence: $R_{X,Y}$ needs to be a product space

$$\begin{aligned}&f_{X,Y}(x,y)=f_X(x)f_Y(y)>0\\&\Rightarrow R_{X,Y}=\{(x,y)|x\in R_X;y\in R_Y\}=R_X\times R_Y\end{aligned}$$

If $R_{X,Y}$ is not a product space, then X and Y are not independent (visually, it's a rectangular space)
Properties:

- if $A\subset\mathbb{R}\wedge B\subset\mathbb{R}$, the events $X\in A$ and $Y\in B$ are independent events in S

$$P(X\in A;Y\in B)=P(X\in A)P(Y\in B)$$

$$P(X\le x;Y\le y)=P(X\le x)P(Y\le y)$$

- for arbitrary functions $g_1(\cdot)$ and $g_2(\cdot)$, $g_1(X)\perp g_2(Y)$
- $f_X(x)>0\Rightarrow f_{Y|X}(y|x)=f_Y(y)$
- $f_Y(y)>0\Rightarrow f_{X|Y}(x|y)=f_X(x)$

Checking independence

Given a joint probability table (for discrete variables), if there are 0 entries in the table, then $R_{X,Y}$ is not a product space, hence $X\not\perp Y$
More generally, both conditions must hold:

1. $R_{X,Y}$ is positive and is a product space
2. for any $(x,y)\in R_{X,Y}$, $f_{X,Y}(x,y)=C\times g_1(x)\times g_2(y)$ (can be decomposed into parts that all do not depend on each other)
   * $g_1(X)$ and $g_2(Y)$ do not need to be probability functions

Joint expectation

$$E(g(X,Y))=\sum_x\sum_y g(x,y)f_{X,Y}(x,y)$$

or

$$E(g(X,Y))=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f_{X,Y}(x,y)dydx$$

Covariance

$$\begin{aligned}\text{cov}(X,Y)&=E[(X-E(X))(Y-E(Y))]\\&=\sum_x\sum_y(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)\\&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_X)(y-\mu_Y)f_{X,Y}(x,y)dxdy\end{aligned}$$

Properties:

- $\text{cov}(X,Y)=E(XY)-E(X)E(Y)$
- if $X\perp Y$, $\text{cov}(X,Y)=0$ (but converse is not true)
- $X\perp Y\Rightarrow E(XY)=E(X)E(Y)$
- $\text{cov}(aX+b,cY+d)=ac\cdot\text{cov}(X,Y)$
  - $\text{cov}(X,Y)=\text{cov}(Y,X)$
  - $\text{cov}(X+b,Y)=cov(X,Y)$
  - $\text{cov}(aX,Y)=a\text{cov}(X,Y)$
- $V(aX+bY)=a^2V(X)+b^2V(Y)+2ab\cdot\text{cov}(X,Y)$
  - $V(aX)=a^2V(X)$
  - $V(X+Y)=V(X)+V(Y)+2\text{cov}(X,Y)$
- $V(a+bX)=b^2V(X)$

Joint variance

- $X\perp Y\Rightarrow V(X\pm Y)=V(X)+V(Y)$
- $V(X_1+X_2+\cdots+X_n)=V(X_1)+V(X_2)+\cdots+V(X_n)+2\sum_{j>i}\text{cov}(X_i,X_j)$
- $X_1\perp X_2\perp\cdots\perp X_n\Rightarrow V(X_1\pm X_2\pm\cdots\pm X_n)=V(X_1)+V(X_2)+\cdots+V(X_n)$