3.1 Expectation and conditional expectation

Given a random variable $X : \Omega \to I$ with $I$ countable we define

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega) = \sum_{i \in I} i P(X = i)$$

and

$$E(f(X)) = \sum_{i \in I} f(i) P(X = i)$$

for a function $f : I \to \mathbb{R}$.

**Theorem 3.1.** Expectation is linear, i.e.,

$$E \left( \sum_{i=1}^{N} X_i \right) = \sum_{i=1}^{N} E(X_i).$$

**Proof.** The left-hand side is

$$E \left( \sum_{i=1}^{N} X_i \right) = \sum_{\omega \in \Omega} \left( \sum_{i=1}^{N} X_i(\omega) \right) P(\omega) = \sum_{i=1}^{N} \sum_{\omega \in \Omega} X_i(\omega) P(\omega) = \sum_{i=1}^{N} E(X_i).$$

Definition 3.2. Two rvs $X$ and $Y$ are independent iff $P(X = i, Y = j) = P(X = i) P(Y = j)$ for all $i, j$.

Note that, $X, Y$ independent implies that $E(XY) = E(X) E(Y)$, but not the converse.

Definition 3.3. Given a rv $X : \Omega \to I$ and an event $B \in \mathcal{F}$ in $\Omega$ with $P(B) > 0$, the conditional expectation of $X$ given $B$ is defined as

$$E(X \mid B) = \sum_{i \in I} i P(X = i \mid B) = \sum_{i \in I} \frac{i P(\{X = i\} \cap B)}{P(B)}.$$

Definition 3.4. Let $X : \Omega \to I$ be a rv and $\mathcal{B} \subseteq \mathcal{F}$ be a $\sigma$-algebra on $\Omega$ generated by a countable partition $\Pi = \{E_i, i \geq 1\}$ of $\Omega$. The conditional expectation of $X$ given $\mathcal{B}$ is a piecewise constant rv defined as

$$E(X \mid \mathcal{B})(\omega) := E(X \mid E_i) \text{ for } \omega \in E_i.$$
Heuristically, this the average of $X$ given the partial information through $B$. Note that, $P(A \mid B)$ is a scalar and so is $E(A \mid B)$. But, both $P(A \mid B)$ and $E(X \mid B)$ are rvs.

**Example 3.5.** 1) The trivial $\sigma$-algebra $B_0 = \{\emptyset, \Omega\}$ has $\Pi = \{\Omega\}$ and thus
\[ E(X \mid B_0) = E(X \mid \Omega) = E(X) \]
as $\omega \in \Omega$ for all $\omega$ and $B$ gives no extra information.

2) The finest $\sigma$-algebra $B_f = 2^\Omega$ has $\Pi = \{E_i = \{\omega\}, \omega \in \Omega\}$ when $\Omega$ is countable and
\[ E(X \mid B_f) = E(X \mid \{\omega\}) = \sum_i iP(X = i \mid \{\omega\}) = X(\omega). \]
Thus, $E(X \mid B_f) = X$ as rvs and $B_f$ has all possible information.

One can check that, if $X$ is measurable w.r.t. $B$, then $E(X \mid B) = X$ and vice versa.

**Theorem 3.6** (Law of total expectation). If $X : \Omega \to I$ is a rv and $B \subseteq F$ is a $\sigma$-algebra on $\Omega$, then
\[ E(E(X \mid B)) = E(X). \]

### 3.2 Filtrations

**Definition 3.7.** Given a probability space $(\Omega, F, P)$, a filtration is a sequence of increasing $\sigma$-algebras on $\Omega$ giving successively more/finer information, i.e., for all $t \geq 0$, $F_t \subseteq F$ is a $\sigma$-algebra on $\Omega$ and $F_s \subseteq F_t$ for all $s \leq t$.

Note that, when $\Omega$ is countable, there is a sequence of increasingly finer partitions corresponding to a filtration.

**Example 3.8.** Toss two fair coins. Thus $\Omega = \{HH, HT, TH, TT\}$. Let
\[ F_0 := \{\emptyset, \Omega\}, \]
\[ F_1 := \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}, \]
\[ F_2 := 2^\Omega. \]
Note that $F_0 \subset F_1 \subset F_2$ is a filtration. Here $F_0$ is the trivial $\sigma$-algebra containing no information, $F_2$ is the finest $\sigma$-algebra containing all information and $F_1$ is the $\sigma$-algebra containing information about the first toss only.

A filtration can be generated by a sequence of rvs $(X_n, n \geq 1)$ from $\Omega$ to $I$ by defining
\[ F_n := \sigma(X_1, X_2, \ldots, X_n) \]
= the smallest $\sigma$-algebra on $\Omega$ such that $(X_1, X_2, \ldots, X_n)$ is jointly measurable.

For the two coin toss example, define $X_1 = 1_{H*} = 1$ if the first toss is $H$ and $0$ if the first toss is $T$; and $X_2 = 1_{*H}$. Then $X_1, X_2$ generate the above $\sigma$-algebras $F_1 \subseteq F_2$.

The partition corresponding to $F_n$ is
\[ \Pi_n = \{E_{(i_1, i_2, \ldots, i_n)}, i_1, i_2, \ldots, i_n \in I\} \]
where
\[ E_{(i_1, i_2, \ldots, i_n)} = \cap_{k=0}^{n} X_k^{-1}(i_k) = \{\omega \mid X_k(\omega) = i_k \text{ for all } k = 1, 2, \ldots, n\}. \]
3.3 Moments of a rv

Definition 3.9. The Moment generating function (mgf) of a rv $X$ is defined as

$$M_X(t) = E(e^{tX}) = \sum_{j \in I} e^{tj} P(X = j).$$

The mgf is multiplicative for $X,Y$ independent, i.e., $M_{X+Y}(t) = M_X(t)M_Y(t)$ for all $t$. The $k$-th moment of $X$ is defined as $E(X^k) = \sum_{j \in I} j^k P(X = j)$. Note that the first moment is the mean.

Theorem 3.10. The mgf generates the moments:

$$E(X^k) = \left. \frac{d^k}{dt^k} M_X(t) \right|_{t=0}.$$

Definition 3.11. We define variance of $X$ as $\text{Var}(X) = E((X - E(X))^2)$.

Definition 3.12. A function $f : \mathbb{R} \to \mathbb{R}$ is convex iff for all $\alpha \in [0,1]$ and $x,y \in \mathbb{R}$ we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

Theorem 3.13 (Jensen’s inequality). For a convex function $f$, we have

$$f(E(X)) \leq E(f(X)).$$

3.4 Basics of linear algebra

A matrix $A$ on $I$ is an array of numbers indexed by $I \times I$, i.e., $A = ((a_{ij}))_{i,j \in I}$. Product of two matrices $A,B$ is defined as the matrix with $ij$-th entry given by $(AB)_{ij} = \sum_{k \in I} a_{ik}b_{kj}$. We define the $k$-th power of $A$ inductively by $A^k = A \cdot A^{k-1}$. A vector $v$ on $I$ is a column vector $v = (v_i)_{i \in I}$. We define left multiplication by a matrix $M$ as $(Mv)_i = \sum_{k \in I} m_{ik}v_j$ and similarly right multiplication by $M$ as $(v^T M)_i = \sum_{k \in I} v_k M_{ki}$.

Definition 3.14. If $v \neq 0$ then $v$ is an eigenvector (right) of $M$ iff $Mv = \mu v$ for some constant $\mu \in \mathbb{C}$. $\mu \in \mathbb{C}$ is called the eigenvalue. Left eigenvalues are equal to right eigenvalue, so they are the same. However, right eigenvector and left eigenvector corresponding to the same eigenvalue need not be same.

Note that, $\mu$ is an eigenvalue iff $M - \mu I$ is not invertible.

Definition 3.15. The characteristic polynomial of $M$ is defined as

$$\mathcal{P}_M(z) := \det(M - zI).$$

Its roots or zeros are the eigenvalues of $M$.

If $|I| = n$ there are exactly $N$ eigenvalues counted with multiplicity.

If $M$ is $N \times N$ and has $N$ distinct eigenvalues, then $M$ is nice in the sense of being diagonalizable, i.e., there exists a diagonal matrix $D$ and an invertible matrix $S$ such that $M = SDS^{-1}$. 
Theorem 3.16. If $M$ has $N$ linearly independent eigenvectors $v_1, v_2, \ldots, v_N$ with eigenvalues $\mu_1, \mu_2, \ldots, \mu_N$ then defining $S := [v_1, v_2, \ldots, v_N]$ and $D := \text{diagonal}(\mu_1, \mu_2, \ldots, \mu_N)$ we have $M = SDS^{-1}$.

Note that, $M^2 = SDS^{-1} \cdot SDS^{-1} = SD^2S^{-1}$ and in general $M^k = SD^kS^{-1}$. Powers of a diagonal matrix are easier to compute.

Definition 3.17. The spectrum of a matrix $M$ is defined as

$$\text{spec}(M) := \{ \mu \mid \mu \text{ is an eigenvalue of } M \}.$$ 

Theorem 3.18 (Gershgorin circle theorem). Let $A$ be an $N \times N$ matrix. For each $i = 1, 2, \ldots, N$ we define

$$R_i := \sum_{j \neq i} |a_{ij}| \text{ and}$$

$$B_i := \overline{B}(a_{ii}, R_i) = \{ z \in \mathbb{C} \mid |z - a_{ii}| \leq R_i \}.$$ 

Then

$$\text{spec}(A) \subseteq \bigcup_{i=1}^{N} B_i.$$ 

Proof. Let $\mu \in \text{spec}(A)$. We want to show that there exists $i$ such that $\mu \in B_i$. Consider an eigenvector $v = (v_i)_{i=1,2,\ldots,N} \neq 0$ associated to $\mu$ so that $Av = \mu v$. Pick the largest magnitude component, i.e., $i := \text{argmax}_{j=1,2,\ldots,N} |v_j|$. Thus $|v_i| > 0$. Wlog we can assume that $v_i > 0$. The $i$-th equation of $Av = \mu v$ is

$$\mu v_i = \sum_{i=1}^{N} a_{ij}v_j = a_{ii}v_i + \sum_{j \neq i} a_{ij}v_j.$$ 

Rearranging we get

$$|\mu - a_{ii}| v_i = \left| \sum_{j \neq i} a_{ij}v_j \right| \leq \sum_{j \neq i} |a_{ij}| |v_j| \leq \sum_{j \neq i} |a_{ij}| v_i = R_i v_i.$$ 

Since $v_i > 0$, we get $\mu \in B_i$. \hfill \blacksquare