

to denote the variance. Thus for instance

$$\mathbf{E}(\mathbf{I}(E)) = \mathbf{P}(E); \quad \mathbf{Var}(\mathbf{I}(E)) = \mathbf{P}(E) - \mathbf{P}(E)^2. \quad (1.1)$$

If F is an event of non-zero probability, we define the conditional probability of another event E with respect to F by:

$$\mathbf{P}(E|F) := \frac{\mathbf{P}(E \wedge F)}{\mathbf{P}(F)}$$

and similarly the conditional expectation of a random variable X by

$$\mathbf{E}(X|F) := \frac{\mathbf{E}(X\mathbf{I}(F))}{\mathbf{E}(\mathbf{I}(F))} = \sum_x x\mathbf{P}(X = x|F).$$

A random variable is *boolean* if it takes values in $\{0, 1\}$, or equivalently if it is an indicator function $\mathbf{I}(E)$ for some event E .

1.1 The first moment method

The simplest instance of the probabilistic method is the *first moment method*, which seeks to control the distribution of a random variable X in terms of its expectation (or first moment) $\mathbf{E}(X)$. Firstly, we make the trivial observation (essentially the pigeonhole principle) that $X \leq \mathbf{E}(X)$ with positive probability, and $X \geq \mathbf{E}(X)$ with positive probability. A more quantitative variant of this is

Theorem 1.1 (Markov's inequality) *Let X be a non-negative random variable. Then for any positive real $\lambda > 0$*

$$\mathbf{P}(X \geq \lambda) \leq \frac{\mathbf{E}(X)}{\lambda}. \quad (1.2)$$

Proof Start with the trivial inequality $X \geq \lambda\mathbf{I}(X \geq \lambda)$ and take expectations of both sides. \square

Informally, this inequality asserts that $X = O(\mathbf{E}(X))$ with high probability; for instance, $X \leq 10\mathbf{E}(X)$ with probability at least 0.9. Note that this is only an *upper tail estimate*; it gives an upper bound for how likely X is to be much larger than $\mathbf{E}(X)$, but does not control how likely X is to be much smaller than $\mathbf{E}(X)$. Indeed, if all one knows is the expectation $\mathbf{E}(X)$, it is easy to see that X could be as small as zero with probability arbitrarily close to 1, so the first moment method cannot give any non-trivial lower tail estimate. Later on we shall introduce more refined methods, such as the second moment method, that give further upper and lower tail estimates.

To apply the first moment method, we of course need to compute the expectations of random variables. A fundamental tool in doing so is *linearity of expectation*, which asserts that

$$\mathbf{E}(c_1X_1 + \cdots + c_nX_n) = c_1\mathbf{E}(X_1) + \cdots + c_n\mathbf{E}(X_n) \quad (1.3)$$

whenever X_1, \dots, X_n are random variables and c_1, \dots, c_n are real numbers. The power of this principle comes from there being no restriction on the independence or dependence between the X_i s. A very typical application of (1.3) is in estimating the size $|B|$ of a subset B of a given set A , where B is generated in some random manner. From the obvious identity

$$|B| = \sum_{a \in A} \mathbf{I}(a \in B)$$

and (1.3), (1.1) we see that

$$\mathbf{E}(|B|) = \sum_{a \in A} \mathbf{P}(a \in B). \quad (1.4)$$

Again, we emphasize that the events $a \in B$ do not need to be independent in order for (1.4) to apply.

A weaker version of the linearity of expectation principle is the *union bound*

$$\mathbf{P}(E_1 \vee \cdots \vee E_n) \leq \mathbf{P}(E_1) + \cdots + \mathbf{P}(E_n) \quad (1.5)$$

for arbitrary events E_1, \dots, E_n (compare this with (1.3) with $X_i := \mathbf{I}(E_i)$ and $c_i := 1$). This trivial bound is still useful, especially in the case when the events E_1, \dots, E_n are rare and not too strongly correlated (see Exercise 1.1.3). A related estimate is as follows.

Lemma 1.2 (Borel–Cantelli lemma) *Let E_1, E_2, \dots be a sequence of events (possibly infinite or dependent), such that $\sum_n \mathbf{P}(E_n) < \infty$. Then for any integer M , we have*

$$\mathbf{P}(\text{Fewer than } M \text{ of the events } E_1, E_2, \dots \text{ hold}) \geq 1 - \frac{\sum_n \mathbf{P}(E_n)}{M}.$$

In particular, with probability 1 at most finitely many of the events E_1, E_2, \dots hold.

Another useful way of phrasing the Borel–Cantelli lemma is that if F_1, F_2, \dots are events such that $\sum_n (1 - \mathbf{P}(F_n)) < \infty$, then, with probability 1, all but finitely many of the events F_n hold.

Proof By monotone convergence it suffices to prove the claim when there are only finitely many events. From (1.3) we have $\mathbf{E}(\sum_n \mathbf{I}(E_n)) = \sum_n \mathbf{P}(E_n)$. If one now applies Markov's inequality with $\lambda = M$, the claim follows. \square