1 Discrete Random Variables and Expectation

A *(real-valued) random variable* X on a sample space Ω is a *measurable function* $X : \Omega \to \mathbb{R}$, and a *discrete random variable* is one which may take on only a countable number of distinct values. "$X = a$" represents the set $\{s \in \Omega \mid X(s) = a\}$, and we denote the probability of that event by $\mathbb{P}(X = a) = \sum_{s\in\Omega:X(s)=a} \mathbb{P}(s)$.

Random variables $X_1, X_2, \cdots, X_n$ are *mutually independent* (simply called *independent* when $k = 2$) if and only if, for any subset $I \subseteq \{1, 2, \cdots, k\}$ and any values x_i $(i \in I)$, $\mathbb{P}(\bigcap_{i\in I}(X_i = x_i)) = \prod_{i\in I} \mathbb{P}(X_i = x_i)$.

The *expectation* of a discrete random variable X, denoted by $\mathbb{E}[X]$, is given by $\mathbb{E}[X] = \sum_i i \cdot \mathbb{P}(X = i)$. Note that the infinite series needs to be *absolutely convergent* (i.e. rearrangements do not change the value of the sum).

Theorem 1.1 (Linearity of expectation) *For discrete random variables $X_1, X_2, \cdots, X_n$ with finite expectations and any contants $c_1, c_2, \cdots, c_n$, we have $\mathbb{E}[\sum_{i=1}^n c_iX_i] = \sum_{i=1}^n c_i\mathbb{E}[X_i]$.*

proof. Observe that we only need to prove the following two cases:

$$\begin{aligned}
\mathbb{E}[X+Y] &= \sum_i \sum_j (i+j) \cdot \mathbb{P}((X=i)\cap(Y=j)) \\
&= \sum_i i \sum_j \mathbb{P}((X=i)\cap(Y=j)) + \sum_j j \sum_i \mathbb{P}((X=i)\cap(Y=j)) = \mathbb{E}[X] + \mathbb{E}[Y], \\
\mathbb{E}[cX] &= \sum_i i \cdot \mathbb{P}(cX = j) = c \cdot \sum_j (j/c) \cdot \mathbb{P}(X = j/c) = c \cdot \sum_k k \cdot \mathbb{P}(X = k) = c \cdot \mathbb{E}[X].
\end{aligned}$$

When there are countably infinite variables, the situation becomes more subtle. We will discuss it later. □

Theorem 1.2 (Jensen's inequality) *If f is a convex function, then $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$.*

proof. Assume that f has a Taylor expansion. Let $\mu = \mathbb{E}[X]$. By Taylor's theorem, there is a value c such that

$$f(x) = f(\mu) + f'(\mu)(x-\mu) + \frac{f''(c)(x-\mu)^2}{2} \geq f(\mu) + f'(\mu)(x-\mu)$$

Taking expectations of both sides

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(\mu) + f'(\mu)(X-\mu)] = \mathbb{E}[f(\mu)] + f'(\mu)(\mathbb{E}[X]-\mu) = f(\mu) = f(\mathbb{E}[X])$$

An alternative proof will be presented in Exercise 2.10. □

Define *conditional expectation* $\mathbb{E}[Y \mid Z = z] = \sum_y y \cdot \mathbb{P}(Y = y \mid Z = z)$ and $\mathbb{E}[Y \mid Z]$ as a random variable $f(Z)$ that takes on the value $\mathbb{E}[Y \mid Z = z]$ when $Z = z$.

Theorem 1.3 (Law of total expectation) *For any random variables X and Y,*

$$\mathbb{E}[X] = \sum_y \mathbb{P}(Y = y) \cdot \mathbb{E}[X \mid Y = y] = \mathbb{E}[\mathbb{E}[X \mid Y]].$$

A *Bernoulli* random variable X takes 1 with probability p and 0 with probability $1-p$. A *binomial* random variable X with parameters n and p, denoted by $B(n,p)$, is defined by *probability distribution* $\mathbb{P}(X = k) = \binom{n}{k} \cdot p^k(1-p)^{n-k}$, $n = 0, 2, \cdots, n$. Its expectation is np.

A *geometric* random variable X with parameter p is defined by probability distribution $\mathbb{P}(X = n) = (1-p)^{n-1}p$, $n = 1, 2, \cdots$. Its expectation is $1/p$. Geometric random variables are *memoryless*, that is, one ignores past failures as distribution does not change. Formally, we have the following statement.

Lemma 1.4 (Memorylessness) *Let X be a geometric random variable with parameter p. Then, for $n > 0$,*

$$\mathbb{P}(X = n + k \mid X > k) = \mathbb{P}(X = n).$$

Lemma 1.5 *Let X be a discrete random variable that takes on only nonnegative integer values. Then,*

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{1\leq i\leq k} \mathbb{P}(X = k) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$