CHAPTER 2

Discrete Random Variables

§ 1. Random Variable and Expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We assume that Ω is discrete and countable (finite or infinite). Note that that $\mathcal{F}$ can be any subset of powerset 2^{Ω}. A random variable X is a "function" $X : \Omega \to \mathbb{R}$ such that $\{w \in \Omega : X(\omega) = r\} \in \mathcal{F}$ for any $r \in \mathbb{R}$. Since Ω is discrete, we can assume, X takes at the most discretely countable values. For most part of this chapter, we take that X takes integer values.

A random variable X is characterised by its distribution i.e., the probabilities $\mathbb{P}(X = x)$ for different values of x. Two random variables are said to be identically distributed if their distributions are same i.e.,

$$\mathbb{P}(X = x) = \mathbb{P}(Y = x)$$

for each $x \in \mathbb{R}$. In such a case we write $X \overset{d}{=} Y$.

Two random variables can be equal as if two X and Y are same as "functions" i.e., $X(\omega) = Y(\omega)$ for each $\omega \in \Omega$. However, the most useful concept, apart from the above distributional equivalence, is the almost sure equivalence. Two random variables X and Y are said to be almost surely equal if

$$\mathbb{P}\{\omega \in \Omega : X(\omega) \neq Y(\omega)\} = 0$$

and we write $X = Y$ a.s.

An important point to note here is that the distributional equivalence does not require that the random variables have to be defined on the same sample space.

The smallest σ-field $\mathcal{F}_X \subseteq \mathcal{F}$ containing the events $\{\omega \in \Omega : X(\omega) = x\}$ for each $x \in \mathbb{R}$ is called the σ-field generated by X (or information space generated by X).

Two random variables X and Y are said to be independent if $\mathcal{F}_X$ and $\mathcal{F}_Y$ are independent i.e., $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for any $A \in \mathcal{F}_X$ and $B \in \mathcal{F}_Y$. In other words,

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

for any $x, y \in \mathbb{R}$.

Independence of random variables extends naturally to a family of random variables. A family of random variables is said to be pairwise independent if any two of the random variables are independent.

Expectation of the random variable X is defined by

$$\mathbb{E}X = \sum_{x \in \mathbb{R}} x\mathbb{P}(X = x)$$

Since X takes at most countably infinite values, the above sum makes "sense". The variance of a random variable is given by

$$Var(X) = \mathbb{E}(X - \mathbb{E}X)^2.$$

(1.1). Theorem. *Let X, Y be two random variables and $c \in \mathbb{R}$. Then $\mathbb{E}(X + Y) = \mathbb{E}X + \mathbb{E}Y$ and $\mathbb{E}cX = c\mathbb{X}$.*

PROOF. Consider

$$\begin{aligned}\mathbb{E}(X+Y) &= \sum_z z\mathbb{P}(X+Y=z) = \sum_z\sum_y z\mathbb{P}(X+Y=z|Y=y)\mathbb{P}(Y=y)\\ &= \sum_z\sum_y z\mathbb{P}(X=z-y)\mathbb{P}(Y=y)\\ &= \sum_z\sum_y (z-y)\mathbb{P}(X=z-y)\mathbb{P}(Y=y) + \sum_z\sum_y y\mathbb{P}(X=z-y)\mathbb{P}(Y=y)\\ &= \sum_y\left[\sum_z (z-y)\mathbb{P}(X=z-y)\right]\mathbb{P}(Y=y) + \sum_y y\left[\sum_z \mathbb{P}(X=z-y)\right]\mathbb{P}(Y=y)\\ &= \sum_y (\mathbb{E}X)\mathbb{P}(Y=y) + \sum_y y\cdot 1\cdot\mathbb{P}(Y=y)\\ &= \mathbb{E}X + \mathbb{E}Y.\end{aligned}$$

Next

$$\begin{aligned}\mathbb{E}cX &= \sum_x x\mathbb{P}(cX=x) = \sum_x c\frac{x}{c}\mathbb{P}(X=\frac{x}{c})\\ &c\mathbb{E}X.\end{aligned}$$

□

An important property of nonnegative integer valued random variable is the following:

(1.2). PROPOSITION. *Let X be nonnegative integer valued random variable. Then*

$$\mathbb{E}(X) = \sum_{n=1}^{\infty}\mathbb{P}(X>n).$$

An important property of convex functions is the following:

(1.3). THEOREM (Jensen's Inequality). *For any convex function f,*

$$f(\mathbb{E}[X]) \leqslant \mathbb{E}f(\mathbb{E}[X]).$$

(1.4). EXAMPLE. Bernoulli random variable: $X \sim Ber(p)$ if

$$X = \begin{cases}1 \text{ with probability } p\\ 0 \text{ with probability } 1-p.\end{cases}$$

Note that $\mathbb{E}X = p$. Find $Var(X)$.

(1.5). EXAMPLE. Binamomial random variable: $X \sim Bin(n,p)$ if

$$X = k \text{ with probability } \binom{n}{k}p^k(1-p)^{n-k}$$

Observe that $X = X_1 + X_2 + \cdots + X_n$ where X_i, $i = 1, 2, \cdots$ are iid (independent and identically distributed with Bernoulli distribution. Find expectation and variance.

An important consequence of the independence of two random variables is the following:

(1.6). THEOREM. *Let X, Y be two independent random variables, then*

$$\mathbb{E}(XY) = \mathbb{E}X\mathbb{E}Y.$$

§ 2. Bernoulli's Law of Large Numbers

Let $X \sim Bin(n,p)$ where $0 < p < 1$. Bernoulli proved the following

$$\lim_{n\to\infty}\mathbb{P}(|\frac{X}{n} - p| > \epsilon) = 0.$$

This is known as Bernoulli's law of large numbers. In the modern language, this is the weak law of large numbers for Bernoulli random variables.

Let $p_X(i) := \mathbb{P}(X = i)$. then

$$\frac{p_X(i)}{p_X(i-1)} = \frac{n-i+1}{i}\frac{p}{q}.$$

Thus $p_X(i) \geqslant p_X(i-1)$ if and only if $(n+1)p \geqslant i$

Let $r \geqslant (n+1)p$ and $k \geqslant 1$, then

$$\frac{p_X(r+k)}{p_X(r+k-1)} \leqslant \frac{n-r}{r+k}\frac{p}{q} \leqslant \frac{n-r}{r}\frac{p}{q}.$$

Set $R = \frac{n-r}{r}\frac{p}{q}$ then $R \leqslant 1$. Now

$$\begin{aligned}
\mathbb{P}(X \geqslant r) &= \sum_{k=0}^{n-r} p_X(r+k) \\
&= \sum_{k=0}^{n-r} p_X(r)\frac{p_X(r+1)}{p_X(r)} \cdots \frac{p_X(r+k)}{p_X(r+k-1)} \\
&\leqslant \sum_{k=0}^{n-r} p_X(r)R^k \\
&\leqslant p_X(r)\frac{1}{1-R}.
\end{aligned}$$

Since $r \geqslant (n+1)p$, $r \geqslant \lfloor (n+1)p \rfloor$ and hence p_X decreases between $[\lfloor (n+1)p \rfloor, r]$. Therefore for $\lfloor (n+1)p \rfloor \leqslant i \leqslant r$, $p_X(r) \leqslant p_X(i)$. Now

$$1 = \sum_{i=0}^{n} p_X(i) \geqslant \sum_{i=\lfloor (n+1)p \rfloor}^{r} p_X(i) \geqslant (r - \lfloor (n+1)p \rfloor + 1)p_X(r).$$

This implies that $p_X(r) \leqslant \frac{1}{r-np}$ and hence $\mathbb{P}(X \geqslant r) \leqslant \frac{rq}{(r-np)^2}$.

For large n, we have $n\epsilon > p$ (by Archemedian property). Take $r = \lceil np + n\epsilon \rceil$. Then

$$\mathbb{P}\big(\frac{X}{n} - p > \epsilon\big) = \mathbb{P}\big(X > np + n\epsilon\big) \leqslant \mathbb{P}\big(X \geqslant \lceil np + n\epsilon \rceil\big) \leqslant \frac{\lceil np + n\epsilon \rceil q}{(\lceil np + n\epsilon \rceil - np)^2} \leqslant \frac{pq}{\epsilon^2 n} + \frac{q}{\epsilon n} + \frac{q}{\epsilon^2 n^2}$$

Hence, $\mathbb{P}\big(\frac{X}{n} - p > \epsilon\big) \to 0$ as $n \to \infty$.

For the other implication, we consider $Y = n - X$. Then $Y \sim Bin(n, q)$ and

$$\mathbb{P}\big(\frac{X}{n} - p < -\epsilon\big) = \mathbb{P}\big(\frac{n-Y}{n} - p < -\epsilon\big) = \mathbb{P}\big(\frac{Y}{n} - q > \epsilon\big)$$

Now applying the above estimate for Y, we obtain

$$\mathbb{P}\big(\frac{X}{n} - p < -\epsilon\big) \leqslant \frac{pq}{\epsilon^2 n} + \frac{p}{\epsilon n} + \frac{p}{\epsilon^2 n^2}.$$

Combining these two estimates, we have

$$\mathbb{P}\big(|\frac{X}{n} - p| > \epsilon\big) \leqslant \frac{2pq}{\epsilon^2 n} + \frac{1}{\epsilon n} + \frac{1}{\epsilon^2 n^2} \to 0 \text{ as } n \to \infty.$$

This proves the Bernoulli's law of large numbers.

§ 3. Moment and Markov Inequality

(3.1). Theorem (Markov's Inequality). *Let X be non-negative random variable. Then*

$$\mathbb{P}(X \geqslant a) \leqslant \frac{\mathbb{E}X}{a}$$

for any $a > 0$.

Proof. For an event A, define the indicator random variable by

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{otherwise.} \end{cases}$$

Let $A = \{X \geqslant a\}$. Then $X = X\mathbf{1}_A + X\mathbf{1}_{A^c}$. Now

$$\begin{aligned}\mathbb{E}X &= \mathbb{E}(X\mathbf{1}_A) + \mathbb{E}(X\mathbf{1}_{A^c}) \\ &\geqslant \mathbb{E}(a\mathbf{1}_A) + \mathbb{E}(X\mathbf{1}_{A^c}) \\ &= a\mathbb{P}(A) + \mathbb{E}(X\mathbf{1}_{A^c}) \\ &\geqslant a\mathbb{P}(A).\end{aligned}$$

The last inequality follows since $\mathbb{E}(X\mathbf{1}_{A^c}) \geqslant 0$ (recall that X is non-negative). And the Markov's inequality follows. □

An important consequence of the Markov's inequality is the Chebyshev's inequality.

(3.2). THEOREM (Chebyshev's Inequality). *Let X be a random variable. Then, for any $a > 0$,*

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant a) \leqslant \frac{Var(X)}{a^2}.$$

PROOF. First note that

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant a) = \mathbb{P}(|X - \mathbb{E}X|^2 \geqslant a^2)$$

Applying Markov's inequality to the non-negative random variable $|X - \mathbb{E}X|^2$, we now hat

$$\mathbb{P}(|X - \mathbb{E}X| \geqslant a) \leqslant \frac{\mathbb{E}|X - \mathbb{E}X|^2}{a^2} = \frac{Var(X)}{a^2}$$

proving the Chebyshev's inequality. □

One important corollary of the Chebyshev's inequality is weak law of large numbers.

(3.3). THEOREM (Weak Law of Large Numbers). *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed (i.i.d.) random variables. Then*

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mathbb{E}X_1\right|\right) \to 0$$

as $n \to \infty$.

PROOF. Since $X_1, X_2, \cdots$ are i.i.d., we have $\mathbb{E}X_i = \mathbb{E}X_1$ and $Var(X_i) = Var(X_1)$ for each $i = 1, 2, \cdots$. Consequently

$$Var(\frac{X_1 + X_2 + \cdots + X_n}{n}) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + \cdots + Var(X_n)) = \frac{Var(X_1)}{n}$$

Now applying Chebyshev's inequality, we have, for every $a > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + X_2 + \cdots + X_n}{n} - \mathbb{E}X_1\right| \geqslant a\right) \leqslant \frac{Var(X_1)}{na^2} \to 0$$

as $n \to \infty$. This proves the weak law. □