3.2 Probability Distributions for Discrete Random Variables

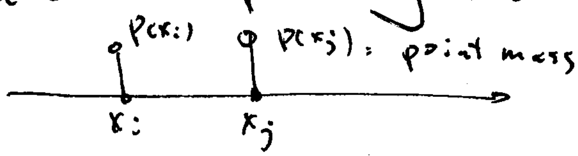
* Definition of probability distribution.

Let X be a discrete r.v. and let $x_1, x_2, \cdots$ be the values which it assumes. Then the function $p(x_j)$ given by

$$p(x_j) := P(X = x_j), \quad j = 1, 2, \cdots$$

is called the probability distribution of X.

* Proposition: The $p(x_j)$'s, which are sometimes called "elementary probabilities" for X, have the following two properties:



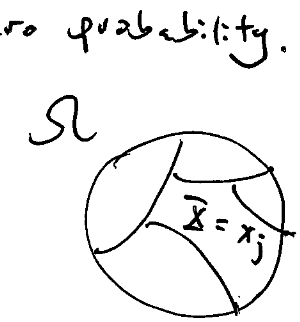
(i) $\forall j, \quad p(x_j) \geq 0$

(ii) $\sum_j p(x_j) = 1$

Rmk: $p(x)$ is also called the probability mass function (pmf).

Proof: (i) It is obvious. Notice that we do not forbid that some x_j may have zero probability.

(ii) We begin with



$$\Omega = \sum_j \{X = x_j\}$$

Hence by countable additivity,

$$1 = P(\Omega) = \sum_j P(X = x_j) = \sum_j p(x_j)$$

- Knowing all the $p(x_j)$'s, we can calculate all probabilities concerning $\underline{X}$.

  — $p(x) = 0$ if $x \notin V_{\underline{X}} = \{x_j\}$

  — $$P(a \le \underline{X} \le b) = \sum_{a \le x_j \le b} p(x_j)$$

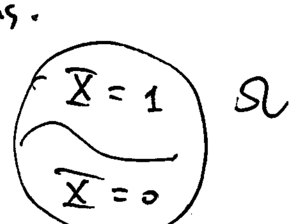
  — $$P(\underline{X} \in A) = \sum_{x_j \in A} p(x_j)$$

  $=$ the sum of $p(x_j)$'s for which the corresponding x_j's belong to A.

<u>Rule</u>: This is a key rule for distributions.

* Example: Bernoulli trial and Bernoulli distributions.

Let $\underline{X}: \Omega \longrightarrow \{0, 1\}$ and



$$p(1) = P(\underline{X} = 1) = p$$

Obviously $\{\underline{X} = 0\}$ and $\{\underline{X} = 1\}$ are complementary, and so by the complement rule

$$p(0) = P(\underline{X} = 0\} = 1 - p = q.$$

where $p + q = 1$, $p \ge 0$, $q \ge 0$.

<u>Bernoulli distribution</u>

$$p(x) = \begin{cases} 1-p, & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

The event $\{\underline{X} = 1\}$ is traditionally known as "Success," and $\{\underline{X} = 0\}$ is known as "failure".

* A Sequence of Bernoulli Trials: by which, we understand a sequence of independent repetitions of an experiment in which the probability of success is the same at each trial. The independence assumption enables us to calculate the probability of any given sequence of successes (S) and failure (F) very easily.

Thus

$$P(SFS) = p\,q\,p = p^2 q$$
$$P(FFFS) = q^3 p$$

* Example 3.10 : Let $p = P(S)$, assume that successive
- Bernoulli Trials

births are independent, and define the r.v.

$\underline{X}$ = number of births observed at a certain hospital until a boy (S) is born.

Then

$$p(1) = P(\underline{X} = 1) = P(S) = p$$
$$p(2) = P(\underline{X} = 2) = P(FS) = p(1-q)$$
$$p(3) = P(\underline{X} = 3) = P(FFS) = (1-q)^2 p$$

In general

$$p(x) = \begin{cases} (1-p)^{x-1} p & x = 1, 2, 3 \ldots \\ 0 & \text{otherwise} \end{cases}$$

* Definition of Distribution Function. Let $\underline{X}$ have distribution $p(x)$. Then the function

$$F(x) = \sum_{y \le x} p(y) = P(\underline{X} \le x)$$

Rmk
$\begin{cases} F(+\infty) = 1 \\ F(-\infty) = 0 \end{cases}$

is called the distribution function of $\underline{X}$.

The value $F(x)$ "picks up" all the probabilities of values of $\underline{X}$ up to x; for this reason the adjective "cumulative" is often added to its name.

* Proposition. (knowing $F(x)$, we can "recover" $p(x)$)

(i) $P(a < \underline{X} \le b) = F(b) - F(a)$

(ii) $P(\underline{X} = x) = \lim_{\varepsilon \to 0} \left[F(x+\varepsilon) - F(x-\varepsilon) \right]$

(iii) $P(a \le X \le b) = F(b) - \lim_{\varepsilon \to 0} F(a+\varepsilon)$

(iv) $P(a < \underline{X} < b) = \lim_{\varepsilon \downarrow 0} F(b-\varepsilon) - F(a)$

Proof: (i)

$$P(a < \underline{X} \le b) = P(\underline{X} \le b) - P(\underline{X} \le a)$$
$$= F(b) - F(a)$$

(ii) - (iv) require more advanced proofs when Ω is countable.

* Proposition. Let X be integer-valued and let F be its distribution function. For any two integers a and b with $a \le b$, we have

(i) $P(a \le X \le b) = F(b) - F(a-1)$

(ii) $P(X = a) = F(a) - F(a-1)$

Proof. (i) $P(a \le X \le b) = P(X = a \text{ or } a+1 \text{ or} \cdots \text{or } b)$

$= F(b) - F(a-1)$

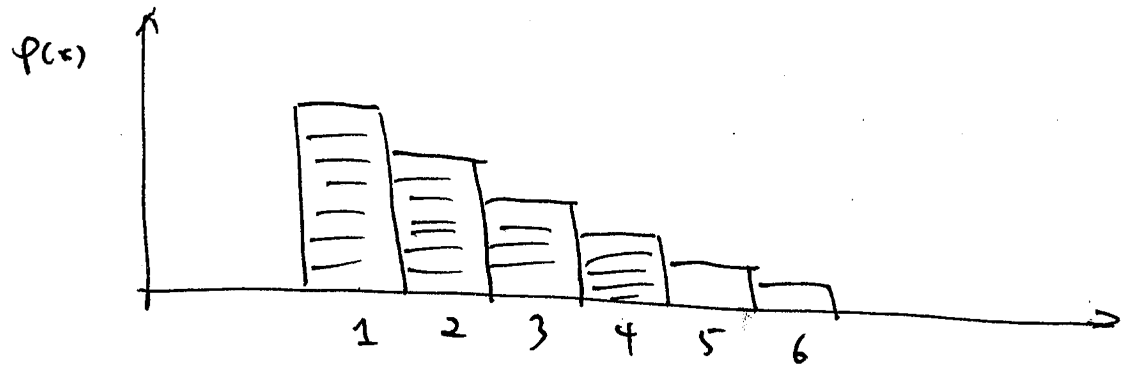
(ii) Taking $a = b$ in (i) yields (ii).

* Probability Histogram. Let X have distribution $p(x)$. The histogram of $p(x)$ can be constructed (Chapter 1):

(1) When possible values of X are equally spaced, the base is chosen as the distance between successive x values.

(2) Above each x with $p(x) > 0$, construct a rectangle centered at x. The bar is the same for all rectangle, the height of each rectangle is proportional to $p(x)$ so that the rectangle has area $p(x)$.

— $F(x)$ is just the area of the histogram to the left of x.



The shaded area is $F(4) = P(X \le 4)$

$= p(1) + p(2) + p(3) + p(4)$

Example 3.12 - Bernoulli Trials

Let $p(x) = \begin{cases} (1-p)^{x-1}\, p, & x = 1, 2, \ldots \\ 0 & \text{otherwise} \end{cases}$

For any positive integer x,

$$F(x) = \sum_{y \le x} p(y) = \sum_{y=1}^{x} (1-p)^{y-1} p = 1 - (1-p)^x$$

Since F is constant between positive integers,

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - (1-p)^{[x]}, & x \ge 1 \end{cases}$$

where $[x]$ is the largest integer $\le x$