Discrete Random Variables
Stat 515H

Let S be a sample space and P a probability on the subsets of S. By definition a **random variable** X is a function mapping S into $\mathbb{R}$. For B a subset of $\mathbb{R}$ we define

$$P(X\in B) = P(\{y\in S: X(y)\in B\}).$$

Let Λ be a subset of $\mathbb{R}$. We say that **X takes values in Λ** if

$$P(X\in\Lambda)=1$$ and Λ is the smallest set w/ this property.

The set Λ is called **discrete** if Λ is finite or countably infinite, which means that there is a one-to-one correspondence mapping $\mathbb{N}$ onto Λ.

We start by considering discrete random variables. By definition, X is a discrete random variable if X takes values in a discrete set Λ.

Examples. (a) Denote the outcomes of a toss of a coin to be $0, 1$. Let $0<p<1$ be given. Define $S=\{0,1\}$ and $P(\{1\})=p$, $P(\{0\})=1-p$. Consider the discrete random variable taking values in $\Lambda=\{0,1\}$ that is defined by $X(0)=0$, $X(1)=1$. Thus $P(X=1)=p$, $P(X=0)=1-p$. X is called a Bernoulli random variable.

(b) Define $S=\{1,2,3,4,5,6\}$ and $P(\{i\})=1/6$ $\forall i\in S$. Consider the discrete random variable taking values in $\Lambda=\{1,2,3,4,5,6\}$ that is defined by $X(i)=i$ for all $i\in S$. Thus $P(X=i)=1/6$ for all $i\in S$. This example corresponds to the toss of a fair die.

(c) Define $S = \{(\alpha,\beta) : 1 \le \alpha \le 6,\ 1 \le \beta \le 6,\ \alpha \text{ and } \beta \text{ integers}\}$ and $P(\{(\alpha,\beta)\}) = 1/36$ for all $(\alpha,\beta) \in S$. Consider the discrete random variable taking values in $\Lambda = \{j \in \mathbb{N} : 2 \le j \le 12\}$ that is defined by $X((\alpha,\beta)) = \alpha + \beta$. For all $j \in \mathbb{N}$

$$P(X = j) = P(\{(\alpha,\beta) \in S : X((\alpha,\beta)) = \alpha+\beta = j\}).$$

We calculated these probabilities in class. This example corresponds to the toss of 2 fair dice.

We return to the general theory. Let X be a discrete random variable taking values in a discrete set Λ.

<u>Definition.</u> For each $x \in \Lambda$ define the <u>probability mass function</u> of X by $p(x) = p_X(x) = P(X = x) = P(\{y \in S : X(y) = x\})$

<u>Theorem 1.</u> (a) For A a subset of Λ

$$P(X \in A) = P(\{y \in S : X(y) \in A\}) = \sum_{x \in A} p(x).$$

(b) For B a subset of $\mathbb{R}$

$$P(X \in B) = P(\{y \in S : X(y) \in B\}) = \sum_{x \in B \cap \Lambda} p(x).$$

<u>Proof.</u> (a) $P(X \in A) = P(\{y \in S : X(y) \in A\}) = P(\{y \in S : X(y) \in \bigcup_{x \in A} \{x\}\})$

$$= P\Big(\bigcup_{x \in A} \{y \in S : X(y) = x\}\Big) = \sum_{x \in A} P(\{y \in S : X(y) = x\})$$

$$= \sum_{x \in A} P(X = x) = \sum_{x \in A} p(x).$$

(b) Since X takes values in Λ, $P(X \in B) = P(X \in B \cap \Lambda)$. We now apply part (a) to $A = B \cap \Lambda$, obtaining

$$P(X \in B) = P(X \in B \cap \Lambda) = \sum_{x \in B \cap \Lambda} p(x). \qquad \text{QED}$$

Example. In Example (b) on page 1,

$$P(X \le 4.1) = P(X \in (-\infty, 4.1]) = P(X \in (-\infty, 4.1] \cap \Lambda)$$
$$= P(X \in \{1,2,3,4\}) = 4/6 = 2/3.$$

We return to the general theory

Theorem 2. The probability mass function of X has the following two properties:

(a) For all $x \in \Lambda$ $p(x) > 0$.

(b) $\sum_{x \in \Lambda} p(x) = 1$.

Proof. (a) This follows from the fact that X takes values in Λ; i.e., $P(X \in \Lambda) = 1$ and Λ is the smallest set with this property. If for some $x \in \Lambda$ we have $p(x) = 0$, then X takes values in $\Lambda \setminus \{x\}$
$P(X \in \Lambda \setminus \{x\}) = P(X \in \Lambda) - P(X = x) = 1 - 0 = 1$.
It would follow that Λ is not the smallest set w/ the property that $P(X \in \Lambda) = 1$.

b) We have from part (a) of Theorem 1 on page 2
$1 = P(X \in \Lambda) = \sum_{x \in \Lambda} p(x)$. QED

Definition. The (cumulative) distribution function of X is defined by $F(z) = F_X(z) = P(X \in (-\infty, z]) = P(X \le z)$ for $z \in \mathbb{R}$, where $(-\infty, z] = \{x \in \mathbb{R} : x \le z\}$

Theorem 2. $F_X(z) = P(X \le z) = \sum_{x \in (-\infty, z] \cap \Lambda} p(x)$

$$= \sum_{\{x \in \Lambda : x \le z\}} p(x).$$

Proof. We apply part (b) of Theorem 1 to $B = (-\infty, z]$. We have

$$F_X(z) = P(X \in (-\infty, z]) = \sum_{x \in (-\infty, z] \cap \Lambda} p(x) = \sum_{\{x \in \Lambda : x \le z\}} p(x). \quad \text{QED}$$

The cumulative distribution function F can be expressed in terms of $p(a)$ by

$$F(a) = \sum_{\text{all } x \le a} p(x)$$

If X is a discrete random variable whose possible values are $x_1, x_2, x_3, \ldots$, where $x_1 < x_2 < x_3 < \cdots$, then the distribution function F of X is a step function. That is, the value of F is constant in the intervals (x_{i-1}, x_i) and then takes a step (or jump) of size $p(x_i)$ at x_i. For instance, if X has a probability mass function given by

$$p(1) = \frac{1}{4} \qquad p(2) = \frac{1}{2} \qquad p(3) = \frac{1}{8} \qquad p(4) = \frac{1}{8}$$

then its cumulative distribution function is

$$F(a) = \begin{cases} 0 & a < 1 \\ \frac{1}{4} & 1 \le a < 2 \\ \frac{3}{4} & 2 \le a < 3 \\ \frac{7}{8} & 3 \le a < 4 \\ 1 & 4 \le a \end{cases}$$

$\Lambda = \{1, 2, 3, 4\}$

$$(-\infty, a] \cap \Lambda = \begin{cases} \emptyset & \text{for } a < 1 \\ \{1\} & \text{for } 1 \le a < 2 \\ \{1, 2\} & \text{for } 2 \le a < 3 \\ \{1, 2, 3\} & \text{for } 3 \le a < 4 \\ \{1, 2, 3, 4\} & \text{for } a \ge 4 \end{cases}$$

This function is depicted graphically in Figure 4.3.

$$F(a) = \begin{cases} 0 & \text{for } a < 1 \\ p(1) & \text{for } 1 \le a < 2 \\ p(1) + p(2) & \text{for } 2 \le a < 3 \\ p(1) + p(2) + p(3) & \text{for } 3 \le a < 4 \\ p(1) + p(2) + p(3) + p(4) & \text{for } a \ge 4 \end{cases}$$

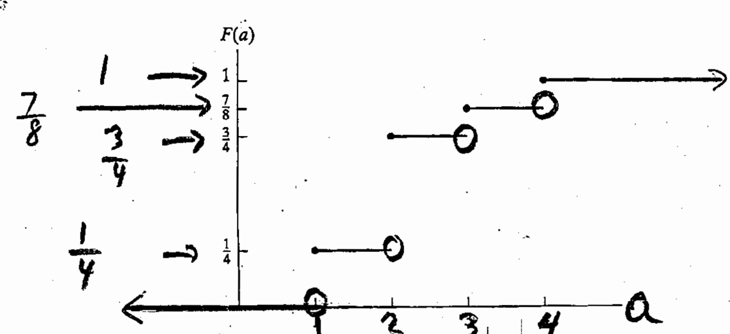

Figure 4.3