STATISTICS
TOPIC III: PROBABILITY

PAUL L. BAILEY

1. Sets and Functions

1.1. **Sets.** A *set* is a collection of objects. The objects in a set are called *elements* of that set. The notation $a \in A$ means "a is an element of the set A". The notation $a \notin A$ means "a is not an element of A". We also use the standard notation whereby $\Rightarrow$ means "implies" and the notation $\Leftrightarrow$ means "if and only if".

Two sets are equal if and only if they contain the same elements;

$$A = B \Leftrightarrow (x \in A \Leftrightarrow x \in B).$$

that is, a set is completely determined by the elements it contains. There is no concept of order or multiplicity when specifying a set:

$$\{1, 1, 2, 2, 3, 4, 4, 5\} = \{1, 2, 3, 4, 5\} \quad \text{and} \quad \{2, 5, 3, 1, 4\} = \{1, 2, 3, 4, 5\}.$$

Set builder notation allows us to specify sets. The notation

$$\{x \mid (\text{some condition on } x)\}$$

denotes the set of all elements x for which the condition is true.

1.2. **Subsets and Set Operations.** We say that B is a *subset* of A, and write $B \subset A$, if every element of B is an element of A:

$$A \subset B \Leftrightarrow (x \in B \Rightarrow x \in A).$$

The *union* of A and B is the set of elements which are in either A or B:

$$A \cup B = \{x \mid x \in A \text{ or } x \in B\}.$$

The *intersection* of A and B is the set of elements which are in both A and B:

$$A \cap B = \{x \mid x \in A \text{ and } x \in B\}.$$

The *complement* of B with respect to A is the set of elements in B but not in A:

$$A \smallsetminus B = \{x \mid x \in A \text{ and } x \notin B\}.$$

A *universal set* is a set U which contains all the elements under consideration in a given context. When a universal set is understood, the *complement* of B is

$$B^{\mathrm{c}} = U \smallsetminus B.$$

The set operations of union, intersection, and complement correspond to the logical operations of OR, AND, and NOT, respectively.

The *empty set*, denoted $\varnothing$, is the set which contains no elements.

We say that A and B are *disjoint* if $A \cap B = \varnothing$.

Date: March 3, 2009.

1.3. **Cartesian Product.** An *ordered pair* consists of two elements in a specified order. The ordered pair containing the elements a and b with a first and b second is denoted (a, b).

Let A and B be sets. The *cartesian product* of A and B is the set

$$A \times B = \{(a, b) \mid a \in A \text{ and } b \in B\}.$$

1.4. **Functions.** A *function* from a set A to a set B is an assignment of each element of A to a unique element of B. The notation $f : A \to B$ mean "f is a function from A to B".

Let $f : A \to B$. If $a \in A$, the unique element of B to which a is assigned is denoted $f(a)$. We call A the *domain* of f, and we call B the *codomain* of f. The function f is completely described by the subset of $A \times B$ given by

$$\{(a, b) \in A \times B \mid b = f(a)\}.$$

If $C \subset A$, the *image of* C *under* f is the subset of the codomain B which consists of all the elements of B to which f assigns some element from C:

$$f(C) = \{b \in B \mid b = f(c) \text{ for some } c \in C\}.$$

The *range* of f is $f(A)$, the image of the domain A.

If $D \subset B$, the *preimage of* D *under* f is the subset of the domain A which consists of all the elements of A which are assigned by f to an element in D:

$$f^{-1}(D) = \{a \in A \mid f(a) \in D\}.$$

1.5. **Collections.** A *collection* is a set whose elements are themselves sets.

The *power set* of a set X is the collection of all subsets of a given set:

$$\mathcal{P}(X) = \{\text{sets } A \mid A \subset X\}.$$

Let X be a nonempty set. A *partition* of X is a collection of nonempty subsets of X, known as *blocks*, such that every element of X is in exactly one block.

2. Probability Spaces

2.1. **Probability Spaces.** A study of probability is on firm ground when it uses the concepts of sets and functions to precisely define its terms. Thus measurement of probabilities takes place in a formal mathematical object known as a probability space.

Definition 1. A *finite probability space* consists of a finite set S together with a function

$$p : S \to [0, 1] \quad \text{satisfying} \quad \sum_{s \in S} p(s) = 1.$$

Let $\mathcal{E}$ denote the set of all subsets of S. Then p determines a function

$$P : \mathcal{E} \to [0, 1] \quad \text{given by} \quad P(E) = \sum_{s \in E} p(s).$$

The elements of S are called *outcomes*. The members of $\mathcal{E}$ are called *events*. The function P is called a *probability measure*. The number $P(E)$ is called *probability* of event E.

Proposition 1. *Let S be a finite probability space. Let $A, B \subset S$. Then*

(a) $P(\varnothing) = 0$;
(b) $P(A^c) = 1 - P(A)$;
(c) $A \subset B \Rightarrow P(A) \leq P(B)$;
(d) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

Corollary 1. Boole's Inequality
Let S be a finite probability space. Let $A, B \subset S$. Then

$$P(A \cup B) \leq P(A) + P(B).$$

Definition 2. Let S be a finite set. The *uniform probability space* on S defined by the function

$$p : S \to [0, 1] \quad \text{given by} \quad p(s) = \frac{1}{|S|}.$$

Then the probability measure on the collection $\mathcal{E}$ of all subsets of S is

$$P : \mathcal{E} \to [0, 1] \quad \text{given by} \quad P(E) = \frac{|E|}{|S|}.$$

Example 1. Let $S = \{H, T\}$. This corresponds to flipping a coin.

Example 2. Let $S = \{1, 2, 3, 4, 5, 6\}$. This corresponds to rolling a fair die.

Example 3. Let $S = R \times U$, where $R = \{2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K, A\}$ is the set of ranks and $U = \{S, H, D, C\}$ is the set of suits. This corresponds to drawing one card from a deck of 52 cards.

2.2. **Disjointness.** We consider the conditions under which the probability of either of two events occurring is the sum of the probabilities of the events.

Definition 3. Let S be a finite probability space. Let $A, B \subset S$.
We say that A and B are *disjoint* (or *mutually exclusive*) if

$$A \cap B = \varnothing.$$

Proposition 2. *Let S be a finite probability space. Let $A, B \subset S$ be disjoint. Then*

$$P(A \cup B) = P(A) + P(B).$$

Proposition 3. *Let S be a finite probability space. Let $\{A_1, \dots, A_n\}$ be a partition of S. Let $E \subset S$. Then*

$$P(E) = \sum_{i=1}^{n} P(E \cap A_i).$$

Suppose the $E = A \cup B \cup C$. Then

$$P(E) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Also,

$$P(E) = P(A) + P(B \cap A^c) + P(C \cap A^c \cap B^c).$$

2.3. **Independence.** We consider the conditions under which the probability of both of two events occurring is the product of the probabilities of the events.

Definition 4. Let S be a finite probability space. Let $A, B \subset S$.

We say that A and B are *independent* if

$$P(A \cap B) = P(A)P(B).$$

Example 4. Let S be a set of 52 cards. Let A be the set of spades and let B be the set of aces. Then

$$P(A \cap B) = \frac{1}{52} = \frac{1}{4}\frac{1}{13} = P(A)P(B).$$

Thus A and B are independent events.

2.4. **Conditioning.** We consider the computation of the probability of an event occurring given that some other event occurred.

Definition 5. Let S be a finite probability space. Let $A, B \subset S$ with $P(B) > 0$. Define the *conditional probability* of A with respect to B to be

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Glorious Conditioning Theorem 1. Multiplication Rule
Let S be a finite probability space and let $A, B \subset S$ with $P(B) > 0$. Then

$$P(A \cap B) = P(A|B)P(B).$$

Glorious Conditioning Theorem 2. Total Probabilities Rule
Let S be a finite probability space and let $A \subset S$. Let $\{B_1, \ldots, B_n\}$ be a partition of S, with $P(B_i) > 0$ for $i = 1, \ldots, n$. Then

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

Glorious Conditioning Theorem 3. Bayes Rule
Let S be a finite probability space and let $A \subset S$. Let $\{B_1, \ldots, B_n\}$ be a partition of S, with $P(B_i) > 0$ for $i = 1, \ldots, n$. Then

$$\begin{aligned} P(B_j|A) &= \frac{A \cap B_j}{A} \\ &= \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)}. \end{aligned}$$

Definition 6. Let S be a finite probability space and let n be a positive integer. Let T be the set of all sequences of length n in S. Define

$$p_T : T \to [0,1] \quad \text{by} \quad p(s_1, \dots, s_n) = \prod_{i=1}^{n} p(s_i).$$

Let $\mathcal{E}_T$ be the collection of subsets of T, which may be viewed as the collection of sequences of events from S. The probability measure on T is

$$P_T : \mathcal{E}_T \to [0,1] \quad \text{given by } P_T(E_1, \dots, E_n) = \prod_{i=1}^{n} p(E_i).$$

We call T the *sequential probability space* of length n over S.

3. Random Variables

3.1. **Random Variables.** The mathematical concept that links probability and statistics is that of a random variable.

Definition 7. Let $(S, \mathcal{E}, P)$ be a probability space. A function $X : S \to \mathbb{R}$ is called a *random variable* if $X^{-1}((-\infty, a]) \in \mathcal{E}$ for all $a \in \mathbb{R}$.

Proposition 4. *Let $(S, \mathcal{E}, P)$ be a probability space and let $X : S \to \mathbb{R}$ be a random variable.*

(a) *if $I \subset \mathbb{R}$ is an interval, then $X^{-1}(I) \in \mathcal{E}$;*
(b) *if $x \in X$, then $X^{-1}(x) \in \mathcal{E}$.*

Proof. Let $I = (a, \infty)$; then $X^{-1}(I) = (X^{-1}((-\infty, a]))^c \in \mathcal{E}$.

Let $I = (a, b]$. Then $X^{-1}(I) = X^{-1}((-\infty, b]) \cap X^{-1}((a, \infty)) \in \mathcal{E}$.

Now $\{b\} = \cap_{n=1}^{\infty}(b - \frac{1}{n}, b]$, so $X^{-1}(b) = \cap_{n=1}^{\infty} X^{-1}((b - \frac{1}{n}, b]) \in \mathcal{E}$. The other forms of intervals are now easily obtained. □

Definition 8. Let $X : S \to \mathbb{R}$ be a random variable.

We say that X is *discrete* if $X(S)$ is countable.

Definition 9. Let $X : S \to \mathbb{R}$ be a discrete random variable. The *density* of X is a function

$$f_X : \mathbb{R} \to [0,1] \text{ given by } f_X(x) = P(X^{-1}(x)).$$

Proposition 5. Dirty Trick Theorem
Let $X : S \to \mathbb{R}$ be a discrete random variable. Then

$$\sum_{x \in \mathrm{img}(X)} f_X(x) = 1.$$

3.2. **Expectation.** Expectation measures the most likely value of a random variable, not in the sense of the mode average, but rather in the sense of the mean average.

Definition 10. Let $X : S \to \mathbb{R}$ be a discrete random variable. The *expectation* of X is a real number

$$E(X) = \sum_{x \in \mathrm{img}(X)} x f_X(x).$$

Proposition 6. *Let $(S, \mathcal{P}(S), P)$ be a uniform probability space, and let $X : S \to \mathbb{R}$ be a random variable. Then*

$$E(X) = \frac{1}{|S|} \sum_{s \in S} X(s).$$

Proof. We have

$$\begin{aligned}
E(X) &= \sum_{x \in \mathrm{img}(X)} x f_X(x) \\
&= \sum_{x \in \mathrm{img}(X)} x \frac{|X^{-1}(x)|}{|S|} \\
&= \frac{1}{|S|} \sum_{x \in \mathrm{img}(X)} x \sum_{s \in X^{-1}(x)} 1 \\
&= \frac{1}{|S|} \sum_{x \in \mathrm{img}(X)} \sum_{s \in X^{-1}(x)} x \\
&= \frac{1}{|S|} \sum_{x \in \mathrm{img}(X)} \sum_{s \in X^{-1}(x)} X(s) \\
&= \frac{1}{|S|} \sum_{s \in S} X(s)
\end{aligned}$$

□

That is, the expectation of a random variable on a finite uniform probability space is the average value of the random variable.

Definition 11. Let S be a finite probability space with $N = |S|$ and let $X : S \to \mathbb{R}$ be a random variable.

The *mean* of X is

$$\mu(X) = \sum_{s \in S} x P(X = x).$$

That is, the mean of a random variable on a uniform probability space equals its expectation.

The *standard deviation* of X is

$$\sigma(X) = \sqrt{\sum_{s \in S} (x - \mu(X))^2 P(X = x)}.$$

3.3. **Distributions.** Distributions describe how the values of a random variable are scattered across the real line.

We now describe the seven great discrete distributions:

(1) Uniform Distribution
(2) Binomial Distribution
(3) Poisson Distribution
(4) Geometric Distribution
(5) Hypergeometric Distribution
(6) Wilcoxon Distribution
(7) Survey Distribution

Great Discrete Distribution 1. Uniform Distribution

Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{|S|} = \frac{|E|}{N}$.

Let $X : S \to \{1, \ldots, N\}$ be a bijective function. Then X is a discrete random variable. We say that X has a *uniform distribution*.

The image of X is $\{1, \ldots, N\}$.

The density of X is

$$f_X(x) = \begin{cases} \frac{1}{N} & \text{if} \quad x = \operatorname{img}(X); \\ 0 & \text{otherwise.} \end{cases}$$

The expectation of X is

$$E(X) = \frac{N+1}{2}.$$

Great Discrete Distribution 2. Binomial Distribution
Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{|S|} = \frac{|E|}{N}$.

Let $R \subset S$ with $|R| = r$ and let $p = P(R) = \frac{r}{N}$.

Define a discrete random variable $Y : S \to \mathbb{R}$ by

$$Y(s) = \begin{cases} 1 & \text{if } s \in R; \\ 0 & \text{if } s \notin R. \end{cases}$$

We say that Y is the *bernoulli* random variable associated to the event R.

The density of Y is

$$f_Y(x) = \begin{cases} p & \text{if } x = 1; \\ 1-p & \text{if } x = 0; \\ 0 & \text{otherwise.} \end{cases}$$

Let n be a positive integer. Let $T = \times_{i=1}^n S$, the cartesian product of S with itself n times. Then $|T| = N^n$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|Q|}{|T|} = \frac{|F|}{N^n}$.

Define a discrete random variable $X : T \to \mathbb{R}$ by

$$X(s_1, \dots, s_n) = \sum_{i=1}^n Y(s_i).$$

We say that X has a *binomial distribution*.

The image of X is

$$\operatorname{img}(X) = \{0, 1, 2, \dots, n\}.$$

The density of X is

$$f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The expectation of X is

$$E(X) = np.$$

Great Discrete Distribution 3. Poisson Distribution
Let T be an infinite probability space and let $X : T \to \mathbb{R}$ be a random variable whose density function satisfying the following.

The image of X is

$$\operatorname{img}(X) = \{0, 1, 2, 3, \dots\}.$$

The density of X is

$$f_X(x) = \begin{cases} e^{-\lambda} \frac{\lambda^x}{x!} & \text{for } x \in \operatorname{img}(X); 0 \quad \text{otherwise.} \end{cases}$$

We say that X has a *Poisson distribution*.

The expectation of X is

$$E(X) = \lambda.$$

Great Discrete Distribution 4. Geometric Distribution
Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{|S|} = \frac{|E|}{N}$.

Let $R \subset S$ with $|R| = r$ and let $p = P(R) = \frac{r}{N}$. Let $Y : S \to \mathbb{R}$ be the bernoulli random variable associated to R, so that

$$Y(s) = \begin{cases} 1 & \text{if } s \in R; \\ 0 & \text{if } s \notin R. \end{cases}$$

Let T be the set of all sequences in S, so that

$$T = \{\sigma : \mathbb{N} \to S\}.$$

We wish to put a probability measure on T; however, T is an uncountable set. Let $\mathcal{E}$ be the sigma algebra generated by the sets

$$E_n(\tau) = \{\sigma \in T \mid \sigma(i) = \tau(i) \text{ for all } i > n\}.$$

Define $Q(E_n(\tau)) = \frac{1}{N^n}$.

Define a discrete random variable $X : T \to \mathbb{R}$ by

$$X(\sigma) = \begin{cases} \min\{i \in \mathbb{N} \mid Y(\sigma(i)) = 1\} & \text{if this set is nonempty;} \\ 0 & \text{otherwise.} \end{cases}$$

We say that X has a *geometric* distribution.

The range of X is

$$\operatorname{img}(X) = \{0, 1, 2, \dots\}.$$

The density of X is

$$f_X(x) = \begin{cases} p(1-p)^{x-1} & \text{if} \quad x \in \{1, 2, \dots\}; \\ 0 & \text{otherwise.} \end{cases}$$

The expectation of X is

$$E(X) = \frac{1}{p}.$$

Okay Discrete Distribution 4. Truncated Geometric Distribution
Let S, R, and Y be as above.

Let T be the cartesian product of S with itself n times. Define a discrete random variable $X : T \to \mathbb{R}$ by

$$X(s_1, \dots, s_n) = \begin{cases} \min\{i \leq n \mid Y(s_i) = 1\} & \text{if this set is nonempty;} \\ 0 & \text{otherwise.} \end{cases}$$

We say that X has a *truncated geometric* distribution.

Great Discrete Distribution 5. Hypergeometric Distribution
Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{N}$.

Let $R \subset S$ with $|R| = r$ and let $p = P(R) = \frac{r}{N}$. Let $Y : S \to \mathbb{R}$ be the bernoulli random variable associated to R, so that

$$Y(s) = \begin{cases} 1 & \text{if } s \in R; \\ 0 & \text{if } s \notin R. \end{cases}$$

The expectation of Y is

$$E(Y) = p.$$

Let n be an integer such that $0 \leq n \leq N$. Set

$$T = \{A \in \mathcal{P}(S) \mid |A| = n\}.$$

Then $|T| = \binom{N}{n}$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|F|}{|T|} = \frac{|F|}{\binom{N}{n}}$.

Define a random variable $X : T \to \mathbb{R}$ by

$$X(A) = \sum_{a \in A} Y(a).$$

Then $X(A) = |A \cap R|$.

The image of X is

$$\operatorname{img}(X) = \{0, 1, \ldots, n\}.$$

The density of X is

$$f_X(x) = \begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & \text{if} \quad x \in \operatorname{img}(X); \\ 0 & \text{otherwise.} \end{cases}$$

The expectation of X is

$$E(X) = \frac{nr}{N} = np.$$

Obtain this as follows. For $a \in S$, the number of sets in T containing a is $\binom{N-1}{n-1}$. Thus

$$\begin{aligned} E(X) &= \frac{1}{|T|} \sum_{A \in T} X(A) \\ &= \frac{1}{|T|} \sum_{A \in T} \sum_{a \in A} Y(a) \\ &= \frac{1}{|T|} \sum_{a \in R} |\{A \in T \mid a \in A\}| \\ &= \frac{1}{|T|} \sum_{a \in R} \binom{N-1}{n-1} \\ &= \frac{\binom{N-1}{n-1} r}{\binom{N}{n}} \\ &= \frac{nr}{N}. \end{aligned}$$

Great Discrete Distribution 6. Wilcoxon Distribution

Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{N}$.

Let $Y : S \to \{1, 2, \ldots, N\}$ be a bijective random variable.

The expectation of Y is

$$E(Y) = \frac{1}{N} \sum_{i=1}^{N} i = \frac{1}{N} \cdot \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

Let n be an integer such that $0 \leq n \leq N$. Set

$$T = \{A \in \mathcal{P}(S) \mid |A| = n\}.$$

Then $|T| = \binom{N}{n}$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|F|}{\binom{N}{n}}$.

Define a random variable $X : T \to \mathbb{R}$ by

$$X(A) = \sum_{a \in A} Y(a).$$

We say that X has a *Wilcoxon distribution*.

The image of X is

$$\text{img}(X) = \{\frac{n(n+1)}{2}, \frac{n(n+1)}{2} + 1, \ldots, \frac{N(N+1)}{2} - \frac{(N-n)(N-n+1)}{2}\}.$$

The density of X is difficult to describe.

The expectation of X is

$$E(X) = \frac{n(N+1)}{2}.$$

Great Discrete Distribution 7. Sample Survey Distribution
Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{N}$.

Let $Y : S \to \mathbb{R}$ be a discrete random variable.

Let n be an integer such that $0 \leq n \leq N$. Set

$$T = \{A \in \mathcal{P}(S) \mid |A| = n\}.$$

Then $|T| = \binom{N}{n}$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|F|}{\binom{N}{n}}$.

Define a random variable $X : T \to \mathbb{R}$ by

$$X(A) = \sum_{a \in A} Y(a).$$

We say that X has a *sample survey* distribution.

The image of X is determined by the image of Y.

The density of X is difficult to describe.

The expectation of X is

$$E(X) = nE(Y).$$

Obtain this as follows.

$$\begin{aligned}
E(X) &= \frac{1}{|T|} \sum_{A \in T} X(A) \\
&= \frac{1}{|T|} \sum_{A \in T} \sum_{a \in A} Y(a) \\
&= \frac{1}{|T|} \sum_{a \in S} |\{A \in T \mid a \in A\}| \cdot Y(a) \\
&= \frac{1}{|T|} \sum_{a \in S} \binom{N-1}{n-1} Y(a) \\
&= \frac{\binom{N-1}{n-1}}{\binom{N}{n}} \sum_{a \in S} Y(a) \\
&= \frac{n}{N} \sum_{a \in S} Y(a) \\
&= nE(Y).
\end{aligned}$$

4. Random Vectors

Definition 12. Let $(S, \mathcal{E}, P)$ be a probability space. A function $\vec{X} : S \to \mathbb{R}^n$ is called a *random vector* if $\vec{X}^{-1}((-\infty, a]^n) \in \mathcal{E}$ for every $a \in \mathbb{R}$.

Proposition 7. *Let $\vec{X} : S \to \mathbb{R}^n$ be a random variable.*

(a) *If $B \subset \mathbb{R}$ is an box, then $X^{-1}(B) \in \mathcal{E}$.*
(b) *If $\vec{x} \in \mathbb{R}^n$, then $\vec{X}^{-1}(x) \in \mathcal{E}$.*

Remark 1. Let $\{A_1, \ldots, A_n\}$ be a collection of sets and let $A = \times_{i=1}^n$ be their cartesian product. Define a function $\pi_i : A \to A_i$ by $\pi_i(a_1, \ldots, a_n) = a_i$. This function is called *projection on the i^{th} component.*

Let $f : B \to A$ be a function. Define a function $f_i : B \to A_i$ by $f_i = \pi_i \circ f$. This function is called the i^{th} *component function* of f. We see that $f(b) = (f_1(b), \ldots, f_n(b))$.

Let $\vec{a} = (a_1, \ldots, a_n) \in A$. Then $f^{-1}(\vec{a}) = \cap_{i=1}^n f_i^{-1}(a_i)$.

Let $A = A_1 \times A_2$. Let $f : B \to A$. Let $\vec{a} = (a_1, a_2)$. Then

(a) $f^{-1}(\vec{a}) = f_1^{-1}(a_1) \cap f_2^{-1}(a_2)$;
(b) $f_1^{-1}(a_1) = \cup_{a_2 \in \mathrm{img}(f_2)} f_2^{-1}(a_2)$.

Proposition 8. *Let $\vec{X} : S \to \mathbb{R}^n$ and let $X_i : S \to \mathbb{R}$ be the i^{th} component function of $\vec{X}$. Then X_i is a random variable.*

Definition 13. Let $\vec{X} : S \to \mathbb{R}^n$ be a random vector.

We say that $\vec{X}$ is *discrete* if $\vec{X}(S)$ is countable.

Definition 14. Let $\vec{X} : S \to \mathbb{R}^n$ be a discrete random vector. The *joint density* of $\vec{X}$ is a function

$$f_{\vec{X}} : \mathbb{R} \to [0, 1] \text{ given by } f_{\vec{X}}(\vec{x}) = P(X^{-1}(\vec{x})).$$

Proposition 9. Dirty Trick Theorem Revisited
Let $\vec{X} : S \to \mathbb{R}^n$ be a discrete random vector. Then

$$\sum_{\vec{x} \in \mathrm{img}(\vec{X})} f_{\vec{X}}(\vec{x}) = 1.$$

Let $[X = x]$ denote the preimage of x under the random variable X.

Proposition 10. *Let $\vec{X} : S \to \mathbb{R}^n$ be a discrete random vector. Let $x \in \mathrm{img}(\vec{X})$. Then $f_{\vec{X}}(x) = P(\cap_{i=1}^n [X_i = x_i])$.*

Proposition 11. *Let $\vec{X} : S \to \mathbb{R}^2$ be a discrete random vector. Let $X, Y : S \to \mathbb{R}$ be the components of $\vec{X}$. Then*

$$f_{X_1}(x) = \sum_{y \in \mathrm{img}(Y)} f_{\vec{X}}(x, y).$$

Multinomial Distribution

Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{|S|} = \frac{|E|}{N}$.

Let $R_1, \ldots, R_n$ be disjoint events.

Let $R_0 = S \smallsetminus \cup_{i=1}^n R_i$, so that $\{R_0, R_1, \ldots, R_n\}$ form a partition of S.

Let $Y_0, Y_1, \ldots, Y_n : S \to \mathbb{R}$ be the corresponding Bernoulli random variables.

Let $p_i = P(R_i)$.

Let n be a positive integer. Let $T = \times_{i=1}^n S$, the cartesian product of S with itself n times. Then $|T| = N^n$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|Q|}{|T|} = \frac{|F|}{N^n}$.

Define discrete random vectors $X_i : T \to \mathbb{R}$ by

$$X(s_1, \ldots, s_n) = \sum_{i=1}^n Y(s_i).$$

Define a discrete random vector $\vec{X} : T \to \mathbb{R}^n$ by $\vec{X} = (X_1, \ldots, X_n)$.

Multivariate Hypergeometric Distribution

Let S be a finite set of cardinality N, and form the uniform probability space $(S, \mathcal{P}(S), P)$, where $P : \mathcal{P}(S) \to [0,1]$ is given by $P(E) = \frac{|E|}{N}$.

Let $R_1, \dots, R_n$ be disjoint events.

Let $R_0 = S \smallsetminus \cup_{i=1}^n R_i$, so that $\{R_0, R_1, \dots, R_n\}$ form a partition of S.

Let $Y_0, Y_1, \dots, Y_n : S \to \mathbb{R}$ be the corresponding Bernoulli random variables.

Let $p_i = P(R_i)$.

Let n be an integer such that $0 \leq n \leq N$. Set

$$T = \{A \in \mathcal{P}(S) \mid |A| = n\}.$$

Then $|T| = \binom{N}{n}$. Form the uniform probability space $(T, \mathcal{P}(T), Q)$, where for $F \subset T$ we have $Q(F) = \frac{|F|}{|T|} = \frac{|F|}{\binom{N}{n}}$.

Define random variables $X_i : T \to \mathbb{R}$ by

$$X_i(A) = \sum_{a \in A} Y_i(a).$$

Then $X_i(A) = |A \cap R|$.

The image of X is

$$\operatorname{img}(X) = \{0, 1, \dots, n\}.$$

The density of X is

$$f_X(x) = \begin{cases} \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}} & \text{if} \quad x \in \operatorname{img}(X); \\ 0 \quad \text{otherwise.} \end{cases}$$

The expectation of X is

$$E(X) = \frac{nr}{N} = np.$$

Obtain this as follows. For $a \in S$, the number of sets in T containing a is $\binom{N-1}{n-1}$. Thus

$$\begin{aligned} E(X) &= \frac{1}{|T|} \sum_{A \in T} X(A) \\ &= \frac{1}{|T|} \sum_{A \in T} \sum_{a \in A} Y(a) \\ &= \frac{1}{|T|} \sum_{a \in R} |\{A \in T \mid a \in A\}| \\ &= \frac{1}{|T|} \sum_{a \in R} \binom{N-1}{n-1} \\ &= \frac{\binom{N-1}{n-1} r}{\binom{N}{n}} \\ &= \frac{nr}{N}. \end{aligned}$$

Example 5. An urn contains 2 red balls, three white balls, and four blue balls. One selects four balls at random from the urn without replacement. Let X_1 denote the number of red balls in the sample, let X_2 denote the number of white balls in the sample, and let X_3 denote the number of blue balls in the sample. Let $\vec{X} = (X_1, X_2, X_3)$.

(a) Find the range of (X, Y, Z).
(b) Find the value of the joint density of (X, Y, Z) at each point in the range.
(c) Find the joint marginal density of (X, Y), (X, Z), and (Y, Z).
(d) Find the three univariate marginal densities.
(e) Find the density of $X + Z$.
(f) Find the expectations of X, Y, Z, $2X + 3Y$.

Solution. Let S be the set of balls in the urn, together with the uniform probability structure.

The range is

$$\{(0,0,3),(0,1,2),(0,2,1),(0,3,0),(1,0,2),(1,1,1),(1,2,0),(2,0,1),(2,1,0)\}.$$

□

Department of Mathematics, University of California, Irvine
E-mail address: pbailey@math.uci.edu