An Example

If a committee of size four is selected at random from 5 Dems., 5 Inds., and 5 Reps, what is the probability that there an equal number of Dems. and Reps.? Let

$$X = \#\text{Dems},$$
$$Y = \#\text{Reps},$$

Then

$$P[X=0, Y=0] = \binom{5}{4} \Big/ \binom{15}{4},$$
$$P[X=1, Y=1] = \binom{5}{1}^2 \binom{5}{2} \Big/ \binom{15}{4},$$
$$P[X=2, Y=2] = \binom{5}{2}^2 \Big/ \binom{15}{4},$$

and

$$P[X=Y] = \frac{5+250+100}{1365} = .260.$$

Jointly Distributed Random Variables

Defs: Given a model, (Ω, P), random variables

$$X, Y, Z, \cdots : \Omega \to I\!R$$

are said to be *jointly distributed.*

Notation: Then, for example,

$$P[X \in A, Y \in B] = P(\omega : X(\omega) \in A \text{ and } Y(\omega) \in B\})$$

and

$$P[(X,Y) \in C] = P(\{\omega : [X(\omega), Y(\omega)] \in C\})$$

for $A, B \subseteq I\!R$ and $C \subseteq I\!R^2$.

The Joint Probability Mass Function

Two RVs: If X and Y are JD discrete RVs, then their *joint probability mass function* is

$$f(x,y) = P[X=x, Y=y]$$

for $x, y \in I\!R$.

Example: *Committees*

$$f(x,y) = \frac{\binom{5}{x}\binom{5}{y}\binom{5}{5-x-y}}{\binom{15}{4}}$$

for integers $x, y \geq 0$ for which $x + y \leq 4$, and $f(x,y) = 0$ otherwise.

Several Random Variables

If $X_1, \cdots, X_m$ are JD discrete RVs, then their *joint probability mass function* is

$$f(x_1, \cdots, x_m) = P[X_1 = x_1, \cdots, X_m = x_m]$$

for $x_1, \cdots, x_m \in I\!R$.

Vector Notation: Let $\mathbf{X} = (X_1, \cdots, X_m)$ and

$$f(\mathbf{x}) = P[\mathbf{X} = \mathbf{x}]$$

for $\mathbf{x} = (x_1, \cdots, x_m) \in I\!R^m$.

Multivariate Hypergeometric Distributions

A box contains N_i tickets of color $i = 1, \cdots, m$; let $N = N_1 + \cdots + N_m$. If n tickets are drawn at random *w.o.r.*, then

$$X_1 = \#\text{tickets of color } 1,$$
$$\cdots,$$
$$X_m = \#\text{tickets of color } m,$$

are JD discrete RVs with joint PMF

$$f(x_1, \cdots, x_m) = \binom{N_1}{x_1} \cdots \binom{N_m}{x_m} / \binom{N}{n}$$

for integers $x_1, \cdots x_m \geq 0$ with $x_1 + \cdots + x_m = n$ and $f(x_1, \cdots, x_m) = 0$ otherwise.

Called *Hypergeometric with parameters* $N_1, \cdots,$ N_m *and* n.

Example: Committees. $N_1 = N_2 = N_3 = 5$, and $n = 4$.

Partitions

Review

If $n \geq 1$ and $n_1, \cdots, n_m \geq 0$ are integers for which

$$n_1 + \cdots + n_m = n,$$

then a set of n elements may be partitioned into m subsets of sizes $n_1, \cdots, n_m$ in

$$\binom{n}{n_1, \cdots, n_m} = \frac{n!}{n_1! \times \cdots \times n_m!}$$

ways.

Example: MISSISSIPPI

$$\binom{11}{4, 1, 2, 4} = \frac{11!}{1! \times 2! \times 4!^2} = 34650.$$

Example

If a balanced (6-sided) die is rolled 12 times, then the probability that each face appears twice is

$$\frac{12!}{2!^6} (\frac{1}{6})^{12} = .0034.$$

For an outcome is

$$\omega = (i_1, \cdots, i_{12}),$$

where $1 \leq i_1, \cdots, i_{12} \leq 6$; there are

$$\#\Omega = 6^{12}$$

such outcomes on

$$\binom{12}{2, 2, 2, 2, 2, 2} = \frac{12!}{2!^6}$$

of which each face appears twice.

Multinomial Distributions

A Loaded Die: Now consider an m-sided, loaded die. Let

$$p_i = Prob[i \text{ spots}]$$

on a single role. So,

$$p_1, \cdots, p_m \geq 0,$$
$$p_1 + \cdots + p_m = 1.$$

Repeated Trials: Suppose that the die is rolled n times, and let

$$X_i = \#\text{roles with } i \text{ spots}$$

for $i = 1, \cdots, m$. Then

$$f(x_1, \cdots, x_m) = P[X_1 = x_1, \cdots, X_m = x_m]$$

is

$$\binom{n}{x_1, \cdots, x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}$$

for integers

$$x_1, \cdots, x_m \geq 0 \qquad (*)$$

with

$$x_1 + \cdots + x_m = n. \qquad (*)$$

For the probability of any sequence with x_i i's is

$$p_1^{x_1} \times \cdots \times p_m^{x_m};$$

and there are

$$\binom{n}{x_1, \cdots, x_m}$$

such sequences.

Thus

$$f(x_1, \cdots, x_m) = \binom{n}{x_1, \cdots, x_m} p_1^{x_1} \times \cdots \times p_m^{x_m}$$

if (*) holds and $f(x_1, \cdots, x_m) = 0$ otherwise.

Def: Called *multinomial with parameters* n, m, *and* $p_1, \cdots, p_m$.

Properties of Multivariate PMFs

If f is the joint PMF of $X_1, \cdots, X_m$, then there is a finite or countably infinite

$$\mathcal{X} \subseteq I\!\!R^m$$

for which

$$f(\mathbf{x}) \geq 0, \text{ for all } \mathbf{x}, \qquad (1)$$

$$f(\mathbf{x}) = 0 \text{ if } \mathbf{x} \notin \mathcal{X}, \qquad (2)$$

and

$$\sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}) = 1. \qquad (3)$$

Also,

$$P[\mathbf{X} \in B] = \sum_{\mathbf{x} \in B \cap \mathcal{X}} f(\mathbf{x})$$

for $B \subseteq I\!\!R^m$.

Conversely, any f that satisfies (1), (2), and (3), is the joint PMF of some random variables $X_1, \cdots, X_m$.

Marginal Distributions

Two Variables

Let X and Y by JD discrete RVs with joint PMF

$$f(x, y) = P[X = x,\ Y = y]$$

and ranges $\mathcal{X}$ and $\mathcal{Y}$. So, $f(x, y) = 0$ unless $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Then X and Y have individual (marginal) PMFs

$$f_X(x) = \sum_{y \in \mathcal{Y}} f(x, y),$$

$$f_Y(y) = \sum_{x \in \mathcal{X}} f(x, y).$$

For

$$\{X = x\} = \bigcup_{y \in \mathcal{Y}} \{X = x,\ Y = y\}$$

and, therefore,

$$P[X = x] = \sum_{y \in \mathcal{Y}} P[X = x,\ Y = y].$$

Example

Two tickets are drawn *w.o.r.* from a box with

1 ticket labelled one,
2 tickets labelled two,
3 ticket labelled three,

Let

$$X = \text{label on first ticket},$$

$$Y = \text{label on second.}$$

Then

Table of $f(x, y)$

x,y	1	2	3	$f_X(x)$
1	0	$\frac{2}{30}$	$\frac{3}{30}$	$\frac{1}{6}$
2	$\frac{2}{30}$	$\frac{2}{30}$	$\frac{6}{30}$	$\frac{2}{6}$
3	$\frac{3}{30}$	$\frac{6}{30}$	$\frac{6}{30}$	$\frac{3}{6}$
$f_Y(y)$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{3}{6}$	

Marginal Distributions

Several Variables

Let

$$X = (X_1, \cdots, X_j)$$

and

$$Y = (Y_1, \cdots, Y_k)$$

be JD discrete RVs with joint PMF

$$f(\mathbf{x}, \mathbf{y}) = P[\mathbf{X} = \mathbf{x},\ \mathbf{Y} = \mathbf{y}]$$

and ranges $\mathcal{X} = \mathbf{X}(\Omega)$ and $\mathcal{Y} = \mathbf{Y}(\Omega)$. Then $\mathbf{X}$ and $\mathbf{Y}$ have individual (marginal) joint PMFs

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{\mathbf{y} \in \mathcal{Y}} f(\mathbf{x}, \mathbf{y})$$

and

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}, \mathbf{y}).$$

Example

Multinomial Distributions

If

$$(X_1, \cdots, X_m) \sim \text{Multinomial}(n, m, p_1, \cdots, p_m),$$

and $1 \le k < n$, then the distribution of

$$(X_1, \cdots, X_k, X_{k+1} + \cdots + X_m)$$

is

$$\text{Multinomial}(n, k+1, p_1, \cdots, p_k, p_{k+1}, \cdots, p_m).$$

In particular,

$$X_i \sim \text{Binomial}(n, p_i).$$

Conditional Distributions

The Discrete Case

Let X and Y have joint PMF f. If $f_X(x) > 0$, then the *conditional PMF of* Y *given* X is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}.$$

Thus,

$$\begin{aligned} f_{Y|X}(y|x) &= \frac{P[X = x,\ Y = y]}{P[X = x]} \\ &= P[Y = y | X = x]. \end{aligned}$$

Note: $f_{Y|X}$ is a PMF, since

$$\sum_{y \in \mathcal{Y}} f_{Y|X}(y|x) = \frac{1}{f_X(x)} \sum_{y \in \mathcal{Y}} f(x,y) = 1.$$

Note: Can reverse the roles of X and Y.

Example

Box Ticket Models

If

$$f(x,y) = \binom{R}{x}\binom{W}{y}\binom{N-R-W}{n-x-y} \Big/ \binom{N}{n},$$

for $x + y \le n$, where $n \le N$, $R, W \ge 1$ and $R + W < N$, then

$$f_X(x) = \binom{R}{x}\binom{N-R}{n-x} \Big/ \binom{N}{n}$$

and

$$f_{Y|X}(y|x) = \binom{W}{y}\binom{N-R-W}{n-x-y} \Big/ \binom{N-R}{n-x}.$$

Note: Intuitive.

Independence

JDRVs X and Y are *independent* if

$$P[X \in A,\ Y \in B] = P[X \in A]P[Y \in B]$$

for all nice subsets $A, B \subseteq \mathbb{R}$ (for example, intervals).

Conditions for Independence

PMFs: If X and Y are discrete, then X and Y are independent iff

$$f(x,y) = f_X(x)f_Y(y) \qquad (*)$$

for all x and y. For if X and Y are independent, then $X = x$ iff $X \in [x, x]$, so that

$$\begin{aligned} f(x,y) &= P[X = x,\ Y = y] \\ &= P[X = x]P[Y = y] \\ &= f_X(x)f_Y(y). \end{aligned}$$

Conversely, if (*) holds, then

$$\begin{aligned} P[X \in A,\ Y \in B] &= \sum_{x \in A \cap \mathcal{X}} \sum_{y \in B \cap \mathcal{Y}} f(x,y) \\ &= \sum_{x \in A \cap \mathcal{X}} \sum_{y \in B \cap \mathcal{Y}} f_X(x)f_Y(y) \\ &= \Big[\sum_{x \in A \cap \mathcal{X}} f_X(x)\Big]\Big[\sum_{y \in B \cap \mathcal{Y}} f_Y(y)\Big] \\ &= P[X \in A]P[Y \in B], \end{aligned}$$

where $\mathcal{X}$ and $\mathcal{Y}$ are the ranges of X and Y.

Example: If E and F are independent events, then $\mathbf{1}_E$ and $\mathbf{1}_F$ are independent random variables. For example,

$$\begin{aligned} P[\mathbf{1}_E = 1,\ \mathbf{1}_F = 1] &= P(E \cap F) \\ &= P(E)P(F) \\ &= P[\mathbf{1}_E = 1]P[\mathbf{1}_F = 1] \end{aligned}$$

Several Variables

Jointly distributed random variables $X_1, \cdots, X_m$ are independent if

$$\begin{aligned} &P[X_1 \in B_1, \cdots, X_m \in B_m] \\ &\qquad = P[X_1 \in B_1] \times \cdots P[X_m \in B_m] \end{aligned}$$

for all $B_i \subseteq \mathbb{R}$. As in the case of two variables, this is equivalent to

$$f(x_1, \cdots, x_m) = f_1(x_i) \times \cdots \times f_m(x_m)$$

for all $x_1, \cdots, X_m$.