Brief Review of Probability with Applications to Dust Bears

This is a brief look at elementary probability using class rank and sex as a example.

Consider the class divided by class rank (sophomores, junior, seniors) and by sex (male and female). Rank and **sex** are independent events.

RANK $P_{So} = 0.1$ $P_J = 0.5$ $P_S = 0.4$ **and** $P_{So} + P_J + P_S = 1.0$

SEX $P_F = 0.6$ $P_M = 0.4$ **and** $P_M + P_F = 1.0$

1. **Rank** and **sex** are independent events. With independent events, the occurrence of one event does not affect the occurrence of the second event. That is, knowledge than an individual is male provides no information about class rank.

   The frequency of any even has a value between 0 and 1. Within **rank** or sex, all frequencies sum to one (1.0).

If simple rules are followed carefully, many calculations can be carried out.

A. Frequency and probability mean the same thing. The frequencies of all possibilities within an independent category (e.g., rank or sex) must sum to *one* (1.0).

B. The probability of having alternate attributes within a single set of possibilities is:

"P(this) or P(that)"

The word or in the equation means that you add the separate possibilities.

Example: What is the probability of being a senior or a sophomore? (where or means add)

ans: $P_{So} + P_S = 0.1 + 0.4 = 0.5$

C. The probability of having a particular attribute from each set when there are more than one set of independent events is:

"P(this) and P(that)"

The word and in the equation means that you multiply the separate probabilities.

Example: What is the probability of being a female senior?

We can restate this question so that we have the *and* or *or* word that indicates multiplication or division:

Restating the question we might say,***"What is the probability of being a female and a senior?***

ans: $P_F \times P_S = 0.6 \times 0.4 = .24$

D. When combining probabilities from two independent sets, **do not double count!** That is, the two attributes must be *mutually exclusive.*

Definition: Things are mutually exclusive if they cannot be both things at one time.

Example. Two events are mutual exclusive:

What is the prob. of being a *senior* or a *junior*? Since *seniors* **cannot** be *juniors* (i.e. mutually exclusive) we can safely add the separate probabilities.

$$P = .4 + .5 = .90$$

*Example. Two events are **not** mutual exclusive:* What is the probability of being a *male* or a *junior*?

Now it is possible to be a *female* and a *junior*, or a *junior* and a *female* so the two attributes are **not** mutually exclusive. If we simply add the probabilities:

$$P_F + P_J = .6 + .5 = 1.1$$

Probabilities can never be greater than 1. There must be some error. This error occurred since we counted *junior females* first as *females* and then counted *junior females a second time as juniors. That is, we double counted junior females. We must subtract the probability of being both a junior and a* female to avoid double counting.

$$P = P_F + P_J -(P_F \text{ x } P_J) = 1.1 - .3 = .80$$

Probability in the dust bear problem.

1. The genotypes of the progeny from each female are independent events. When looking at the probability of obtaining a particular genotype (or phenotype) from among the progeny of more than one female, you add up the separate probabilities.

2. Dust bears always produce 64 males and 64 females. Furthermore sex and genotype are independent events. Thus, the probability of getting a male and a female of the specified genotype are also independent events.

3. In general, you need to obtain at least one male and at least one female of the desired genotype. Within a single cross, you need to know the probability of getting the undesired result. That is of obtaining **zero** males of the correct genotype <u>or</u> **zero** females <u>or</u> both **zero** males and **zero** females. In this case, it is possible to double count.

Binomial Expansion

Often in genetics we need to deal with the probability of a combination of events happening. It quickly becomes unwieldy to do all the calculations and then add up the desired probabilities.

For example, a couple wants to have a family of three children, two girls and a boy. Assuming the probability of having either sex is 50%, how likely is it that they will have their desired family?

There are eight possible families of three children if we look at birth order.

First Child		Second Child		Third Child		Probability this family occurs
B		B		B		1/2 * 1/2 * 1/2 = 1/8
B		B		G		1/8
B		G		B		1/8
B		G		G		1/8 **YES!**
G		B		B		1/8
G		B		G		1/8 **YES!**
G		G		B		1/8 **YES!**
G		G		G		1/8

There are three families that meet the condition of two girls and a boy (without considering birth order). The probability of two girls and a boy is 1/8+ 1/8+ 1/8 = 3/8.

It quickly becomes unwieldy to write out all possible combinations and sum up the ones that meet our criteria. The alternate method is to use the binomial expansion:

$$Pr = \frac{n!}{w!x!} p^w q^x$$

where:
***n* = the total number events (*w* + *x* = *n*)**
***w* = the total number of events of one type**
***x* = the total number of events of the other type**
***p* = probability of event w**
***q* = probability of event x (p + q = 1)**

Example: In rumbunnies (cute animal found only in genetics problem sets and tests) bloodshot eyes (B) are dominant to clear (eyes). In a monohybrid F_2cross, what is the probability of getting 6 bloodshot and 2 clear eyed rumbunnies (i.e., a 3:1 ratio)?

n = 8	**Given**
w = # bloodshot = 6	**Given**
x = # clear eye = 2	**Given**
p = 3/4	**Derived from Mendelian Principles**
q = 1/4	**Derived from Mendelian Principles**

$$Pr = \frac{n!}{w!x!} p^w q^x$$

$$Pr = \frac{8!}{6!\,2!} (3/4)^6 * (1/4)^2$$

$$Pr = \frac{8*7*6*5*4*3*2*1}{6*5*4*3*2*1*2*1} (729/4096)(1/16)$$

$$Pr = 28(729/4096)(1/16)$$

$$Pr = 20412/65536 = .311$$

With eight progeny, the chance of getting the 3:1 Mendelian ratio exactly is about 31%.

Multinomial Expansion

The binomial expansion can be extended to cover any number of alternative events. For three alternative probabilities the equation would be:

$$Pr = \frac{n!}{w!x!y!} p^w q^x r^y$$

For four alternative probabilities the equation would be:

$$Pr = \frac{n!}{w!x!y!z!} p^w q^x r^y s^z$$

Breeding Dust Bears of a Known Homozygous Genotype.

The probability of obtaining any particular combination of progeny phenotypes can be obtained a multinomial expansion. In the first F_2 cross there are four expected phenotypes (Cross 2). In the second F_2 cross there are eight expected phenotypes (Cross 4).

However, in the dust bears example, we are only interested in getting a single rare genotype. All the other possible phenotypes are incorrect. We can treat the all the incorrect phenotypes as a single wrong phenotype. This reduces the complex multinomial expansion to the binomial expansion given below:

$$Pr = \frac{n!}{w!x!} p^w q^x$$

For the dust bears, among the F_2 progeny of the initial cross of **aaBBCCDD** x **AAbbCCDD** there should be four phenotypes in a 9:3:3:1 ratio. Since each dust bear produces 64 males and 64 females we expect the following ***numbers*** of dust bears of each type. (The expected number of progeny of each sex is determined by multiplying the expected frequency by the total number of progeny)

Progeny	**Exp. Freq.**	**Exp. No. females**	**Exp. No. males**
Purple Eye, striped, green pellicle and two terminal claws	**9/16**	**36**	**36**
Blue Eye, striped, green pellicle and two terminal claws	**3/16**	**12**	**12**
Purple Eye, **solid**, green pellicle and two terminal claws	**3/16**	**12**	**12**
Blue Eye, solid, green pellicle and two terminal claws	**1/16**	**4**	**4**

However, it is possible that there are greater or less than 4 males of the desired genotype produced from this cross. Since it is essential to obtain *at least* one male dust bear with the phenotype: **Blue**

Eye, solid, green pellicle and two terminal claws, it is important to know the probability of obtaining **zero** males of the correct genotype. This can be done with the binomial equation.
We can assign the 5 variables that go into the binomial expansion based on information given in the problem and our knowledge of Mendelian genetics.

***n* = the total number events (*w* + *x* = *n*)**
***w* = the total number of events of one type**
***x* = the total number of events of the other type**
***p* = probability of event w**
***q* = probability of event x**

n	the total number of events is **64**. This is given in the problem.
w	is the total number of male dust bears with the phenotype, **Blue Eye, solid,** green pellicle and two terminal claws We have already decided that the number of those dust bears should be **zero (0).**
x	is the total number of male dust bears that do not have the phenotype **Blue Eye, solid,** green pellicle and two terminal claws among the F_2 progeny (cross 2). Since there are 64 total males and none of them have the desired phenotype, this value is **64.** *(remember n = w + w , therefore 64 = 0 + 64)*
p	is the probability of obtaining a male dust bear with the phenotype **Blue Eye, solid,** green pellicle and two terminal claws among the F_2 progeny (cross 2). **p** is **1/16**
q	is the probability of obtaining a male dust bear that does not have the phenotype **Blue Eye, solid,** green pellicle and two terminal claws among the F_2 progeny (cross 2). **q** is **15/16**.

CALCULATIONS

$$Pr = \frac{n!}{w!x!} p^w q^x$$

$$Pr = \frac{64!}{0!64!} (1/16)^0 (15/16)^{64}$$

Pr = .016 or 1.6%

We could also estimate the probability of obtaining 1, 2 , 3, etc. males in the F_2 cross using the binomial expansion. For one male, w = 1 and x - 63, for two males w = 2 and x = 63, etc).

Probability of obtaining males with the desired phenotype in a single F_2 cross:	
Number	**Probability**
0	1.6%
1	6.9%
2	14.4%
3	19.8%
4	20.2%
5	16.1%

The chance is very good that one or more males will be obtained in a single F_2 cross. However, many geneticists would rather spend money than time so they might set up 2 or more F_2 crosses rather than risk a 1.6% failure rate. For to crosses the change of getting zero males in the first cross and zero males in the second cross is .016 x .016 or about .026%. With two or three crosses the risk of getting no males becomes quite small.

Probability of getting zero males (or females) of the desired phenotype	
Number of F_2 crosses	Probability
1	1.6%
2	.026%
3	.0004%

The number of F_2 crosses (Cross 4) needed to obtain *at least one male* with the final target phenotype of **Blue Eye, solid,** green pellicle and **four** terminal claws is more important. The critical values for the binomial expansion are:

n	the total number of events is **64**. This is given in the problem.
w	is the total number of dust bears with the phenotype **Blue Eye, solid,** green pellicle and **four** terminal claws. The number of these dust bears is **zero (0)** **(**because we wish to know how unlikely this is).
x	is the total number of dust bears that do not have the phenotype **Blue Eye, solid,** green pellicle and **four** terminal claws. Since there are 64 total and none of them have the desired phenotype this value is **64.** *(remember n = w + w , therefore 64 = 0 + 64)*
p	The probability of obtaining a dust bear with the phenotype: **Blue Eye, solid,** green pellicle and **four** terminal claws is **1/64**
q	The probability of obtaining a dust bear that does not do not have the phenotype **Blue Eye, solid,** green pellicle and **four** terminal claws is **63/64**.

$$Pr = \frac{n!}{w!x!} p^w q^x$$

$$Pr = \frac{64!}{0!64!} (1/64)^0 (63/64)^{64}$$

Pr = .365 or 36.5%

That is, there is a greater than 1/3 chance of obtaining no males (or females) of the desired phenotype!!

Probability of getting **zero** males (or females) of the desired phenotype in the F_2 generation	
Number of different F_2 crosses	Probability
1	36.5%
2	13.3%
3	4.86%
4	1.77%
5	0.65%
6	0.24
7	0.086%
8	0.032%
9	0.012%
10	0.004%

Practically speaking, the number of crosses to be carried out is a function of the cost in terms of time and money *if one fails to obtain the animals with the correct genotype*. Dust bears are inexpensive to rear, have large number of progeny and have a short generation time. A mistake would be trivial. Mice on the other hand have generation time of 3-4 months, are expensive to rear and have litters of 4 to 10. Mistakes with mice could delay an experiment for more that a year.

It is not enough to get a males of the desired genotype if you don't also get a females of the desired phenotype. Since both sex and genotype are independent events, the probability of not getting both a male and a female is given below.

Probability of getting no males or no females or neither males nor females for the desired genotype from a single F_2 cross (Cross 4)
Prob. zero males **and** Prob. of zero females **but not** both zero males and zero females (double counting).
Pr. = .365 +.365 - (.365 X.365).
Pr. = .598

In this dust bear example, one would have to exceptionally lucky to get both a male and a female of the desired phenotype from a single F_2 cross (Cross 4). If one set up 5 F_2 crosses, there is still greater than a 1% chance that one fails to obtain a male **or** a female **or** both. With 10 crosses the probability that one fails to obtain a male or a female *or* both is .008%.