Mathematical methods in biology
Eva Kisdi, University of Helsinki

Expectations and variances

We have derived the following equations during the lecture. This note simply summarizes the most important definitions and formulas for expectations and variances, given full information about the distribution of the random variables.

Let p_i denote the probability that random variable ξ takes the value x_i, i.e., $P(\xi = x_i) = p_i$. Here we assume that p_i are known (see lectures for sampling). Then we have the basic identity

$$\sum_i p_i = 1 \qquad (1)$$

where the summation is for all possible values of i.

The <u>expectation</u> (or mean) of ξ is given by

$$E(\xi) = \sum_i p_i x_i \qquad (2)$$

(note that $E(\xi)$ is simply a number!). The linear function $a + b\xi$ has expectation

$$E(a + b\xi) = a + bE(\xi) \qquad (3)$$

and the expectation of the sum of two random variables (ξ and η) is the sum of the expectations,

$$E(\xi + \eta) = E(\xi) + E(\eta) \qquad (4)$$

The <u>variance</u> of ξ is defined as

$$V(\xi) = E\left[(\xi - E(\xi))^2\right] \tag{5}$$

and this is equivalent to

$$V(\xi) = E(\xi^2) - [E(\xi)]^2 \tag{6}$$

The standard deviation is the square root of the variance:

$$D(\xi) = \sqrt{V(\xi)} \tag{7}$$

The standard deviation has the same units as ξ or $E(\xi)$, and is therefore "comparable" to $E(\xi)$. In contrast, the variance has unit^2.

The variance of the linear function $a + b\xi$ is

$$V(a + b\xi) = b^2 V(\xi) \tag{8}$$

(note that *a* does not influence the variance and *b* is squared in (8)). The variance of the sum of two random variables (ξ and η) is in general NOT the sum of the variances but

$$V(\xi + \eta) = V(\xi) + V(\eta) + 2COV(\xi, \eta) \tag{9}$$

where COV is the covariance, defined as

$$COV(\xi, \eta) = E\left[(\xi - E(\xi))(\eta - E(\eta))\right] = E(\xi\eta) - E(\xi)E(\eta) \tag{10}$$

If ξ and η are <u>independent</u>, then COV(ξ,η)=0 and the variance of the sum is the sum of variances.

Note that $COV(\xi,\xi) = V(\xi)$, i.e., the covariance of a random variable with itself is the variance. The scaled quantity

$$r = \frac{COV(\xi,\eta)}{D(\xi)D(\eta)} \tag{11}$$

(where D is from (7)), which takes values between -1 and 1, is called the correlation coefficient between ξ and η.

[The last page of this pdf is a collection of only the formulas described above.]

Expectation

$$E(\xi) = \sum_i p_i x_i$$

$$E(a + b\xi) = a + bE(\xi)$$

$$E(\xi + \eta) = E(\xi) + E(\eta)$$

Variance

$$V(\xi) = E\left[(\xi - E(\xi))^2\right] = E(\xi^2) - \left[E(\xi)\right]^2 ,$$

$$D(\xi) = \sqrt{V(\xi)}$$

$$V(a + b\xi) = b^2 V(\xi)$$

$$V(\xi + \eta) = V(\xi) + V(\eta) + 2COV(\xi, \eta)$$

$$(COV(\xi, \eta) = 0 \text{ if independent})$$

$$r = \frac{COV(\xi, \eta)}{D(\xi) D(\eta)}$$