Expectations, Variances and Covariances of one or more variables - week one

Luke Lloyd-Jones

Centre for Neurogenetics and Statistical Genomics, Queensland Brain Institute, University of Queensland, Brisbane, QLD

April 27, 2016

Overview

- Review some key definitions and operations for expectations, variances, and covariances of random variables
- Look at Taylor series and why they are useful
- Initial look at the Taylor series and expectations of functions of random variables
- Derive an unbiased estimator of the sample (independent) variance

1 Random variables, definitions and operations of expectations, variances, and covariances

The work that we will cover rests heavily on the notion of *random variables*. A more rigorous definition of a *random variable* requires the notions of the sample space, collection of events, and a probability measure. The sample space, let's denote it Ω, is the set of possible outcomes of an experiment.

Example. If we toss a coin twice then $\Omega = \{HH, HT, TH, TT\}$

Random variables (RVs) link the concept of sample spaces to data.

Definition. A *random variable* is a mapping $X : \Omega \to \mathbb{R}$ that assigns a real number $X(\omega)$ to each of the realisations in Ω denoted by ω.

Example.

Flip a coin twice and let X be the number of heads. Then,

$$\mathbb{P}(X = 0) = \mathbb{P}(TT) = 1/4$$
$$\mathbb{P}(X = 1) = \mathbb{P}(HT, TH) = 1/2$$
$$\mathbb{P}(X = 2) = \mathbb{P}(HH) = 1/4$$

This is an example of a discrete random variable i.e., X takes countably many values. Assigned to each random variable is a *cumulative distribution function* (CDF) and a *probability density function* (PDF). We will not explore these concepts but they are lurking in the background. **Notation** – we tend to use upper case letters like X, Y, Z to denote random variables and their realisations with x, y, z.

Depending on the processes being modelled/generated by a random variable there appears to be best mappings from sample space to the real line. For example, above is an instance of a *Binomial* random variable, which is one of the set of *discrete random variables*. There exists other random variables that are *continuous*, which possess different properties to discrete RVs. These notions have implications for how we define the *expectation* for a random variable.

Definition. The expectation (or expected value, mean, or first moment) of a random variable X is defined to be

$$\mathbb{E}(X) = \sum_{x} x f(x)$$

in the discrete case and

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx$$

in the continuous case. Sometimes denoted μ_X.

Once we have this definition we can leave the concept of the random variable behind and work with the abstraction that requires us to adhere to a set of rules derived from the constraints of summation and integration set out in the original definitions.

Definition. The variance of a random variable X, denoted by $\text{Var}(X)$ is defined to be

$$\text{Var}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2.$$

The square root of the variance is called the standard deviation. Variance is sometimes denoted σ_X^2.

The following key properties of expectation and variance hold for both discrete and continuous random variables.

(1) $$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b$$

(2) $$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

(3) $$\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$$

Definition. The covariance of two random variables X and Y, denoted by $\mathrm{Cov}(X, Y)$ is defined to be

$$\mathrm{Cov}(X, Y) = \mathbb{E}\{[X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]\},$$

which is a measure of the amount of linear dependence between two variables. Scaling the covariance by the product of the standard deviations of X and Y gives you the **correlation coefficient**.

The following key properties of variance and covariance hold for both discrete and continuous random variables.

$$\text{(4)} \qquad \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\text{(5)} \qquad \mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$$

$$\text{(6)} \qquad \mathrm{Cov}(aX+bY,Z) = a\mathrm{Cov}(X,Z) + b\mathrm{Cov}(Y,Z)$$

$$\text{(7)} \qquad \mathrm{Cov}(X,X) = \mathrm{Var}(X)$$

$$\text{(8)} \qquad \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X,Y)$$

$$\text{(9)} \qquad \text{If } X \text{ and } Y \text{ are independent} \rightarrow \mathrm{Cov}(X,Y) = 0$$

$$\text{(10)} \qquad \text{If sequence } X_n \text{ of RVs are independent} \rightarrow \mathrm{Var}(a + \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} b_i^2 \mathrm{Var}(X_i)$$

2 Taylor series and expectation

Taylor series (Brook Taylor 1685-1731) come out of the study of power series, which are functions of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + \dots.$$

These functions are infinite in nature but their properties make them easy to work with in circumstances that involve differentiation and integration. Often these operations are difficult with functions such as $f(x) = \arctan(x)$ so if we can find a power series representation then we may be able to do some more work with this function. So we ask, given $\arctan(x)$ can we find coefficients such that

$$\arctan(x) = a_0 + a_1 x + a_2 x^2 + \dots$$

for at least some nonzero values of x.

We require an infinitely differentiable function $f(x)$ defined on some interval and suppose that it has a power series expansion, what would be the coefficients? So

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

Now Brook Taylor noticed

$$f(0) = a_0$$
$$f'(0) = a_1$$
$$f''(0) = 2a_2$$
$$f'''(0) = 3 \times 2 \times a_3$$

and so on and so forth. The notation $'$ denotes the derivative of $f(x)$ with respect to x. Rearranging these expressions in terms of the coefficients then we get the following representation

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2}x^2 + \frac{f''(0)}{3!}x^3 + \dots,$$

which is Taylor's formula.

Why is that interesting? Well we now have a way of represent a function in a different manner. We may be able to exploit this for approximations if we ignore higher order terms.

3 Expectations of functions of random variables

Let's move back to statistics and have a look at expectations of complex variables.

We have a random variable X and its expectation $\mathbb{E}(X)$. Now what happens when we have a function of X for example $f(X) = X^2$ or $\log(X)$. Function of random variables are interestingly random variables themselves. Let's try and approximate the expectation by expanding the function around its 'true' expectation denoted μ_X.

$$\begin{aligned}\mathbb{E}[f(X)] &= \mathbb{E}\left[f(\mu_X) + (X-\mu_X)f'(\mu_X) + (X-\mu_X)^2\frac{f''(\mu_X)}{2} + \dots\right] \\ &= f(\mu_X) + \mathbb{E}[(X-\mu_X)]f'(\mu_X) + \frac{\mathbb{E}[(X-\mu_X)^2]}{2}f''(\mu_X) + \dots\end{aligned}$$

The last step follows because the derivative terms, evaluated at μ_X, are really just constants. Let's make a few observations

$$\mathbb{E}[(X-\mu_X)] = \mathbb{E}(X) - \mathbb{E}(\mu_X) = \mu_X - \mu_X = 0$$

and

$$\mathbb{E}[(X-\mu_X)^2]$$

is the definition of the variance. Therefore, ignoring third and higher-order terms,

$$\mathbb{E}[f(X)] \approx f(\mu_X) + \sigma_X^2\frac{f''(\mu_X)}{2}.$$

This is very powerful because we now have an approximation to the expectation of any function of any random variable provided the function meets a few criteria.

Example. If $f(X) = X^2$ then a *second* order Taylor series approximation to the expectation is

$$\mathbb{E}[f(X)] \approx \mu_x^2 + \sigma_X^2.$$

Let's use these concepts to investigate the estimator for the variance. Similar to the example in the Appendix of Lynch et al. [1998] but with different notation.

By definition the variance σ_X^2 is characterised by the expected value of $(X-\mu_x)^2$. Now suppose we have a sequence of random variables $X_1, \dots, X_n$ and we will say that they are independent and identically distributed (i.i.d) with mean μ and variance σ^2 i.e., that each random variable has the same distribution and is independent $\rightarrow \text{Cov}(X_i, X_j) = 0$.

An intuitive estimator for the variance of this sequence of random variables is

$$S^2(X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2,$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$. We are estimating the mean and variance from the data. One property of this estimator that we may like it to have is that on average it gives us back the truth i.e., $\mathbb{E}[S^2(X)] = \sigma^2$. This the classical notion of *unbiasedness* in statistics.

So $S(X_1, \ldots, X_n)$ looks an awful lot like a function of random variables. We may think to ourself that the Taylor series trick may give us an avenue to find the expected value and see whether our estimator is unbiased. Now we have only seen Taylor series for one variable and here we have a function of multiple random variables. There is an analogous version for a function of multiple random variables

$$\mathbb{E}[f(X_1, \ldots, X_n)] \approx f(\mu_1, \ldots, \mu_n) + \frac{1}{2} \sum_{i=1}^{n} \sigma_i^2 \frac{\partial^2 f(\mu_1, \ldots, \mu_n)}{\partial X_i^2} + \sum_{i=1}^{n} \sum_{j>i}^{n} \sigma^2(X_i, X_j) \frac{\partial^2 f(\mu_1, \ldots, \mu_n)}{\partial X_i \partial X_j}.$$

Our sequence has some nice properties that include $\mu_1 = \mu_2 = \ldots = \mu_n = \mu$, $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$, and thankfully $\sigma^2(X_i, X_j) = 0$ for all i and j. We can therefore ignore the final sum in the above expression. So

$$\mathbb{E}[f(X_1, \ldots, X_n)] \approx f(\mu) + \frac{1}{2} \sum_{i=1}^{n} \sigma^2 \frac{\partial^2 f(\mu_1, \ldots, \mu_n)}{\partial X_i^2}.$$

We are interested to see if $\mathbb{E}[S^2(X)] = \frac{1}{n} \mathbb{E}[\sum_{i=1}^{n} (X_i - \bar{X})^2] = \sigma^2$ Let's just work with the part $D(X_1, \ldots, X_n) = \sum_{i=1}^{n} (X_i - \bar{X})^2$ as the $1/n$ does not contribute much to operations as it is a scalar for fixed n. To fill out the above expression we will need the following componentry

$$
\begin{aligned}
\frac{\partial D(X_1,\dots,X_n)}{\partial X_i} &= \frac{\partial}{\partial X_i}\sum_{i=1}^{n}(X_i-\bar{X})^2 = \frac{\partial}{\partial X_i}\sum_{i=1}^{n}[X_i^2-2X_i\bar{X}+\bar{X}^2] \\
&= \frac{\partial}{\partial X_i}[\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+n\bar{X}^2] \\
&= \frac{\partial}{\partial X_i}[\sum_{i=1}^{n}X_i^2-2\bar{X}n\bar{X}+n\bar{X}^2] \\
&= \frac{\partial}{\partial X_i}[\sum_{i=1}^{n}X_i^2-n\bar{X}^2] \\
&= 2X_i-2\bar{X} \\
\frac{\partial^2 D(X_1,\dots,X_n)}{\partial X_i^2} &= \frac{\partial}{\partial X_i}2X_i-2\bar{X} = 2-\frac{2}{n} = 2\frac{n-1}{n}
\end{aligned}
$$

Plug these into above

$$
\mathbb{E}[D(X_1,\dots,X_n)] \approx D(\mu)+\frac{1}{2}\sum_{i=1}^{n}\sigma^2 2\frac{n-1}{n}
$$

which is not approximate anymore as the higher order partial derivatives are 0. Another plus of the Taylor series sometimes. There is no i in this expression now so

$$
\mathbb{E}[D(X_1,\dots,X_n)] = D(\mu)+\sigma^2\frac{n}{n}(n-1)
$$

simplifying

$$
\mathbb{E}[D(X_1,\dots,X_n)] = D(\mu)+\sigma^2\,(n-1)
$$

and $D(\mu)=0$ because

$$
D(\mu) = [\sum_{i=1}^{n}X_i^2-n\bar{X}^2] = [n\mu^2-n[\frac{1}{n}\sum_{i=1}^{n}X_i]^2] = [n\mu^2-n[\frac{n}{n}\mu^2]] = 0
$$

Finally,

$$\mathbb{E}[S^2(X_1,\dots,X_n)] = \frac{(n-1)}{n}\sigma^2,$$

which is not what we want. For the final line to equal σ^2 the original expression for S^2 should have been divided by $n-1$.

Exercises

1. Show (3) from properties of expectation and variances

2. Show (8) from properties of variances and covariances

3. Provide a *third* order approximation to the expectation of any random variable

4. Given your approximation in question (3) write down the expectation of $\log(X)$

5. If X is normally distributed what is the $\mathbb{E}(\log(X))$? If X is gamma distributed what is the $\mathbb{E}(\log(X))$?

6. Review the final derivation and contrast it with
   `https://www.youtube.com/watch?v=D1hgiAla3KI`

References

Michael Lynch, Bruce Walsh, et al. *Genetics and analysis of quantitative traits*, volume 1. Sinauer Sunderland, MA, 1998.