BTRY 4830/6830: Quantitative Genomics and Genetics

Lecture 4: Expectations, variances and covariances of random variables and random vectors

Jason Mezey
jgm45@cornell.edu
Sept. 10, 2013 (T) 8:40-9:55
Announcements

• Class videos are posted - for future lectures “split screen” versions will be posted

• Homework #1 is posted:
  • There is an error in this version (2b and 2c are the same) - a correction “V2” will be posted later today (we will send out an email when it is up
  • If you are in Ithaca: you must email your answers to Amanda by 11:59PM, Mon. (!!) yg246@cornell.edu
  • If you are in NYC: you must email your answers to Jin by 11:59PM, Mon. (!!) jj328@cornell.edu

• Reminder: Office hours begin this week
  • Both Ithaca and Weill, 101 Biotech (Ithaca), main conference room, Dept. of Genetic Med. (NYC - please go the long way!)
  • This week only (!!) 4-5PM (usually 3-5PM)

• Note supplementary material #2 will be posted (matrix basics) that will be posted this afternoon

• Last reminder: make sure you are on the email list!!
Summary of lecture 4:

• Last lecture, we introduced the critical concept of a random variable and discussed how these functions are what we generally work with in prob. / statistics.

• In this lecture, we will complete our discussion of random vectors (the generalization of random variables) and associated concepts.

• We will also discuss expectations and variances (covariances) of random variables / vectors.
Conceptual Overview

System → Question → Sample → Prob. Models → Assumptions

Experiment → Question → Sample → Prob. Models → Assumptions

Inference → Question → Sample → Prob. Models → Assumptions

Statistics → Question → Sample → Prob. Models → Assumptions
Random Variables

\[ X = x, \ Pr(X) \]

\[ \mathcal{E}(\Omega) \]

Experiment

\[ \Omega \]

(Sample Space)

\[ \mathcal{F} \]

(Sigma Algebra)

Random Variable

\[ X(\Omega) \]

\[ Pr(\mathcal{F}) \]
Experiments and Sample Spaces

- **Experiment** - a manipulation or measurement of a system that produces an outcome we can observe
- **Experimental trial** - one instance of an experiment
- **Sample outcome** - a possible outcome of the experiment
- **Sample** - the results of one or more experimental trials
- **Sample Space** \((\Omega)\) - set comprising all possible outcomes associated with an experiment
- **Sigma Algebra** \((\mathcal{F})\) - a collection of events (subsets) of interest with the following three properties: 1. \(\emptyset \in \mathcal{F}\), 2. \(A \in \mathcal{F}\) then \(A^c \in \mathcal{F}\), 3. \(A_1, A_2, \ldots \in \mathcal{F}\) then \(\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}\)

Note that we are interested in a particular Sigma Algebra for each sample space...
Random variables review

- **Random variable** - a real valued function on the sample space:

\[ X(\Omega) : \Omega \to \mathbb{R} \]

- A critical point to note: because we have defined a probability function on the sample space \( S \), this induces a probability function on the random variable \( X \):

\[ Pr(\mathcal{F}) \xrightarrow{\text{map}} Pr(X) \]

- We considered examples of discrete and continuous random variables defined on the two-coin flip and reals (heights) where this relationships results in a pmf and pdf respectively:

![Graphs showing pmf and pdf](image)
Random vectors

- We are often in situations where we are interested in defining more than one r.v. on the same sample space.
- When we do this, we define a random vector.
- Note that a vector, in its simplest form, may be considered a set of numbers (e.g. [1.2, 2.0, 3.3] is a vector with three elements).
- Also note that vectors (when a vector space is defined) have similar properties to numbers in the sense that we can define operations for them (e.g. addition, multiplication), which we will use later in this course.
- Beyond keeping track of multiple r.v.’s, a random vector works just like a r.v., i.e. a probability function induces a probability function on the random vector and we may consider discrete or continuous (or mixed!) random vectors.
- Finally, note that while we can define several r.v.’s on the same sample space, we can only define one true probability function (why!?)
Example of a discrete random vector

- Consider the two coin flip experiment and assume a probability function for a fair coin: \( Pr(HH) = Pr(HT) = Pr(TH) = Pr(TT) = 0.25 \)

- Let’s define two random variables: “number of Tails” and “first flip is Heads”

\[
\begin{align*}
X_1(\Omega) &= \begin{cases} 
X_1(HH) = 0 \\
X_1(HT) = X_1(TH) = 1 \\
X_1(TT) = 2
\end{cases} \\
X_2(\Omega) &= \begin{cases} 
X_2(TH) = X_2(TT) = 0 \\
X_2(HH) = X_2(HT) = 1
\end{cases}
\end{align*}
\]

- The probability function induces the following pmf for the random vector \( \mathbf{X} = [X_1, X_2] \), where we use bold \( \mathbf{X} \) do indicate a vector (or matrix):

\[
Pr(\mathbf{X}) = Pr(X_1 = x_1, X_2 = x_2) = P_{\mathbf{X}}(\mathbf{x}) = P_{X_1, X_2}(x_1, x_2)
\]

\[
Pr(X_1 = 0, X_2 = 0) = 0.0, Pr(X_1 = 0, X_2 = 1) = 0.25 \\
Pr(X_1 = 1, X_2 = 0) = 0.25, Pr(X_1 = 1, X_2 = 1) = 0.25 \\
Pr(X_1 = 2, X_2 = 0) = 0.25, Pr(X_1 = 2, X_2 = 1) = 0.0
\]
Example of a continuous random vector

- Consider an experiment where we define a two-dimensional Reals sample space for “height” and “IQ” for every individual in the US (as a reasonable approximation)

- Let’s define a bivariate normal probability function for this sample space and random variables $X_1$ and $X_2$ that are identity functions for each of the two dimensions

- In this case, the pdf of $X = [X_1, X_2]$ is a bivariate normal (we will not write out the formula for this distribution - yet):

$$\Pr(X) = \Pr(X_1 = x_1, X_2 = x_2) = f_X(x) = f_{X_1, X_2}(x_1, x_2)$$

Again, note that we cannot use this probability function to define the probabilities of points (or lines!) but we can use it to define the probabilities that values of the random vector fall within (square) intervals of the two random variables (!) $[a,b], [c,d]$

$$\Pr(a \leq X_1 \leq b, c \leq X_1 \leq d) = \int_a^b \int_c^d f_{X_1, X_2}(x_1, x_2) \, dx_1, dx_2$$

For example, if we were interested in the genetics of human height, our experiment might be to measure both the height of each individual. This produces two sample spaces.

Before we leave the concept of joint pdf’s, let’s consider a conceptual extension. Note that for the specific probability function and random variables we have defined $p$, this produces $m$ induced a pdf $p$ on a single random variable $X$. Just as a probability function could also be a vector $r$. A random vector is therefore a vector-valued function $\mathbf{v}$.
Random vectors: conditional probability and independence

- Just as we have defined conditional probability (which are probabilities!) for sample spaces, we can define conditional probability for random vectors:

\[ Pr(X_1 | X_2) = \frac{Pr(X_1 \cap X_2)}{Pr(X_2)} \]

- As a simple example (discrete in this case - but continuous is the same!), consider the two flip sample space, fair coin probability model, random variables: “number of tails” and “first flip is heads”:

<table>
<thead>
<tr>
<th></th>
<th>(X_2 = 0)</th>
<th>(X_2 = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1 = 0)</td>
<td>0.0</td>
<td>0.25</td>
</tr>
<tr>
<td>(X_1 = 1)</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>(X_1 = 2)</td>
<td>0.25</td>
<td>0.0</td>
</tr>
</tbody>
</table>

\[ Pr(X_1 = 0 | X_2 = 1) = \frac{Pr(X_1 = 0 \cap X_2 = 1)}{Pr(X_2 = 1)} = \frac{0.25}{0.5} = 0.5 \]

- We can similarly consider whether r.v.’s of a random vector are independent, e.g. \( Pr(X_1 = 0 \cap X_2 = 1) = 0.25 \neq Pr(X_1 = 0)Pr(X_2 = 1) = 0.25 \times 0.5 = 0.125 \)
Marginal distributions of random vectors

- As a final concept for today, we will consider **marginal distributions** of random vectors, where these are the probability of a r.v. of a random vector after summing (discrete) or integrating (continuous) over all the values of the other random variables:

\[
P_{X_1}(x_1) = \sum_{x_2=\min(X_2)}^{\max(X_2)} Pr(X_1 = x_1 \cap X_2 = x_2) = \sum Pr(X_1 = x_1 | X_2 = x_2) Pr(X_2 = x_2)
\]

\[
f_{X_1}(x_1) = \int_{-\infty}^{\infty} Pr(X_1 = x_1 \cap X_2 = x_2) dx_2 = \int_{-\infty}^{\infty} Pr(X_1 = x_1 | X_2 = x_2) Pr(X_2 = x_2) dx_2
\]

- Again, as a simple illustration, consider our two coin flip example:

<table>
<thead>
<tr>
<th></th>
<th>(X_2 = 0)</th>
<th>(X_2 = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(X_1 = 0)</td>
<td>0.0</td>
<td>0.25</td>
</tr>
<tr>
<td>(X_1 = 1)</td>
<td>0.25</td>
<td>0.25</td>
</tr>
<tr>
<td>(X_1 = 2)</td>
<td>0.25</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>
Three last points about random vectors

• Just as we can define cmf’s / cdf’s for r.v.’s, we can do the same for random vectors:

\[ F_{X_1,X_2}(x_1, x_2) = Pr(X_1 \leq x_1, X_2 \leq x_2) \]

\[ F_{X_1,X_1}(x_1, x_2) = \int_{-\infty}^{i} \int_{-\infty}^{j} f_{X_1,X_2}(i, j) \]

• We have been discussing random vectors with two r.v.’s, but we can consider any number n of r.v.’s:

\[ Pr(X) = Pr(X_1 = x_1, X_2 = x_2, ..., X_n = x_n) \]

• We refer to probability distributions defined over r.v. to be univariate, when defined over vectors with two r.v.’s they are bivariate, and when defined over three or more, they are multivariate
Expectations and variances

- We are now going to introduce fundamental functions of random variables / vectors: **expectations** and **variances**
- These are functions can be thought of as having the following form:

  \[ f(\mathbf{X}(\mathcal{F}), \Pr) : \{\mathbf{X}, \Pr(\mathbf{X})\} \rightarrow \mathbb{R} \]

- These are critical concepts for understanding the structure of probability models where the interpretation of the specific probability model under consideration
- They also have deep connections to many important concepts in probability and statistics
- Note that these are distinct from functions (Transformations) that are defined directly on \( X \) and not on \( \Pr(X) \), i.e. \( Y = g(X) \):

  \[ g(\mathbf{X}) : X \rightarrow Y \]

  \[ g(\mathbf{X}) \rightarrow Y \Rightarrow \Pr(X) \rightarrow \Pr(Y) \]
Expectations I

- Following our analogous treatment of concepts for discrete and continuous random variables, we will do the same for expectations (and variances), which we also call expected values.

- Note that the interpretation of the expected value is the same in each case.

- The expected value of a discrete random variable is defined as follows:

\[ \text{EX} = \sum_{i=\min(X)}^{\max(X)} X_i \text{Pr}(X_i) \]

- For example, consider our two-coin flip experiment / fair coin probability model / random variable “number of tails”:

\[ \text{EX} = (0)(0.25) + (1)(0.5) + (2)(0.25) = 1 \]
Expectations II

- The expected value of a continuous random variable is defined as follows:

\[ E(X) = \int_{-\infty}^{+\infty} x f_X(x) \, dx \]

- For example, consider our height measurement experiment / normal probability model / identity random variable:
Expectations III

• In the discrete case, this is the same as adding up all the possibilities that can occur and dividing by the total number, e.g. \((0+1+1+2) / 4 = 1\) (hence it is often referred to as the mean of the random variable)

• An expected value may be thought of as the “center of gravity”, where a median (defined as the number where half of the probability is on either side) is the “middle” of the distribution (note that for symmetric distributions, these two are the same!)

• The expectation of a random variable \(X\) is the value of \(X\) that minimizes the sum of the squared distance to each possibility

• For some distributions, the expectation of the random variable may be infinite. In such cases, the expectation does not exist
Variances I

- We will define variances for discrete and continuous random variables, where again, the interpretation for both is the same.

- The variance of a discrete random variable is defined as follows:

\[
\text{Var}(X) = VX = \sum_{i=\text{min}(X)}^{\text{max}(X)} (X_i - EX)^2 Pr(X_i)
\]

- For example, consider our two-coin flip experiment / fair coin probability model / random variable “number of tails”:

\[
\text{Var}(X) = (0 - 1)^2(0.25) + (1 - 1)^2(0.5) + (2 - 1)^2(0.25) = 0.5
\]
Variances II

• The variance of a continuous random variable is defined as follows:

\[ \text{Var}(X) = VX = \int_{-\infty}^{+\infty} (X - EX)^2 f_X(x) dx \]

• For example, consider our height measurement experiment / normal probability model / identity random variable:
### Variances III

- Intuitively, the variance quantifies the “spread” of a distribution.
- The squared component of variance has convenient mathematical properties, e.g. we can view them as sides of triangles.
- Other equivalent (and often used) formulations of variance:
  
  \[
  \text{Var}(X) = \mathbb{E} [(X - \mathbb{E}X)^2] \\
  \text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2
  \]

- Instead of viewing variance as including a squared term, we could view the relationship as follows:

  \[
  \text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)]
  \]
Generalization: higher moments

- The expectation of a random variable is the “first” moment and we can generalize this concept to “higher” moments:

\[ \text{EX}^k = \sum_{i=\min(X)}^{\max(X)} X_i^k \Pr(X_i) \]

\[ \text{EX}^k = \int_{-\infty}^{+\infty} x^k f_X(x)dx \]

- The variance is the second “central” moment (i.e. calculating a moment after subtracting off the mean) and we can generalize this concept to higher moments as well:

\[ \text{C}(X^k) = \sum_{i=\min(X)}^{\max(X)} (X_i - \text{EX})^k \Pr(X_i) \]

\[ \text{C}(X^k) = \int_{-\infty}^{+\infty} (x - \text{EX})^k f_X(x)dx \]
Random vectors: expectations and variances

- Recall that a generalization of a random variable is a random vector, e.g.
  \[ X = [X_1, X_2] \quad \text{or} \quad P_{X_1,X_2}(x_1, x_2) \text{ or } f_{X_1,X_2}(x_1, x_2) \]

- The expectation (a function of a random vector and its distribution!) is a vector with the expected value of each element of the random vector, e.g.
  \[ \mathbf{E}X = [\mathbf{E}X_1, \mathbf{E}X_2] \]

- Variances also result in variances of each element (and additional terms... see next slide!!)

- Note that we can determine the conditional expected value or variance of a random variable conditional on a value of another variable, e.g.
  \[
  \begin{align*}
  \mathbf{E}(X_1 | X_2) &= \sum_{i=\min(X_1)}^{\max(X_1)} X_1(i) \cdot Pr(X_1(i) | X_2) \\
  \mathbf{V}(X_1 | X_2) &= \sum_{i=\min(X_1)}^{\max(X_1)} (X_1(i) - \mathbf{E}X_1)^2 \cdot Pr(X_1(i) | X_2) \\
  \mathbf{E}(X_1 | X_2) &= \int_{-\infty}^{+\infty} X_1 f_{X_1 | X_2}(x_1 | x_2) \, dx_1 \\
  \mathbf{V}(X_1 | X_2) &= \int_{-\infty}^{+\infty} (X_1(i) - \mathbf{E}X_1)^2 f_{X_1 | X_2}(x_1 | x_2) \, dx_1
  \end{align*}
  \]
Random vectors: covariances

- Variances (again a function!) of a random vector are similar producing variances for each element, but they also produce **covariances**, which relate the relationships between random variables of a random vector!!

\[
\text{Cov}(X_1, X_2) = \sum_{i=\min(X_1)}^{\max(X_1)} \sum_{j=\min(X_2)}^{\max(X_2)} (X_{i,1} - \mathbb{E}X_1)(X_{j,2} - \mathbb{E}X_2) P_{X_1,X_2}(x_1, x_2)
\]

\[
\text{Cov}(X_1, X_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (X_1 - \mathbb{E}X_1)(X_2 - \mathbb{E}X_2) f_{X_1,X_2}(x_1, x_2) dX_1 dX_2
\]

- Intuitively, we can interpret a positive covariance as indicating “big values of \(X_1\) tend to occur with big values of \(X_2\) AND small values of \(X_1\) tend to occur with small values of \(X_2\)”

- Negative covariance is the opposite (e.g. “big \(X_1\) with small \(X_2\) AND small \(X_1\) with big \(X_2\)”)

- Zero covariance indicates no relationship between big and small values of \(X_1\) and \(X_2\)
An illustrative example

- For example, consider our experiment where we have measured “height” and “IQ” / bivariate normal probability model / identity random variable:
Notes about covariances

- Covariance and independence, while related, are NOT synonymous (!!), although if random variables are independent, then their covariance is zero (but necessarily vice versa!)

- Covariances are symmetric: $\text{Cov}(X_1, X_2) = \text{Cov}(X_2, X_1)$

- Other equivalent (and often used) formulations of covariances:

  $\text{Cov}(X_1, X_2) = E[(X_1 - EX_1)(X_2 - EX_2)]$

  $\text{Cov}(X_1, X_2) = E(X_1X_2) - EX_1EX_2$

- From these formulas, it follows that the covariance of a random variable and itself is the variance:

  $\text{Cov}(X_1, X_1) = E(X_1X_1) - EX_1EX_1 = E(X_1^2) - (EX_1)^2 = \text{Var}(X_1)$
Covariance matrices

- Note that we have defined the “output” of applying an expectation function to a random vector but we have not yet defined the analogous output for applying a variance function to a random vector.

- The output is a covariance matrix, which is square, symmetric matrix with variances on the diagonal and covariances on the off-diagonals.

- For example, for two and three random variables:

  \[
  \text{Var}(\mathbf{X}) = \begin{bmatrix}
  \text{Var}X_1 & \text{Cov}(X_1, X_2) \\
  \text{Cov}(X_1, X_2) & \text{Var}X_2
  \end{bmatrix}
  \]

  \[
  \text{Var}(\mathbf{X}) = \begin{bmatrix}
  \text{Var}X_1 & \text{Cov}(X_1, X_2) & \text{Cov}(X_1, X_3) \\
  \text{Cov}(X_1, X_2) & \text{Var}X_2 & \text{Cov}(X_2, X_3) \\
  \text{Cov}(X_1, X_3) & \text{Cov}(X_2, X_3) & \text{Var}(X_3)
  \end{bmatrix}
  \]

- Note that not all square, symmetric matrices are covariance matrices (!!), technically they must be positive (semi)-definite to be a covariance matrix.
Covariances and correlations

• Since the magnitude of covariances depends on the variances of $X_1$ and $X_2$, we often would like to scale these such that “1” indicates maximum “big with big / small with small” and “-1” indicates maximum “big with small” (and zero still indicates no relationship)

• A correlation captures this relationship:

$$\text{Corr}(X_1, X_2) = \frac{\text{Cov}(X_1, X_2)}{\sqrt{\text{Var}(X_1)} \sqrt{\text{Var}(X_2)}}$$

• Where we can similarly calculate a correlation matrix, e.g. for three random variables:

$$\text{Corr}(X) = \begin{bmatrix} 1 & \text{Corr}(X_1, X_2) & \text{Corr}(X_1, X_3) \\ \text{Corr}(X_1, X_2) & 1 & \text{Corr}(X_2, X_3) \\ \text{Corr}(X_1, X_3) & \text{Corr}(X_2, X_3) & 1 \end{bmatrix}$$
Algebra of expectations and variances

- If we consider a transformation on $X$ (a function on the random variable but not on the probabilities directly), recall that this can result in a different probability distribution for $Y$ and therefore different expectations, variances, etc. for $Y$ as well.

- We will consider two types of transformations on random variables and the result on expectation and variances: sums $Y = X_1 + X_2 + \ldots$ and $Y = a + bX_1$ where $a$ and $b$ are constants.

- For example, for sums, $Y = X_1 + X_2$ we have the following relationships:

\[
E(Y) = E(X_1 + X_2) = EX_1 + EX_2
\]

\[
Var(Y) = Var(X_1 + X_2) = VarX_1 + VarX_2 + 2Cov(X_1, X_2)
\]

- As another example, for $Y = X_1 + X_2 + X_3$ we have:

\[
E(Y) = E(X_1 + X_2 + X_3) = EX_1 + EX_2 + EX_3
\]

\[
Var(Y) = Var(X_1 + X_2 + X_3) = VarX_1 + VarX_2 + VarX_3 + 2Cov(X_1, X_2) + 2Cov(X_1, X_3) + 2Cov(X_2, X_3)
\]
Algebra of expectations and variances

- For the transformation $Y = a + bX$ we obtain the same relationships for (univariate) random variables and random vectors $Y = a + bX$

- For example, for a vector with two elements $Y = a + bX$:
  
  $Y = [Y_1, Y_2] = [a + bX_1, a + bX_2]$
  
  $EY = E [Y_1, Y_2] = [a + bEX_1, a + bEX_2]$
  
  $\text{Var}Y = \text{Var} [Y_1, Y_2] = b^2 \text{Var}X = [b^2 \text{Var}X_1, b^2 \text{Var}X_2]$

- Finally, note that if we were to take the covariance (or correlation) of two random variables $Y_1$ and $Y_2$ with the relationship:

  $Y_1 = a_1 + b_1X_1, \ Y_2 = a_2 + b_2X_2$

  $\text{Cov}(Y_1, Y_2) = b_1b_2\text{Cov}(X_1, X_2)$

  $\text{Corr}(Y_1, Y_2) = \text{Corr}(X_1, X_2)$
That’s it for today

• Next lecture, we will introduce parameterized probability models, samples, and begin our discussion of inference