Quantitative Genomics and Genetics
BTRY 4830/6830; PBSB.5201.01

Lecture 5: Probability distribution functions, expectations, variances and covariances of random variables and random vectors

Jason Mezey
jgm45@cornell.edu
Feb. 4, 2020 (T) 8:40-9:55
Announcements

• Everyone up on Piazza / (Almost) everyone up on CMS!
• Registration for “not” graduate students in NYC: almost there… hope to have information today
• Homeworks:
  • Homework #1 due tomorrow night (11:59PM)!
  • Homework #2 will be posted tomorrow afternoon on CMS only (!!) and will (likely) be due 11:59PM, Tues., Feb. 11)
Summary of lecture 5:

• 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 variables and introduce random vectors (the generalization of random variables) and associated concepts.

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

- System
- Question
- Sample
- Prob. Models
- Inference
- Assumptions
- Statistics
- Experiment
Review

- **Experiment** - a manipulation or measurement of a system that produces an outcome we can observe
- **Sample Space** (\(\Omega\)) - set comprising all possible outcomes associated with an experiment
- **Sigma Algebra** or **Sigma Field** (\(\mathcal{F}\)) - a collection of events (subsets) of the sample space of interest
- **Probability Measure** (=**Function**) - maps a Sigma Algebra of a sample to a subset of the reals
- **Random Variable** - (measurable) function on a sample space
**Review: random variables**

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

\[ X : \Omega \rightarrow \mathbb{R} \]

- Intuitively:

\[ \Omega \rightarrow [X(\omega), \omega \in \Omega] \rightarrow \mathbb{R} \]

- Note that these functions are not constrained by the axioms of probability, e.g. not constrained to be between zero or one (although they must be measurable functions and admit a probability distribution on the random variable!)

- We generally define them in a manner that captures information that is of interest

- As an example, let’s define a random variable for the sample space of the “two coin flip” experiment that maps each sample outcome to the “number of Tails” of the outcome:

\[ X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2 \]
Review: random variables

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

Random Variable

\[ X(\omega), \omega \in \Omega \]

Experiment

\[ \mathcal{X} \]

(Sample Space)

\[ \Omega \]

(Sigma Algebra)

\[ \mathcal{F} \]
Review: Discrete random variables / probability mass functions (pmf)

- If we define a random variable on a discrete sample space, we produce a discrete random variable. For example, our two coin flip / number of Tails example:

\[ X(HH) = 0, X(HT) = 1, X(TH) = 1, X(TT) = 2 \]

- The probability function in this case will induce a probability distribution that we call a **probability mass function** which we will abbreviate as pmf.

- For our example, if we consider a fair coin probability model (assumption!) for our two coin flip experiment and define a “number of Tails” r.v., we induce the following pmf:

\[ Pr(HH) = Pr(HT) = Pr(TH) = Pr(TT) = 0.25 \]

\[ P_X(x) = Pr(X = x) = \begin{cases} Pr(X = 0) = 0.25 \\ Pr(X = 1) = 0.5 \\ Pr(X = 2) = 0.25 \end{cases} \]
Review: Discrete random variables / cumulative mass functions (cmf)

• An alternative (and important!) representation of a discrete probability model is a **cumulative mass function** which we will abbreviate (cmf):

\[ F_X(x) = Pr(X \leq x) \]

where we define this function for \( X \) from \(-\infty\) to \(+\infty\).

• This definition is not particularly intuitive, so it is often helpful to consider a graph illustration. For example, for our two coin flip / fair coin / number of Tails example:

![Graph Illustration](image)
Review: Continuous random variables / probability density functions (pdf)

- For a continuous sample space, we can define a discrete random variable or a continuous random variable (or a mixture!)
- For continuous random variables, we will define analogous “probability” and “cumulative” functions, although these will have different properties
- For this class, we are considering only one continuous sample space: the reals (or more generally the multidimensional Euclidean space)
- Recall that we will use the reals as a convenient approximation to the true sample space
- For the reals, we define a probability density function (pdf): \( f_X(x) \)
- A pdf defines the probability of an interval of a random variable, i.e. the probability that the random variable will take a value in that interval

\[
Pr(a \leq X \leq b) = \int_a^b f_X(x) \, dx
\]
Mathematical properties of continuous r.v.'s

- The pdf of $X$, a continuous r.v., does not represent the probability of a specific value of $X$, rather we can use it to find the probability that a value of $X$ falls in an interval $[a,b]$: 

$$Pr(a \leq X \leq b) = \int_{a}^{b} f_X(x)dx$$

- Related to this concept, for a continuous random variable, the probability of specific value (or point) is zero (why is this!?)

- For a specific continuous distribution the cdf is unique but the pdf is not, since we can assign values to non-measurable sets

- If this is the case, how would we ever get a specific value when performing an experiment!?
Probability density functions (pdf): normal example

• To illustrate the concept of a pdf, let’s consider the reals as the (approximate!) sample space of human heights, the normal (also called Gaussian) probability function as a probability model for human heights, and the random variable $X$ that takes the value “height” (what kind of function is this?)

• In this case, the pdf of $X$ has the following form:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
Continuous random variables / cumulative density functions (cdf)

- For continuous random variables, we also have an analog to the cmf, which is the **cumulative density function** abbreviated as cdf:

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

- Again, a graph illustration is instructive

- Note the cdf runs from zero to one (why is this?)
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) **ARE NOT REALLY NUMBERS** although 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.

- Note that we can define several r.v.’s on the same sample space (= a random vector), but this will result in one probability distribution 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”

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

- 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):

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

\[
Pr(X_1 = 0, X_2 = 0) = 0.0, \quad Pr(X_1 = 0, X_2 = 1) = 0.25 \\
Pr(X_1 = 1, X_2 = 0) = 0.25, \quad Pr(X_1 = 1, X_2 = 1) = 0.25 \\
Pr(X_1 = 2, X_2 = 0) = 0.25, \quad Pr(X_1 = 2, X_2 = 1) = 0.0
\]
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 \(\mathbf{X}=[X_1, X_2]\) is a bivariate normal (we will not write out the formula for this distribution - yet):

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

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 (!) \([a,b], [c,d]\)

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

- Note that **marginal distributions** of random vectors 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:

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

\[ F_{X_1,X_2}(x_1,x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(x_1,x_2) \, dx_1 \, dx_2 \]

• 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 **functionals** - map a function to a scalar (number)
- These intuitively (but not rigorously!) these may be thought of as “a function on a function” with the following form:

\[ f(X(\Omega), Pr(X)) : \{X, Pr(X)\} \rightarrow \mathbb{R} \]

- 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. \(Y = g(X)\):

\[ g(X) : X \rightarrow Y \]

\[ g(X) \rightarrow Y \Rightarrow Pr(X) \rightarrow Pr(Y) \]
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:

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

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

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

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

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

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

\[ X = \begin{cases} X = t, & \text{if } X = \text{HH} \\ X = u, & \text{if } X = \text{TH} \\ X = v, & \text{if } X = \text{TT} \end{cases} \]

The expected value of this random variable is:

\[ E(X) = t \cdot \frac{1}{2} + u \cdot \frac{1}{2} + v \cdot \frac{1}{2} \]

We may similarly define the expectation of a continuous random variable:

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

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

- 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:

A few comments about expectations of random variables:

- An intuitive interpretation of the variance is that it summarizes the ‘spread’ of a distribution. That is, if you were to look at a pmf or pdf, the wider the distribution of probability along the X-axis, the greater the variance and vice versa.

- There are other ways besides equations like and like to write the formula for the variance including:

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

\[
\text{Var}(X) = \text{E}[X^2] - \text{E}[X]^2
\]
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:

\[
\text{Var}(X) = \mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)]
\]
That’s it for today

• Next lecture, we will begin our introduction to parameterized probability models and samples