Topics We Touched On, From Chapters 6 - 8

Math/Stat 394B Summer 2010

1 Chapter 6

1.1 Joint Distribution Functions (6.1)

When dealing with more than one Random variable at a time, it makes sense to think of them as a *vector random variable*, and use the corresponding multi-variable language from calculus. Thus, for example, given a 2-component RV, (X, Y), we will think of their *joint density function* (we are restricting ourselves to these nice situations), $f_{X,Y}(x, y)$, such that

$$P[(X, Y) \in A] = \int_A f_{X,Y}(x, y)\, dxdy$$

(where, of course, that is a double integral of the 'nice enough' domain A)

1.2 Independent Random Variables (6.2)

In the *very special* case when $f_X(x, y) = g(x)h(y)$, the two RVs are said to be independent. It is easy to see how this is perfectly consistent with our discussion of independence in the discrete case. Please, check out the more detailed discussion in the other files here. When we are talking about these functions, we are thinking of them in their precise sense, including domain. Thus, for example,

$$f_{X,Y}(x, y) = \begin{cases} 2 & 0 \le x < y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

is a legitimate density for the pair (X, Y), but the two components are *not independent*, as can be argued in any number of ways

1.3 Sums of Independent Random Variables (6.3)

If we have a collection of RVs $X_1, X_2, \dots, X_n$, it is *always true* that

$$E\left[\sum_{k=1}^{n} a_k X_k\right] = a_k \sum_{k=1}^{n} E\left[X_k\right]$$

(the expectation operator is *linear*). In most any other situation, linearity is obviously absent. However, in the special case when the variables are also *independent*[1], it also happens that

$$Var\left[\sum_{k=1}^{n} a_k X_k\right] = a_k^2 \sum_{k=1}^{n} Var\left[X_k\right]$$

The proof is easy: we can argue (take $n = 2$, and, if you want to be really precise, use induction) that

$$Var\left[aX + bY\right] = E\left[(aX + bY)^2 - (aEX + bEY)^2\right] =$$

$$= E\left[a^2\left(E\left[X^2\right] - (EX)^2\right) + b^2\left(E\left[Y^2\right] - (EY)^2\right) + 2ab\left(E\left[XY\right] - E\left[X\right]E\left[Y\right]\right)\right] =$$

$$= a^2 Var\left[X\right] + b^2 Var\left[Y\right] + 2abCov\left[X, Y\right]$$

The last term has been summarized as "$Cov\left[X, Y\right]$", to follow common usage which calls it the *covariance of X and Y*. Using the definition of independence, it is very easy to see that for independent X and Y. $Cov\left[X, Y\right] = 0$ (the reverse is not true at all!). Thus for independent RVs (and, in fact, under the much weaker condition of "zero-correlation"), the variance of the sum is the sum of the variances.

1.4 The rest of Chapter 6

We don't have time to address sections 6.4 and 6.5. Some comments on the topic are in the additional material files. It is an exceedingly interesting and useful topics, but we will have to leave it for your next probability course.

[1] actually, the specific condition required is weaker – we only need the variables to be *uncorrelated* – but we are not going to discuss this issue, which you will definitely address in your next classes.

2 Chapter 7

2.1 Sections 7.2 and 7.3

We have discussed the core of this topic in the previous section 1.3. The main issue here is how to handle sums of independent (or, at least, uncorrelated) RVs, in view of the next Chapter.

2.2 Normal RVs (section 7.8)

This is an obviously important special case. It has vast ramifications, but these will have to be left to your future probability and statistics courses.

2.3 Multidimensional Normal Distributions

Even though there is, technically, not a great deal of extra work involved to deal with this case, we will have to leave it for future developments.

Still, if you have some familiarity with conic sections – i.e., multi-dimensional quadratic functions – the work involved here is not difficult. Basically, much like in the 1-dimensional case, you are looking at a multi-dimensional "negative definite" quadratic polynomial in the exponent (we need the density to be integrable).

Just looking at the exponent (and concentrate on the case of dimension 2 to keep some visual intuition handy), you will see that the "iso-lines" (the curves on which the exponent is constant) have to be ellipses, and, by completing the square (plus the extra tricks involved when your ellipse has axes that are not parallel to the coordinate axes), you can bring the exponents into a standard form.

Also, you will notice that a linear change of variables (in fact, a *rotation of the axes*), will change the exponent so that the mixed terms (in xy, as opposed to x^2, y^2, x, y) disappear, and the quadratic polynomial can be written as a sum of a polynomial in x and a polynomial in y. Since this is an exponent, the function can can be written as a product of a density in x with a density in y. In other words, any multidimensional Gaussian, can be transformed, via a linear transformation of the variables, into a collection of independent Gaussians. For example, starting from

$$e^{3x^2-2xy+3y^2+2x+2y+1}$$

changing to new variables $u = \frac{x+y}{2}, v = \frac{x-y}{2}$ (a rotation of the axes by $\frac{\pi}{4}$), we arrive at an expression like

$$e^{u^2+2u+1+2y^2} = e^{(u+1)^2}e^{2y^2}$$

3 Limit Theorems (Chapter 8)

3.1 The "Weak" Law of Large Numbers (Section 8.2)

In class, we have looked at an even (apparently) weaker result, based on convergence of Moment Generating Functions (one can repeat the argument for characteristic Functions, but to do it properly, we need a little familiarity with complex valued functions of a complex variable). The standard proof is in the book, and in the additional material on the web. It does provide insight, but we'd also like to stress how the MGF proof (OK, actually the Characteristic Function proof, but we have to make do) stresses the idea of the LLN as a "lowest order approximation" theorem for fairly general distribution.

3.2 The Central Limit Theorem (Section 8.3)

This is such an important theorem, that we had to look at it, even if under a serious technical limitation (we could only handle MGFs, wile the full-blown theorem actually uses Characteristic Functions). The proof in the book is repeated in the additional material on line. What you might want to keep out of the proof is the intuition that, in some sense (that can be made rigorous, under appropriate assumptions), Normal distributions are a natural "first order" approximation to "general" distributions.

The other points you should remember about this all-important theorem are

- We do need moments to be well defined for this theorem to hold (no variance, no CLT)
- The *speed of convergence* to the normal distribution (whatever measure you want to use) is never addressed in the proof. Thus, whether it is appropriate or not to use a normal approximation to the distribution of the sum of RVs depends heavily on the properties of the distribution of these RVs. In particular, for heavily skewed distributions, the convergence can be very slow, while for very symmetric distributions it can be surprisingly fast. This is a serious issue in many statistical problems.