Intro to Ch 5 – Sampling Theory

Section 5.1 Joint probabilities

This chapter starts to describe how you set up the mathematics of samples or measurements (measurements). As you will see, the independence of successive sample values (or successive measurements) simplifies things considerably. When you see the triple integrals, you will be glad that this independence allows us to avoid these complications.

You can ignore the double and triple integrals – we don't really need this in this intro course. But you need to have the idea that, while with one random variable you integrate the density function to get probabilities for an interval of values, with two random variables, you still integrate to get probabilities, but this time it is a double integral and it is the volume under the density surface that gives the probability. By analogy, any number of random variables can have a joint density (a hypersurface) and probabilities from multiple integration are provided by a hypervolume. Fortunately, this complexity collapses for our purposes because in the special case that our many variables are IID (independent and identically distributed), we never have to do this multiple integration. As we will see, the IID case describes a random sample (of size n say), since each sampled value comes from the same population, and what happens in selecting sample i does not affect the selection of sample j (i.e. independence).

The boxes on p 206-207 show you how to write down **joint** and **marginal** probabilities for two discrete RVs. Turning the summation signs into integral operators provides the equivalent definitions for continuous RVs – see boxes on p 208-209.

Is it obvious that $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$?

Think in terms of probabilities. If X=x, it must have happened in conjunction with some value of Y. The integral sums up the probabilities for all possible values of y. (Actually, not "probabilities" but densities, but the analogy is very close.

Remember from Ch 2 (p 87), if A and B are independent events, $P(A \cap B) = P(A)*P(B)$. It is also true that, if X and Y are independent RVs,

p(x,y) = p(x).p(y) when X,Y discrete and independent

and

f(x,y)=f(x).f(y) when X,Y continuous and independent

(p 211)

When these relationships are not true for **all** x,y, X,Y are called "dependent".

box on p 212 extends all this in an obvious way to n variables.

One interesting extension of the Binomial model is the Multinomial model.
Recall we had a fixed number of Bernoulli trials with p(success) constant leading to a Binomial RV as the number of successes. Note that we could consider this model leading to **two** outcomes (#successes, #failures). We chose to focus on # successes only since we could always compute #failures = n - # successes. But if we allow the **bi**nomial to have **two** outcomes, then the **tri**nomial would have **three** outcomes: say type1, type2 and type3. The trinomial RV would be reported as three counts, one for each type. Similarly there is a multinomial RV – see p 213.

If you rolled a dice 25 times, the outcome might be recorded as something like (4,3,6,2,4,6) meaning 4 1s, 3 2s, etc. The probability of this particular outcome would be, according to the formula on p 213,

$$\frac{25!}{4!3!6!2!4!6!}(1/6)^{25}$$

Note the independence of successive trials (dice rolls) made the calculation of the joint probability easy.

For an example of a similar computation for continuous RVs, see Example 5.11. Note the fact (worth remembering) that for an exponential RV X, P(X>t) = $e^{-\lambda t}$.

In addition to **joint** and **marginal** probabilities, we need **conditional** probabilities. We need to explain how the idea of conditioning is represented mathematically when the random variables involved are continuous.

We had in Ch 2 P(A|B)=P(A ∩ B)/P(B) p 78.
If we define an event A as X=x and B as Y=y,
we have P(X=x|Y=y) = P(X=x,Y=y)/P(Y=y)

So it is not surprising that for continuous variables:

$f_{Y|X}(y \mid x) = f_{X,Y}(x,y)/f_X(x)$, or in words,

conditional density = joint density / marginal density

Section 5.2 – Covariance and Correlation

A useful concept in data analysis is to describe the intensity of the relationship between two variables. One way to do this is the correlation coefficient – it measures the intensity of a particular kind of relationship, a **linear** relationship. It is sometimes called the linear correlation coefficient to emphasize this emphasis. Look at Fig 5.4 on p 221. If the two axes plot the data values of two variables X and Y, Fig 5.4 a) and b) show X and Y having a close linear relationship, while 5.4 c) shows no such relationship. As we shall see the correlation coefficient for Fig 5.4 a,b,and c are approx 0.9, -0.9, 0 respectively.

To describe the correlation coefficient, and at the same time provide a bit more machinery to work with jointly distributed random variables, we need the detail of section 5.2 (pp 219-224).

The box on p 219 is just saying that if you want to know the average value of a function of X and Y, (that's the h(X,Y)), then of course you need to weight each possible value by the frequency with which that possible value occurs. It should make sense if you really understood E(X) from Ch 2.

Now the function h() that gives us what we need on the way to the correlation coefficient is the function $h(X,Y) = (X-\mu_x)(Y-\mu_y)$. Notice that if X and Y tend to be on the same side of their means at the same time, h(X,Y) will be positive. If on opposite sides, h(X,Y) will be negative. We define $E((X-\mu_x)(Y-\mu_y))$ as the **covariance** of X and Y. We write COV(X,Y). See box p 220 for definitions for discrete and continuous RVs. The box on p 221 may be a little easier to remember, but the one on p 220 is more informative about what it actually calculates.

Now, except for the sign of it, the covariance is not very informative about the strength of the relationship. The problem is that the scale of the covariance depends directly on the particular units used for X and Y. The covariance of height and weight in cms and grams would be much bigger than if the same data was expressed in meters and kilograms. So for describing the intensity of linear relationship between X and Y, we use the **correlation** coefficient – it is proportional to the covariance but always is between -1 and +1.

Correlation (X,Y) = Covariance (X,Y) /(SD_X SD_Y) and in symbols = $\sigma_{X,Y}/\sigma_X\sigma_Y$

There is a close relationship between correlation of 0 and independence. See box p 223.