CS 5961 Computational Statistics

WEB 1248
Tu/Th 14:00 – 15:20

Lecture Notes for Monty Hall, Joint Probability and Convolution
3 Feb 2009

1. Monty Hall Problem
    a. Demo game with some runs through this nice applet: www.nytimes.com/2008/04/08/science/08monty.html
    b. Explain the game. Discuss the common (false) intuition.
        i. How can switching *possibly* improve odds, much less *double* them?
    c. Consider the analysis
        i. Combinatorics: enumerating the event space reveals that switching has better possible outcomes
            1. 2 of 3 (versus 1 of 3).
        ii. Look at the conditional probably argument.
    d. Study some statistical runs using applet: http://www.shodor.org/interactivate/activities/GeneralizedMontyHall/
        i. This lets you run a batch and collects stats
        ii. Set n=600, say, and observe a few runs with "stay" and then with "switch"
        iii. Questions:
            1. How many in class were they familiar with this game?
            2. How many in class were they surprised?
            3. Does this make any sense "in your gut?"
2. Expectation (or Expected Value) of a random variable X, denoted, $E(X)$. This is sort of the "theoretical mean," or what you should expect a long run of trails to tend toward.
    a. $E(X) = \sum_{i=1}^{n} P(X_i) X_i$,

       where $\{X_1, \ldots, X_n\}$ is the set of possible outcomes of X, and $P(X_i)$ is the respective likelihood that event X_i will occur.

b. The *Law of Large Numbers* basically says that, given enough set of trials or a large enough sample the "observed mean" will tend to the Expected Value).

c. First we consider the expected value $E(X)$ for a single die:

$$E(X) = \frac{1}{6}(1) + \frac{1}{6}(2) + \frac{1}{6}(3) + \frac{1}{6}(4) + \frac{1}{6}(5) + \frac{1}{6}(6)$$
$$= \frac{1}{6}(1+2+3+4+5+6)$$
$$= \frac{1}{6}(21) = \frac{7}{2}$$
$$\boxed{E(X) = 3.5}$$

d. Now go through the hand-out examining the outcome space and probabilities for rolling 2 (fair) dice: X and Y

e. With regard to the aforementioned definition above, note that, $E(X+Y) = E(X) + E(Y)$

i. Work out this rather straightforward property.

ii. Make a note of the particular structure of the problem, i.e., how can $E(X+Y) = t$? Observe that there is a distinct pattern in the analysisi.

3. The distribution that governs $Z = X + Y$ is called the *joint distribution* of X *and* Y. That is, given 2 random variables X, Y of known distributions, how is the sum distributed? For simplicity, we take the random variable to be identically distributed, but they do not have to be.

4. Now, consider the *joint distribution of two continuous random variable*s X, Y distributed according to density functions $f(x), g(y),$ respectively.

a. This means that $P(x \le a) = \int_{-\infty}^{a} f(t)dt$,

$$for \;\; f(t) \ge 0 \quad and \quad \int_{-\infty}^{+\infty} f(t) = 1.$$

and, analogously, for $g(y)$.

b. What then is the distribution function that governs the behavior of $z = x + y$, where x *and* y are defined as above?

c. With a little investigation, we can see the formal definition of *convolution* of $f * g$ reveals an intriguingly related structure,

$$(f*g)(t) = \int_{-\infty}^{+\infty} f(t-\tau)g(\tau)d\tau$$
$$= \int_{-\infty}^{+\infty} f(\tau)g(t-\tau)d\tau$$

reveals an intriguingly related structure. Let us investigate further.

5. Looking at matters from the view of the *joint distribution of two continuous random variable*s X, Y distributed according to density functions $f(x), g(y)$, respectively, we can also get some insights.
   a. Let us become convinced that convolution $f * g$ is just the right construction for defining the joint distribution of $f \text{ and } g$.
6. We return to the discrete situation where there is a natural analog of the continuous convolution, namely, the *discrete convolution*. Observe that *discrete convolution* is also just the right mechanism, e.g., for describing the *joint distribution* of 2 dice. Refer to the handout enumerating the expected value of a pair of dice:
   <eng.utah.edu/~cs5961/lec_Notes/EVof2Dice.pdf>
7. While we are on the subject, we can see that the discrete convolution also can pertain to the task of *data smoothing*. Consider a doubly infinite set of measured observations $\{x_i\}_{i=-\infty}^{+\infty}$. (We are simply trying to push out the boundary conditions to infinity so that they do not enter into our consideration.)
   a. Consider an example, say, of the successive daily readings of temperature measured in centigrade at the same time (at high noon) and location each day, ignoring daylight saving time, etc. There will be variations due to local influences and measurement errors, but over a longer interval there will be clear trends.
   b. What about the Dow Jones Index, assuming it has been measured from the beginning to the end of time…? There will be clear trends, but also much local fluctuation and *noise*. We can view this as a pure signal with noise *superimposed*. We want to elucidate (extract) the underlying phenomenon. This is what a keen stock investor would want to do.
   c. Same for ocean waves, right? There are often steady fundamental frequencies and then local wind perturbations superimposed on them, as in the North Sea, for example, which is relatively shallow.
8. Now, we exploit what we have established to introduce the notion of a (data) *filter*. as An example, define a new sequence derived from the original sequence $\{x_i\}_{i=-\infty}^{+\infty}$ in the following way, $\hat{x}_i = \frac{1}{4}\left(x_{i-1} + 2x_i + x_{i+1}\right)$.

a. What will $\{\hat{x}_i\}_{i=-\infty}^{+\infty}$ look like compared to the original data $\{x_i\}_{i=-\infty}^{+\infty}$?

b. Why do we divide by 4? What happens if the weights, the data coefficients, do not add to the denominator?

i. It biases the entire sequence, but still smoothes the data fluctuations.

9. Jumping back, we look again at the continuous case and ask how to define a continuous filter.

a. We need the concept of a *sliding data window*.

b. We may think of a window and a signal function as performing distinct roles, but they are fundamentally (formally) interchangeable in the mathematical definition.

c. What are some common characteristics of a filter?

i. Does it have to be everywhere nonnegative?

ii. Other examples of a filter that you might know about or concoct?

1. $$\hat{x}_i = \frac{1}{7}\left(\frac{1}{2}x_{i-2} + \frac{3}{2}x_{i-1} + 3x_i + \frac{3}{2}x_{i+1} + \frac{1}{2}x_{i+2}\right)$$

2. $$\hat{x}_i = \frac{1}{4}\left(x_{i-2} - \frac{1}{2}x_{i-1} + 3x_i - \frac{1}{2}x_{i+1} + x_{i+2}\right)$$

iii. Or should the denominator above be $\frac{1}{6}$? Argue which is the proper denominator value by considering the filter's action on the constant sequence $\{x_i = 1\}_{i=-\infty}^{+\infty}$. What should the filter produce as a resulting sequence $\{\hat{x}_i\}_{i=-\infty}^{+\infty}$?

iv. What does a filter do in circuit theory? What does a power filter do for a media device? It is simply a hardware realization of the concept.

v. A convolution is a *linear filter*. What does that mean intuitively? The effect of the filter does not vary over time. Data in the beginning of the sequence is treated in the same manner as data toward the end.

vi. What do you suppose are the characteristics of a "good" filter? How should a "good" filter perform? Does this depend on the application needs?

10. Generally, this quickly gets into a very sophisticated and elaborate but fundamentally important theory. However, the basic idea of the convolution mechanism, and related concepts, is tractable. That is all that we want to accomplish here, just an exposure to the concept. This is a very powerful construction that has pervasive application.

a. Filtering is an essential component of nearly all communications schemes, and many other common devices.

b. This idea is related to a common band pass filter presented in many sound systems.

c. Filters are used pervasively in image enhancement, satellite images, and the like.

d. What is a good way to expand a lower resolution image to a higher resolution definition? How should the extra, “in between,” pixels be determined?

e. The human ear applies a natural version of this.

    i. How do we detect sound frequency, when we try to play an instrument “in tune?”