

Lecture Notes: *Convolution*

1. The distribution that governs $Z = X + Y$ is called the *joint distribution* of X and Y . That is, given 2 random variables X, Y distributed according to $f(x), g(y)$, respectively, how is the sum $Z = X + Y$ distributed? For simplicity, we take the random variable to be identically distributed, although they do not have to be.
2. Now, consider the *joint distribution of two continuous random variables* X, Y distributed according to density functions $f(x), g(y)$, respectively.

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

where, $f(t) \geq 0$ and $\int_{-\infty}^{+\infty} f(t) = 1$,

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,

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

Let us investigate further.

3. Looking at matters from the view of the *joint distribution of two continuous random variables* density functions $f(x), g(y)$, respectively, we can also get some insights.
 - a. Let us become convinced that the convolution $f * g$ is just the right construction for defining the joint distribution of f and g .

4. 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 exactly 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>
5. 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 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.
6. 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, $\hat{x}_i = \frac{1}{4}(x_{i-1} + 2x_i + x_{i+1})$.
 - 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. If the weights do not sum to 1, the entire sequence becomes biased by a constant, but the data fluctuations are still smoothed. Just consider the sequence $\{x_i = 1\}_{i=-\infty}^{+\infty}$ to see that the entire sequenced is scaled by a factory equal to the sum of all the weights.
7. Returning, 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, as their roles are quite symmetric 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 same 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?

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

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

b. This idea is related to a common band pass filter present 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 sound filtering.

- i. How do we detect sound frequency, when we try to play an instrument “in tune?”
9. Here we review the definition of discrete convolution, something we encountered in analyzing the distribution of two dice.

- a. Consider two doubly infinite sequences $\mathcal{A} = \{a_i\}_{i=-\infty}^{+\infty}$ and $\mathcal{B} = \{b_i\}_{i=-\infty}^{+\infty}$. The discrete convolution $\mathcal{A} * \mathcal{B} = \mathcal{C} = \{c_k\}_{k=-\infty}^{+\infty}$,

$$\text{where, } c_k = \sum_{i=-\infty}^{+\infty} a_i b_{k-i} .$$

- b. That is, an element the new of the sequence c_k is the sum of all product terms a_i and b_{k-i} whose indices sum to k . We saw an example of this when we were looking at how two dice could yield a particular sum like 8.