Lecture 6

• Comments on FFTs and applications
• Probability and stochastic processes (start)
Comments on FT Apps

• Spectral analysis of time series and image formation in aperture synthesis are analogs:
  – time series:spectrum = visibilities:image

• The pulsar applications discussed largely represent implementations of linear systems for data reduction

• We will see other applications where the FT is used to characterize signals.
Summary of Results on DFT of Complex Exponential + Noise

Time Series: \( \frac{S}{N} \sim \frac{A}{\sigma_n} \)

DFT (N-pt): \( \frac{S}{N} \sim \sqrt{N} \left( \frac{A}{\sigma_n} \right) \)

Estimation Error (if \( S/N \ll 1 \)) \( \varepsilon_k = 1 \) (100% error)

Uncorrelated DFT amplitudes \( \langle \tilde{X}_k \tilde{X}^*_k \rangle = \frac{\sigma^2}{N} \delta_{kk'} \)

\[ \tilde{X}_k = N^{-1} \sum_{n=0}^{N-1} x_n e^{-2\pi i nk/N} \]
\( \tilde{X}_k \rightarrow \text{Gaussian complex noise} \)

(CLT as \( N \rightarrow \infty \) if \( N \gg 1 \))

PDF of \( S_k = |\tilde{X}_k|^2 \)

(null hypothesis) \( \langle S \rangle^{-1} e^{-S/\langle S \rangle} U(S) \) (if noise only in \( x_n \))

- When signal is present, PDF of spectrum has a different form.
- Unaveraged or unsmoothed spectral estimator has two degrees of freedom.
- Summing \( M \) spectra from \( M \) realizations of the measurements increases the estimate to \( 2M \) d.o.f.
- Smoothing by \( m \) samples in a single spectrum increases the estimate to \( 2m \) d.o.f.
- For a given length data set \( T = N \Delta t \), the same number of degrees of freedom is gotten with an \( N \)-pt DFT followed by smoothing over \( M \) samples, or by averaging the spectra from data blocks of length \( T/M \). Note that the frequency resolution is the same in the two cases.
A guided tour of the fast Fourier transform

The fast Fourier transform algorithm can reduce the time involved in finding a discrete Fourier transform from several minutes to less than a second, and also can lower the cost from several dollars to several cents.

G. D. Bergland  
Bell Telephone Laboratories, Inc.

For some time the Fourier transform has served as a bridge between the time domain and the frequency domain. It is now possible to go back and forth between waveform and spectrum with enough speed and economy to create a whole new range of applications for this classic mathematical device. This article is intended as a primer on the fast Fourier transform, which has revolutionized the digital processing of waveforms. The reader's attention is especially directed to the IEEE Transactions on Audio and Electroacoustics for June 1969, a special issue devoted to the fast Fourier transform.

This article is written as an introduction to the fast Fourier transform. The need for an FFT primer is apparent from the barrage of questions asked by each new person entering the field. Eventually, most of these questions are answered when the person gains an understanding of some relatively simple concept that is taken for granted by all but the uninstructed. Here the basic concepts will be introduced by the use of specific examples. The discussion is centered around these questions:

1. What is the fast Fourier transform?
2. What can it do?
3. What are the pitfalls in using it?
4. How has it been implemented?

Representative references are cited for each topic covered so that the reader can conveniently interrupt this fast guided tour for a more detailed study.

What is the fast Fourier transform?

The Fourier transform has long been used for characterizing linear systems and for identifying the frequency components making up a continuous waveform. However, when the waveform is sampled, or the system is to be analyzed on a digital computer, it is the finite, discrete version of the Fourier transform (DFT) that must be understood and used. Although most of the properties of the continuous Fourier transform (CFT) are retained, several differences result from the constraint that the DFT must operate on sampled waveforms defined over finite intervals.

The fast Fourier transform (FFT) is simply an efficient method for computing the DFT. The FFT can be used in place of the continuous Fourier transform only to the extent that the DFT could before, but with a substantial reduction in computer time. Since most of the problems associated with the use of the fast Fourier transform actually stem from an incomplete or incorrect understanding of the DFT, a brief review of the DFT will first be given. The degree to which the DFT approximates the continuous Fourier transform will be discussed in more detail in the section on "pitfalls."

The discrete Fourier transform. The Fourier transform pair for continuous signals can be written in the form

\[ X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} \, dt \]  

(1)

\[ x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} \, df \]  

(2)

for \(-\infty < f < \infty, \) \(-\infty < t < \infty,\) and \(i = \sqrt{-1}.\) The uppercase \(X(f)\) represents the frequency-domain function; the lowercase \(x(t)\) is the time-domain function.

The analogous discrete Fourier transform pair that applies to sampled versions of these functions can be written in the form

\[ X(j) = \frac{1}{N} \sum_{k=0}^{N-1} x(k) e^{-j2\pi jk/N} \]  

(3)

\[ x(k) = \sum_{j=0}^{N-1} X(j) e^{j2\pi jk/N} \]  

(4)

for \(j = 0, 1, \ldots, N - 1;\) \(k = 0, 1, \ldots, N - 1.\) Both

Efficacy of FFT
FIGURE 7. The Fourier coefficients of the discrete Fourier transform viewed as a corrupted estimate of the continuous Fourier transform.
DFT Properties

- Frequency mapping:

\[ x(t) = e^{i\omega_0 t} = e^{2\pi if_0 t} \]

- in an N-point DFT, in which bin does the signal fall (mostly)?
- \( T = N\delta t \)
- \( \delta f = 1/T \)
- define \( f_j = j\delta f \) for \( j = 0, N/2 \)
  \[ = (N-j)\delta f \] for \( j = N/2+1, N-1 \)
- Nyquist frequency
- \( f_N = N\delta f/2 = N/2T = N/2N\delta t = 1/2\delta t \)
- For \( f_0 \leq f_N \) we have \( j_0 = f_0N\delta t \)
- The frequency bin \( j_0 \):
  - varies with \( N \) for fixed \( f_0 \) and \( \delta t \)
  - varies with \( \delta t \) for fixed \( f_0 \) and \( N \)
Symmetry Properties of DFT

• DFT:

\[ x_n = \sum_{k=0}^{\infty} e^{2\pi ink/N} \tilde{X}_k \]

\[ \tilde{X}_k = \frac{1}{N} \sum_{n=0}^{N-1} e^{-2\pi ink/N} x_n \]

• periodic in \( N \) in both domains
• discrete functions in time and frequency \( \delta t \) and \( \delta f \)
• \( T = \) time series length = \( N \delta t \)
• \( \delta f = 1/T \)
• Nyquist frequency = \( N \delta f/2 = 1/2 \delta t \)
Symmetry Properties of DFT

- **Hermitian**
  \[ x_n^* \iff \tilde{X}_{N-k}^* \]
  Show by substituting \( N-k \) into DFT
  - If \( x_n \) is real then
    \[ x_n \iff \tilde{X}_{N-k}^* \]
    \[ \tilde{X}_{N-k}^* = \tilde{X}_k \]

- The symmetry properties tell us how to fill an array with data to achieve specific results

- What are the symmetry properties of the 2D DFT?

\[ \tilde{X}_{kl} = \frac{1}{NM} \sum_n \sum_m x_{nm} e^{-2\pi i (nk+ml)/NM} \]
How do we fill an array to get a real signal in the other domain?

- \( k = 0, 1, \ldots, N-1 \)
- Need to know \( \frac{N}{2}+1 \) unique values

\[
\tilde{X}_0 \equiv \tilde{X}_N \\
\tilde{X}_1 \equiv \tilde{X}_{N-1} \\
\tilde{X}_2 \equiv \tilde{X}_{N-2}
\]
Gaussian Example

• How do we insert a Gaussian function into an array in order to get a real DFT?
• Recall that a Gaussian function centered on $t=0$ has a real FT:
\[ e^{-\pi t^2} \iff e^{-\pi f^2} \]
• What is the equivalent placement for the DFT?
• Answer: see previous page!
Cyclical vs Non-cyclical
Convolution & Correlation

**FIGURE 15.** The cyclical convolution of two finite signals analogous to that performed by the FFT algorithm.

**FIGURE 14.** The procedure for performing noncyclical convolution on two finite signals.

Usually we want

- Usually filtered by
- Accomplished by sliding and summing
Noncyclical Procedures

**Figure 16.** Noncyclical convolution of two finite signals analogous to that performed by the FFT algorithm.

**Figure 17.** A method for convolving a finite impulse response with an infinite time function by performing a series of fast Fourier transforms.

**Figure 18.** A method of using the fast Fourier transform algorithm to compute $N$ lags of the autocorrelation function of an $M$-term series.
Simulating power-law noise

• Applications: many phenomena in nature are processes with spectra that are power-law in form (temporally or spatially)
  \[ S(f) \propto f^{-\alpha} \quad f_0 \leq f \leq f_1 \]

• Can use a linear filter with impulse response \( h(t) \):

\[
\begin{align*}
x(t) & \rightarrow h(t) \rightarrow y(t) \\
y(t) &= x(t) * h(t) \\
\tilde{Y}(f) &= \tilde{X}(f) \times \tilde{H}(f) \\
\langle |\tilde{Y}(f)|^2 \rangle &= \langle |\tilde{X}(f)|^2 \rangle |\tilde{H}(f)|^2
\end{align*}
\]

• Let \( x(t) = \) realization of white noise and \( H(f) = \sqrt{\text{shape}} \) of spectrum that is wanted
Procedure

• Generate spectrum in frequency domain
• Generate white noise realizations for real and imaginary parts of $H(f)$
  – random number generator: gaussian, uniform
• Multiply by $f^{-\alpha/2}$ for $f_0 \leq f \leq f_1$
• Fill vector to force Hermiticity
• Do inverse DFT to get real time-domain realization of the noise
• Can be applied to any shape of spectrum of course
• What will the statistics be in the time domain?
slope = 0

slope = 1
slope = 2

slope = 3
slope = 4

slope = 5
slope = 6
Notions of Probability

**Qualitative:** (Knowledge)

As a measure of the degree to which propositions, hypotheses, or quantities are known. The measure can be an intrinsic property of an event in and of itself or relative to other events.

eg. “It is not probable that the sun will explode as a supernova” is an absolute statement based on our background knowledge of stellar evolution.

eg. “Horse A is more probable to win, place, or show than horse B” is a relative statement that includes the constraint that one horse wins to the exclusion of any other horse in the race.

Bayesian inference makes use of this qualitative form of probability along with quantitative aspects discussed below. In this sense, Bayesian methods attribute probabilities to more entities than do some of the more formalistic approaches.
Quantitative: (Frequentist approach: random variables and ensembles)

Classical: Probabilities of events $\zeta$ in an event space $S$ are found, a priori, by considerations of possible outcomes of experiments and the manner in which the outcomes may be achieved. For example, a single-die yields probabilities of $1/6$ for the outcomes of a single toss of the die. A problem with the classical definition (which often calculates probabilities from line segment lengths, areas, volumes, etc.) is that they cannot be evaluated for all cases (eg. an unfair die).

Relative Frequency: The probability of an event $\zeta$ is defined to be the limit of the frequency of occurrence of the event in $N$ repeated experiments as $N \to \infty$:

$$P\{\zeta\} = \lim_{N \to \infty} \frac{N_\sigma}{N}.$$  

The problem with frequencies is that in real experiments $N$ is always finite and therefore probabilities can only be estimated. Such estimates are a poor basis for deductive probabilistic theories.
Axiomatic: A deductive theory of probability can be based on axioms of probability for allowed events in an overall event space. The axioms are:

For \( \zeta \) = an element in the event space \( S \) and \( P(\zeta) \) = probability of the event \( \zeta \) the following hold:

i) \( 0 \leq P(\zeta) \leq 1 \)

ii) \( P\{S\} = 1 \) (\( S \) = set of all events, so \( P\{S\} \) is the probability that any event will be the outcome of an experiment.

iii) If \( \zeta_1 \) and \( \zeta_2 \) are mutually exclusive then the probability of \( \zeta_1 + \zeta_2 \) (\( \equiv \) the event that \( \zeta_1 \) or \( \zeta_1 \) occurs) is

\[
P(\zeta_1 + \zeta_2) = P(\zeta_1) + P(\zeta_2).
\]

These axioms imply further that:

iv) If \( \bar{\zeta} \) = event that \( \zeta \) does not occur, then \( P(\bar{\zeta}) = 1 - P(\zeta) \).

v) If \( \zeta_1 \) is a sufficient condition for \( \zeta_2 \), then \( P(\zeta_1) \leq P(\zeta_2) \). Equality holds when \( \zeta_2 \) is also a necessary condition for \( \zeta_2 \).

vi) Let \( P(\zeta_1 \zeta_2) \) = probability of the event that \( \zeta_1 \) and \( \zeta_2 \) occurs (overlap in Venn diagram). Mutually exclusive events have \( P(\zeta_1 \zeta_2) = 0 \) while generally \( P(\zeta_1 \zeta_2) \geq 0 \). In general we also have

\[
P(\zeta_1 + \zeta_2) = P(\zeta_1) + P(\zeta_2) - P(\zeta_1 \zeta_2) \leq P(\zeta_1) + P(\zeta_2).
\]
Conditional Probabilities:

Also satisfying the axioms are *conditional probabilities*. Consider the probability that an event $\zeta_2$ occurs given that the event $\zeta_1$ occurs. It may be shown that this probability is

$$P\{\zeta_2|\zeta_1\} = \frac{P\{\zeta_1\zeta_2\}}{P\{\zeta_1\}}.$$ 

Similarly,

$$P\{\zeta_1|\zeta_2\} = \frac{P\{\zeta_2\zeta_1\}}{P\{\zeta_2\}}.$$ 

Bayes’ Theorem:

Solving for $P\{\zeta_1\zeta_2\}$ from the two preceding equations and setting these solutions equal yields *Bayes’ theorem*

$$P\{\zeta_2|\zeta_1\} = \frac{P\{\zeta_1|\zeta_2\}P\{\zeta_2\}}{P\{\zeta_1\}}.$$
\[ P(\Psi | S) = \frac{P(\Psi S)}{P(S)} \]

\[ \text{no overlap } \Rightarrow P(\Psi | S) = 0 \quad \& \quad P(\Psi S) = 0 \]

\[ \text{Complete overlap: } P(\Psi | S) = \frac{P(\Psi S)}{P(S)} = \frac{P(S)}{P(S)} = 1 \]
**Bayesian Inference:**

Bayesian inference is based on the preceding equation where, with relaxed definitions of the event space, hypotheses and parameters are attributed probabilities based on knowledge before and after an experiment is conducted. Bayes’ theorem combined with the qualitative interpretation of probability therefore allows the sequential acquisition of knowledge (i.e., learning) to be handled. The implied temporal sequence of events, by which data are accumulated and the likelihood of a hypothesis being true increases or decreases, represents the power of the Bayesian outlook. Moreover, with Bayesian inference, assumptions behind the inference are often brought “up front” as conditions upon which probabilities are calculated.
Basic tools and jargon

- Random variables, event space
- PDF, CDF, characteristic function
- Median, mode, mean
- Comparing PDFs
- Moments and moment tests
- Sums of random variables and convolution theorem
- Central Limit Theorem
- Changes of variable
- Functions of random variables
- Sequences of random variables
- Stochastic processes
- Power spectrum, autocorrelation, autocovariance, and structure functions
- Bispectra
- Random walks, shot noise, autoregressive, moving average, Markov processes
Probability and Random Processes

Experiments

Set up certain conditions to which the possible outcomes are called events.

The event space $S$ is the set of all outcomes or events $\zeta_i, i = 1, N$.

Events may or may not be quantitative e.g. the experiment may consist of choosing colored marbles from a hat.

Detections are experiments designed to answer the question: Is (effect) present in this physical system?

Measurements are experiments designed to yield quantitative measures of some physical phenomenon. Measurements are simply a highly structured form of interaction with a physical system. As such, they are never precise. Estimation of physical parameters is the best one can do, even for values of fundamental constants.

Probability

The notion of probability arises when we wish to consider the likelihood of a given event occurring or if we wish to estimate the number of times an identifiable event will occur if we repeat a given experiment $N$ times.

Event space:

Events $\zeta_i \in S$:

Events are possible outcomes of experiments.

Events can be combined to define new events.

The set of all events is the event space.
As such, probability is a theoretical quantity and is not the same as the frequency of occurrence of an event in repeated trials of an experiment. Of course, one can estimate the probabilities from repeated trials.

We will consider probability to be the underpinning of experiments and we will require it to behave according to three axioms:

Let \( \zeta \) be an event in \( S \), then

i) \( 0 \leq P(\zeta) \leq 1 \)

ii) \( P(S = \text{space of all events}) = 1 \)

iii) If two events are mutually exclusive [i.e. the occurrence of \( \zeta \) does not influence the occurrence of \( \psi \)], then the probability of the event \( \zeta + \psi = \text{event that } \zeta \text{ or } \psi \text{ occurs} \) is \( P(\zeta + \psi) = P(\zeta) + P(\psi) \)

(+ means ‘or’)

From the axioms, one can construct such results as:

1. \( \bar{A} = \text{event that } A \text{ does not occur} \)
   \[ P(\bar{A}) = 1 - P(A) \]

2. If the occurrence of \( A \) is a sufficient condition for \( B \) occurring, \([A \Rightarrow B \text{ but } B \text{ may occur when } A \text{ does not}] \text{ then } i \)
   \[ P(A) \leq P(B) \]

3. \( P(A + B) = P(A) + P(B) - P(AB) \) where \( P(AB) = \text{probability that both } A \text{ and } B \text{ occur} \). \( P(AB) \geq 0 \), with equality when \( A, B \) are mutually exclusive.
   \[ \Rightarrow P(A + B) \leq P(A) + P(B). \]
I. Mutually exclusive events:

If $a$ occurs then $b$ cannot have occurred.

Let $c = a + b$  \[ \Rightarrow \text{“or” (same as } a \cup b) \]

$P(c) = P\{a \text{ or } b \text{ occurred} \} = P(a) + P(b)$

Let $d = a \cdot b$  \[ \Rightarrow \text{“and” (same as } a \cap b) \]

$P(d) = P\{a \text{ and } b \text{ occurred} \} = 0$  \text{if mutually exclusive}

II. Non-mutually exclusive events:

$P(c) = P\{a \text{ or } b \} = P(a) + P(b) - P(ab)$

III. Independent events:

$P(ab) \equiv P(a)P(b)$
Examples

I. Mutually exclusive events

toss a coin once:
2 possible outcomes H & T
H & T are mutually exclusive
H & T are not independent because \( P(HT) = P\{\text{heads} \& \text{tails}\} = 0 \) so \( P(HT) \neq P(H)P(T) \).

II. Independent events

toss a coin twice = experiment
The outcomes of the experiment are

<table>
<thead>
<tr>
<th>1st toss</th>
<th>2nd toss</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H_1)</td>
<td>(H_2)</td>
</tr>
<tr>
<td>(H_1)</td>
<td>(T_2)</td>
</tr>
<tr>
<td>(T_1)</td>
<td>(H_2)</td>
</tr>
<tr>
<td>(T_1)</td>
<td>(T_2)</td>
</tr>
</tbody>
</table>

events might be defined as:
\(H_1H_2 = \text{event that H on 1st toss, H on 2nd}\)
\(H_1T_2 = \text{event that H on 1st toss, T on 2nd}\)
\(T_1H_2 = \text{event that T on 1st toss, H on 2nd}\)
\(T_1T_2 = \text{event that T on 1st toss, T on 2nd}\)

note \( P(H_1H_2) = P(H_1)P(H_2) \) [as long as coin not altered between tosses]
Random Variables

Of interest to us is the distribution of probability along the real number axis:

Random variables assign numbers to events or, more precisely, map the event space into a set of numbers:

\[ a \rightarrow X(a) \]

\[ \text{event} \rightarrow \text{number} \]

The definition of probability translates directly over to the numbers that are assigned by random variables. The following properties are true for a real random variable.

1. Let \( \{X \leq x\} \) = event that the r.v. \( X \) is less than the number \( x \); defined for all \( x \) [this defines all intervals on the real number line to be events]

2. the events \( \{X = +\infty\} \) and \( \{X = -\infty\} \) have zero probability. (Otherwise, moments would not be finite, generally.)
Distribution function: (CDF = Cumulative Distribution Function)

\[ F_X(x) = P\{X \leq x\} \equiv P\{\text{all events } A : X(A) \leq x\} \]

properties:

1. \( F_X(x) \) is a monotonically increasing function of \( x \).
2. \( F(-\infty) = 0, F(+\infty) = 1 \)
3. \( P\{x_1 \leq X \leq x_2\} = F(x_2) - F(x_1) \)

Probability Density Function (PDF)

\[ f_X(x) = \frac{dF_X(x)}{dx} \]

properties:

1. \( f_X(x) \, dx = P\{x \leq X \leq x + dx\} \)
2. \( \int_{-\infty}^{\infty} f_X(x) \, dx = F_X(\infty) - F_X(-\infty) = 1 - 0 = 1 \)

Continuous r.v.’s: derivative of \( F_X(x) \) exists \( \forall x \)

Discrete random variables: use delta functions to write the PDF in pseudo continuous form.

e.g. coin flipping

Let \( X = \begin{cases} 1 & \text{heads} \\ -1 & \text{tails} \end{cases} \)

then

\[ f_X(x) = \frac{1}{2}\left[ \delta(x + 1) + \delta(x - 1) \right] \]

\[ F_X(x) = \frac{1}{2}\left[ U(x + 1) + U(x - 1) \right] \]
All three measures are localization measures.

Other quantities are needed to measure the width and asymmetry of the PDF, etc.
**Functions of a random variable:**

The function \( Y = g(X) \) is a random variable that is a mapping from some event \( A \) to a number \( Y \) according to:

\[
Y(A) = g[X(A)]
\]

**Theorem.** If \( Y = g(X) \), then the PDF of \( Y \) is

\[
f_Y(y) = \sum_{j=1}^{n} \frac{f_X(x_j)}{|dg(x)/dx|_{x=x_j}},
\]

where \( x_j, j = 1, n \) are the solutions of \( x = g^{-1}(y) \). Note the normalization property is conserved (unit area).

This is one of the most important equations! ♻

**Example**

\[
Y = g(X) = aX + b
\]

\[
\frac{dg}{dx} = a
\]

\[
g^{-1}(y) = x_1 = \frac{y - b}{a}
\]

\[
f_Y(y) = \frac{f_X(x_1)}{|dg(x)/dx|_{x=x_1}} = a^{-1}f_X\left(\frac{y - b}{a}\right).
\]

To check: show that \( \int_{-\infty}^{\infty} dy \ f_Y(y) = 1 \)
Example Suppose we want to transform from a uniform distribution to an exponential distribution:

We want \( f_Y(y) = \exp(-y) \). A typical random number generator gives \( f_X(x) \) with

\[
    f_X(x) = \begin{cases} 
    1, & 0 \leq x < 1; \\
    0, & \text{otherwise}. 
    \end{cases}
\]

Choose \( y = g(x) = -\ln(x) \). Then:

\[
    \frac{dg}{dx} = -\frac{1}{x}
\]

\[
    x_1 = g^{-1}(y) = e^{-y}
\]

\[
    f_Y(y) = \frac{f_X[\exp(-y)]}{| -1/x_1 |} = x_1 = e^{-y}.
\]

Factoid: Poission events in time have spacings that are exponentially distributed.
Moments

We will always use angular brackets $\langle \rangle$ to denote average over an ensemble (integrating over an ensemble); time averages and other sample averages will be denoted differently.

Expected value of a random variable with respect to the PDF of $x$:

$$E(X) \equiv \langle X \rangle = \int dx \ x f_X(x)$$

Arbitrary power:

$$\langle X^n \rangle = \int dx \ x^n f_X(x)$$

Variance:

$$\sigma_x^2 = \langle X^2 \rangle - \langle X \rangle^2$$

Function of a random variable: If $y = g(x)$ and $\langle Y \rangle \equiv \int dy \ y f_Y(y)$ then it is easy to show that $\langle Y \rangle = \int dx \ g(x) f_X(x)$.

Proof:

$$\langle Y \rangle \equiv \int dy \ f_Y(y) = \int dy \sum_{j=1}^n \frac{f_X[x_j(y)]}{|dg[x_j(y)]/dx|}$$

A change of variable: $dy = \frac{dg}{dx} \ dx$ yields the result.

Central Moments:

$$\mu_n = \langle (X - \langle X \rangle)^n \rangle$$
Moment Tests:

Moments are useful for testing hypotheses such as whether a given PDF is consistent with data:

E.g. Consistency with Gaussian PDF:

\[
kurtosis \quad k = \frac{\mu_4}{\mu_2^{3/2}} - 3 = 0
\]

\[
skewness \quad \gamma = \frac{\mu_3}{\mu_2^{3/2}} = 0
\]

\[k > 0 \Rightarrow 4\text{th moment proportionately larger} \Rightarrow \text{larger amplitude tail than Gaussian and less probable values near the mean.}\]

\[\alpha_3 > 0 = \text{positively skewed} \rightarrow\]

\[\alpha_3 > 0 = \text{negatively skewed} \rightarrow\]

\[\alpha_3 = 0 = \text{symmetric} \rightarrow\]

\[\alpha_3 > 3 = \text{leptokurtic = highly-peeked} \rightarrow\]

\[\alpha_3 > 3 = \text{platykurtic = flat-topped} \rightarrow\]

Figure 5.3 Single peak distributions with different coefficients of skewness and kurtosis.
Uses of Moments:

Often one wants to infer the underlying PDF of an observable, e.g. perhaps because determination of the PDF is tantamount to understanding the underlying physics of some process.

Two approaches are:

1. construct a histogram and compare the shape with a theoretical shape.
2. determine some of the moments (usually low-order) and compare.

Suppose the data are \( \{x_j, j = 1, N\} \)

1. One could form bins of size \( \Delta x \) and count how many \( x_j \) fall into each bin. If \( N \) is large enough so that \( n_k = \# \) points in the \( k \)-th bin is also large, then a reasonably good estimate of the PDF can be made. (But beware of dependence of results on choice of binning.)

2. However, often times \( N \) is too small or one would like to determine only basic information about the shape of the distribution (is it symmetric?), or determine the mean and variance of the PDF or test whether the data are consistent with a given PDF (hypothesis testing).
Some of the typical situations are:

i) assume the data were drawn from a Gaussian parent PDF; estimate the mean and \( \sigma \) of the Gaussian [parameter estimation]

ii) test whether the data are consistent with a Gaussian PDF [moment test]

note that if the r.v. is zero mean then the PDF is determined solely by one parameter: \( \sigma \)

\[
f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}
\]

The moments are

\[
\langle x^n \rangle = \begin{cases} 
1 \cdot 3 \cdot \ldots \cdot (n-1) \sigma^n \equiv (n-1)!! \sigma^n & \text{n even} \\
0 & \text{n odd}
\end{cases}
\]

Therefore, the \( n = 2 \) moment = 1st non-zero moment \( \Rightarrow \) all other moments.

This statement remains for more multi-dimensional Gaussian processes:

Any moment of order higher than 3 is redundant ... or can be used as a test for gaussianity.
Characteristic Function:

Of considerable use is the characteristic function

\[ \Phi_X(\omega) \equiv \langle e^{i\omega X} \rangle \equiv \int dx \ f_X(x) e^{i\omega x}. \]

If we know \( \Phi_X(\omega) \) then we know all there is to know about the PDF because

\[ f_X(x) = \frac{1}{2\pi} \int d\omega \ \Phi_X(\omega) \ e^{-i\omega x} \]

is the inversion formula.

If we know all the moments of \( f_X(x) \), then we also can completely characterize \( f_X(x) \). Similarly, the characteristic function is a moment-generating function:

\[ \Phi_X(\omega) = \langle e^{i\omega X} \rangle \equiv \left( \sum_{n=0}^{\infty} \frac{(i\omega X)^n}{n!} \right) = \sum_{n=0}^{\infty} \frac{(i\omega)^n}{n!} \langle X^n \rangle \]

because the expectation of the sum = sum of the expectations.

By taking derivatives we can show that

\[
\left. \frac{\partial \Phi}{\partial \omega} \right|_{\omega=0} = i \langle X \rangle \\
\left. \frac{\partial^2 \Phi}{\partial \omega^2} \right|_{\omega=0} = i^2 \langle X^2 \rangle \\
\left. \frac{\partial^k \Phi}{\partial \omega^k} \right|_{\omega=0} = i^k \langle X^k \rangle
\]

or

\[
\langle X^n \rangle = i^{-n} \left. \frac{\partial^n \Phi}{\partial \omega^n} \right|_{\omega=0} = (-i)^n \left. \frac{\partial^n \Phi}{\partial \omega^n} \right|_{\omega=0} \text{ Price's theorem}
\]

Characteristic functions are useful for deriving PDFs of combinations of r.v.'s as well as for deriving particular moments.
Joint Random Variables

Let $X$ and $Y$ be two random variables with their associated sample spaces. The actual events associated with $X$ and $Y$ may or may not be independent (e.g. throwing a die may map into $X$; choosing colored marbles from a hat may map into $Y$). The relationship of the events will be described by the joint distribution function of $X$ and $Y$:

$$F_{XY}(x, y) \equiv P\{X \leq x, Y \leq y\}$$

and the joint probability density function is

$$f_{XY}(x, y) \equiv \frac{\partial^2 F_{xy}(x, y)}{\partial x \partial y} \quad \text{(a two dimensional PDF)}$$

Note that the one dimensional PDF of $X$, for example, is obtained by integrating the joint PDF over all $y$:

$$f_X(x) = \int dy \ f_{XY}(x, y)$$

which corresponds to asking what the PFF of $X$ is given that the certain event for $Y$ occurs.

Example: flip two coins $a$ and $b$. Let heads = 1; tails = 0. Define 2 r.v.'s: $X = a + b; Y = a$. With these definitions $X + Y$ are statistically dependent.

Characteristic function of joint r.v.'s:

$$\Phi_{XY}(\omega_1, \omega_2) = \langle e^{i(\omega_1 X + \omega_2 Y)} \rangle = \iint dx \ dy \ e^{i(\omega_1 x + \omega_2 y)} f_{XY}(x, y).$$

For $x, y$ independent

$$\Phi_{XY}(\omega_1, \omega_2) = \left[ \int dx \ f_X(x) \ e^{i\omega_1 x} \right] \left[ \int dy \ f_Y(y) \ e^{i\omega_2 y} \right] = \Phi_X(\omega_1) \Phi_Y(\omega_2).$$

Example for independent r.v.'s: flip two coins $a$ and $b$. As before, heads = 1 and tails = 0, let $x = a, y = b$ ($x$ and $y$ are independent).
Independent random variables

Two random variables are said to be independent if the events mapping into one r.v. are independent of those mapping into the other.

In this case, joint probabilities are factorable so that

\[ F_{XY}(x, y) = F_X(x) F_Y(y) \]
\[ f_{XY}(x, y) = f_X(x) f_Y(y). \]

Such factorization is plausible if one considers moments of independent r.v.’s:

\[ \langle X^n Y^m \rangle = \langle X^n \rangle \langle Y^m \rangle \]

which follows from

\[ \langle X^n Y^m \rangle \equiv \iint dx\, dy\, x^n y^m \, f_{XY}(x, y) = \left[ \int dx\, x^n f_X(x) \right] \left[ \int dy\, y^m f_Y(y) \right]. \]
Convolution theorem for sums of independent RVs

If $Z = X + Y$ where $X, Y$ are independent random variables, then the PDF of $Z$ is the convolution of the PDFs of $X$ and $Y$:

$$f_Z(z) = f_X(x) * f_Y(y) = \int dx \ f_X(x) \ f_Y(z-x) = \int dx \ f_X(z-x) \ f_Y(x).$$

**proof:** By definition,

$$f_Z(z) = \frac{d}{dz} F_Z(z)$$

Consider

$$F_z(z) = P\{Z \leq z\}$$

Now, as before, this is

$$F_z(z) = P\{X + Y \leq z\} = P\{Y \leq z - X\}.$$

To evaluate this, first evaluate the probability $P\{Y \leq z - x\}$ where $x$ is just a number.

Now

$$P\{Y \leq z - x\} \equiv F_Y(z - x) \equiv \int_{-\infty}^{z-x} dy \ f_Y(y)$$

but $P\{Y \leq z - X\}$ is the probability that $Y \leq z - x$ for all values of $x$ so we need to integrate over $x$ and weight by the probability of $x$:

$$P\{Y \leq z - X\} = \int_{-\infty}^{\infty} dx \ f_X(x) \ \int_{-\infty}^{z-x} dy \ F_Y(y)$$

that is, $P\{Y \leq z - X\}$ is the expected value of $F_Y(z - x)$. By the Leibniz integration formula

$$\frac{d}{db} \int_a^{g(b)} d\omega \ h(\omega) \equiv h(g(b)) \frac{dg(b)}{db}$$

we obtain the convolution results.
Characteristic function of $Z = X + Y$

For $X, Y$ independent we have

$$f_Z = f_X * f_Y \quad \Rightarrow \quad \Phi_Z(\omega) = (e^{i\omega z}) = \Phi_X(\omega) \Phi_Y(\omega)$$

Variance of $Z$: if variance of $X$ and $Y$ are $\sigma_X^2, \sigma_Y^2$, then variance of $Z$ is $\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$.

Assume $X$ and $Y$ and hence $Z$ are zero mean r.v.’s, then we have

$$\sigma_X^2 = \langle x^2 \rangle \quad = \quad i^{-2} \frac{\partial^2 \phi_x}{\partial \omega^2}(\omega = 0) \quad = \quad -\frac{\partial^2 \phi_x}{\partial \omega^2}(\omega = 0)$$

$$\sigma_Y^2 = \langle y^2 \rangle \quad = \quad -\frac{\partial^2 \phi_y}{\partial \omega^2}(\omega = 0)$$

Using Price’s theorem:

$$\sigma_Z^2 = \langle Z^2 \rangle \quad = \quad -\frac{\partial^2 \phi_z}{\partial \omega^2}(\omega = 0)$$

$$= \quad -\frac{\partial^2}{\partial \omega^2} \left[ \phi_X(\omega) \phi_Y(\omega) \right]_{\omega=0}$$

$$= \quad -\frac{\partial}{\partial \omega} \left[ \phi_X \frac{\partial \phi_Y}{\partial \omega} + \phi_Y \frac{\partial \phi_X}{\partial \omega} \right]_{\omega=0}$$

$$= \quad -\left[ \phi_X \frac{\partial^2 \phi_Y}{\partial \omega^2} + \phi_Y \frac{\partial^2 \phi_x}{\partial \omega^2} + 2 \frac{\partial \phi_X}{\partial \omega} \cdot \frac{\partial \phi_Y}{\partial \omega} \right]_{\omega=0}.$$ 

We have ”discovered” that variances add (independent variables only):

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2.$$
Multivariate random variables:  $N$ dimensional

The results for the bivariate case are easily extrapolated. If

$$Z = X_1 + X_2 + \ldots + X_N = \sum_{j=1}^{N} X_j$$

where the $X_j$ are all independent r.v.'s, then

$$f_Z(z) = f_{X_1} * f_{X_2} * \ldots * f_{X_N}$$

and

$$\Phi_Z = \prod_{j=1}^{N} \Phi_{X_j}(\omega)$$

and

$$\sigma^2_Z \equiv \sum_{j=1}^{N} \sigma^2_{X_j}.$$
Central Limit Theorem:

Let

\[ Z_N = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} X_j \]

where the \( X_j \) are independent r.v.’s with means and variances

\[ \mu_j \equiv \langle X_j \rangle \]

\[ \sigma^2_j = \langle X_j^2 \rangle - \langle X_j \rangle^2. \]

and the PDFs of the \( X_j \)’s are almost arbitrary. Restrictions on the distributions of each \( X_j \) are that

i) \( \sigma^2_j > m > 0 \quad m = \text{constant} \)

ii) \( \langle |X|^n \rangle < M = \text{constant for } n > 2 \)

In the limit \( N \to \infty \), \( Z_N \) becomes a Gaussian random variable with mean

\[ \langle Z_N \rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \mu_j \]

and variance

\[ \sigma^2_Z = \frac{1}{N} \sum_{j=1}^{N} \sigma^2_j. \]

Example: suppose the \( X_j \) are all uniformly distributed between \( \pm \frac{1}{2} \), so

\[ f_X(x) = \Pi(x) \Leftrightarrow \frac{\sin \pi f}{\pi f} = \frac{\sin \frac{\omega}{2}}{\omega/2} \]
Thus the characteristic function is

\[ \Phi_j(\omega) = \langle e^{i\omega x} \rangle = \frac{\sin \omega/2}{\omega/2} \]

Graphically:

<table>
<thead>
<tr>
<th>( N = 2 )</th>
<th>( N = 3 )</th>
<th>( N = \infty )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e^{-x^2} )</td>
<td>( (\sin \omega/2)^2 )</td>
<td>( (\sin \omega/2)^3 )</td>
</tr>
</tbody>
</table>

From the convolution results we have

\[ \phi_{\sqrt{N}Z_N}(\omega) = \left(\frac{\sin \omega/2}{\omega/2}\right)^N \]

From the transformation of random variables we have that

\[ f_{Z_N}(x) = \sqrt{N} f_{\sqrt{N}Z_N}(\sqrt{N}x) \]
and by the scaling theorem for Fourier transforms

$$\phi_{Z_N}(\omega) = \phi_{\sqrt{N}Z_N} \left( \frac{\omega}{\sqrt{N}} \right) = \left( \frac{\sin(\omega / 2\sqrt{N})}{\omega / 2\sqrt{N}} \right)^N.$$ 

Now

$$\lim_{N \to \infty} \phi_{Z_N}(\omega) = e^{-\frac{1}{2}\omega^2 \sigma_Z^2}$$

or

$$f_{Z_N}(x) = \frac{1}{\sqrt{2\pi \sigma_Z^2}} e^{-x^2 / 2\sigma_Z^2}.$$  

Consistency with this limiting form can be seen by expanding $\phi_{Z_N}$ for small $\omega$

$$\phi_{Z_N}(\omega) \approx \left( \frac{\omega/2\sqrt{N} - \frac{1}{12}(\omega/2\sqrt{N})^3}{\omega/2\sqrt{N}} \right)^N \approx 1 - \frac{\omega^2}{24}$$

that is identical to the expansion of $\exp (-\omega^2 \sigma_Z^2 / 2)$. 

21
CLT: Example of a PDF that does not work

The Cauchy distribution and its characteristic function are

\[ f_X(x) = \frac{\alpha}{\pi \alpha^2 + x^2} \]
\[ \Phi(\omega) = e^{-\alpha |\omega|} \]

Now

\[ Z_N = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} x_j \]

has a characteristic function

\[ \Phi_N(\omega) = e^{-N\alpha |\omega|/\sqrt{N}} \]

By inspection the exponential will not converge to a Gaussian. Instead, the sum of \( N \) Cauchy RVs is a Cauchy RV.

Is the Cauchy distribution a legitimate PDF? No!

The variance diverges:

\[ \langle X^2 \rangle = \int_{-\infty}^{\infty} dx \frac{1}{\pi \alpha^2 + x^2} \rightarrow \infty. \]