A6523
Signal Modeling, Statistical Inference and Data Mining in Astrophysics
Spring 2011

Reading
• Chapter 5 of Gregory (Frequentist Statistical Inference)

Lecture 7
• Examples of FT applications
  – Simulating processes with power-law spectra
  – wave propagation through screen
  – dechirping filter (dedispersion)
• Probability and Stochastic Processes
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

\[
X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} \, dt \tag{1}
\]

\[
x(t) = \int_{-\infty}^{\infty} X(f) e^{j2\pi ft} \, df \tag{2}
\]

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

\[
X(k) = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N} \tag{3}
\]

\[
x(k) = \sum_{n=0}^{N-1} X(n) e^{j2\pi nk/N} \tag{4}
\]

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

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

(a) $s(t)$

(b) $w(t)$

(c) $s(t) - w(t)$

(d) $c(t)$

(e) $s(t) - w(t) - c(t)$

(f) $\hat{s}(k)$
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.
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.

Partial correlation

Digital filtering

2N samples

2N terms

2T seconds
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 the 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 Hermaticity
• 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

Default solid = pre-fit  Red: solid = residuals  dotted = fit

slope = 5

Default solid = pre-fit  Red: solid = residuals  dotted = fit
slope = 6
Examples II. Fresnel Diffraction Patterns

In particular, let’s consider a 1D problem where a plane wave passes through a ‘screen’ described by $S(x)$.

![Diagram of a plane wave and screen](image)

The wavefield is given by the Kirchoff diffraction integral

$$A(z) = I_0^{1/2} \int dx' K_0(x - x') S(x')$$

where the wave propagator is

$$K_0(x) = r_P^{-1} e^{i\pi(x/r_P)^2}$$

and the Fresnel scale is $r_P = \sqrt{\lambda z}$.

Note this is in the form of a convolution. Thus, the wavefield may be solved using Fourier transforms

$$A(z) = \text{IFT}\{\text{FT}[K_0(x)] \cdot \text{FT}[S(x)]\}.$$ 

For phase changing screens only we have $S(x) \equiv e^{i\phi(x)}$, an excellent model for the turbulent atmosphere, the ionosphere, the solar wind, the interstellar medium, etc. It can also be used to model gravitational bending of light.
Thin Screens with Turbulent Refractive Index Fluctuations

Optical speckle interferogram of Gamma Ori

Object: Gamma Ori (unresolved star)
Telescope: 6 m SAA telescope
Filter: center wavelength 500 nm
bandwidth 15 nm
Exposure time: 20 ms
Field of view: 1.84 arcsec x 1.84 arcsec
Fresnel Function

- The $\text{FT}\{\text{Fresnel function}\} = \mathcal{F}$
Application to Dechirping

• Numerous applications involve “chirped” signals
  • radar swept frequency signals
  • plasma dispersion
• e.g. impulse passed through a swept-frequency system:

\[ s_i(t) = \delta(t) \]

\[ s_i(t) \rightarrow e^{i\alpha t^2} \rightarrow s(t) \]

• Deconvolution is trivial in this case: pass the measured signal \( s(t) \) through the inverse filter:

\[ s(t) \rightarrow e^{-i\alpha t^2} \rightarrow s_i(t) \]

• The procedure applies for an arbitrary-shaped pulse:

\[ S_i(t) = \sum_j a_j \delta(t - t_j) \]

\[ S(t) = \sum_j a_j e^{i\alpha(t-t_j)^2} \]
Basic data unit = swept frequency pulses

$10^6 - 10^8$ samples x 64 $\mu$s

Fast-dump spectrometers:

- Analog filter banks
- Correlators
- FFT (hardware)
- FFT (software)
- Polyphase filter bank
- Alternative: Baseband Sampling to get complex voltage (much faster sample rate)
Dedispersion
Two methods:

**Coherent:**
- operates on the voltage proportional to the electric field accepted by the antenna, feed and receiver
- computationally intensive because it requires sampling at the rate of the total bandwidth
- “exact”

**Post-detection:**
- operates on intensity $= |voltage|^2$
- computationally less demanding
- an approximation
Postdetection Dedispersion:

Sum intensity over frequency after correcting for dispersion delay

\[ \Delta t = \left[ \Delta t_{DM}^2 + \left(1/\Delta \nu \right)^2 \right]^{1/2} \]

\[ = \left[ (a\Delta \nu)^2 + (\Delta \nu)^{-2} \right]^{-1/2} \]

Residual time smearing: minimum smearing time across a channel when

\[ \Delta \nu = \left[ 8.3 \, \mu s \, DM \nu^{-3} \right]^{-1/2} \]
Coherent Dedispersion
pioneered by Tim Hankins (1971)

Dispersion delays in the time domain represent a phase perturbation of the electric field in the Fourier domain:

\[ \tilde{E}_{\text{measured}}(\omega) = \tilde{E}_{\text{emitted}}(\omega)e^{ik(\omega)z} \]

Coherent dedispersion involves multiplication of Fourier amplitudes by the inverse function,

\[ e^{-ik(\omega)z} \]

For the non-uniform ISM, we have

\[ k(\omega)z \rightarrow \int dz k(\omega) \propto \omega^2 DM + \text{constant} \]

which is known to high precision for known pulsars. The algorithm consists of

\[ \text{IFFT}\{\text{FFT}[E_{\text{measured}}(t)] \times e^{-ik(\omega)z}\} \approx E_{\text{emitted}}(t) \]

Application requires very fast sampling to achieve usable bandwidths.
6-Bit Quadrature Digitizer

General Description

The MAX2101 6-bit quadrature digitizer combines quadrature demodulation with analog-to-digital conversion on a single bipolar silicon die. This unique RF-to-Bits™ function bridges the gap between existing RF downconverters and CMOS digital signal processors (DSPs).

The MAX2101’s simple receiver subsystem is designed for digital communications systems such as those used in DBS, TVRO, WLAN, and other applications. The MAX2101 accepts input signals from 400MHz to 700MHz and applies adjustable gain, providing at least 40dB of dynamic range.

Each baseband is filtered by an on-chip, 5th-order Butterworth lowpass filter, or the user can select an external filter path. Baseband sample rate is 60Msp/s. The MAX2101 is available in a commercial temperature range, 100-pin MQFP package.

Applications

- Recovery of PSK and QAM Modulated RF Carriers
- Direct-Broadcast Satellite (DBS) Systems
- Television Receive-Only (TVRO) Systems
- Cable Television (CATV) Systems
- Wireless Local Area Networks (WLANs)

Features

- ADCs Provide Greater than 5.5 Effective Bits at fS = 60Msp/s, fIN = 15MHz
- Fully Integrated Lowpass Filters with Externally Variable Bandwidth (10MHz to 30MHz)
- 40dB Dynamic Range
- Integrated VCO and Quadrature Generation Network for I/Q Demodulation
- Divide-by-16 Prescaler for Oscillator PLL
- Programmable Counter for Variable Sample Rates
- Signal-Detection Function
- Selectable Offset Binary or Twos-Complement Output Data Format
- Automatic Baseband Offset Cancellation

Ordering Information

<table>
<thead>
<tr>
<th>PART</th>
<th>TEMP. RANGE</th>
<th>PIN-PACKAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAX2101CMQ</td>
<td>0°C to +70°C</td>
<td>100 MQFP</td>
</tr>
</tbody>
</table>

Typical Application Circuit

TM RF-to-Bits is a registered trademark of Tektronix, Inc.
Micropulses coherently dedispersed (Hankins 1971)
Extra Slides
Coherent Dedispersion
pioneered by Tim Hankins (1971)

Coherent dedispersion works by explicit deconvolution:

\[ E_{\text{measured}}(t) = E_{\text{emitted}}(t) \ast \text{FFT}\{e^{ik(\omega)z}\} \]

\[ \Rightarrow E_{\text{emitted}}(t) \approx E_{\text{measured}}(t) \ast \text{FFT}\{e^{-ik(\omega)z}\} \]

Comments and Caveats:

• Software implementation with FFTs to accomplish deconvolution (Hankins 1971)

• Hardware implementations: real-time FIR filters (e.g. Backer et al. 1990s-present)

• Resulting time resolution: \( 1 / (\text{total bandwidth}) \)

• Requires sampling at Nyquist rate of 2 samples \( \times \) bandwidth

  \[ \Rightarrow \text{Computationally demanding} \]

• Actual time resolution often determined by interstellar scattering (multipath)

• Most useful for low-DM pulsars and/or high-frequency observations
Nanostructure in Crab pulsar giant pulses
from T. Hankins 2005
Crab Pulsar
2-ns resolution

Total intensity (Jy)

Time (microseconds)

Polarized flux (Jy)

LCP a.  b.  c.  d.  e.  f.

RCP 0  0.1
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 $N \rightarrow \infty$:

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

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 $\zeta_1 + \zeta_2$ (≡ 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

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

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 (ie. 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.
\[ P(\psi | \sigma) = \frac{P(\psi \sigma)}{P(\sigma)} \]

No overlap \( \Rightarrow P(\psi | \sigma) = 0 \) \& \( P(\psi \sigma) = 0 \)

Complete overlap: \[ P(\psi | \sigma) = \frac{P(\psi \sigma)}{P(\sigma)} = \frac{P(\sigma)}{P(\sigma)} \equiv 1 \]
Binomial Distribution and Repeated Trials

Suppose an experiment has only 2 outcomes (e.g. coin flipping) with probabilities \( p \) and \( q = 1 - p \), respectively, corresponding to heads (H) and tails (T).

Now suppose the experiment is repeated \( n \) times with heads occurring \( k \) times and tails occurring \( n - k \) times. Then the probability of obtaining that particular sequence of heads and tails is

\[
p^k(1-p)^{n-k}.
\]

Consider now that we want to know the probability of obtaining \( k \) heads and \( n - k \) tails in any order.

Now the number of ways that we can choose \( k \) heads out of \( n \) trials is

\[
\binom{n}{k} = \frac{n!}{(n-k)!k!}
\]

\[
P(k|n,p) = \binom{n}{k} p^k(1-p)^{n-k} \quad \text{Binomial Distribution}
\]

e.g. for coin flipping with fair coin:

\[
p = q = \frac{1}{2} \Rightarrow P(k) = \binom{n}{k} 2^{-n}
\]

Expect most probable value to be for \( k = n/2 \), therefore

\[
P\left(\frac{n}{2}\right) = \binom{2k}{k} 2^{-2k} = \frac{(2k)!}{k!k!} 2^{-2k}
\]

Repeated trials

- Gaussian [large numbers]
- binomial
- Poisson [small numbers]
E.g. throw a point onto time line between 0 and T

i) Assume uniform probability of point in [0,T].

ii) Assume more than one point can occupy same subinterval (independence).

The two events are: 1. point in \( \Delta t \); 2. point not in \( \Delta t \).

one trial: \( P\{ \text{point in } \Delta t \} = p = \Delta t/T \)

\( n \) trials: \( P\{k \text{ points in } \Delta t \} = \binom{n}{k} p^k(1-p)^{n-k} \)

Poisson limit: \( n \to \infty \) \( p \to 0 \) \( \lambda \equiv np \neq 0 \) and not large, e.g. expected number in the bin \( \Delta t \) is not large.

Then

\[
P\{k\} = \frac{\lambda^k}{k!} e^{-\lambda}
\]

\[
\langle k \rangle = \sum_k k P\{k\}
\]

\[
\langle k^2 \rangle = \sum_k k^2 P\{k\}
\]

\[
\lambda = \langle k \rangle = \langle k^2 \rangle - \langle k \rangle^2
\] (1)

Note that the mean = variance for a Poission random variable.

Gaussian limit: \( n \to \infty \) \( \mu = np = \text{constant} \gg 1 \). Then

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

\[
\mu = \langle x \rangle = np
\] (3)

\[
\sigma^2 = np(1-p).
\] (4)
Example:

The Poisson distribution describes photon counting and other processes for which the occurrence of a “count” or an event is independent of any other occurrence.

Photons gathered in bins of size $\Delta t$.

Let $\lambda = $ photon rate (change in notation).

Then $\langle k \rangle = \lambda \Delta t$

If the total number of photons is large [all bins] but

a) the number per bin is small, then the number per bin is described by the Poisson p.d.f.

b) the number per bin is large, then the number per bin follows a Gaussian p.d.f.

Other examples:

e.g. collisions between gas molecules, collisions giving rise to Brownian motion, a bad volleyball game, etc.

e.g. Shot noise
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.

The set of all events is the 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 ζ 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 ζ does not influence the occurrence of ψ], then the probability of the event $\zeta + \psi = \text{event that ζ or ψ occurs}$ is $P(\zeta + \psi) = P(\zeta) + P(\psi)$

(+ means ‘or’)

From the axioms, one can construct such results as:

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

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

3. $P(A + B) = P(A) + P(B) - P(\bar{A}B)$ where $P(AB) = \text{probability that both A and B occur. } P(AB) \geq 0$, 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 $c = a + b \quad \Rightarrow \text{“or” (same as } a \cup b)$

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

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

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

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 $P(HT) = P\{\text{heads & tails}\} = 0$ 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. \( \int_{-\infty}^{\infty} dx \, f_X(x) = F_X(\infty) - F_X(-\infty) = 1 - 0 = 1 \)
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

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

$$F_X(x) = \frac{1}{2} [U(x + 1) + U(x - 1)]$$

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**

$$Y = g(X) = aX + b$$

$$\frac{dg}{dx} = a$$

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

$$f_Y(y) = \frac{f_X(x_1)}{|dg(x_1)/dx|} = a^{-1}f_X\left(\frac{y - b}{a}\right).$$
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 ant \( f_Y(y) = \exp(-y) \). 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:

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

**Moments**

Note: we will always use brackets to \( \Rightarrow \) average w.r.t. pdf or over an ensemble; time averages and other sample averages will be denoted differently.

Expected value of a random variable:

\[
  E(X) \Rightarrow \langle X \rangle = \int dx \, x f_X(x)
\]

\( \langle \cdot \rangle \) denotes expectation w.r.t. the pdf of \( x \)

**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 \( \langle Y \rangle \equiv \int dy \, f_Y(y) \) then it is easy to show that \( \langle Y \rangle = \int dx \, g(x) f_X(x) \).

Proof:

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