UCLA STAT 19
Order & Organization in the Stochastic Universe
A *Fiat Lux* Course

•**Instructor:** **Ivo Dinov,**

Asst. Prof. In Statistics and Neurology

University of California, Los Angeles, Fall 2004

http://www.stat.ucla.edu/~dinov/courses_students.html

Course Organization

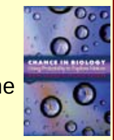
Software: **SOCR resource,**

http://socr.stat.ucla.edu

Texts: *A New Kind of Science* by Stephen Wolfram (2002)

Chance in Biology: Using Probability to Explore Nature

by Mark Denny and Steven Gaines (2002)



Course Description, Class homepage, online supplements, VOH's, etc.

http://www.stat.ucla.edu/~dinov/courses_students.html

Topics

- The Nature of Chance
- Determinism versus Chance
- Chaos vs. Order
- Experiments, Observations & Distributions
- Types & Causes of Variation, Entropy
- Discrete & Continuous Patterns of Disorder
- The Normal Distribution
- Central Limit Theory
- Randomness in Biology, Genetics, Eng. & Physics
- Random Walks (may be Dynamical Systems, Fractals)

Topics (cont.)

- Duality Principles: The Uncertainty Principle (momentum vs. position)
- Balancing Quality and Volume of Information
- Statistical vs. Practical Significance
- Statistics of Extremes
- Intra- vs. Extra-polation
- Noise and Perception
- Bayesian Theory

The Nature of Chance

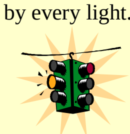
- ***The Nature of Chance*** extends far beyond the random events that shape human existence. Chance is ubiquitous, and its role in Life and the Universe is the subject of this Course.
- ***Examples:*** mountain stream sound, arrival times, MPG mileage, wind patterns (vorticity), (thermal molecular) diffusion/movements/walks, material strength, the drift of genes in a population, longevity of phytoplankton, etc.
- Is **chance** in life **unavoidable**?

The Nature of Chance

- **(Classical) Traditional approach** is to *view chance as a necessary evil* that can be tamed via application of clever techniques (filtering) or inferential statistics.
- **Alternatively**, if chance is a given in life, why not *use it to our advantage*? In other words, if we know that a system will behave in a random (***chaotic***) fashion in the short term and at small scale (as with the random thermal motions of protein chains in spider silk), we can use this information to make accurate predictions as to how the system will behave in the long run and on a larger scale.

The Nature of Chance

- ***Probability theory*** was originally devised to predict the outcomes in games of chance, but its utility has been extended far beyond games.
- ***Life*** itself is a chancy proposition, a fact apparent in our daily lives. Some days you are lucky (every stoplight turns green as you approach) and other days, just by chance, you are stopped by every light.



Determinism versus Chance

- **Deterministic Processes**: One of Sir Isaac Newton's grand legacies is the idea that much about how the Universe works can be precisely described. *I.e. given sufficient knowledge of the initial state of a system, its future can be determined exactly.*
- **Examples**: If we know the exact masses of the Moon and Earth and their current speed relative to each other, **Newtonian mechanics** and the law of gravitation should be able to tell us the exact position of the Moon relative to Earth at any future time (e.g., predict solar and lunar eclipses, comet arrivals, etc.).
- **Euclidean Geometry**: Given 2 sides of a right-triangle we can determine exactly the length of the 3rd size.

Determinism versus Chance

- Good examples of **real-world** deterministic processes are difficult to find. Many of the processes that seem simple when described abstractly are exceedingly complex in reality. Details inevitably intrude, bringing with them an **element of unpredictability**.
- **Approximations**: In some cases, the amount of variability associated with a process is sufficiently small that we are willing to view the system as being *approximately deterministic*, and accept as fact predictions regarding its behavior.
- The physics of a **pendulum clock**, for instance, is so straightforward that we are content to use these machines as an accurate means of measuring time.

Determinism versus Chance

- In biology, few systems are so reliable, and deterministic behavior can be viewed at best as a wishful-thinking.
- **Fidelity, Imprecision, Repetition and Uncertainty** is present in virtually all biological processes – read the online article by Miroslav Radman in Nature 413 (Class notes online).
- Perfection of organisms and the accuracy of biological processes are still used in religious explanations of the origin of life. However, **in real life it is survival, not fidelity, that is the ultimate virtue.**

Determinism versus Chance

- Because adaptability involves exploration of genetic possibilities to fit ecological *niches*, **molecular infidelity and repetition are more likely to succeed than a precise, non-repetitive processes**.
- Only a tiny fraction of antibodies produced will ever be useful; the rest can be considered as mistakes.
- At least **half of all human embryos fail** during development.
- During chromosome segregation from a mother cell into two daughters, the polymerizing fibers (microtubules) do not know the exact location of the chromosomal target (the centromere) — they **shoot and miss until one hits**.

Determinism versus Chance

- A precise, **single shot would often miss a target** of uncertain position, whereas **successive, imprecise firing will eventually lead to a hit**.
- Selection at the level of molecules, cells and organisms may give the impression of designed perfection, but **life's structures do not emerge by a fully deterministic design**.
- Errors, infidelity & wastefulness, can cause individual failure, but also provide innovation and robustness, ensuring the perpetuation of life. Nature does not exhaust itself for the sake of fidelity and perfectionism. Rather, errors are made, often repaired or discarded, but always tested as the **source of blind innovation** during the continuous adaptation to unpredictable environmental changes and challenges.

Efficiency vs. Precision

- What about the **fidelity of enzymatic reactions**? In the very precise process of DNA replication, accuracy is achieved by using a ***proofreading*** system to remove erroneously inserted nucleotides, and then by ***quality-checking*** the synthesized DNA using a mismatch-repair system that removes virtually all remaining mistakes.
- ***It would take too long to get it exactly right in the first place***. DNA replication is efficient and therefore relatively imprecise, leaving mistakes to error-correction enzymes which are themselves efficient because their substrates are specific mistakes made by other enzymes.

Determinism versus Chance

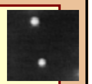
- If a system or process is not deterministic, it is by definition **stochastic**. Even if we know exactly the state of a stochastic system at one time, we can never predict exactly what its state will be in the future.
- **Stochasticity** can manifest itself to a variable degree (**wind speed could easily double** in 1 second in a turbulent environment, whereas **random washer-thickness variations** in a precise engineering design could often be within 10^{-10} m).

Determinism versus Chance

- Many stochastic processes are approximately predictable with just a minor overlay of random behavior.
- **Example**: The light intensity reflected from a mountain stream is chaotic. Yes, there are minor random short-term light fluctuations. However, if we were to take 5-minute averages of the light level (long-time picture exposure) this intensity could be predicted fairly accurately. Note the word **average** above, this is a key concept we'll discuss later in stochastic process modeling.
- In other cases, the predictability of a system is negligible, and chance alone governs its behavior (the movement of molecules in a gas at room-temperature).

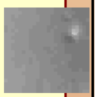
Determinism versus Chance

- Minor random fluctuations.



Demo of 2 micrometer diameter particles in pure water. As can be seen, each particle is constantly moving, and its motion is uncorrelated with the other particles.

- Chance alone governs its behavior



A bead-labeled RecBCD molecule translocating along a single DNA molecule. Initially, the free diffusion of beads in solution is seen. After a few seconds, one of the bead-labeled enzyme molecules attaches to the end of a DNA molecule at the center of the field (arrows); attachment is detected as the cessation of free diffusion and the commencement of characteristic tethered-particle Brownian motion in the vicinity of a single point on the microscope slide. Subsequent translocation of the enzyme along the DNA molecule is visualized as a gradual decrease in the spatial range of the Brownian motion; this decrease continues until the beads ceases visible movement altogether. The video is real time; the frame size is 6.5 μm wide by 6.6 μm tall.

Determinism versus Chance

- Practically, the dividing line between deterministic & stochastic is open to interpretation.
- **Example**, usually (both in games and texts on probability theory) we accept the P(Coin = Head) as a chance proposition, a stochastic process. But if you know enough about the height above the ground at which the coin is flipped, the angular velocity initially imparted to the coin, and the effects of air resistance, it may be possible to decide in advance whether the coin will land heads up.
- Indeed, much of what we accept as stochastic may well be deterministic given sufficient understanding of the experiment design.

Determinism versus Chance

- So, the line between **deterministic** & **stochastic** is often drawn as a matter of convenience.
- If the precise predictions that are possible in theory are too difficult to carry out in practice, we shift the line a bit and think of the process as being stochastic.
- This is not to imply that all processes are deterministic, however. As far as physicists have been able to divine, there are aspects of nature, encountered at very small scales of time and space, that are unpredictable even in theory (will talk about the **Uncertainty Principle**, Heisenberg's inequality, later).

Determinism versus Chance

- There are limits to the precision with which you can know both the velocity and the location of an object.
- If you could know exactly where an electron is at some point in time, you couldn't know what its velocity is.
- Conversely, if you know exactly what its velocity is, you can't know exactly its position.
- This is the strange realm of quantum/statistical mechanics, where chance reigns and human intuition is of little use.

Determinism versus Chance

- Returning to Einstein's nagging doubts about quantum mechanics, Nobel laureate Gerard 't Hooft of Utrecht University (NL) has begun to outline a way in which its apparent play of chance might be underpinned by precise physical laws that describe the way the world works.
- The physicists' most fundamental theory of the properties of matter and energy, quantum mechanics holds that there are things we just cannot know. For example, it forbids us from knowing everything about a subatomic particle: its exact speed, position, mass and energy.

Nature, Jan. 08, 2003, by Philip Ball

Determinism versus Chance

- Einstein did not like this idea, and suspected that another theory - another layer of reality - might underlie quantum mechanics, in which everything is spelled out precisely. These deeper properties of objects became known as *hidden variables*. According to this view, our ignorance about the nature of a quantum object is illusory; we just haven't found the right theory to describe it yet.
- Today, most physicists adhere to a different reading of quantum theory, called the Copenhagen Interpretation, as advocated by the Danish nuclear physicist of the 1940s, Niels Bohr.

Determinism versus Chance

- This says that there is no deeper reality, that hidden variables don't exist and that the world is simply probabilistic. It holds that we are not ignorant about quantum objects, it's just that there is nothing further to be known.
- Indeed, in the 1980s, the Copenhagen Interpretation was put to an experimental test based on a theorem devised by the Irish physicist John Bell - and it stood up. *Hidden variables* had to go.
- The key is information loss. At the smallest conceivable size scale - the Planck (**$h=6.6\times10^{-34}\ kg \cdot m^2 \cdot s^{-1}$**) Scale, many trillions of times smaller than the nucleus of an atom - there exists complete information about the world. This information gets lost very quickly, however.

Example – Stochastic Gene Expression?

- ***Stochastic Gene Expression in a Single Cell* by M Elowitz, A Levine, E Siggia, P Swain, Science, 297(5584), 2002, pp. 1183-1186**
- Living cells possess very **low copy numbers** of many components, including DNA and important regulatory molecules. Thus, **stochastic** effects in gene expression may account for the **large amounts of cell-cell variation** observed in isogenic populations. Such effects can play crucial roles in biological processes, such as development, by establishing initial asymmetries that, amplified by feedback mechanisms, determine cell fates.

Example – Stochastic Gene Expression?

- ***Stochastic Gene Expression in a Single Cell* by M Elowitz, A Levine, E Siggia, P Swain, Science, 297(5584), 2002, pp. 1183-1186**
- For any particular gene, it remains unknown whether **cell-cell variation** in the abundance of its product is set by **noise in expression of the gene itself** **or** by fluctuations in the amounts of other cellular components.
- The difficulty of experimentally distinguishing between these two possibilities has thus far precluded detection of **intrinsic noise** in living cells. The magnitude of the noise intrinsic to gene expression, and its relative importance compared with other sources of cell-cell variability, are fundamental characteristics of the cell that require measurement.

Example – Stochastic Gene Expression?

- In general, the amount of protein produced by a particular gene varies from cell to cell. The noise (σ/μ) in this distribution is labeled η_{tot} and can be divided into two components. Because expression-rates of each gene is controlled by the concentrations, states, and locations of molecules such as regulatory proteins and polymerases, fluctuations in the amount or activity of these molecules cause corresponding fluctuations in the output of the gene. Therefore, they represent sources of **extrinsic noise** (η_{ext}) that are global to a single cell, but vary from one cell to another.
- On the other hand, consider a population of cells identical not just genetically but also in the concentrations and states of their cellular components.

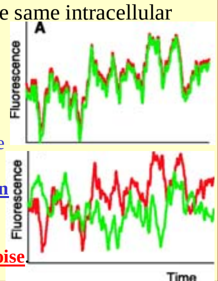
Example – Stochastic Gene Expression?

- Even in such a (hypothetical) population, the rate of expression of a particular gene would still vary from cell to cell because of the random microscopic events that govern which reactions occur and in what order.
- This inherent stochasticity, or **intrinsic noise,** η_{int}, is that remaining part of the total noise arising from the discrete nature of the biochemical process of gene expression.
- No matter how accurately the levels of regulatory proteins are controlled, intrinsic noise fundamentally limits the precision of gene regulation.
  - $\eta_{tot} = \eta_{extr} + \eta_{int}$

Example – Stochastic Gene Expression?

- In practice, intrinsic noise for a given gene may be defined as the extent to which the activities of two identical copies of that gene, in the same intracellular environment, fail to correlate.

- *Escherichia coli* stains incorporate the distinguishable cyan (*cfp*) and yellow (*yfp*) alleles of green fluorescent protein in the chromosome. In the absence of intrinsic noise, the two fluorescent proteins **fluctuate in a correlated fashion** over time in a single cell. Expression of the two genes may become **uncorrelated** In individual cells because of **intrinsic noise**.



Chaos

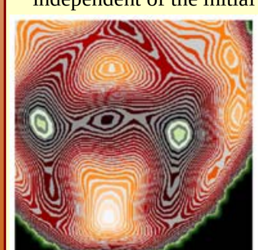
- The deterministic/stochastic dividing line may become even fuzzier.
- Lately, a wide variety of physical systems that should behave deterministically were found in fact to behave unpredictably. These systems are said to exhibit **deterministic chaos**, or just chaos, for short. But if they are deterministic, how can they be unpredictable?
- This apparent conflict is solved by the fact that chaotic systems are very sensitive to the state in which they are started (**initial conditions**).
- **Ex.** A Contractive Fractal Dynamical System is independent of the initial conditions (mriFracDecSeqA1.pdf).

Chaos – Contractive Fractal Systems

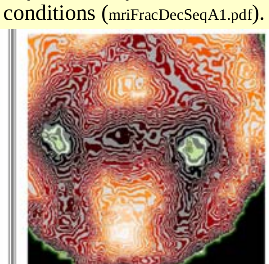
- **Ex.** A Contractive Fractal Dynamical System is independent of the initial conditions (mriFracDecSeqA1.pdf).

Chaos – Contractive Fractal Systems

- **Ex.** A Contractive Fractal Dynamical System is independent of the initial conditions (mriFracDecSeqA1.pdf).



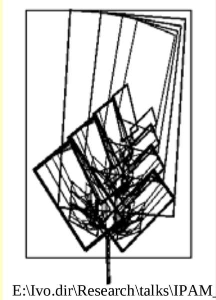
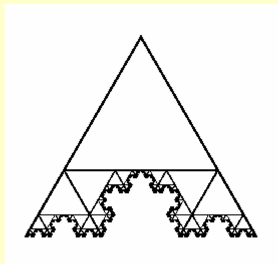
Native Data



Fractal Representation – Fixed Point

Chaos – Contractive Fractal Systems

- **Ex.** Barnsley's Fern & von Koch's curve

E:\Ivo.dir\Research\talks\IPAM_BrainMapping04\images\fernAll.pdf

Chaos

- Example, the **flight path of a baseball** through still air. In this case, if we know the initial speed of the ball (20 m/s), its initial location, and its direction of motion (45º to the horizontal), we may **predict where the ball will land**.
- A **small error in the measurement** of any of these initial conditions (say, the ball is moving at 20.001 rather than 20.000 m/s), the error in predicting the ball's landing is concomitantly small.
- If the motion of a baseball were chaotic, however, its flight would be quite different. Every time a deterministic ball were launched at exactly 20 m/s and an angle of exactly 45º from the center of home plate, it would land in the same spot near second base;

Chaos

- If, however, a chaotic ball were launched with even a slight error in any of these initial conditions, its eventual landing spot would be drastically different. A shift from 20.000 to 20.001 m/s might cause it to land outside the stadium.
- Granted, every time the ball is launched at exactly 20.001 m/s, it ends up in the same place, so the system is still deterministic, but it is extremely sensitive to the initial conditions.
- This is related to the three **types of randomness** which we'll discuss later on.

Chaos

- The primary **disadvantage** of treating a chaotic system as if it were **just stochastic** is a loss of insight. Once a process is stamped with the title **random** – it is easy to stop looking for a mechanistic cause for its behavior.
- It is interesting to note that the motion of the planets, which has long been cited as the **classical example of deterministic mechanics, is in fact chaotic**. Because each planet is subject to a gravitational pull from all other planets and other external forces (e.g., comets!), there is the possibility that at some time in the future the alignment of the solar system may be such that one of the planets could be thrown substantially off its present orbit, and this potential makes it virtually impossible to predict accurately where the planets will be at a given date in the future.

Order vs. Chaos

- A classical rotor, subject to deterministic and random forces, shows a broad spectrum of motion including deterministic, deterministic chaos, random and, erratic behavior.
- One of the main achievements of twentieth-century physics has established that **deterministic and random phenomena complement rather than contradict each other**.

Ordered vs. Chaotic motion

- *Deterministic motion.* If θ is the particle position and t is the time, there are PDE's that describe the motion in terms of the time. If the initial condition $\theta_0 = \theta(t = 0)$ uniquely determines the values of θ for all t, i.e. the motion is fully deterministic.
- *Random motion.* If one adds a random force to the equation, a rotor is capable of overcoming barriers at $\theta = m\pi$, and the motion becomes random (noise-induced instability).
- *Deterministic chaos.* Initial conditions play a major role.

Matter vs. Patterns

- *What is the Universe?*
  - *A pool of particles (stuff)?*
  - *Or a collection of patterns (of particles)?*
- If I ask the question, **Who am I?** I could conclude that, perhaps I am this stuff here, i.e., the ordered and chaotic collection of atoms & molecules (particles) that comprise my body.
- However, the specific set of particles that comprise my body are completely different from the atoms and molecules than comprised me only a few weeks ago.

Matter vs. Patterns

- So, I am a completely different set of stuff than I was a month ago. All that persists is the pattern of organization of that stuff. The pattern changes also, but slowly and in a continuum from my past self.
- From this perspective I am rather like the pattern that water makes in a stream as it rushes past the rocks in its path. The actual molecules (of water) change every millisecond, but the pattern persists for hours or even years. Even atomic structures are rearranged with time.
- It is patterns (e.g., people, ideas, objects, not elementary particles) that persist, and constitute the foundation of what fundamentally exists. The view of the Universe ultimately is a pattern of information.

Matter vs. Patterns

- The information is not just embedded as properties of some other substrate (as in the case of conventional computer memory) but rather **information is the ultimate reality**.
- What we perceive as matter and energy are simply abstractions, i.e., properties of patterns. As a further motivation for this perspective, it is useful to consider that the vast majority of processes underlying human intelligence are based on the recognition of patterns.
- **AI** !?!

Three Types of Randomness (S. Wolfram)

- **Randomness from environment** (continuous random feedback from the ambient environment) – this is like considering the random movement of a boat in the ocean. The randomness of boat movement is apparent from the randomness of sea surface movement at each time point.
- **Randomness from initial condition** – which is like rolling a pair of dice in a controlled environment, where only the random momentum and direction of the force (initial conditions) determine the outcome.
- **Intrinsic Randomness** – which is acclaimed as the only pure random phenomenon. Example – Geiger counter in vacuum.

Types of Randomness

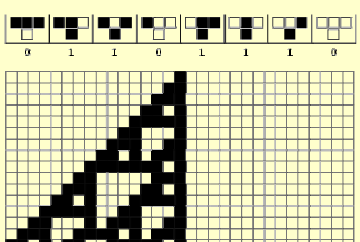
- Wolfram considered that intrinsic randomness could emerge from very simple rules and very simple initial conditions. Still there should be something at the start.
  Something should come from something else & every outcome should have input beforehand.

Wolfram's Rule 110

$110_{(10)} = 01101110_{(2)}$

Wolfram 1983, 2002

Random Outcomes →

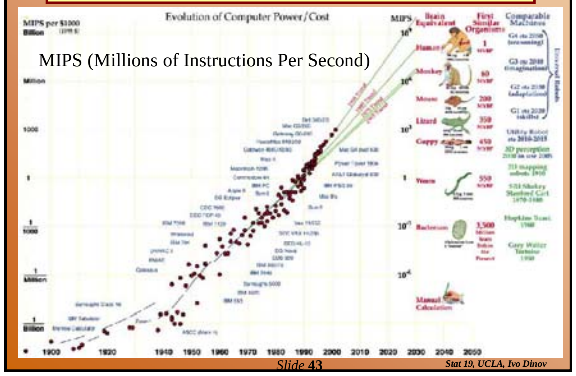


Artificial Intelligence – past century

- 1st Generation Robots – like instinct-ruled reptiles, will handle only those contingencies explicitly covered in their programs. MIPS=**Millions of Instr's per Sec.**
- 2nd Generation – like mouse 300,000 MIPS, which could be available within about 30 years, will be able to adapt, and even be trainable.
- 3rd Generation – a monkeylike 10^6 MIPS will be available within about 40 years, learn quickly from mental rehearsals in simulations modeling physical, cultural, and psychological factors.
- 4th Generation – a humanlike 300-million MIPS, within 50 years, able to abstract and generalize.

Hans Moravec, Comm. of the ACM Volume 46, Number 10 (2003), Pp. 90-97
http://www.acm.org/cacm

Artificial Intelligence – past century



Information & Entropy

- A sequence of N coin tosses (of a fair coin) has 2^N possible outcomes: there is an uncertainty in the outcome that we measure by the **entropy**, <u>which is the logarithm of the number of possible equally likely outcomes</u>:
  - $S = \log_2 2^N = N \log_2 2 = N$
- We could use the sequences e.g. {H, H, T, …, H} to send a message, and we could send 2^N different messages. The **information capacity** of this scheme is again measured by the logarithm of the number of possible messages
  - $I = \log_2 2^N = N \log_2 2 = N$

Information & Entropy

- In this context we would often use base 2 for the log and say there are N bits of information.
- An alternative point of view is that the measurement of a particular result i.e., sequence of heads and tails has told us something about the system and we <u>have learned $N \log_2 2$ bits of information</u>.
- Uncertainty of the outcomes (**entropy**) and what has been learned from a measurement (**information**) are complementary.
- Generalizing these ideas to a system of N possible results with **different independent probabilities p_i** gives the expression for the entropy of the system S or the information learned by finding a particular result.

Kolmogorov Complexity – Entropy

- In Shannon's information theory, the degree of randomness of a finite sequence of discrete values can be quantified by calculating the **entropy** (<u>amount of information</u>) as

$$(entropy)\ H = -\sum_{p_k \neq 0} p_k \log_2(p_k)$$

- where p_k is the probability of occurrences of value i. Using this criterion, the **higher the entropy, the more the randomness**. For instance, the sequence 00100010 (entropy= –0.25*log(0.25) – 0.75*log(0.75) = 0.81) is less random than 01010101 (entropy = – 0.5*log(0.5) – 0.5*log(0.5) = 1).

Information Theory - Frequentist

- A ***frequentist*** version of probability: In this version, we assume we have a set of possible events, each of which we assume occurs some number of times. Thus, if there are N
  - distinct possible events $(x_1, x_2, \dots, x_N)$, no two of which can occur simultaneously, and
  - The events occur with frequencies $(n_1, n_2, \dots, n_N)$, we say that the probability of event xi is given by
  - $P(x_i) = \frac{n_i}{\sum_{j=1}^{N} n_j}$.
- This definition has the nice property that:

$$\sum_{i=1}^{N} P(x_i) = 1$$

Information Theory - Bayesian

- An *observer relative* (***Bayesian***) version of probability: In this version, we take a statement of *probability* to be an assertion about the <u>**belief**</u> that a specific observer has of the occurrence of a specific event.
- Note that in this version of *probability*, <u>it is possible that two different observers may assign different probabilities to the same event</u>. Furthermore, the *probability* of an event is likely to change as we learn more about the event, or the context of the event.

$$P(A|B) \times P(B) = P(B|A) \times P(A),$$

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

Axioms of Information Theory

- The following represent a set of reasonable axioms for an ***information*** measure $I(p)$:

- Information is a non-negative quantity: $I(p) >= 0$.
- If an event has probability 1, we get no information from the occurrence of the event: $I(1)=0$.
- If two independent events occur (whose joint probability is the product of their individual probabilities), then the information we get from observing the events is the sum of the two informations:
  $I(p1*p2) = I(p1) + I(p2)$.
- We also want our *information* measure to be a **continuous** (and, in fact, monotonic) function of the probability (slight changes in probability should result in slight changes in *information*).

Information Theory

- From these, we can derive the nice property of *information* measure:

$$I(p) = -logb(p) = logb(1/p)$$

Thus, using different bases for the logarithm results in information measures which are just constant multiples of each other, corresponding with measurements in different units:

- $\log_2$ units are bits (from binary)
- $\log_3$ units are trits (from trinary)
- $\log_e$ units are nats (from natural logarithm) (We commonly use $\ln(x)=\log_e(x)$)
- $\log_{10}$ units are Hartleys, after R.V.L. Hartleys, 1942.

Information Theory

- Suppose now that we have n symbols $\{a_1, a_2, \ldots, a_n\}$, and some source is providing us with a stream of these symbols.
- Suppose further that the source emits the symbols with probabilities $\{p_1, p_2, \ldots, p_n\}$, respectively.
- For now, we also assume that the symbols are emitted **independently** (successive symbols do not depend in any way on past symbols).
- What is the average amount of *information* we get from each symbol we see in the stream? What we really want here is a weighted average. If we observe the symbol a_i, we will be getting $log(1/p_i)$ *information* from that particular observation.

•In a long run of (say N) observations, we will see (approximately) $N * p_i$ occurrences of the symbol a_i (in the frequentist sense, that's what it means to say that the probability of seeing a_i is p_i).

Information Theory

- Thus, in the N (independent) observations, we will get total information

$$\text{Total Information (TI)} = \sum_{i=1}^{n} N \times p_i \times \log\left(\frac{1}{p_i}\right)$$

And therefore, the **average information** we get per symbol observed will be

$$I = \frac{1}{N}\sum_{i=1}^{n} N \times p_i \times \log\left(\frac{1}{p_i}\right) = \sum_{i=1}^{n} p_i \times \log\left(\frac{1}{p_i}\right)$$

Note that $p_i*log(1/p_i) \rightarrow 0$, as $p_i \rightarrow 0$, so we can, for our purposes, define $pi*log(1/pi)$ to be 0, when $p_i = 0$. This brings us to a fundamental definition. This definition is essentially due to Shannon in 1948, in the seminal papers in the field of information theory. As we have observed, we have defined *information* strictly in terms of the probabilities of events. Therefore, let us suppose that we have a set of probabilities (a probability distribution $P = \{p_1, p_2, \ldots, p_n\}$).

Information Theory

- Suppose that we have a set of probabilities (a probability distribution $P = \{p_1, p_2, \ldots, p_n\}$).

Definition: We define the (Shannon-Wiener) *entropy* of the distribution P by:

$$H(P) = \sum_{k=1}^{n} p_k \times log\left(\frac{1}{p_k}\right) = -\sum_{k=1}^{n} p_k \times log\left(p_k\right)$$

There is an obvious generalization of the entropy for continuous, rather than discrete, probability distribution $P(x)$:

$$H(P) = -\int P(x) \times log\left(P(x)\right)dx$$

Information Theory

- With this definition, we have that: $H(P) = E(I(p))$.

- In other words, the *entropy* of a probability distribution is just the *expected value* of the *information* measure of that distribution. For more discussion the following **important points**:

•What properties does the function $H(P)$ have? For example, does it have extrema, and if so where?
•Is *entropy* a reasonable name for this? In particular, the name *entropy* is already in use in physics/thermodynamics.
•How are these uses of the term related to each other?
•What can we do with this new tool?

•See my Spring 2004 Class notes on Medical Imaging Online
  •http://www.stat.ucla.edu/~dinov/courses_students.html