Due: Thu, Mar 8, 2012

Why did you choose that article? Why are you interested in that topic?

I chose the article "The sands of time: does history obey the rules of statistics?" for three reasons. First of all, it reminded me of "Prince of Persia: The Sands of Time", a game that I have previously played and particularly enjoyed. So the article caught my attention. Upon opening it, I noticed that Mark Buchanan wrote it. I have read an excerpt from his book called "The Social Atom", which was a very interesting read as it attempted to model human interactions and the stock market to a system that could be studied in detail. Finally and most importantly, this article seemed the most interesting to me out of all of the other ones because I am very intrigued by events of both small and large significance and how they shape my life and lives of other people around me.

What did you learn, and how does it connect to what you have learned in class?

After reading the article, I was surprised to learn that events, which seemingly have obvious explanations and causes, may not be so obvious at all. The problem of predicting when and where an avalanche will occur after sprinkling grains of rice on a table is quite fascinating. It reminded me of the Poisson and Gamma distributions which we learned about in class and how this problem can be turned into a random experiment. If we randomly sprinkle grains of rice on a table of fixed size with a constant rate, we can observe when avalanches occur. We can gather enough statistical data to compute a correlation of the size of the table and rate with which the rice is sprinkled. Then we can predict out the number of avalanches that occur over a given period of time using the Poisson distribution. We can also predict out how long it will take for the next avalanche to occur using the Gamma distribution.

I learned that several seemingly natural events map to a statistical distribution called the power law. Similar to the way we have identified the probability distribution / mass functions and the appropriate parameters for various real-life situations, the author identifies the distribution and parameters of avalanches. He also points out how the power law has been mapped to several other situations such as the number of deaths in wars, the magnitude of price changes causing market fluctuations, and the extinction of certain species over time.

All of these events seem very unrelated and it is very non intuitive that they can be studied using the same mathematical tools and methodologies. This resonates with what I have learned in class about statistics and how my intuition can often be misleading. It also emphasizes the importance of an empirical approach and the transformation of a real life situation into a mathematical problem.

What questions were you left with?

One of the major questions that I was left with was the implication of the power law. I am curious about what other events and situations that can be related with the power law. When the author mentioned the use of artificial intelligence agents that played as traders in a market to predict its outcome, I wonder if we can do the same with political stances of countries and try to predict future conflicts and resolve them before they occur.

STAT 310 March 8, 2012

The Best Questions For a First Date

Here it is: the secrets of the universe – or at least, the secrets to dating success. All the answers (actually the questions) are right here. No longer will I spend the whole night anxious of if I'm going to get lucky. No longer will I wonder if my date really believes that the Earth is 6000 years old. I have a powerful arsenal of tell-tale questions to ask any potential mate. Or do I?

I chose this article because of the relevance to the life of a lonely, desperate college sophomore. I want to find more effective ways to meet new people, especially members of the opposite sex, so scientific research like this to back up pick-up artists' tales is very welcome. But of course, we need tools from our statistical toolbox to show us how useful these tools really are!

Say I want to see how compatible I'd be with someone (C). The article cites three questions that 'couples' (people who have left OkCupid because they have met someone on OkCupid – which we'll assume as the article does that this equates to long-term compatability) most often agreed on (Q). The article gives the probability that two people agree on these three questions given that they are a couple, $P(Q|C) = 0.32$. We are also given that the probability of agreeing on these questions by chance, $P(Q) = 0.075$.

Given this, we can use Bayes' Rule to calculate $P(C|Q)$, the probability that a person is a good long-term match for me given that we agree on all three questions. Bayes' Rule states that $P(C|Q) = P(Q|C)P(C)/P(Q)$. We need one more crucial datum: $P(C)$, the probability that two people are long term compatible by chance – and keeping consistent with our previous assumptions, this would be what fraction of OkCupid's users leave because they find someone. We can assume this is small (<0.1: after all, true love is rare!), then we can calculate $P(C|Q) = 0.32 * 0.1 / 0.075 = 0.43$. Slightly less than half, assuming the upper bound of the maximum compatibility rate. A lot better than our assumed base rate (the chance that you meet someone special by pure chance), but nowhere near a sure-thing.

Of course, in using these statistics for personal matters means that you are assuming that OkCupid users are representative of your pool of potential dates, when in fact this is not a valid assumption: for the question about political preferences, the article notes that this is on a national level, and areas with more concentrated views (like Rice) do not have this property. Likewise, we can question our assumption that people who end up dating and quitting OkCupid can be assumed to be indicators of long-term compatibility.

Regardless, this article once again shows that the results of statistical research can defy intuition. The best indicators of compatibility (given our assumptions) are not "important" personal questions about God, sex, and health, but seemingly trivial questions about movie genres and hypothetical travel.

This makes me wonder what sort of hidden connections one can make in other personal relationships. If someone makes their bed regularly, does that mean they're a good promise keeper? Can you trust a friend to pay you back if he likes the color red? Could employers discern information about interviewees based on casual, non-work-related discussion? Maybe these statistics do hold the secrets of the universe, after all.

Stats In Practice

The article "Spoilt for Choice" discusses whether organic milk is truly healthier than non-organic. Findings seem to indicate that the answer is yes; beneficial fatty acids were found in much higher quantities in organic milk.

My personal interest in this result stems from my roommate's obsession with organic milk. She insists that it is superior, not for its health benefits, but because it spoils more slowly. Unfortunately, she drinks it too fast for me to ever tell if it spoils sooner than any non-organic milk I buy. Since organic milk tends to be more expensive than non-organic, I was hoping to find data on whether the extra amount she spends is worth it. While the article said nothing about the spoiling rate of organic vs. non-organic milk, it does seem to indicate there is some health benefit to organic milk. Unfortunately, my roommate couldn't care less about that, so I will have to return to observing how fast the milk turns on my own.

What I find most interesting about this article is that the connection between health and the organic label seems to be primarily related to the diet of the cows. This seems to be a direct connection to the idea of independence. The study uncovered a relationship between the season and the quality of the milk; in the winter, the milk is less healthy. These two variables are quite clearly connected to each other. However, their connection does not mean that cold weather causes cows to produce worse milk. The researchers realized that the diet of the cows would change as the grass died, and further related the breakdown of food to the quality of the milk. Cows which grazed primarily on fresh grass produced the best milk, those that ate preserved grass had second-best, and those which ate cereal mixes instead had the worst milk. Diet and milk quality are clearly not independent. Weather and milk quality are also not independent; however, this does not mean weather is directly responsible for the quality of milk produced. This is an example of a third variable affecting two other related variables. If one wanted to define the relationship, a bivariate distribution of some type would be needed, with weather and food growth as inputs and the quality of milk as the output.

The other idea discussed in the study regarding statistics was the idea of significance. When the occurrence of certain vitamins and acids is too different between the various samples, it is incredibly unlikely to have occurred by chance. The effect of the farming system (and thus the diet) is shown to be more significantly related to the milk quality in the warmer area sampled, where the date seems to generally have a greater effect in the colder area. As the season affects how cold the temperature gets, even if a farmer had cows entirely grazing, they would still have to get cereal or preserved grass for the months when the grass was dead, reducing the quality of the milk. Likewise, if the temperature is generally warm enough to support the grass growth, the farming system will have the greatest impact on the milk quality, as what the farmer has the cows eat will directly affect the milk.

What I remain curious about is the effect of this study. While diet has clearly been shown to impact milk quality, consumers who care about the quality of their milk will likely already be purchasing organic milk. Those who do not care as strongly will be less affected by the findings, as I doubt any industries which use the cheaper cereals to feed their cattle will change their cattle's diet unless it would positively affect their sales. I am also curious how the samples themselves were taken. Were cows chosen at random from groups that would be considered organic and non-organic? What other factors affect milk quality? Antibiotics are briefly mentioned in the study; do these impact the quality of the milk? Is that impact over time or nearly instantaneous? Does the healthiness of the cow in question matter significantly? Diet is one factor, but I doubt it is the only thing to affect the healthiness of milk.

Stats in Practice #1
STAT 310

The article *London murders: a predictable pattern?* was an insightful read on how something that doesn't seem to be related at all can not only be organized together but also predicted in the future. In this case, the unpredictable events were murders in London. Statisticians were interested in the "sudden" uprisings in murders and wanted to see if they were just an anomaly, a sign of increasing violence, or something that could have been predicted using statistics.

The article seemed interesting because murder seems to be something completely random and unpredictable. Every time a murder occurs, it is put on the front page of the local newspaper and made to seem like an event that came out of nowhere. People also become more worried if there seems to be an uprising of murders in their community. This article explained in great detail how murders can actually be predicted and that these isolated incidents of multiple murders should not be a cause for alarm. For example, by using statistics and data of the past years murders, it was reported that about 93 murders would have occurred by July 28th. On this date, the newspapers announced that 90 murders had already occurred.

This proved two things. It proved that by using statistics we are able to predict the number of murders that will occur over a span of time or the probability that they will occur on any given day. It also proved that even though we know the predictability of murders, they still create concern among the community even when there is no reason for it. Another example that clearly showed how predictable murders could be was when four murders occurred on the same day. This created a lot more concern because of how unusual it was for it to happen. However, statisticians had predicted that this was bound to happen once a year and it did. Now if this had occurred several more times, it would have been cause for concern.

In class we have discussed random variables and the different distributions that can be associated with them. Even though we call them random variables, we have usually dealt with events with some predictability. These events have included how many free throws a certain athlete is supposed to make or waiting times between events. In our minds these events seem more predictable if we take into consideration how talented the athlete is or how long we end up waiting. However, it defies human intuition to believe that murders would be predictable, yet they can also be considered a random variable. The article does just that and uses the poisson distribution to find the probabilities of certain events occurring. The use of this distribution means that each event occurs independently of one another and that it occurs at an average rate. Even if this distribution works pretty well, it may still have flaws. One major flaw is that not all murders occur independently. One person can murder more than one person or there can be a reason several people are getting murdered at one point.

It was interesting to learn that even data that seems difficult to learn can be described using a simple method. If murders can be predicted then there are several other things that could be predicted as well such as how many car crashes or rainy days there will be in a year. Even though we could estimate these numbers using road conditions and weather patterns, it can still be treated as a random variable.

The article did a great job summarizing how murders can be predicted. Unfortunately, it did not mention how this could help with the safety of the community. Being informed about the trends does help to see if there should be any alarm for a sudden uprising in murders, or if it was predicted to happen. However, it would have also been insightful to learn if they recorded the reasons and locations behind the murders and predicted the "triggers" that made them occur.

I choose this article out of interest in the patterns of criminal action. I have always enjoyed television shows like *Numbers* and *Sherlock Holmes* (especially the new BBC series). The idea of criminal patterns emerging from data through analysis using game theory and statistics is really intriguing to me. Personally, I am interested in the topic due to my caution and desire to stay safe. In life, I feel that it is almost always advantageous to have as much information or data available in order to make informed decisions.

One of the most interesting discussions within the paper concerned was the fitting of the homicide data to a distribution. According to the article, the data (occurrences versus number of homicides per day) should fit a poisson distribution with mean of 0.44 very well. It turns out this is true when one uses the police data. To me this is interesting, because it is a real-life example of the usefulness of mapping data to distributions and random variables. Comparing real-life data to known distributions is something we have done before in class. Unlike in class, where we often sought reasons why the data might not perfectly fit the distribution, it seems that the authors accept it just because it looks right; because of our heavy scrutiny in class and their glossing over of this part, I wish they had offered some insight into the many variables that might affect the distribution's correctness.

To me, another interesting component of this analysis is that the data is spread out quite uniformly over the months of the entire year - there does not seem to be a month that is more popular with murders over any other. Using my intuition I'd expect some difference between colder months and warmer months (perhaps fewer murders in winter months due to fewer people on the streets at night; perhaps more murders in the summer due to unruly adolescents off from school). It would seem my intuition here is incorrect, and any such factors (if they exist) are either minor or counter each other out.

While the uniformity between months and frequency matching poisson is interesting, to me the most interesting part of the authors' analysis concerns the number of murders per day, by day of the week. According to the authors of the paper, there is a "'Saturday effect' of around a 60% increase in homicide rate compared with all other days of the week combined" (pg. 7). While anyone could speculate on the reason for this, it is an interesting and intuitive fact that the authors derived from the data.

A poignant discussion within the article was the authors' cautioning against misusing or misunderstanding the data (especially in the media); I feel that this is relevant across all fields of study. They cautioned against stating, "we haven't seen so few (murders) for several years" (pg. 8). Even if crime looks like it is decreasing, one should be cautious to whether that is just a normal fluctuation in these "random" events or whether crime is actually decreasing. Unfortunately, this is not always intuitive and I feel that the authors' warning is in vain; it seems that people in the media consistently prove that statistical methods are not always well understood in the general public.

While this was a very interesting and relevant article, I feel that a more in-depth analysis along with more data from around the world would be very interesting. Comparison and contrasts between other large cities would be interesting. The fact that this article was so easy for me to understand means that either the analysis could be more sophisticated (and people would still be able to understand it). This article definitely left me wanting more. Until then I guess I'm left with *Numbers* and *Sherlock Holmes*.

Statistics in Practice

Although I have been exhorted by my parents repeatedly to drink organic milk, I've never bothered to learn what truly made one type of milk better than another. Thus, the article "Spoilt for choice: is organic milk really any different?" appealed to me on a personal level. Also, I've always been interested in the role of statistics in substantiating claims and attacking assumptions. Statistics and mathematics, if applied properly, hold true, no matter what common sense, popular culture, or an industry might dictate.

The article investigates the differences in milk quality between organic and non-organic and how those differences in milk quality arise from a combination of diet, health, and breed. Organic milk is indeed healthier than conventionally produced, non-organic milk. However, the actual reason why organic milk is healthier is solely because the cows are grazed longer than their non-organic, high-input counterparts, rather than any other reason why the milk is categorized as organic. Diet was found to be the strongest predictor of milk quality, and thus cows that grazed longer on fresh grass in both low-input and organic farms had similar, healthier milk compositions than cows raised in high-input, conventional farms.

To me, this article exemplifies the modern role of statistics in attacking assumptions, in that a common perception is taken, mathematically defined, and judged based on solid mathematical evidence. Additionally, this paper adheres to the general problem-solving approach introduced in stat310. For one, the authors defined upfront a linear mixed-effect model, a model they used to investigate the differences in the quality of the milk between the different production systems. When the authors discussed the "fixed effects" and "random variables" of the mixed-effect models, which affect the mean and variance of the response variable respectively, I immediately recalled how the different distributions can be altered through manipulation of mean and variance. For each measurement taken, the authors characterized vague, qualitative attributes as quantifiable measurements. For example, the vague term "which milk is better for you?" is translated into that milk's concentration of beneficial compounds, such as vitamin E and carotenoids. Also, the vague question "how does varying each component of a cow's lifestyle affect what the milk's components?" was replaced by multivariate analysis which numerically defined such relationships between each factor in a cow's lifestyle and concentration of each milk component.

After reading this article, I was left with several questions about the statistical analyses and models used into the study. Firstly, I'd like to learn more about the linear mixed-effect statistical model, which was used in the study to determine differences in milk qualities between various systems of milk production, so that I can fully appreciate the study's design. Secondly, I'm curious about the multivariate redundancy analysis used to determine the relationships between the factors of diet, health, and breed and the cow's milk components. Although I've taken linear algebra, the significance of the eigenvalues in that analysis was lost upon me. Lastly, the final question raised by this article for me is not one of statistics, but one of concern – I now wonder how the food industry manages to be so successful in marketing labels such as "organic" without having to prove their benefits.

_ STAT 310 Essay _ Dr. Hadley _ Feb 25, 2012

Randomness and IPOD Shuffle

Froelich et. al.'s "Does Your iPod Really Play Favorites?"[1] gives a detailed argument that IPOD Shuffle does play songs randomly. The authors recognize that people have a fallible intuition about randomness, and demonstrate Shuffle's randomness using probability.

As an IPOD user myself, I'm always confused about the "shuffle" feature of IPOD. According to Apple, "shuffle" is a software which plays the songs in a playlist in a random order. I have 633 songs in my iTunes library, one third of which is from the same band. But when I play "shuffle", I feel that the songs from that band do not get to play as often as I expected. Is the rearrangement in "shuffle" truly random? Does it attempt to rearrange the songs so that different genres and singers' songs get to play at the same frequency?

So, when I saw the paper "Does Your iPod Really Play Favorites?", I was so glad someone actually tested it. I learnt from the paper how to use probability concept to prove randomness in real life and how my misunderstanding of randomness fools me. Take my IPOD shuffle as an example. The "shuffle" feature acts like a sampling without displacement, so IPOD has 633 P 633 ways to play my songs. This is a gigantic number at a level of 10^{1500}. Among all of these permutations, how often does the songs from my favorite band (let's call it A) occur? Is my feeling that shuffle doesn't play A's songs so often right?

Think of the number of A's songs played as a Binomial distribution. Set A={number of A's songs played}. Since one third of my songs are from A, the probability of the next song being from A is 1/3. The distribution is thus A~Bin(n, 1/3). For every 10 songs, the E(A)=10/3 (it doesn't matter which n songs we considered since the order of songs in shuffle is independent to each other.). This is not a big number considering how much I love the band and always want it to be the next song. In fact, if the "shuffle" feature is truly random, the probability of all 10 songs in a row are from A is just P(X=10)=0.0017%! It even has a much higher probability to play none of the 10 songs from A since P(X=0)=1.734%! How I feel I don't have enough A's songs in shuffle actually illustrates IPOD shuffle plays randomly indeed.

✓ How my feeling of IPOD shuffle's randomness deviates from the fact shows an interesting insight on product design. People don't understand randomness by intuition. When you watch a dance and see all the dancers spread out on the stage, you think they are doing this "randomly". But if this is truly random, how can there be people at every corner of the stage? What you think is randomness is actually a planned evenness. Similarly, for a product like shuffle, what the company needs to achieve might not be the statistical randomness, but what the randomness is in people's mind. Do people expect to hear songs from different genres when they play shuffle? Do people expect the next song to be from a different singer? To achieve people's expectation of random order, shuffle might need a careful formula for the order of songs, which is not statistically random at all. However, what really matters to a company is the consumers' feedback, not the math. Although shuffle is random as proved in the paper, the controversy on this topic gives us an idea on what people actually want from a "random" shuffle.

1. Amy G. Froelich, William M. Duckworth, Jessica Culhane. The American Statistician. August 1, 2009, 63(3): 263-268. doi:10.1198/tast.2009.07073.

What questions were you left with?

I read "The Best Questions for a First Date" by Christian Rudder, and this indeed was a highly entertaining article. The reason I chose this article is because I am absolutely fascinated with relationships. I am pursuing a second degree in Psychology, and my favorite topics to study are gender differences and family dynamics, including parent-child relationships and marriage/dating relationships. I think my interest stemmed mostly from reflecting on my own childhood and observing my friends' relationships with their parents. Growing up, my mother would always give me advice on what sort of man I should marry in the future. On several occasions, I realized that my own preferences were different from her suggestions; thus, I became more and more interested in studying family and relationships. I always want to learn more about what makes marriages "work" and all the stages of dating that come before marriage.

This article offered a great deal of insight, almost a formulaic method of selecting a mate from OkCupid. I think most people would agree that in order to be successful in a relationship, the couple must have similar beliefs or interests in some very specific fields. Using statistics, Rudder has discovered a way to find out what a person is really like by asking a few simple questions. I think this is an ingenious way to figure out those burning questions you have on that first date. However, as a Psychology major, I would feel like I was deceived if this method was used on me. Asking a girl if she like the taste of beer or not should not be taken as a go-ahead signal to sleep with her. Although this may be a correlation, and a very strong correlation at that, it is not a fool-proof way to figure out one's date. I believe that time and patience will always give you a better answer.

However, for those of us who are impatient or just a little socially shy, Rudder has done an excellent job of collecting data and forming the conclusions he has. I have to admit that if people actually used his methods, there might be a decrease in misunderstandings and an increase in better relationships. If the correlations are that strong, chances are, you could be spot on in your inference of your date, and there could seriously be some potential in your future with this person.

In order to collect and analyze the data, Rudder must have used the Binomial Distribution. Additionally, he must have used the multiplication principle to figure out the probability of randomly choosing answers for the example with questions looking for similar answers. He also used conditional probability in order to draw conclusions based off of correlational data. For example, he had to consider the probability that someone will have sex given that they like beer, as well as the probability that someone will not have sex given that they like beer. This also involves the Law of Total Probability. All of these concepts are things we learned in Stat 310 this semester.

Lastly, I still want to know more. His conclusions were only drawn from the OkCupid website. What about those people who do not utilize this website? If he only considers those using OkCupid, hasn't he fallen victim of convenience sampling bias as well as voluntary response bias? What are some other interesting deceiving questions you can ask on a first date and form a conclusion from a correlation? Why is it that atheists write better than those who consider themselves religious? What does religion have to do with any of that? It's going to take actual explanations and reasoning to convince me; correlations are great but that's not enough to explain why it works.

All in all, I found this article to be extremely interesting. I think this may be an area of statistics that I would want to pursue further, because I think it can incorporate my psychological interest with mathematical foundation. I think it is brilliant how useful probability and statistics can be in any field. This definitely makes me more excited to continue pursuing Statistics and see where it can take me in the future.

Cat Lady No Longer

Too many times do individuals run through a gauntlet of dates only to find at the end of the fifth or sixth date that the person in interest is yet again another incompatible disappointment. It's hard to shift through reality and (for lack of a better word) bullsh** especially during the first-date when everyone is obviously on his and her best behavior. Asking the essential but obviously too personal questions such as "Do you want to get married?" or "Are you religious?" may give Prince or Princess Charming the wrong idea on the first-date. Furthermore, most people will obviously try to give a more diplomatic answer that will appease their date than the honest answer. So as a college student at Rice, running through the seemingly endless gauntlet of dates, how do I escape the cycle of incompatible disappointments and find my Prince Charming? Let's be honest, I may be 19 and still have plenty of time, but at the end of so many disappointing dates, I might spontaneously combust into a cynical 76 year-old cat lady. However, after reading "The Best Questions for First Dates" by Christian Rudder, my days as a future old maid may be numbered after all. I can use statistical correlations during first-dates to cut down on time and cut through (for lack of a better word, again) the bullsh** in order to see if the boy or girl in question is compatible. Wanting to know if my date is religious, I can simply ask if grammar mistakes bother him or her. Wanting to know to his or her political ideology, I can instead ask if he prefers a simple or complex life.

By utilizing the power of statistics, Rudder essentially conducts a random experiment in order to find statistical correlations between the answers of personal questions and "easy to bring up" questions". Since the answers to the questions can only be "yes" or "no," the data collected corresponds to the Bernoulli distribution, which can only have two outcomes. Furthermore, correlations represent a basic concept learned in Stats 310 that basically refers to the degree the measurements for two elements vary together. Rudder sets out to define which questions are "easy to bring up" by using a process similar to the one used in Stats 310 which involves a concrete definition to the intangible words in a math problem. Rudder defines "easy to bring up" as those questions where the individuals sampled didn't check a privacy box. Using the database of questions and answers from the dating website OkCupid, Rudder divides the questions between weighty topics and casual topics and also between questions that tell you a little and those that tell you a lot. By finding the intersection between questions people feel comfortable publically discussing and those mathematically able to tell you something, Rudder is concerned with a specific subset from the experiment and the event of shallow first-date questions that hold statistical correlation and that actually give real insight on your date. Thus, wanting to know if our date is a sexually open individual, we can ask if they enjoy the taste of beer since "beer-lovers are 60% more likely to be okay sleeping with someone" after the first date. Wanting to know if your date is liberal, ask a question about burning flags. The numerical percentages given represent probabilities, another concept for Stat 310 and math machinery used to answer questions about uncertain events. However, a question to keep in mind is how reliable Rudder's numerical facts are. A random experiment must be repeatable and the figures reliable. Obviously, the real-life question here involves finding love. Utilized with the power of correlations, the problem of unrevealing first-date ends. And this leads to possibility of further using the power of correlations to solve other real-life problems. Trying to figure out the likelihood someone suffers from depression, ask about family discipline. Trying to figure out the sexual preference someone possesses, ask about bed sheets. The list of real-life problem solvable with correlation continues on but personally, I'm going to use the power of correlation to solve the mystery of life: love.

Statistics in Practice: "On the need to use error-correcting memory"

In less than fifty years, the computer has transformed from a slow, inefficient, and limited novelty, useful for only a few highly specialized tasks into a cornerstone of modern society. Computers in our cars are now helping us not only get where we're going but also get us there safely and efficiently, nearly everyone uses a computer at work, most of us don't know what we would do without the internet, and a massive portion of humanity's knowledge is now stored electronically. Personally, I am interested in computers because nearly all of the work currently being done in physics today requires the aid of extremely advanced computing systems. Also, I am extremely concerned that as the computations for advanced physics become increasingly complex, there is a definite possibility that the advance of physics may actually be limited by the power of our modern computers.

For these reasons, I found this article concerning how cosmic rays can corrupt the memory of our computers today exceptionally interesting. In this article, the author, Berke Durak, draws the attention of the reader to the fact that cosmic radiation from space can have a corruptive effect on the RAM memory of the computer systems on earth. By the use of relatively simple statistics—he does little more than calculate probabilities based on given rates and time intervals—well within the scope of Stat 310, Durak illustrates the dangers in the seemingly complex interaction between cosmic rays and computer RAM memories. Although the statistical methods he uses are not extremely complicated, it should be noted that the principles of statistics are what allowed Durak to accurately demonstrate the importance of error-correcting memory. I can think of no other way to arrive at the conclusion that without error correction capabilities 4GB of RAM has a ninety-six percent chance of being corrupted in seventy-two hours while the same amount of RAM with error correcting capabilities only has only one chance in six billion of being corrupted in this same period other than by using the principles that we have learned in Stat 310.

maybe briefly explain how

As technology advances, the application of statistics to computers will become increasingly important. This increasing necessity will occur for two primary reasons. First, computers will become increasingly integrated into the various parts of our lives, and second, the capacity and complexity of computers is rapidly increasing. As computers become increasingly integrated into our lives, applying statistical analysis will become increasing important because there will be more at stake. For example, as computers become increasingly important components of our cars, it will be more important to understand the rates of failure and error since these errors can directly lead to the loss of human life through car crashes. Second, as computers become more complex and drastically increase their capacities, statistical analysis will become even more suited to predicting rates of error and failure due to the simple fact that the number of circuits and bits is increasing. This second connection is derived merely from the fact that statistics is better suited to describing large systems than to describing small systems.

how does this directly connect to the article?

a little redundant

In conclusion, Berke Durak uses simple probability calculations—most no more complex than those taught in Stat 310—to illustrate the importance of error correcting capabilities in computer RAM memory systems. This use of statistics foreshadows the growing application of statistics to computers as they play larger roles in the different facets of our lives, and become increasingly complex and large. Despite all this, Durak does not take the next step after demonstrating the number of memory errors that error correcting capabilities can find and correct. He does not explain why this is important. He fails to examine the severity of the problems prevented by particular error correcting programs. Durak leaves the reader still asking the question: so what if a couple of my RAM bits get corrupted?

Interesting point

he outlines the gravity of how a single bit can kill a video, cause software to malfunction etc.

http://lambda-diode.com/opinion/ecc-memory