

On Forecasting Solar Flares via. the Bayesian Blocks Method using Flare Data from  
Terrestrial Solar Observatories vs. Extraterrestrial Detectors

---

A Thesis  
Presented to  
The Division of Mathematics and Natural Sciences  
Reed College

---

In Partial Fulfillment  
of the Requirements for the Degree  
Bachelor of Arts

---

Tenzin Sangpo

December 2019



Approved for the Division  
(Mathematics)

---

Andrew P. Bray



# Acknowledgements

I want to thank a few people.



# Preface

This is an example of a thesis setup to use the reed thesis document class.





# List of Abbreviations

You can always change the way your abbreviations are formatted. Play around with it yourself, use tables, or come to CUS if you'd like to change the way it looks. You can also completely remove this chapter if you have no need for a list of abbreviations. Here is an example of what this could look like:

<b>CME</b>	Coronal Mass Ejection
<b>SEP</b>	Solar Energetic Particle



# Contents

<b>Introduction</b>	<b>1</b>
0.1 Why use it?	1
0.2 Who should use it?	1
<b>Chapter 1: On Solar Flares</b>	<b>3</b>
1.1 Understanding Solar Flares	3
1.2 Observing Solar Flares	4
1.3 Classifying Solar Flares	4
1.4 How Flares Become	6
1.5 Why They Matter	7
<b>Chapter 2: Applied Notions of Statistics</b>	<b>9</b>
2.1 Probability	9
2.2 Conditional Probability	10
2.3 Random Variables Distributions	12
2.4 Bayesian Statistics	15
<b>Chapter 3: Wheatland's Bayesian Approach</b>	<b>17</b>
3.1 The Phenomenological Rules	17
3.2 Previous Method	19
3.3 New Method	19
3.3.1 Process	19
3.3.2 Estimating the Power Law Index - ' $\gamma$ '	19
3.3.3 Estimating the Rate of Occurrence - ' $\lambda$ '	19
3.3.4 Estimating the Probability that At least One Large Flare Event Occurs - ' $\epsilon$ '	19
3.4 Simulation	19
3.4.1 Exploratory Data Analysis	19
3.4.2 Applying the Changepoint Analysis	19
3.4.3 Simulation Outcome	19
<b>Conclusion</b>	<b>21</b>
<b>Appendix A: The First Appendix</b>	<b>23</b>
<b>Appendix B: The Second Appendix, for Fun</b>	<b>25</b>

Bibliography . . . . .	27
------------------------	----

# List of Tables

1.1	H-alpha spectral observations in terms of flare brightness. . . . .	5
1.2	H-alpha spectral observations in terms of flare size. . . . .	5
1.3	Flare classes on the basis of X-ray peak flux. . . . .	5



# List of Figures

1.1	Solar flares detected by NOAA satellites in July 2000 . . . . .	6
1.2	Magnetic field structure around an active region of the sun . . . . .	7





# Abstract

The Bayesian blocks approach to solar flare prediction relies not only on the optical classification of sunspots and the past flaring rate of a given cluster, but also the flaring record of an active region for events small and large. Studies implementing the refinement, however, run programs written in the Interactive Data Language (IDL), a proprietary programming language. This dissertation aims to employ the novel approach to forecast large flare occurrences via. Changepoint Analysis using the Pruned Exact Linear Time (PELT) algorithm in R, a more accessible platform. Available scientific literature on flares is discussed, the applied essentials in statistics, the theories of the respective methods are outlined, and simulations for the latter is presented to demonstrate how the method works in practice. Furthermore, forecasts from historical data acquired from terrestrial solar telescopes and the Geostationary Operational Environmental Satellites (GOES) under the National Oceanographic and Atmospheric Administration are compared with actual observations to ascertain both, the validity of the approach and the necessity of equipping satellites with flare detectors.



# Dedication

You can have a dedication here if you wish.



# Introduction

Welcome to the  $\text{\LaTeX}$  thesis template. If you've never used  $\text{\TeX}$  or  $\text{\LaTeX}$  before, you'll have an initial learning period to go through, but the results of a nicely formatted thesis are worth it for more than the aesthetic benefit: markup like  $\text{\LaTeX}$  is more consistent than the output of a word processor, much less prone to corruption or crashing and the resulting file is smaller than a Word file. While you may have never had problems using Word in the past, your thesis is going to be about twice as large and complex as anything you've written before, taxing Word's capabilities. If you're still on the fence about using  $\text{\LaTeX}$ , read the Introduction to LaTeX on the CUS site as well as skim the following template and give it a few weeks. Pretty soon all the markup gibberish will become second nature.

## 0.1 Why use it?

$\text{\LaTeX}$  does a great job of formatting tables and paragraphs. Its line-breaking algorithm was the subject of a PhD. thesis. It does a fine job of automatically inserting ligatures, and to top it all off it is the only way to typeset good-looking mathematics.

## 0.2 Who should use it?

Anyone who needs to use math, tables, a lot of figures, complex cross-references, IPA or who just cares about the final appearance of their document should use  $\text{\LaTeX}$ . At Reed, math majors are required to use it, most physics majors will want to use it, and many other science majors may want it also.



# Chapter 1

## On Solar Flares

### 1.1 Understanding Solar Flares

Solar flares are localized sudden and intense bursts of radiation from the sun that can last from several seconds to a few hours. Their concurrence with solar energetic particles (SEPs) and coronal mass ejections (CMEs), as well as their proximity to clusters of sunspots is well documented. In fact, the first solar flare was observed on September 1, 1805 by the astronomer Richard C. Carrington, and then later that year by Richard Hodgson, while monitoring sunspots. The flares manifested themselves in the middle of a sunspot group as two blotches of white-light rapidly brightening across the spectrum of visible electromagnetic (EM) radiation. They then dimmed as they moved before finally disappearing in a matter of minutes. But solar flares are more than mere flashes; they are tremendous explosions that occur on the surface of our resident star. The energy liberated from the said events can range from  $10^{23}$  to  $10^{25}$  joules. That is as much clout as a billion megatons of TNT which dwarfs the 50 million tons of TNT yield of 'Tsar Bomba', considered among the mightiest thermonuclear weapon ever conceived. Solar flares are the most powerful phenomenon in our solar system.

And the released energy manifests itself in various forms. Firstly, the materials encompassing the regions of interest can expect to be heated up to 20 million kelvins. The sun's photosphere i.e. the outer shell from which light is emitted, sits at about 6,000 kelvins. Secondly, accelerated particles therein such as electrons, protons and ions can gain energy up to several giga electron-volts (GeV). And finally, there is strong plasma motion i.e. movement of said materials down a temperature or pressure gradient. The accelerated charged particles and heated matter motion in turn cause the emission of electromagnetic radiation across the entire electromagnetic spectrum i.e. radio, microwaves, visible light, ultraviolet, X-ray and gamma ray.

## 1.2 Observing Solar Flares

The National Oceanographic and Atmospheric Administration (NOAA) oversees the satellites charged with detecting solar flares. The Geostationary Orbital Environmental Satellite – R Series (GOES-R) orbits the Earth on the geosynchronous orbit, about 22,236 miles above the ‘Blue Planet’, where it matches the revolution of the Earth about its axis. It is equipped with a Solar Ultraviolet Imager (SUVI) telescope and an X-ray spectrometer to constantly monitor the sun for flares. The telescope enables researchers to locate the flare on the sun while the spectrometer affords an estimate as to the magnitude of the flare strength. With this information, NOAA is able to determine if the explosion is well positioned and severe enough to approach Earth. The Deep Space Climate Observatory (DSCOVR) satellite positioned at a point of neutral gravity between the Earth and the sun can detect the approaching storm and issue space-weather warning about 15 to 60 minutes before it reaches Earth. Upon arrival, GOES-R can measure the magnetic fields as well as the radiation levels from the particles ejected from the sun at the geosynchronous orbit above the Earth due to this flare or explosion.

But solar telescopes around the globe have been making Hydrogen-alpha (H-alpha) spectral observations from Earth long before the first artificial satellite was even considered as a notion. Hydrogen – the most abundant element in the Sun – can only exist in its atomic configuration in the sun’s outermost and visible regions where it is cool enough. The electron of a given hydrogen atom here can absorb energy and attain a higher energy level ( $n$ ). But when the said electron drops to a lower energy level, it relinquishes the spare energy in the form of electromagnetic radiation. The emission in the visible spectrum occurs if electron falls to the energy level with  $n = 2$  and is referred to as the Balmer Series. The H-alpha light specifically has a reddish hue and is emitted by hydrogen atoms when the electron falls from the energy level  $n = 3$  to  $n = 2$ . Solar telescopes use H-alpha filters to isolate this light. Observatories may also capture images of the Sun at regular intervals to visually monitor and document solar flares.

## 1.3 Classifying Solar Flares

Solar flares were previously classified on the basis of their H-alpha spectral observations in terms of brightness and size. For the former, a qualitative cataloguing of H-alpha intensity was employed with the designations to report flare brightness as shown in Table 1.1. As for the latter, a measure of then emitting surface area in terms of *Millionth of Solar Hemispherical Area* was considered to record flare size with the rubric shown in Table 1.2.

However, since solar flares also release their energy in the form of X-rays, they are also categorized on the basis of the brightness in their X-rays. The categorization of



Category	Non-abbreviated
<i>F</i>	Faint
<i>N</i>	Normal
<i>B</i>	Bright

Table 1.1: H-alpha spectral observations in terms of flare brightness.

Category	Size (Millionths of Solar Hemispherical Area)
<i>S</i>	Less than 100 units
<i>1</i>	Between 100 & 250 units
<i>2</i>	Between 250 & 600 units
<i>3</i>	Between 100 & 1200 units
<i>4</i>	More than 1200 units

Table 1.2: H-alpha spectral observations in terms of flare size.

flares based on X-ray (100 - 800 pico-meters wavelength range) Peak Flux in *Watts per Square-Meters* is as shown in Table 1.3.

Class	Size (Watts per Square-Meters)
<i>A</i>	Less than $10^{-7}$ units
<i>B</i>	Between $10^{-7}$ & $10^{-6}$ units
<i>C</i>	Between $10^{-6}$ & $10^{-5}$ units
<i>M</i>	Between $10^{-5}$ & $10^{-4}$ units
<i>X</i>	More than $10^{-4}$ units

Table 1.3: Flare classes on the basis of X-ray peak flux.

The A-class flares are near background levels. Followed by B, C, M and X, each letter represents a 10-fold increase in energy output. So, an X is ten times an M and 100 times a C. Within each letter class there is a finer scale from 1 to 9. C-class and smaller flares are too weak to noticeably affect Earth. M-class flares can cause brief radio blackouts at the poles and minor radiation storms that might endanger astronauts. And then come the X-class flares. Although X is the last letter, there are flares more than 10 times the power of an X1, so X-class flares can go higher than 9. The most powerful flare measured with modern methods was in 2003, during the last solar maximum, and it was so powerful that it overloaded the sensors measuring it. The sensors cut out at X2.

The measure of the latter – also called X-ray flux – is pivotal in characterizing the magnitude of the overall flare energy emitted. As was previously noted, GOES-R satellite orbiting the Earth are equipped with telescopes to locate the flares and with

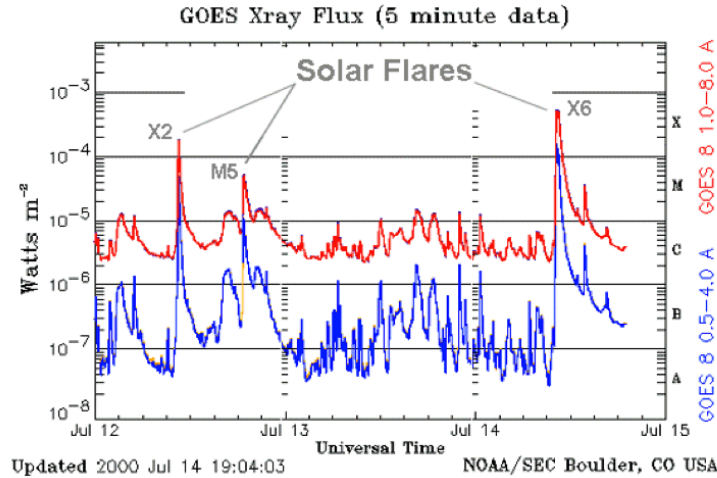


Figure 1.1: Solar flares detected by NOAA satellites in July 2000

detectors to detect the X-rays in it. If the flare is severe enough to cause a geomagnetic storm, the DSCOVR satellite can issue a 15 to 60 minutes warning.

## 1.4 How Flares Become

Like the Earth, the magnetic field of the sun is generated by the motion of electrically conducting fluids – plasma in case of the sun – within it. With the preponderance of evidence favoring the proximity of sunspots with solar flares, determining the structure of magnetic field around the former is pivotal to better comprehend and forecast the latter. It has been measurably observed that if and when the magnetic field is twisted or sheared, the otherwise separated field lines can cross over each other. The resulting short-circuiting causes a tremendous release of magnetic energy in the form of explosive flares.

This had been demonstrated by measurements on the orientation of magnetic field in active regions from the NASA-MSFC (Marshall Space Flight Center) Vector Magnetograph.

In the image above, the solid blue lines account for the neutral lines in between areas of oppositely charged magnetic fields. Its direction and magnitude – the magnetic field vector components – are denoted by these many line segments. As is conventional understanding, the line segments loop directly across the blue neutral lines from an outward pointing i.e. positive magnetic region to an inward pointing i.e. negative magnetic region. And the length of an individual segment corresponds to the magnetic field strength of the given space. The solid blue neutral lines and the magnetic field line segments happens to fall on a sunspot cluster. And the most interesting find by the NASA-MSFC Vector Magnetograph is that the flares – the

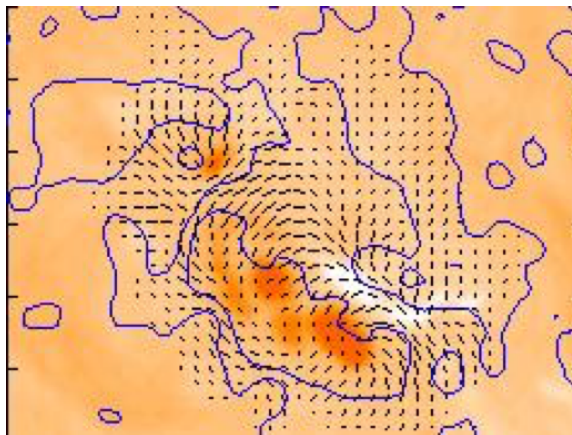


Figure 1.2: Magnetic field structure around an active region of the sun

red regions – evidently falls on the areas where the magnetic field – the line segments orientation – is twisted in such a way that they point along the solid blue neutral lines.

## 1.5 Why They Matter

Life on Earth has evolved to benefit from the sun as a primary source of energy. It thus cannot be absurd to avow that occurrences on our nearest star can impact Earth dwellers. This must certainly be the case with the most powerful of all solar phenomena, with solar flares. The intense X-ray emissions in a particularly large flare lasting for hours and recurring a number of times a day risks a communication black-out. This is because during such a flare, all high-frequency (HF) radio broadcasts such as shortwave Voice of America (VOA) and the British Broadcasting Corporation (BBC) traversing the ionosphere can be absorbed. This is especially so with radio wave frequencies ranging from 3 MHz (Mega-Hertz) to 30 MHz.

However, it is often not just the flares themselves, but also that they serve a visible harbinger producing highly energetic particles from the sun.

Astronauts beyond the protective magnetic field of the Earth would have to contend with the radiation hazards of chromosome damage and even cancer upon prolonged exposure. A significantly high doses could result in instant death. Similarly, satellites encircling the Earth in and above the low-Earth orbit could find their logic circuits for computer memory or control mechanisms upset. Solid state electronics high up could be severely impacted as energetic particles can also damage their space materials. Also, because these energetic particles cannot effectively be shielded off by the magnetic field of the Earth around its poles, they can reach the ionosphere in the regions causing positioning errors on transpolar pathways to the order of kilometers.

As for the situation within the magnetic field of the Earth, there is the possibil-

ity of a geomagnetic storm occurring. Such storms are caused by inhomogeneities in the density, speed and magnetic field of the solar wind particles interacting with the magnetic field of the Earth. Some of these charged particles trapped above the atmosphere can flow around the Earth. This current would generate a magnetic field of its own that can be felt on the Earth. Terrestrial power grids have been disabled for extended periods of time by such storms. Should a satellite fall in their circuit, these could charge the satellite surface coating. The physical properties of the surface materials could be altered. Also, the energy of these particles deposited in the upper-atmosphere of the earth during a geomagnetic storm in turn can cause an ionospheric storm. This would again impact all radio communication, now in a larger frequency bracket, ranging from 3 KHz (Kilo-Hertz) to 30 GHz (Giga-Hertz). Many of our satellite communication systems operate in this range.

These vulnerabilities in our global infrastructure upon which our reliance has steadily grown warrant predicting said flares to be one of the most important goals of solar physics.

# Chapter 2

## Applied Notions of Statistics

### 2.1 Probability

The probability of obtaining a given outcome from a process can be interpreted to mean the relative frequency with which the said outcome would occur should the process be repeated under similar conditions and for a large number of times. But it is generally understood as a measure of the likelihood of an outcome being obtained when an experiment is performed. The probability measure of any event is generally assigned a value between 0 and 1 corresponding to the said event's likelihood. The aforementioned process in which all the possible outcomes have already been identified is defined to be an experiment, and an event is the set for some of its possible outcomes. The experiment in our endeavor would thus be observing and measuring the radiation intensity from the sun thorough terrestrial observatories and artificial satellites. The event of interest then is the occurrence of solar flares.

Given  $S$  as our sample space i.e. the collection of all possible outcomes of an experiment, the probability of some event  $A$ , denoted  $Pr(A)$ , must obey three specific axioms.

1. The probability of every event must be nonnegative,  $Pr(A) \geq 0$ .
2. The probability of a certain event is 1,  $Pr(S) = 1$ .
3. Events that cannot occur on the same time are defined to be disjoint or mutually exclusive events. For any sequence of such events  $\{A_1, A_2, A_3, \dots\} \subset S$ , the probability of one or any other event in the sequence occurring is the sum of their individual probabilities,  $Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$ .

It follows from the third axiom that the probability of an impossible event is 0, i.e.  $Pr(\phi) = 0$ . And based on these axioms and their corollaries, we can have other general properties of probability.

1. For any event  $A$ ,  $Pr(\text{not event } A) = Pr(A^c) = 1 - Pr(A)$ .

2. If event A implies event B, i.e.  $A \subseteq B$ , then  $Pr(A) \leq Pr(B)$ .
3. For every event A,  $0 \leq Pr(A) \leq 1$ .
4. For any two events A and B,  
 $Pr(A \text{ and not } B) = Pr(A \cap B^c) = Pr(A) - Pr(A \cap B)$ .
5. For any two events A and B,  
 $Pr(A \text{ or } B) = Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$ .

## 2.2 Conditional Probability

The probability of an event A after event B has occurred – denoted  $Pr(A|B)$  – is defined to be the conditional probability of A given B. Provided  $Pr(B) > 0$ , it is determined as

$$Pr(A|B) = \frac{Pr(A \cap B)}{Pr(B)}.$$

Conditional probabilities, just like probabilities, obey all of the above stated axioms. The conditional probability is not defined when  $Pr(B) = 0$ .

From this definition we derive the *Multiplication Rule for Conditional Probabilities*. It states that if  $Pr(B) > 0$ , then

$$Pr(A \cap B) = Pr(B) Pr(A|B),$$

and if  $Pr(A) > 0$ , then

$$Pr(A \cap B) = Pr(A) Pr(B|A).$$

Now suppose we consider k disjoint events  $\{B_1, B_2, B_3, \dots, B_k\}$  such that they form a partition on of our sample space S i.e.  $\bigcup_{i=1}^k B_i = S$ , and  $Pr(B_i) > 0$  for all  $i = 1, 2, 3, \dots, k$ . Then the *Law of Total Probability* states that for every event A in S,

$$Pr(A) = \sum_{i=1}^k Pr(B_i) Pr(A|B_i).$$

From this, we can also ascertain the *Conditional Version of the Law of Total Probability* for another event C which states,

$$Pr(A|C) = \sum_{i=1}^k Pr(B_i|C) Pr(A|B_i \cap C).$$

Sometimes, the occurrence of event B does not impact the probability of event A. This occurs when events A and B are independent and

1.  $Pr(A|B) = Pr(A)$ .

2.  $Pr(A \cap B) = Pr(A) Pr(B)$ .

More generally, if we have a number of mutually independent events  $B_1, B_2, B_3, \dots, B_k$ , then for every corresponding subset of  $j$  ( $j = 2, 3, 4, \dots, k$ ) of these events  $B_{i,1}, B_{i,2}, \dots, B_{i,k}$ ,

$$Pr(B_{i,1} \cap \dots \cap B_{i,j}) = Pr(B_{i,1}) \dots Pr(B_{i,j}).$$

For instance, for three independent events A, B and C, all of the following conditions are satisfied:

1.  $Pr(A \cap B) = Pr(A) Pr(B)$ .
2.  $Pr(A \cap C) = Pr(A) Pr(C)$ .
3.  $Pr(B \cap C) = Pr(B) Pr(C)$ .
4.  $Pr(A \cap B \cap C) = Pr(A) Pr(B) Pr(C)$ .

Again, the conditional probability for such larger collections of independent events holds. For  $B_1, B_2, B_3, \dots, B_k$  events such that  $Pr(B_{i,1} \cap \dots \cap B_{i,j}) > 0$ . Then  $B_1, B_2, B_3, \dots, B_k$  are independent if and only if for every two disjoint subsets  $i_1, \dots, i_m$  and  $j_1, \dots, j_l$  of  $1, \dots, k$ , then

$$Pr(B_{i,1} \cap \dots \cap B_{i,m} | B_{j,1} \cap \dots \cap B_{j,l}) = Pr(B_{i,1} \cap \dots \cap B_{i,m}).$$

The said  $k$  events are independent if and only if knowledge of some events occurring does not in any way impact the probability of any combination of the other events occurring.

To a lesser extent, the events  $B_1, B_2, B_3, \dots, B_k$  are *Conditionally Independent* given A if for its every subset  $B_{i,1}, \dots, B_{i,j}$  where  $j = 1, 2, \dots, k$  provided  $Pr(A) > 0$  is

$$Pr(B_{i,1} \cap \dots \cap B_{i,m} | A) = Pr(B_{i,1} | A) \dots Pr(B_{i,m} | A).$$

The definition also accordingly implies that events  $B_1$  and  $B_2$  are conditionally independent if and only if  $Pr(B_2 | B_1 \cap A) = Pr(B_2 | A)$  given  $Pr(B_1 | A) > 0$ .

To determine the probability for some of these mutually independent events, say  $B_1, B_2, B_3, \dots, B_k$ , that partition the sample space S such that  $Pr(B_j) > 0$  for  $j = 1, \dots, k$ , and A be an event such that  $Pr(A) > 0$ , then for  $i = 1, 2, 3, \dots, k$ ,

$$Pr(B_i | A) = \frac{Pr(B_i)Pr(A|B_i)}{\sum_{i=1}^k Pr(B_i)Pr(A|B_i)}.$$

This is according to *Bayes' Theorem* and can be derived from the very definition of conditional probability.

Now  $Pr(B_i)$  is the Prior Probability because it is the probability of the event  $B_i$  before the occurrence and impact of event A has been take into consideration.

Correspondingly, the probability  $Pr(B_i|A)$  would be the Posterior Probability of event  $B_i$  since it is the probability of the event after the consequences of event  $A$  have also been taken into consideration. In our project, the phenomenological rules that govern solar flare occurrence. The posterior is then ascertained after taking all the additional data into consideration so as to obtain better flare predictability.

## 2.3 Random Variables Distributions

A random variable is a real-valued function defined on a sample space that assigns a numerical value to the outcome of a random experiment. For each random variable  $X$  and each subset  $R$  of real numbers, we can determine the probability that  $X$  takes a value in  $R$ . The collection of all these probabilities is the distribution of  $X$ . The two categories of distributions and random variables we are interested in is either *Discrete* or *Continuous*.

A random variable  $X$  has a *discretedistribution* if it can only assume a finite or countably infinite positive values.  $X$  is then referred to as a discrete random variable. The probability (mass) function or p.m.f. of  $X$  is a real valued function  $f$  for every real number  $x$  where,

$$f(x) = Pr(X = x).$$

The *support of (distribution of)  $X$*  is the set of values in  $X$  that have non-zero probability values. For any value that is not contained in  $X$  i.e.  $y \notin X$ ,  $Pr(y) = 0$ . Also, if  $X = \{x_1, x_2, x_3, x_4, \dots\}$ ,

$$\sum_{i=1}^{\infty} f(x_i) = 1.$$

Similarly, a random variable  $X$  is believed to have a *continuous distribution* if there is a non-negative function  $f$ , defined on the real line, where for every bounded or non-bounded interval of real values, the probability that  $X$  takes a value in the interval is the integral of  $f$  over the interval.  $X$  is then referred to as a *continuousrandom* variable. Say there is a bounded and closed interval in  $X$  i.e.  $[a, b] \subseteq X$ , then

$$Pr(aXb) = \int_{x=a}^{x=b} f(x)dx.$$

Similarly,

$$Pr(xb) = \int_{x=-\infty}^{x=b} f(x)dx,$$

and,

$$Pr(xa) = \int_{x=a}^{x=\infty} f(x)dx.$$



The function  $f$  from above is called the *probability density function (p.d.f.)* of  $X$ . The *support of (the distribution of)*  $X$  is the closure of the set  $\{x : f(x) > 0\}$ . One thing that sets a continuous distribution apart from a discrete distribution is that the former assigns probability 0 to any individual value.

In either case, we define the Cumulative Distribution Function (c.d.f.)  $F$  of a random variable  $X$  is the function,

$$F(x) = Pr(X \leq x) \text{ for, } -\infty < x < \infty.$$

The distributions discussed so far have one random variable  $X$ . It is possible to generalize this concept to an arbitrary finite number  $n$  of random variables  $X_1, X_2, X_3, \dots, X_n$ . This joint distribution of more than two random variables is called *multivariate distribution*, if  $n = 2$  the it is referred to as *bivariate distribution*. The *joint distribution function (c.d.f.)* of  $n$  random variables  $X_1, X_2, X_3, \dots, X_n$  is the function  $F$  whose value at every point  $(x_1, x_2, \dots, x_n)$  in  $n$ -dimensional space  $R^n$  specified as

$$F(x_1, \dots, x_n) = Pr(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n).$$

If the  $n$  random variables  $X_1, X_2, X_3, \dots, X_n$  have a *discrete joint distribution*, the *joint probability (mass) function* or *p.m.f.* of  $X_1, X_2, X_3, \dots, X_n$  is the function  $f$  such that for every point  $(x_1, x_2, \dots, x_n) \in R^n$  defined as,

$$f(x_1, \dots, x_n) = Pr(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

or,

$$f(x) = Pr(X = x)$$

in vector notation where  $X = (X_1, X_2, X_3, \dots, X_n)$ , and  $x = (x_1, x_2, x_3, \dots, x_n)$ . Then for every subset  $A \subset R^n$ ,

$$Pr(X \in C) = \sum_{x \in C} f(x).$$

Similarly suppose the  $n$  random variables  $X_1, X_2, X_3, \dots, X_n$  have a *continuous joint distribution*, then there exists a nonnegative function  $f$  defined on  $R^n$  such that for every subset  $C \subset R^n$ ,

$$Pr[(X_1, \dots, X_n) \in C] = \int \cdots \int f(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

provided the integral exists. The function  $f$  serves as the *joint probability density function (p.d.f.)* of  $X_1, \dots, X_n$ . And similarly as was the case with the vector notation for the joint discrete distribution, the continuous joint distribution in vector notation can be expressed as,

$$Pr(X \in C) = \int \cdots \int f(x) dx.$$

And if the joint distribution of  $n$  random variables  $X_1, \dots, X_n$  is known, then the *Marginal Distribution* – the probability distribution of each variable contained in the subset  $(X_1, \dots, X_n)$  of  $R^n$  – can be derived from the joint distribution above. For instance, the marginal p.d.f.  $f_i$  of  $X_i$  is specified at every value  $x_i$  by the equation

$$f_i(x_i) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

integrated over  $n - 1$  variables except  $x_i$ .

If the  $n$  random variables  $X_1, \dots, X_n$  are independent if, for every  $n$  sets  $C_1, C_2, \dots, C_n$  of real numbers,

$$Pr(X_1 \in C_1, X_2 \in C_2, \dots, X_n \in C_n) = Pr(X_1 \in C_1) Pr(X_2 \in C_2) \dots Pr(X_n \in C_n).$$

And if  $F_i$  were to denote the marginal c.d.f. of  $X_i$  for  $i = 1, 2, \dots, n$ , then the variables  $X_1, \dots, X_n$  are *independent* if for any point  $(x_1, \dots, x_n) \in R^n$ ,

$$F(x_1, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n).$$

Similarlly, if the random variables  $X_1, \dots, X_n$  have a continuous, discrete or a mixed joint distribution for which the joint p.d.f., joint p.m.f. or joint p.d.f. is  $f$ , and if  $f_i$  is the marginal univariate p.d.f. or p.m.f. of  $X_i$  ( $i = 1, 2, \dots, n$ ), then  $X_1, \dots, X_n$  are independent if and only if their joint p.d.f./p.m.f. is the product of their  $n$  individual p.d.f.'s or p.m.f.'s. So then for any point  $(x_1, \dots, x_n) \in R^n$ ,

$$f(x_1, \dots, x_n) = f_1(x_1) f_2(x_2) f_3(x_3) \dots f_n(x_n).$$

Suppose the random variables vector  $X = (X_1, \dots, X_n)$  is divided into a  $k$ -dimensional random variable vector  $Y$ , and a  $(n - k)$ -dimensional variable random vector  $Z$ . Suppose also that the joint p.d.f. or p.m.f. of  $(Y, Z)$  is  $f_{y,z}$ , and that the marginal distributions of  $Y$  and  $Z$  are  $f_y$  and  $f_z$  respectively, then their relationship with the conditional p.d.f. or p.m.f.  $f_{y|z}$  of  $Y$  given  $Z = z$  is defined as:

$$f_{z|y}(zy) = \frac{f_{(y|z)}(y|z) f_z(z)}{(f_y(y))}.$$

The *Multivariate Law of Total Probability*, *Bayes' Theorem* and the conditions given above provided  $Z$  has a continuous joint distribution, then the marginal p.d.f. of  $Y$  is

$$f_y(y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{y|z}(y|z) f_z(z) dz.$$

And suppose we have another random vector  $W = w$  satisfying the prerequisite conditions, then

$$f_y(y|w) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{y|z,w}(y|z, w) f_z(z|w) dz,$$

where  $f_{y|w}(y|w)$  is the conditional p.d.f. or p.m.f. of  $Y$  given  $W = w$ ,  $f_{y|z,w}(y|z, w)$  is the conditional p.d.f. or p.m.f. of  $Y$  given  $(Z, W) = (z, w)$  and  $f_z(z|w)$  is the conditional p.d.f. or p.m.f. of  $Z$  given  $W = w$ . Then again we have the following result:

$$f_{(z|y,w)}(z|y, w) = \frac{f_{(y|z,w)}(y|z, w)f_z(z)}{f_{(y|w)}(y|w)}.$$

And for the random variable vector  $Z$  with joint p.d.f. or p.m.f. given by  $f_z(z)$ , suppose the random variables  $X_1, \dots, X_n$  are all *conditionally independent* given  $Z$ , then for all  $z \in Z$  such that  $f_z(z) > 0$ , we have:

$$f_{x,z}(x, z) = \prod_{i=1}^n f_i(x_i|z),$$

where  $f_{x,z}(x, z)$  is the conditional multivariate p.d.f. or p.m.f. of  $X$  given  $Z = z$  and  $f_i(x_i|z)$  is the conditional univariate p.d.f. or p.m.f. of  $X_i$  given  $Z = z$ .

Now suppose  $X$  has a discrete distribution with p.m.f.  $f$ , and let the other random variable be  $Y = r(X)$ , where  $r$  is some function defined on the set of all permissible values for  $X$ . Then for any  $y \in Y$ , the p.m.f. of  $Y$ ,  $f_y(y)$  is defined as:

$$f_y(y) = Pr(Y = y) = Pr[r(X) = y] = \sum_{x:r(x)=y} f(x).$$

The result can be generalized to  $m$  functions  $Y_1, Y_2, \dots, Y_m$  where  $Y_i = r_i(X)$  for  $i = 1, 2, \dots, m$ . Then the joint p.d.f. of  $Y_1, Y_2, \dots, Y_m$  for any specified point  $y = (y_1, y_2, \dots, y_m)$  is determined by,

$$f_y(y) = \sum_{x \in A} f(x_1, x_2, \dots, x_n),$$

where  $A$  is the set of all possible points  $x = (x_1, x_2, \dots, x_n)$  satisfying the  $m$  functions.

Now suppose  $X$  has a continuous distribution with p.d.f.  $f$  and let  $Y = r(X)$ . Then for each real value  $y$ , the c.d.f.  $G(y)$  of  $Y$  is determined via.

$$G(y) = Pr(Y \leq y) = Pr[r(X) \leq y] = \int_{x:r(x) \leq y} f(x)dx.$$

And should the random variable  $Y$  itself also have a continuous distribution, then its p.d.f. at every point  $y \in Y$  where  $G$  is differentiable can be obtained by from the expression:

$$g(y) = G'(y) = \frac{d}{dy}G(y).$$

## 2.4 Bayesian Statistics

Sometimes you'd like to refer to a table or figure, e.g. you can see in Figure ?? that you can rotate figures . Start by labeling your figure or table with the label command

(`\label{labelvariable}`) below the caption (see the chapter on graphics and tables for examples). Then when you would like to refer to the table or figure, use the `ref` command (`\ref{labelvariable}`). Make sure your label variables are unique; you can't have two elements named "default." Also, since the reference command only puts the figure or table number, you will have to put "Table" or "Figure" as appropriate, as seen in the following examples:

# Chapter 3

## Wheatland's Bayesian Approach

### 3.1 The Phenomenological Rules

By phenomenological rules we are referring specifically to the laws governing the various aspects of solar flare occurrence based on observations from solar telescopes. The first rule proposed by James F. Drake in 1971 suggests that flare events follow a power-law size distribution where by ‘size’ refers to the peak flux in soft X-ray or extreme ultraviolet in terms of estimated energy. Drake’s flare *’Frequency-Size Distribution’* would translate as:

$$N(S) = AS^{-\gamma} \quad (3.1)$$

where ‘ $N(S)$ ’ is the distribution for the number of flare events per unit size ‘ $S$ ’ and per unit time, with constants ‘ $A$ ’ and ‘ $\gamma$ ’. The constant ‘ $A$ ’ is the overall flaring rate, and the exact power-law index ‘ $\gamma$ ’ is a constant estimated by Aschwanden et al. in 2000 to be 1.8 and certainly less than 2. But it was already understood to have a constant value lying somewhere in between 1.5 and 2 according to Crosby et al. in 1993. This was so that the mean flare size is finite. Bai in 1993 uncovered some indication that the power-law index varies with the solar cycle. But Wheatland in 2000 revealed its value to be mostly constant irrespective of the active region under consideration.

There are however numerable inconsistencies in the model proposed above.

As noted above, Bai in 1993 showed that the occurrence frequency of flares as a function of hard X-ray peak count rate varies. The power-law index of size distribution was found to change with time and with the phase of a 154-day periodicity. Kucera et al. in 1997 found fewer high count rate flares (i.e.  $\leq 104$  counts per second

and size above 60 keV) from regions with small sunspot areas (ranging from 0 to 500 microhemispheres) than the power-law extrapolation from the frequency distribution of flares with peak rates higher than 50 counts per second above 60 keV. A lower proportion low-energy events was seen in large complex regions than in smaller, simpler regions. Furthermore, Ian Sammis in 1999 discovered that all active regions do not share a common power-law distribution i.e. local distribution was not consistent with the global distribution. Finally, Wheatland in 2000 studied the small and highly active region 11029 and noted its departure from the power-law behavior above  $10Wm^{-2}$  to  $6Wm^{-2}$ . Wheatland attributed this to the small active region having a finite amount energy.

Another candidate governs the '*Waiting-Time Distribution*' i.e. the way flares occur in time. It was first modeled as a Poisson process in time by Rosner and Vaiana in 1978. The model assumes individual flares to be independent of each other, and the distribution for events occurring between times  $\tau$  and  $\tau + d\tau$  i.e.  $P(\tau)d\tau$ , where

$$P(\tau) = \lambda e^{-\lambda\tau} \quad (3.2)$$

where ' $\lambda$ ' is the constant mean rate of occurrence or frequency. We note here that the waiting-times of the Poisson process follows an exponential distribution.

But several observations were shown to conflict with the given poisson model. Biesecker in 1994 found the waiting-time distribution for hard X-rays to be consistent with a time-dependent Poisson process. Whereas the waiting-time distribution was shown to obey follow the power law via. statistical analysis of observational data by Dennis in 1985, and by Crosby et al. and Pearce et al. independently 1993. Furthermore, using data observed by GOES sensors between 1976 and 1996, Boffetta et al. in 1999 revealed that for waiting times longer than 10 hours, it follows a power-law. But then Wheatland in 2000 showed that the distribution of GOES flares is consistent with a time-dependent Poisson process. A new phenomenological rule for solar flares, that the time distribution of rates of the flares averaged over several solar cycles is approximately exponential, was also presented.

The Poisson model was then restructured into a time-dependent Poisson process by Wheatland and Litvinenko in 2002. Independent events with time varying mean rate ' $\lambda(\tau)$ ' was accounted for. And for a varying mean rate given by some distribution, say ' $f(\tau)$ ', the corresponding waiting-time distribution between times  $\tau$  and  $\tau+d\tau$  i.e.  $P(\tau)d\tau$ , where

$$P(\tau) = \frac{1}{\lambda} \int_0^{\infty} f(\lambda) \lambda^2 e^{-\lambda\tau} d\lambda \quad (3.3)$$

The GOES data found the distribution for the slowly varying ‘ $f(\lambda)$ ’ approximation to be valid. But then again, they also noted the deviation in the power-law index with the solar cycle variation. That very year, Moon et al. studied sympathetic Coronal Mass Ejections. Sympathetic events are solar eruptions and flares that occur almost simultaneously in one complex region or in different active regions with a physical link. Although the results indicated that sympathetic events are rarer than their anti-sympathetic counterparts, they did find a candidate sympathetic event pair where the first event may trigger the second flare occurrence. This starkly contradicts the central tenet of a Poisson process where events are assumed to be independent of each other.

## 3.2 Previous Method

## 3.3 New Method

### 3.3.1 Process

### 3.3.2 Estimating the Power Law Index - ‘ $\gamma$ ’

### 3.3.3 Estimating the Rate of Occurrence - ‘ $\lambda$ ’

### 3.3.4 Estimating the Probability that At least One Large Flare Event Occurs - ‘ $\epsilon$ ’

## 3.4 Simulation

### 3.4.1 Exploratory Data Analysis

### 3.4.2 Applying the Changepoint Analysis

### 3.4.3 Simulation Outcome





# Conclusion

Here's a conclusion, demonstrating the use of all that manual incrementing and table of contents adding that has to happen if you use the starred form of the chapter command. The deal is, the chapter command in  $\text{\LaTeX}$  does a lot of things: it increments the chapter counter, it resets the section counter to zero, it puts the name of the chapter into the table of contents and the running headers, and probably some other stuff.

So, if you remove all that stuff because you don't like it to say "Chapter 4: Conclusion", then you have to manually add all the things  $\text{\LaTeX}$  would normally do for you. Maybe someday we'll write a new chapter macro that doesn't add "Chapter X" to the beginning of every chapter title.



# Appendix A

## The First Appendix



## Appendix B

The Second Appendix, for Fun



# Bibliography

- Angel, E. (2000). *Interactive Computer Graphics : A Top-Down Approach with OpenGL*. Boston, MA: Addison Wesley Longman.
- Angel, E. (2001a). *Batch-file Computer Graphics : A Bottom-Up Approach with QuickTime*. Boston, MA: Wesley Addison Longman.
- Angel, E. (2001b). *test second book by angel*. Boston, MA: Wesley Addison Longman.
- Deussen, O., & Strothotte, T. (2000). Computer-generated pen-and-ink illustration of trees. *“Proceedings of” SIGGRAPH 2000*, (pp. 13–18).
- Fisher, R., Perkins, S., Walker, A., & Wolfart, E. (1997). *Hypermedia Image Processing Reference*. New York, NY: John Wiley & Sons.
- Gooch, B., & Gooch, A. (2001a). *Non-Photorealistic Rendering*. Natick, Massachusetts: A K Peters.
- Gooch, B., & Gooch, A. (2001b). *Test second book by gooches*. Natick, Massachusetts: A K Peters.
- Hertzmann, A., & Zorin, D. (2000). Illustrating smooth surfaces. *Proceedings of SIGGRAPH 2000*, 5(17), 517–526.
- Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Molina, S. T., & Borkovec, T. D. (1994). The Penn State worry questionnaire: Psychometric properties and associated characteristics. In G. C. L. Davey, & F. Tallis (Eds.), *Worrying: Perspectives on theory, assessment and treatment*, (pp. 265–283). New York: Wiley.
- Noble, S. G. (2002). *Turning images into simple line-art*. Undergraduate thesis, Reed College.
- Reed College (2007). Latex your document. <http://web.reed.edu/cis/help/LaTeX/index.html>
- Russ, J. C. (1995). *The Image Processing Handbook, Second Edition*. Boca Raton, Florida: CRC Press.

- Salisbury, M. P., Wong, M. T., Hughes, J. F., & Salesin, D. H. (1997). Orientable textures for image-based pen-and-ink illustration. *“Proceedings of” SIGGRAPH 97*, (pp. 401–406).
- Savitch, W. (2001). *JAVA: An Introduction to Computer Science & Programming*. Upper Saddle River, New Jersey: Prentice Hall.
- Wong, E. (1999). *Artistic Rendering of Portrait Photographs*. Master’s thesis, Cornell University.