October 20: Zhi Chen (Graduate Research Assistant, Civil & Environmental Engineering Department)
A multivariate Poisson-lognormal regression model for crashes by vehicle travelling direction
Abstract: Crash modeling is an important technique to apply statistical methods for crash analysis and safety countermeasure identification. Recent research shows that it may be more appropriate to model the crash frequency at an intersection by vehicle travelling directions because this classification distinguishes various vehicular conflicts. Since crashes of different types can be affected by a variety of factors unavailable in the model due to data collection limitation such as weather, driver population and land use characteristic, a univariate model that considers each crash type separately is unable to account for the correlation between crash types. In this study, a multivariate Poisson-lognormal (MVPLN) regression model has been employed to simultaneously model four collision types: opposite-direction, same-direction, intersection-direction crashes, and single-vehicle crashes. The MVPLN model assumes the heterogeneous multivariate Poisson means follow lognormal distribution. The estimation results show the significant correlations between collision types and data overdispersion in each collision type. To assess the performance of the MVPLN model, the results were compared with its univariate counterpart such as the univariate Poisson model (UVP) and the univariate negative binomial model (UVNB).

October 27: Chris Saunders (Assistant Professor, Mathematics and Statistics)
An Overview of the Forensic Identification of Source Problem
Abstract: A goal in the forensic interpretation of scientific evidence is to make an inference about the source of a trace of unknown origin. This inference process usually considers two alternative and mutually exclusive, but non-exhaustive, propositions. The first proposition usually corresponds to the prosecution hypothesis and states that a given specific source is the actual source of the trace of unknown origin. The second proposition usually corresponds to the defense hypothesis and states that the actual source of the trace is not the one considered under the prosecution hypothesis, but that it originates from another, unrelated, source in a specified relevant alternative population of sources.

The evidence a forensic scientist is given for deciding between these two propositions usually takes the following form: (1) the trace of unknown origin, (2) a sample from the specific source specified by the prosecution hypothesis, and (3) a collection of samples from the alternative source population specified by the defense proposition. In certain applications, the choice of the alternative source population will be mandated by available databases or in extreme situations there will be no such samples available.

In this presentation and discussion we will review some of the common statistical approaches that forensic scientists use to characterize the support that the evidence provides for deciding between the prosecution and defense propositions.

November 3: Danica Ommen (PhD student, Computational Science and Statistics)
Computational and Statistical Aspects of the Forensic Identification of Source Problem
Abstract: The main goal of most legal trials is to evaluate all the evidence in order to determine the guilt or innocence of the suspect. Bayesian methods of forensic evidence interpretation require the statement of two mutually exclusive, but non-exhaustive, models about how the evidence in a given situation has arisen. One model corresponds to the defense and the other corresponds to the prosecution. Commonly, the information that a forensic scientist has available to decide between the two models is called the evidence. The forensic scientist is required to interpret and present the value of the evidence, typically in the form of a Bayes Factor or Bayesian likelihood ratio, to aid the decision-maker in choosing between the two competing models.

There are many computational difficulties in calculating the Bayes Factor in real-world situations. The Bayes Factor is a ratio of two marginal likelihoods, which are defined in terms of integrals. It is often the case that the integral cannot be evaluated in closed form, and must be computed numerically. In this case, the marginal
likelihoods are approximated using Monte Carlo integration techniques. However, well-known research in computational statistics and physics has illustrated the inefficiency of the Monte Carlo methods in approximating the Bayes Factor. This inefficiency has a significant negative impact on the implementations of these methods in real world applications.

In addition, we must be able to sample from the distribution for the parameters of interest. If the distribution is unknown, a Gibbs sampler will be used to sample values from the approximate distribution. This will also lead to a reduction in the efficiency of the integration technique. Finally, to demonstrate computing the Bayes Factor where the posterior density is unknown, we will look at a more complex, but well-known, example relating to glass fragments.

November 10: Chris Saunders (Assistant Professor, Mathematics and Statistics)
An Overview of the Forensic Identification of Source Problem, cont.

November 17: Vicki Geiser (Assistant Professor of Biology, Director of Medical Laboratory Sciences, Mount Marty College)
Incorporation of the Department of Energy “Interpret a GEBA Genome” Education Program into an Undergraduate Molecular Biology Course
Abstract: Bioinformatics is widely used in molecular biology and proficiency in this discipline has been linked to success in other areas of biology and chemistry. However, acquiring expertise in bioinformatics requires experiential learning with access to a broad spectrum of bioinformatics tools in order for students to comprehend the complexity and value associated with bioinformatics analysis. To increase undergraduate student aptitude in bioinformatics applications, Mount Marty College established collaborations with the Department of Energy Joint Genome Institute (JGI). Faculty collaborators in the “Interpret a GEBA Genome” education program receive training and have access to the Integrated Microbial Genomes Annotation Collaboration Toolkit (IMG-ACT). This program provides a faculty portal which permits the inclusion of a variety of bioinformatics tools along with virtual notebooks and reports that are utilized by the JGI to update their annotation databases based on the manual annotations performed by undergraduate students. In this report, we describe the incorporation of the “Interpret a GEBA Genome” program into an undergraduate molecular biology course and the subsequent increase in student learning in both bioinformatics and associated content. The laboratory component of the molecular biology course was modified such that genome analysis and gene annotation were the predominant content for the laboratory sessions. Students were assessed for knowledge in bioinformatics tools selection, data interpretation, and associated molecular biology content prior to participation and periodically throughout the program. Overall, inclusion of the program has increased student learning in bioinformatics and molecular biology.

November 24: Doug Armstrong (PhD student, Computational Science and Statistics)
The qualification of the weight of fingerprint evidence when the suspect has been generated by an AFIS
Abstract: The evaluation of the weight of complex evidential forms by means of likelihood ratios (LR), such as fingerprint evidence, requires sophisticated procedures due to the usually large number of dimensions of the variables characterizing the samples. Several models have been proposed that invoke an ad hoc proximity measure between pairs of fingerprint impressions. Unfortunately, the use of proximity measures (or distances) in LR is not straightforward. An additional level of complexity lives in the manner in which the suspect has been generated. The denominator of the LR will be significantly different whether the suspect has been generated through a police investigation or the search of a trace impression in a large database. In this talk, we present a novel model based on multidimensional scaling or large distance matrices obtained by comparing trace, control (i.e. suspect) and reference (database) impressions.

December 1: Ben Roycraft (MS student, Mathematics)
Applications of Algebra in Statistics
Abstract: The possible applications of algebraic techniques to statistics problems are little explored. The field of Algebraic Statistics seeks to apply the tools and language of algebra to solve practical problems. For discrete
problems, in particular, algebra offers great insight. Problems include: Bayesian integration via MCMC, maximum likelihood estimation, experimental design, and dynamic programming. In this talk, we will cover the basics of Algebraic Statistics, along with research topics.

December 8: **Don Vestal** (Associate Professor, Mathematics)

*STAT 792 (Probabilistic Graphical Modeling) – A Preview*

Abstract: Next semester, I’ll be teaching a course on Probabilistic Graphical Modeling (although the class has morphed into a somewhat more general course). In this talk, I’ll give an overview of some of the topics that will be covered in the course. We’ll keep the math to a minimum so as not to reveal any spoilers.