**Astronomy 34000**
**Spring Quarter 2012**

Problem Set 1

Due: Wednesday, April 18

If you have any questions, please contact me *before* the due date by e-mail: frieman@fnal.gov.

1. The skewness and kurtosis of a distribution are defined by $s_3 = \mu_3/(\mu_2)^{3/2}$ and $s_4 = (\mu_4/(\mu_2)^2) - 3$, where $\mu_r = \langle (X-\mu)^r \rangle$ is the $r$th central moment of the distribution. Suppose you flip a coin $n$ times (independently each time) with probability $p$ of getting heads on each flip (i.e., it may not be an unbiased coin). Derive expressions for the skewness and kurtosis of the distribution of the number of heads, in terms of $n$ and $p$.

2. Suppose we have $n$ independent random variables $X_1, ..., X_n$, each with the same mean, i.e., $\langle X_i \rangle = \mu$ for all $i$, but with *different* variances $\sigma_1^2, ..., \sigma_n^2$, where $\sigma_i^2 = \langle X_i^2 \rangle - \mu^2$. This situation often arises when one has $n$ different measurements of the same quantity, each with different errors. Previously, we defined the sample mean by $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$.

   (a) Define a new estimator of the mean by $\tilde{X}_n = \sum_{i=1}^{n} X_i w_i$, where $w_i$ is a weight for variable (or measurement) $i$. Requiring $\tilde{X}_n$ to be an unbiased estimate of $\mu$, derive expressions for the weights $w_i$ which minimize the variance $V(\tilde{X}_n)$ of this estimator, in terms of the mean and variances of the random variables.

   (b) Give expressions for the variances $V(\tilde{X}_n)$ and $V(\bar{X}_n)$ in terms of the variances of the random variables. Compare the results for $n = 2$, with $\sigma_1 = 1, \sigma_2 = 2$.

   (c) Now drop the assumption that the $X_i$ are independent. Letting $e_i = X_i - \mu$, the error covariance matrix is $C_{ij} = \langle e_i e_j \rangle$, with inverse $K = C^{-1}$. For this case, repeat the calculation of part (a), i.e., find the minimum variance expression for the weights $w_i$ and give the answer in terms of the inverse of the error covariance matrix. What is the corresponding expression for the variance of the new estimator of the mean?

3. In a court case in the 1960's, the prosecution was having a difficult time establishing the identity of a couple accused of a burglary in a large city in southern California. The prosecutor had established from witnesses that the couple who had committed the crime had a number of identifying features: the woman had (a) blond hair, (b) a ponytail, (c) their car was partly yellow, (d) the man had a mustache, (e) a beard, (f) they were an interracial couple, etc. Assuming (without justification) that each of these features is independent, the prosecutor

called a mathematics instructor to the stand, had him multiply together the probabilities of each of these features in the general population (probabilities which the prosecutor also estimated without justification) and claimed that the total probability of there being a couple with all of these features was only 1 in 12 million. The prosecutor claimed that this was the chance that the accused couple, which had all these features, was innocent, and they were convicted. The conviction was later overturned on appeal by a judge with better statistical sense.

Even ignoring the prosecutor's unjustified assumption of independent probabilities for these different features, it is worth looking at this issue in more depth. Let $p$ be the probability that a certain distinctive combination of characteristics, designated $C$, will occur jointly in a random couple.

(a) Find the probability that $C$ will occur in none of $N$ couples chosen at random.

(b) Find the probability of $C$ occurring in at least one of $N$ random couples.

(c) Given a particular couple selected from a random set of $N$ couples, find the probability that $C$ will occur in the selected couple and in no other couple.

(d) Find the probability that $C$ will occur in exactly one of the $N$ random couples, without regard to which one.

(e) Find the probability that $C$ will occur in more than one couple.

(f) Find the probability that $C$ will occur more than once in a group of $N$ couples in which $C$ occurs at least once.

(g) Let $\mu = Np$, and for this last probability take the limit $N \to \infty$ but keeping $\mu$ fixed.

(h) Assume $p$ is indeed 1 in 12 million and that the number of couples who conceivably could have been near the scene at the time was 12 million (it's a large city), i.e., $\mu = 1$. What is the probability that another couple with the same distinguishing features as above could have been in the area at the time of the crime?

4. Here is an example of Bayesian inference. Ralph gets tested for a certain disease. Denote Ralph's state of health by the variable $a$: if Ralph has the disease, then $a = 1$, if Ralph does not have the disease, then $a = 0$. Let $b$ denote the test result: the test result is either 'positive' ($b = 1$) or 'negative' ($b = 0$). We are told that the test is 95% reliable: in 95% of cases of people who really have the disease, a positive result is returned, and in 95% of cases of people who do not have the disease, a negative result is obtained. The final piece of information we have is that 1% of people of Ralph's age and background have the disease.

   Ralph has the test, and the result is positive. What is the probability that he has the disease?