ASTP 611-01: Statistical Methods for Astrophysics

Problem Set 4

Assigned 2014 February 20
Due 2014 February 27

Show your work on all problems! Be sure to give credit to any collaborators, or outside sources used in solving the problems. Note that if using an outside source to do a calculation, you should use it as a reference for the method, and actually carry out the calculation yourself; it's not sufficient to quote the results of a calculation contained in an outside source.

1 Poisson Distribution

Consider a Poisson process with rate r. Let k_1 be the number of events occurring between $t = 0$ and $t = T_1$ and k_2 be the number of events occurring between $t = 0$ and $t = T_2$, where $0 < T_1 < T_2$.

a) Use the Poisson distribution to find the following probability mass functions (where I is the background information involved in setting up the problem but does *not* include the specification of the values of r, k_1 or k_2):

   i) The pmf $p(k_1|r, I)$ for the number of events between $t = 0$ and $t = T_1$. For which values of k_1 is it non-zero?

   ii) The pmf $p(k_2|r, I)$ for the number of events between $t = 0$ and $t = T_2$. For which values of k_2 is it non-zero?

   iii) The pmf $p([k_2 - k_1]|r, I)$ for the number of events between $t = T_1$ and $t = T_2$. For which values of $k_2 - k_1$ is it non-zero?

   iv) The conditional pmf $p(k_2|k_1, r, I)$ for the number of events between $t = 0$ and $t = T_2$, given that k_1 events occurred between $t = 0$ and $t = T_1$, where k_1 is a non-negative integer. For which values of k_2 is it non-zero?

b) Use Bayes's theorem to find the conditional pmf $p(k_1|k_2, r, I)$ for the number of events between $t = 0$ and $t = T_1$, given that k_2 events occurred between $t = 0$ and $t = T_2$, where k_2 is a non-negative integer. Simplify your result as much as possible. For which values of k_1 is it non-zero?

c) Your result for part b) should have the form of a binomial distribution. Describe an alternate derivation of the pmf $p(k_1|k_2, r, I)$ using the properties of a Poisson distribution which leads directly to the binomial form.

2 Gaussian Approximation

Many distributions, for certain values of their parameters, can be approximated by a Gaussian distribution with the same mean and variance.

a) The ipython notebook `http://ccrg.rit.edu/~whelan/courses/2014_1sp_ASTP_611/data/ps04.ipynb` produces plots of the probability density or mass functions for various distributions and compares them to the corresponding Gaussian distribution. For the parameters chosen, the approximation is not very good. Execute each cell of the notebook, and see this Make sure you understand what is being done at each step, but don't turn in this version.

b) Modify the parameters of each distribution to find values for which the actual distribution and the Gaussian approximation look identical to the eye. (You may need to change the location of the legend to make the plot visible.) In each case, you should be able to acheive this by modifying one of the parameters, while changing any other parameters doesn't make the approximation any better or worse. Turn in this modified version of the notebook, with all the cells executed. For the distributions listed, indicate which parameter you changed to make the Gaussian approximation valid:

   i) Binomial with n trials and probability α of success on each trial.
   ii) Poisson with mean μ.
   iii) Chi-square with n degrees of freedom.
   iv) Gamma with parameters α and β.

   Why is it not possible to make an exponential distribution look like a Gaussian by changing the parameter λ?

c) For the Poisson and Binomial distributions, we compared the probability mass function $p(x) = P(X = x)$ for a discrete random variable X to the probability density function $f(x)$ for a continuous random variable $\mathcal{X}$. To justify this, work out

   i) The probability $P(k - \frac{1}{2} < X < k + \frac{1}{2})$ where k is an integer, assuming $p(x)$ is non-zero only for integer values.
   ii) The probability $P(k - \frac{1}{2} < \mathcal{X} < k + \frac{1}{2})$, assuming that $f(x)$ is approximately constant on the interval $x \in (k - \frac{1}{2}, k + \frac{1}{2})$.

3 Chi-Square, Gamma and Exponential Distributions

a) We've seen that the exponential distribution with rate parameter λ, which has pdf $f(x|\lambda) = \lambda e^{-\lambda x}$ and cdf $F(X|\lambda) = P(X \leq x|\lambda) = 1 - e^{-\lambda x}$, is a special case of the Gamma distribution with parameters $\alpha = 1$ and $\beta = 1/\lambda$, while the chi-square distribution with ν degrees of freedom is a special case of the Gamma distribution with parameters $\alpha = \nu/2$ and $\beta = 2$. This means that there is a choice of α and β for which the Gamma distribution is simultaneously an exponential distribution *and* a chi-square distribution. Find this α and β and the corresponding parameters λ and ν.

b) Let $\{D_i|i = 1, \ldots n\}$ be a set of n independent Gaussian random variables each with zero mean $\langle D_i \rangle = 0$ and the same variance $\mathrm{Var}(D_i) = \sigma^2$. Let $Y = \sum_{i=1}^{n} D_i^2$, and show that $W = Y/\sigma^2$ is a chi-squared random variable with n degrees of freedom.

c) Show that Y follows a Gamma distribution, and find the parameters α and β. Use this to write the mean $\langle Y \rangle$, variance $\mathrm{Var}(Y)$, and standard deviation $\sqrt{\mathrm{Var}(Y)}$.

d) Show that if $n = 2$, $Y = (D_1)^2 + (D_2)^2$ obeys an exponential distribution, and find the parameter λ. Use this to write the probability that $Y \leq a\sigma^2$ where a is any positive number and $\sigma^2 = \mathrm{Var}(D_1) = \mathrm{Var}(D_2)$.