1060-710
Mathematical and Statistical Methods for Astrophysics

Problem Set 8

Assigned 2009 November 5
Due 2009 November 12

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 Binomial Distribution

Consider a random event that has a probability of α of occurring in a given trial (e.g., detection of a simulated signal by an analysis pipeline, where α is the efficiency), so that

$$p(1|\alpha,1) = p(Y|\alpha) = \alpha \tag{1.1a}$$

$$p(0|\alpha,1) = p(N|\alpha) = 1-\alpha \tag{1.1b}$$

We write $p(k|\alpha,n)$ as the probability that if we do n trials, we will find a "yes" result in k of them. For n trials, there are 2^n possible sequences of yes and no results. The probabilty of a particular sequence of k yes and $n-k$ no results is $\alpha^k(1-\alpha)^{n-k}$, and the number of such sequences for a given k and n is "n choose k", $\binom{n}{k} = \frac{n!}{k!(n-k)!}$, so the probability of exactly k "yes" results in n trials is

$$p(k|\alpha,n) = \binom{n}{k}\alpha^k(1-\alpha)^{n-k} \tag{1.2}$$

a) Show that $p(k|\alpha,n)$ is properly normalized, i.e., that

$$\sum_{k=0}^{n} p(k|\alpha,n) = 1 \tag{1.3}$$

(Note that the sum is from 0 to n rather than from 0 to $n-1$, because the number of "yes" trials k can be any integer between zero and the total number of trials, inclusive.)

b) Show that the expectation value of the number of yes results is $n\alpha$.

$$\langle k \rangle = \sum_{k=0}^{n} k\, p(k|\alpha, n) = n\alpha \tag{1.4}$$

(Hint: factor an $n\alpha$ out of the sum and then change variables in what remains to $n' = n - 1$ and $k' = k - 1$ and show that it sums to unity.)

c) Show that the expected variance in the number of yes results is

$$\left\langle k^2 \right\rangle - \langle k \rangle^2 = n\alpha(1 - \alpha) \tag{1.5}$$

d) Evaluate the expected mean $\langle k/n \rangle$ and standard deviation $\sqrt{\langle (k/n)^2 \rangle - \langle k/n \rangle^2}$ of the fraction k/n of yes trials. (This is not a trick question; n is not a random variable, so you're really just adjusting the scale to get a fraction.)

e) Now consider the Bayesian perspective, where we have done n trials and found a total of k "yes" results, and wish to say something about the efficiency α. Assume a uniform prior on α so that

$$p(\alpha|k, n) \propto p(k|\alpha, n) \propto \alpha^k (1 - \alpha)^{n-k} \tag{1.6}$$

i) Construct $L(\alpha) = \ln p(\alpha|k, n)$ up to an overall constant, calculate $L'(\alpha)$ and find the "maximum posterior" estimate $\hat{\alpha}$ defined by $L'(\hat{\alpha}) = 0$.

ii) Calculate $L''(\alpha)$ and find the error $1/\sqrt{-L''(\hat{\alpha})}$ associated with this estimate of α. Express this first in terms of k and n, and then in terms of n and $\hat{\alpha}$.

iii) Using the exact posterior, find the expectation value

$$\langle \alpha \rangle = \int_0^1 \alpha\, p(\alpha|k, n)\, d\alpha \tag{1.7}$$

(To do this part, you need to work out the normalization of $p(\alpha|k, n)$, which wasn't necessary before.)

iv) Evaluate $\langle \alpha^2 \rangle$ and find the standard deviation

$$\sqrt{\langle \alpha^2 \rangle - \langle \alpha \rangle^2} \tag{1.8}$$

associated with the posterior $p(\alpha|k, n)$, expressed in terms of k and n.

f) Extra credit: explain the significance of your error estimates in the case where $\alpha = 0$ or $\alpha = 1$ in the frequentist case, and $k = 0$ or $k = n$ in the Bayesian case.

2 Marginalization and the Inverse Fisher Matrix

Consider two variables x_1 and x_2 whose joint pdf is a Gaussian with zero mean:

$$p(\mathbf{x}) = \frac{\sqrt{\det \mathbf{F}}}{2\pi} \exp\left[-\frac{1}{2}\mathbf{x}^{\mathrm{T}}\,\mathbf{F}\,\mathbf{x}\right] = \frac{\sqrt{F_{11}F_{22} - F_{12}^2}}{2\pi} \exp\left[-\frac{F_{11}}{2}(x_1)^2 - F_{12}x_1x_2 - \frac{F_{22}}{2}(x_2)^2\right] \tag{2.1}$$

where $\mathbf{F}$ is some symmetric, positive definite matrix.

a) Show that $\mathbf{F}$ is indeed the Fisher matrix.

b) Marginalize over x_2 and show that the resulting pdf for x_1 is a Gaussian whose variance is the $1,1$ component of the inverse Fisher matrix $\mathbf{F}^{-1}$:

$$p(x_1) = \int_{-\infty}^{\infty} p(x_1, x_2)\, dx_2 = \frac{1}{\sqrt{2\pi\,(F^{-1})_{11}}} \exp\left(-\frac{x_1^2}{2(F^{-1})_{11}}\right) \tag{2.2}$$

3 Central Limit Theorem

Consider a uniformly-distributed random variable, with the pdf

$$p(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

a) Calculate

$$\mu_x = \langle x \rangle \tag{3.2a}$$

$$\sigma_x^2 = \langle x^2 \rangle - \mu_x^2 \tag{3.2b}$$

b) What is the pdf $p(z)$ of the random variable $z = (x - \mu_x)/\sigma_x$?

c) Consider

$$X = \sum_{k=0}^{N-1} x_k\ , \tag{3.3}$$

the sum of N independent random variables, each distributed according to (3.1). Calculate

$$\mu_X = \langle X \rangle \tag{3.4a}$$

$$\sigma_X^2 = \langle X^2 \rangle - \mu_X^2 \tag{3.4b}$$

d) Use the Central Limit Theorem to write an approximation for the pdf $p(X)$, valid for large N.

e) Extra credit: experimentally check the validity of this approximation for $N = 20$ by randomly generating a large number of X values (each being the sum of twenty uniform random deviates) and plotting their histogram on the same set of axes as the approximate pdf arising from the Central Limit Theorem.