ASTP 611-01: Statistical Methods for Astrophysics

Problem Set 7

Assigned 2017 October 19
Due 2017 October 31

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 Least Squares and Chi-Squared

Consider measurements $\{x_i\}$ taken at times $\{t_i\} = \{-1, 0, 1, 2\}$. We wish to fit these measurements with a straight-line model with predicted expectation values $\mu_i = \lambda_1 + \lambda_2 t_i$. The model predicts measurments which differ from μ_i by uncorrelated Gaussian errors with standard deviations $\{\sigma_i\} = \{\sqrt{2}, 1, \sqrt{2}, \sqrt{3}\}$.

a) Find the matrix $\mathbf{A}$ describing the linear relationship $\boldsymbol{\mu} = \mathbf{A}\boldsymbol{\lambda}$, i.e.,

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \mathbf{A} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \tag{1.1}$$

b) Since the errors are uncorrelated, the standard deviations are described by a matrix

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_1 & 0 & 0 & 0 \\ 0 & \sigma_2 & 0 & 0 \\ 0 & 0 & \sigma_3 & 0 \\ 0 & 0 & 0 & \sigma_4 \end{pmatrix} . \tag{1.2}$$

Construct the matrix $\mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2}\mathbf{A}$ and find its inverse $\left[\mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2}\mathbf{A}\right]^{-1}$. (Since it is a 2×2 matrix, you should actually be able to invert it by hand.)

c) In class we showed that if the measured values are $\mathbf{x}$, the maximum likelihood estimates of the parameters will be $\hat{\boldsymbol{\lambda}}(\mathbf{x}) = \left[\mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2}\mathbf{A}\right]^{-1} \mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2}\mathbf{x}$. Work out the elements of the matrix appearing for this problem in

$$\begin{pmatrix} \hat{\lambda}_1 \\ \hat{\lambda}_2 \end{pmatrix} = \left[\mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2}\mathbf{A}\right]^{-1} \mathbf{A}^{\mathrm{T}}\boldsymbol{\sigma}^{-2} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \tag{1.3}$$

d) Suppose we measure $\{x_i\} = \{1.07241020, 0.40438919, 2.89906726, 8.98526374\}$. Calculate, to three significant figures,

   i) The best-fit parameters $\hat{\lambda}_1$ and $\hat{\lambda}_2$

   ii) The χ^2 value relating the data to the best-fit model,

$$\chi^2 = (\mathbf{x} - \mathbf{A}\hat{\boldsymbol{\lambda}})^{\mathrm{T}} \boldsymbol{\sigma}^{-2} (\mathbf{x} - \mathbf{A}\hat{\boldsymbol{\lambda}}) \tag{1.4}$$

   iii) The p value, i.e., probability that data generated according to the model would have a χ^2 equal to or higher than the one observed.

2 Student t-distribution from a Bayesian perspective

Consider a random sample of size n drawn from a Gaussian distrition with mean μ and variance σ^2. The joint pdf is

$$f(\{x_i\}|\mu, \sigma, I) = (\sigma\sqrt{2\pi})^{-n} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right) \tag{2.1}$$

In this problem, we will show that, given certain assumptions about the prior probability distribution $f(\mu, \sigma|I)$, the marginalized posterior $f(\mu|\{x_i\}, I) = \int_0^\infty f(\mu, \sigma|\{x_i\}, I)\, d\sigma$ obeys a Student's t-distribution.

a) One sensible prior distribution is a pdf uniform in μ and $\ln\sigma$, which would have the form

$$f(\mu, \sigma|I) \propto \sigma^{-1} \qquad -\infty < \mu < \infty;\ 0 < \sigma < \infty \tag{2.2}$$

In principle, we should actually restrict the ranges of μ and σ to finite intervals like $-A < \mu < B$ and $C^{-1} < \sigma < D$, so that the normalization constant implied by the "$\propto$" above would be finite, but in practice the results we'd get in the limit that A, B, C, and D all went to infinity would be the same as we'll get here by assuming the simpler form above and sweeping the normalization constants under the rug. With that in mind, use Bayes's theorem to write an expression for $f(\mu, \sigma|\{x_i\}, I)$ of the form

$$f(\mu, \sigma|\{x_i\}, I) \propto g(\mu, \sigma; \{x_i\}) \tag{2.3}$$

You may collect any factors independent of μ and σ into the implied proportionality constant. This includes expressions like $f(\{x_i\}|I)$. At the end of the day, we won't need them because normalization requires

$$f(\mu, \sigma|\{x_i\}, I) = \frac{g(\mu, \sigma; \{x_i\})}{\int_0^\infty \int_{-\infty}^\infty g(\mu', \sigma'; \{x_i\})\, d\mu'\, d\sigma'} \tag{2.4}$$

You may also find it convenient for future calculations to use the abbreviation

$$\mathcal{Q} = \sum_{i=1}^{n}(x_i - \mu)^2 \tag{2.5}$$

b) Calculate, up to a proportionality constant, the marginalized pdf

$$f(\mu|\{x_i\}, I) = \int_0^\infty f(\mu, \sigma|\{x_i\}, I)\, d\sigma \tag{2.6}$$

You may find it useful to make the change of variables $\tau = \mathcal{Q}^{1/2}/\sigma$ to do the integral. You may also once again collect any factors in the result independent of μ (and $\mathcal{Q}$, which depends on μ), into an implied proportionality constant. This includes any integrals which do not contain any functions of μ or $\mathcal{Q}$ (e.g., because they've been factored out).

c) Show that

$$\mathcal{Q} = \sum_{i=1}^{n} (x_i - \mu)^2 = n(\overline{x} - \mu)^2 + (n-1)s^2 \tag{2.7}$$

where $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$.

d) By making the substitution (2.7) show that

$$f(\mu|\{x_i\}, I) \propto \left(1 + \frac{n}{n-1}\frac{(\overline{x}-\mu)^2}{s^2}\right)^{-n/2} \tag{2.8}$$

i.e., that $t = (\overline{x} - \mu)/\sqrt{s^2/n}$ satisfies a Student t distribution with $n-1$ degrees of freedom,

$$f(t|\{x_i\}, I) \propto \left(1 + \frac{t^2}{n-1}\right)^{-n/2} \tag{2.9}$$

3 Chi-Squared from Multinomial Distribution

Carry out the following steps to demonstrate that if $\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_k \end{pmatrix}$ is a multinomial random vector with n trials, probabilities $p_1, \dots, p_k$ where $\sum_{i=1}^{k} p_i = 1$, and all of the $\{np_i\}$ are large enough that we can approximate $\mathbf{X}$ as a multivariate normal random vector with the same mean and variance-covariance matrix, then the statistic $Q = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$ approximately obeys a $\chi^2(k-1)$ distribution.

a) As a warmup, consider the case $k = 2$, a binomial distribution. In this case, $X_1 \sim \text{Bin}(n, p_1)$ and $X_2 = n - X_1$. We know that $E[X_1] = np_1$ and $\text{Var}(X_1) = np_1(1-p_1) = np_1p_2$ where we have defined $p_2 = 1 - p_1$ in analogy to the multinomial distribution. If n is large enough that np_1 and np_2 are not too small, we can approximate the distribution of X_1 as a Gaussian distribution with mean np_1 and variance np_1p_2. We can thus make an approximately $\chi^2(1)$-distributed random variable

$$Q = \frac{(X_1 - np_1)^2}{np_1p_2} \tag{3.1}$$

i) Show that this statistic can also be written

$$Q = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} \tag{3.2}$$

(hint: show $|X_1 - np_1| = |X_2 - np_2|$)

ii) Calculate the covariance $\mathrm{Cov}(X_1, X_2)$ and thus the variance-covariance matrix $\mathrm{Cov}(\mathbf{X})$.

b) Now consider a multinomial distribution which has joint pmf

$$p(x_1, x_2, \ldots, x_k) = \begin{cases} \frac{n!}{x_1!x_2!\cdots x_n!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k} & \begin{array}{l} x_1 = 0, 1, 2, \cdots, n;\ x_2 = 0, 1, 2, \cdots, n - x_1; \\ x_{k-1} = 0, 1, 2, \cdots, n - x_1 - x_2 - \cdots - x_{k-2}; \\ x_k = n - x_1 - x_2 - \cdots - x_{k-1} \end{array} \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

and joint mgf for all k multinomial random variables is $M_{\mathbf{X}}(t_1, \ldots, t_k) = (p_1\, e^{t_1} + \cdots + p_k\, e^{t_k})^n$.

i) Use the mgf to find the mean $\boldsymbol{\mu} = E[\mathbf{X}]$ and variance-covariance matrix $\boldsymbol{\sigma}^2 = \mathrm{Cov}(\mathbf{X})$. (The latter can be found by calculating a typical diagonal element like $\mathrm{Var}(X_1)$ and a typical off-diagonal element $\mathrm{Cov}(X_1, X_2)$ and generalizing.)

ii) Assume that we can treat $\mathbf{X}$ approximately as a $N_k(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ multivariate normal random vector. Define a random vector $\mathbf{Y} = \begin{pmatrix} \frac{X_1 - np_1}{\sqrt{np_1}} \\ \vdots \\ \frac{X_k - np_k}{\sqrt{np_k}} \end{pmatrix}$, and by writing $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$, show that $\mathbf{Y}$ can be treated approximately as a $N_k(\mathbf{0}, \mathbf{1} - \mathbf{w}\mathbf{w}^{\mathrm{T}})$ multivariate normal random vector, where $\mathbf{w} = \begin{pmatrix} \sqrt{p_1} \\ \vdots \\ \sqrt{p_k} \end{pmatrix}$

iii) Show that $\mathbf{w}^{\mathrm{T}}\mathbf{w} = 1$ and use this to show that $\mathbf{1} - \mathbf{w}\mathbf{w}^{\mathrm{T}}$ is a projector onto the $k-1$ dimensional subspace perpendicular to $\mathbf{w}$. Use this to show that, analogous to the proof of point 2 of Student's theorem given in the lecture notes, $Q = \mathbf{Y}^{\mathrm{T}}\mathbf{Y}$ can be treated approximately as a $\chi^2(k-1)$ random variable.