Solutions to the Homework in Lecture 6
Fundamentals of Observational Cosmology (II)

These notes provide worked solutions to the homework problems in Lecture 6. We use the same conventions as in the lecture notes, in particular the sample-statistics formulas with the factor $1/(N-1)$.

Problem 1. Bayes practice

Question. Reproduce the phone-factory calculation in the lecture notes. Then change the defect rate of factory C from 0.020 to 0.012 and recompute $P(A \mid D)$, $P(B \mid D)$, and $P(C \mid D)$.

Solution. Let D denote the event that a phone is defective. The factory fractions are

$$P(A) = 0.35, \qquad P(B) = 0.35, \qquad P(C) = 0.30.$$

The original defect rates are

$$P(D \mid A) = 0.015, \qquad P(D \mid B) = 0.010, \qquad P(D \mid C) = 0.020.$$

Therefore the total defect probability is

$$P(D) = \sum_i P(D \mid i)P(i) = 0.015 \times 0.35 + 0.010 \times 0.35 + 0.020 \times 0.30 = 0.01475.$$

Bayes' theorem gives

$$P(A \mid D) = \frac{P(D \mid A)P(A)}{P(D)} = \frac{0.015 \times 0.35}{0.01475} \approx 0.356,$$

$$P(B \mid D) = \frac{0.010 \times 0.35}{0.01475} \approx 0.237, \qquad P(C \mid D) = \frac{0.020 \times 0.30}{0.01475} \approx 0.407.$$

So the original posterior probabilities are

$$\boxed{P(A \mid D) \approx 0.356, \qquad P(B \mid D) \approx 0.237, \qquad P(C \mid D) \approx 0.407.}$$

Now change the factory-C defect rate to

$$P(D \mid C) = 0.012.$$

Then

$$P(D) = 0.015 \times 0.35 + 0.010 \times 0.35 + 0.012 \times 0.30 = 0.00525 + 0.00350 + 0.00360 = 0.01235.$$

Hence

$$P(A \mid D) = \frac{0.015 \times 0.35}{0.01235} = \frac{0.00525}{0.01235} \approx 0.425,$$

$$P(B \mid D) = \frac{0.010 \times 0.35}{0.01235} = \frac{0.00350}{0.01235} \approx 0.283,$$

$$P(C \mid D) = \frac{0.012 \times 0.30}{0.01235} = \frac{0.00360}{0.01235} \approx 0.291.$$

Therefore, after lowering the defect rate of factory C, we obtain

$$\boxed{P(A \mid D) \approx 0.425, \qquad P(B \mid D) \approx 0.283, \qquad P(C \mid D) \approx 0.291.}$$

This makes sense: once factory C becomes less defective, a defective phone is less likely to have come from C.

Problem 2. Sample statistics

Question. For the data vectors

$$X = (2, 4, 5, 7, 8), \qquad Y = (1, 3, 4, 6, 8),$$

compute the sample mean, variance, covariance, and correlation coefficient.

Solution. There are $N = 5$ data points. The sample means are

$$\bar{X} = \frac{2+4+5+7+8}{5} = \frac{26}{5} = 5.2, \qquad \bar{Y} = \frac{1+3+4+6+8}{5} = \frac{22}{5} = 4.4.$$

Sample variances

For X, the deviations from the mean are

$$X - \bar{X} = (-3.2, -1.2, -0.2, 1.8, 2.8).$$

So

$$\sum_{i=1}^{5} (X_i - \bar{X})^2 = 10.24 + 1.44 + 0.04 + 3.24 + 7.84 = 22.80.$$

Hence

$$s_X^2 = \frac{1}{N-1} \sum_{i=1}^{5} (X_i - \bar{X})^2 = \frac{22.80}{4} = 5.70.$$

For Y, the deviations are

$$Y - \bar{Y} = (-3.4, -1.4, -0.4, 1.6, 3.6),$$

so

$$\sum_{i=1}^{5} (Y_i - \bar{Y})^2 = 11.56 + 1.96 + 0.16 + 2.56 + 12.96 = 29.20.$$

Therefore

$$s_Y^2 = \frac{29.20}{4} = 7.30.$$

Thus

$$\boxed{\bar{X} = 5.2, \quad \bar{Y} = 4.4, \quad s_X^2 = 5.7, \quad s_Y^2 = 7.3.}$$

Covariance

The products of deviations are

$$(-3.2)(-3.4) = 10.88, \quad (-1.2)(-1.4) = 1.68, \quad (-0.2)(-0.4) = 0.08,$$
$$(1.8)(1.6) = 2.88, \quad (2.8)(3.6) = 10.08.$$

Summing them gives

$$\sum_{i=1}^{5}(X_i - \bar{X})(Y_i - \bar{Y}) = 10.88 + 1.68 + 0.08 + 2.88 + 10.08 = 25.60.$$

Therefore

$$\mathrm{Cov}(X, Y) = \frac{25.60}{4} = 6.40.$$

So

$$\boxed{\mathrm{Cov}(X, Y) = 6.4.}$$

Correlation coefficient

The sample standard deviations are

$$s_X = \sqrt{5.7}, \qquad s_Y = \sqrt{7.3}.$$

Hence

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{s_X s_Y} = \frac{6.4}{\sqrt{5.7 \times 7.3}} \approx 0.992.$$

Therefore

$$\boxed{\mathrm{Corr}(X, Y) \approx 0.992.}$$

The two datasets are therefore very strongly positively correlated.

Problem 3. From likelihood to χ^2

Question. Starting from the one-dimensional Gaussian likelihood, show that maximizing the likelihood is equivalent to minimizing

$$\chi^2 = \sum_i \frac{(d_i - m_i)^2}{\sigma_i^2}$$

when the data points are independent.

Solution. Suppose the data points d_i are independent and each has a Gaussian error with standard deviation σ_i. If the model prediction for the ith point is m_i, then the likelihood for one point is

$$p(d_i \mid m_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(d_i - m_i)^2}{2\sigma_i^2}\right].$$

Because the measurements are independent, the full likelihood is the product

$$\mathcal{L} = \prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(d_i - m_i)^2}{2\sigma_i^2}\right].$$

Using the fact that the exponential of a sum is the product of exponentials, we can write

$$\mathcal{L} = \left[\prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}}\right] \exp\left[-\frac{1}{2}\sum_i \frac{(d_i - m_i)^2}{\sigma_i^2}\right].$$

Define

$$\chi^2 \equiv \sum_i \frac{(d_i - m_i)^2}{\sigma_i^2}.$$

Then the likelihood becomes

$$\mathcal{L} = \left[\prod_i \frac{1}{\sqrt{2\pi\sigma_i^2}}\right] e^{-\chi^2/2}.$$

It is usually easier to work with the log-likelihood:

$$\ln \mathcal{L} = -\frac{1}{2}\sum_i \ln(2\pi\sigma_i^2) - \frac{1}{2}\chi^2.$$

If the σ_i do not depend on the model parameters, then the first term is a constant. Therefore maximizing $\ln \mathcal{L}$ is exactly the same as minimizing χ^2. So we conclude that

$$\boxed{\text{for independent Gaussian errors, maximizing } \mathcal{L} \text{ is equivalent to minimizing } \chi^2.}$$

Problem 4. Correlated data

Question. Consider two measurements with covariance matrix

$$\mathbf{C} = \begin{pmatrix} 4 & 1.2 \\ 1.2 & 1 \end{pmatrix}.$$

Compute the correlation coefficient and write the full matrix expression for χ^2.

Solution. The variances are read from the diagonal:

$$\sigma_1^2 = 4, \qquad \sigma_2^2 = 1.$$

Therefore

$$\sigma_1 = 2, \qquad \sigma_2 = 1.$$

The covariance is the off-diagonal element,

$$\mathrm{Cov}_{12} = 1.2.$$

So the correlation coefficient is

$$\rho = \frac{\mathrm{Cov}_{12}}{\sigma_1\sigma_2} = \frac{1.2}{2 \times 1} = 0.6.$$

Hence

$$\boxed{\rho = 0.6.}$$

This is a moderately strong positive correlation.

Now define the residual vector

$$\boldsymbol{\Delta} = \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix} = \begin{pmatrix} d_1 - m_1 \\ d_2 - m_2 \end{pmatrix}.$$

Then the correct correlated chi-square is

$$\chi^2 = \mathbf{\Delta}^T \mathbf{C}^{-1} \mathbf{\Delta}.$$

The determinant of $\mathbf{C}$ is

$$|\mathbf{C}| = 4 \times 1 - 1.2^2 = 4 - 1.44 = 2.56.$$

So the inverse matrix is

$$\mathbf{C}^{-1} = \frac{1}{2.56} \begin{pmatrix} 1 & -1.2 \\ -1.2 & 4 \end{pmatrix}.$$

Therefore

$$\chi^2 = \begin{pmatrix} \Delta_1 & \Delta_2 \end{pmatrix} \frac{1}{2.56} \begin{pmatrix} 1 & -1.2 \\ -1.2 & 4 \end{pmatrix} \begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}.$$

Expanding this gives

$$\chi^2 = \frac{1}{2.56} \left(\Delta_1^2 - 2.4\Delta_1\Delta_2 + 4\Delta_2^2 \right).$$

So a convenient final form is

$$\boxed{\chi^2 = \frac{1}{2.56} \left[(d_1 - m_1)^2 - 2.4(d_1 - m_1)(d_2 - m_2) + 4(d_2 - m_2)^2 \right].}$$

Problem 5. Confidence regions

Question. Explain in a few sentences why the 68% contour in two parameters corresponds to $\Delta\chi^2 = 2.30$ rather than $\Delta\chi^2 = 1$.

Solution. For one parameter, the quantity $\Delta\chi^2$ behaves like a χ^2 variable with one degree of freedom, so the central 68.3% interval corresponds to $\Delta\chi^2 = 1$. For *two* parameters, however, the probability inside a contour depends on the area enclosed in the two-dimensional parameter plane. In that case, $\Delta\chi^2$ follows a χ^2 distribution with $k = 2$ degrees of freedom, whose cumulative distribution is

$$F(x; 2) = 1 - e^{-x/2}.$$

To enclose 68.3% of the total probability, we must solve

$$1 - e^{-x/2} = 0.683.$$

This gives

$$x = -2\ln(1 - 0.683) \approx 2.30.$$

Therefore the 68% joint confidence contour for two parameters is

$$\boxed{\Delta\chi^2 \approx 2.30.}$$

By contrast, in two dimensions the contour $\Delta\chi^2 = 1$ would enclose only

$$1 - e^{-1/2} \approx 0.393,$$

that is, only about 39% of the probability, not 68%. So the change from 1 to 2.30 is simply a consequence of moving from a one-dimensional to a two-dimensional confidence region.

Problem 6. Optional coding task (Julia)

Question. Draw 10^5 realizations of the sum of 20 uniformly distributed random numbers and show numerically that the histogram is close to a Gaussian.

Solution. Let

$$S = U_1 + U_2 + \cdots + U_{20},$$

where each U_i is independent and uniformly distributed on $[0, 1]$. For one uniform random variable,

$$\mathbb{E}[U_i] = \frac{1}{2}, \qquad \mathrm{Var}(U_i) = \frac{1}{12}.$$

Therefore

$$\mathbb{E}[S] = 20 \times \frac{1}{2} = 10, \qquad \mathrm{Var}(S) = 20 \times \frac{1}{12} = \frac{5}{3}, \qquad \sigma_S = \sqrt{\frac{5}{3}} \approx 1.291.$$

By the central limit theorem, the distribution of S should be close to a Gaussian with mean 10 and variance $5/3$.

A compact Julia program is:

```
using Random, Statistics, Printf, Plots
Random.seed!(2026)
nreal, nterm = 100_000, 20
sums = vec(sum(rand(nreal, nterm), dims = 2))
mu_theory = nterm / 2
sigma_theory = sqrt(nterm / 12)

@printf("sample mean = %.4f\n", mean(sums))
@printf("sample std  = %.4f\n", std(sums))
@printf("theory mean = %.4f\n", mu_theory)
@printf("theory std  = %.4f\n", sigma_theory)

gauss(x) = exp(-(x - mu_theory)^2 / (2 * sigma_theory^2)) /
           (sqrt(2pi) * sigma_theory)
xs = range(minimum(sums), maximum(sums); length = 500)

histogram(sums; bins = 80, normalize = :pdf,
          xlabel = "sum of 20 U(0,1)",
          ylabel = "probability density",
          label = "simulation")
plot!(xs, gauss.(xs); linewidth = 2, label = "Gaussian N(10, 5/3)")
savefig("lecture6_clt_hist.pdf")
```

For 10^5 realizations, the sample mean should be very close to 10 and the sample standard deviation should be very close to

$$\sqrt{5/3} \approx 1.291.$$

The histogram should look smooth and bell-shaped, and the overplotted Gaussian curve should match it closely, up to small Monte Carlo fluctuations in the tails.