Midterm: Solutions

1. An oceanographic experiment measures 3 variables, x, y, and ϕ, with measurement errors σ_x, σ_y, and σ_ϕ respectively. From these variables, you wish to compute:

$$u = x\sin(\phi) + y^2\cos(\phi). \tag{1}$$

What is σ_u, the estimated error in u? What assumptions have you used to derive σ_u?

solution: Following standard rules for propagating errors:

$$\begin{aligned}\sigma_u^2 &= \left(\frac{\partial u}{\partial x}\sigma_x\right)^2 + \left(\frac{\partial u}{\partial y}\sigma_y\right)^2 + \left(\frac{\partial u}{\partial \phi}\sigma_\phi\right)^2 \\ &= \sin^2(\phi)\sigma_x^2 + 4y^2\cos^2(\phi)\sigma_y^2 + (x\cos(\phi) - y^2\sin(\phi))^2\sigma_\phi^2\end{aligned}$$

so

$$\sigma_u = \sqrt{\sin^2(\phi)\sigma_x^2 + 4y^2\cos^2(\phi)\sigma_y^2 + (x\cos(\phi) - y^2\sin(\phi))^2\sigma_\phi^2}$$

This solution depends on an assumption that errors are unbiased, small relative to the mean, and uncorrelated. Some interpretations may require that the errors be normally distributed.

2. Consider the following pdf:

$$F_Y(y)dy = \begin{cases} \alpha(1+y)dy & \text{for } -1 < y \le 0 \\ \alpha dy & \text{for } 0 < y \le 1 \\ \alpha(2-y)dy & \text{for } 1 < y \le 2 \\ 0. & \text{otherwise} \end{cases} \tag{2}$$

a. What should α be?
b. What are the moments of the pdf F_Y (about the mean)?
c. How do you generate data with a distribution $F_Y(y)$ from a uniform distribution?

solution:
a. The integral of the pdf should be 1, and

$$\int_{-1}^{2} F_y(y)\,dy = 2\alpha,$$

so $\alpha = 1/2$.

b.

$$\begin{aligned}\mu_1 &= \int_{-\infty}^{\infty} yF_y(y)\,dy \\ &= \frac{1}{2}\left[\int_{-1}^{0}(y+y^2)dy + \int_0^1 y dy + \int_1^2 (2y-y^2)dy\right]\end{aligned}$$

$$
\begin{aligned}
&= \left(\frac{y^2}{4} + \frac{y^3}{6}\right)\Bigg|_{-1}^{2} + \frac{y^2}{4}\Bigg|_0^1 + \left(\frac{y^2}{2} - \frac{y^3}{6}\right)\Bigg|_1^2 \\
&= \frac{1}{2} \\
\mu_2 &= \int_{-\infty}^{\infty} (y - \langle y \rangle)^2 F_y(y)\, dy \\
&= \frac{5}{12} \\
\mu_3 &= \int_{-\infty}^{\infty} (y - \langle y \rangle)^3 F_y(y)\, dy \\
&= 0 \\
\mu_4 &= \int_{-\infty}^{\infty} (y - \langle y \rangle)^4 F_y(y)\, dy \\
&= \frac{91}{240}
\end{aligned}
$$

You could derive μ_3 by inspection, by noting the symmetry of the pdf. The exact numerical value for μ_4 was a lengthy calculation, and correct final answers scored bonus points.

c. We require that the lowest quarter of the uniform distribution matches the segment of F_y between -1 and 0, the middle half matches F_y between 0 and 1, and the upper quarter matches F_y between 1 and 2. Thus,

$$
\begin{aligned}
\int_0^{x_o} dx &= \int_{-1}^{y_o} \frac{1+y}{2} dy \text{ for } x < 1/4 \\
\int_{1/4}^{x_o} dx &= \int_{0}^{y_o} \frac{1}{2} dy \text{ for } 1/4 \le x < 3/4 \\
\int_{3/4}^{x_o} dx &= \int_{0}^{y_o} \frac{2-y}{2} dy \text{ for } 3/4 \le x.
\end{aligned}
$$

This yields the substitutions

$$
\begin{aligned}
y &= -1 + 2\sqrt{x} \text{ for } x < 1/4 \\
y &= 2\left(x - \frac{1}{4}\right) \text{ for } 1/4 \le x < 3/4 \\
y &= 2 - 2\sqrt{1-x} \text{ for } 3/4 \le x
\end{aligned}
$$

3. The quantity y is defined so that:

$$
y(n) = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3}
$$

a. The variables x are uniformly distributed. What is the distribution of $y(n)$?

b. When n is large, if x has standard deviations σ_x, what is the standard deviation of y?

solution:
a. $y(n)$ is uniform when $n = 1$. As n increases, $y(n)$ represents a convolution of n uniform distributions. This converges to a Gaussian distribution for large n.
b. The second moment of x about the mean is σ_x. The second moment of y about the mean is

$$\sigma_y^2 = \sum_{i=1}^{n} \int_0^1 \left(\frac{x - \overline{x}}{n}\right)^2 dx = \frac{n\sigma_x^2}{n^2} = \frac{sigma_x^2}{n}$$

so, $\sigma_y = \sigma_x/\sqrt{n}$. You can derive the same result using simple error propagation.

4. On the course web site (under homework), you will find 3 data files containing the u and v components of wind at 3 locations in the tropical Pacific.

a. Compute the covariance matrix for these winds.
b. What is the correlation between these winds?
c. What fraction of the variance in u can be explained by v? What fraction of the variance in u at one location can be explained by u at a different location? Could you use this to fill gaps in the time series?

solution:
a. The covariance matrix for the 6 wind components is a 6 by 6 symmetric matrix containing covariances of each of the wind components relative to each other. Here the winds are arranged $u_1, u_2, u_3, v_1, v_2, v_3$.

$$C_{\mathbf{uu}} = \begin{bmatrix} 11.8005 & 7.7261 & 5.0892 & -3.5688 & -3.2513 & -2.2293 \\ 7.7261 & 10.3586 & 6.7736 & -1.6780 & -1.8993 & -1.7873 \\ 5.0892 & 6.7736 & 8.1965 & 0.1726 & -0.3117 & -0.8409 \\ -3.5688 & -1.6780 & 0.1726 & 9.3824 & 6.1257 & 3.6438 \\ -3.2513 & -1.8993 & -0.3117 & 6.1257 & 7.7472 & 4.7103 \\ -2.2293 & -1.7873 & -0.8409 & 3.6438 & 4.7103 & 6.2074 \end{bmatrix}$$

b. The correlations can also be expressed as a symmetric matrix.

$$\rho_{\mathbf{uu}} = \begin{bmatrix} 1.0000 & 0.7148 & 0.5090 & -0.3392 & -0.3449 & -0.2592 \\ 0.7148 & 1.0000 & 0.7613 & -0.1726 & -0.2120 & -0.2287 \\ 0.5090 & 0.7613 & 1.0000 & 0.0196 & -0.0409 & -0.1179 \\ -0.3392 & -0.1726 & 0.0196 & 1.0000 & 0.7226 & 0.4814 \\ -0.3449 & -0.2120 & -0.0409 & 0.7226 & 1.0000 & 0.7044 \\ -0.2592 & -0.2287 & -0.1179 & 0.4814 & 0.7044 & 1.0000 \end{bmatrix}$$

Notice that the zonal winds are correlated with each other, but anti-correlated with the meridional winds.
c. The fraction of variance explained by one field is simply the squared correlation coefficient,

ρ^2.

$$\rho^2_{\mathbf{uu}} = \begin{bmatrix} 1.0000 & 0.5109 & 0.2591 & 0.1150 & 0.1189 & 0.0672 \\ 0.5109 & 1.0000 & 0.5795 & 0.0298 & 0.0450 & 0.0523 \\ 0.2591 & 0.5795 & 1.0000 & 0.0004 & 0.0017 & 0.0139 \\ 0.1150 & 0.0298 & 0.0004 & 1.0000 & 0.5221 & 0.2317 \\ 0.1189 & 0.0450 & 0.0017 & 0.5221 & 1.0000 & 0.4962 \\ 0.0672 & 0.0523 & 0.0139 & 0.2317 & 0.4962 & 1.0000 \end{bmatrix}$$

The fraction of u explained by v at the same location varies between 1.4% and 11.5%. The fraction of u explained by u at a different location varies between 58% and 26%. The correlations are statistically significant (using a test that we did not discuss in class—see *Numerical Recipes*), so the decision about whether one time series can be used to substitute for another depends on the application, and in this case on the details of the data gaps. In this case the vast majority of the data gaps are the same in all 3 data sets, so correlation coefficients could prove to be a very unsatisfactory way to fill out the time series.