Problem Set 2 Solutions

1. A random variable x has probability density $F_x = \frac{1}{2}(\delta(x-\alpha)+\delta(x+\alpha))$. A new random variable is defined as $y = \frac{1}{3}(x_1 + x_2 + x_3)$ where the x_i may be assumed to be independent realizations of x. What is the probability density for y? What are its first three moments?

solution: Solve using the product of the inverse Fourier transform.

$$\begin{aligned}
\mathcal{F}(F_y) &= \mathcal{F}(F_x)^3 = \left[\frac{1}{2}\int_{-\infty}^{\infty} e^{ikx}\left(\delta(x-\alpha)+\delta(x+\alpha)\right)\, dx\right]^3 \\
&= \frac{1}{2^3}\left[e^{ik\alpha} + e^{-ik\alpha}\right]^3 \\
&= \frac{1}{8}\left[e^{3ik\alpha} + 3e^{ik\alpha} + 3 + e^{-ik\alpha} + e^{-3ik\alpha}\right] \qquad (1)
\end{aligned}$$

Thus the pdf of y is:

$$\begin{aligned}
F_y &= \frac{1}{8}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iky}\left(e^{3ik\alpha} + 3e^{ik\alpha} + 3 + e^{-ik\alpha} + e^{-3ik\alpha}\right) dk \\
&= \frac{1}{8}\left[\delta(y-\alpha) + 3\delta(y-\alpha/3) + 3\delta(y+\alpha/3) + \delta(y+\alpha)\right] \qquad (2)
\end{aligned}$$

The pdf is clearly symmetric about zero, so the first and 3rd moments are easily shown to be zero. The second moment is:

$$\begin{aligned}
\mu_2 &= \frac{1}{8}\int_{-\infty}^{\infty} y^2\left[\delta(y-\alpha) + 3\delta(y-\alpha/3) + 3\delta(y+\alpha/3) + \delta(y+\alpha)\right] dy \\
&= \frac{1}{8}\left[\alpha^2 + 3\left(\frac{\alpha}{3}\right)^2 + 3\left(\frac{\alpha}{3}\right)^2 + \alpha^2\right] \\
&= \frac{\alpha^2}{3} \qquad (3)
\end{aligned}$$

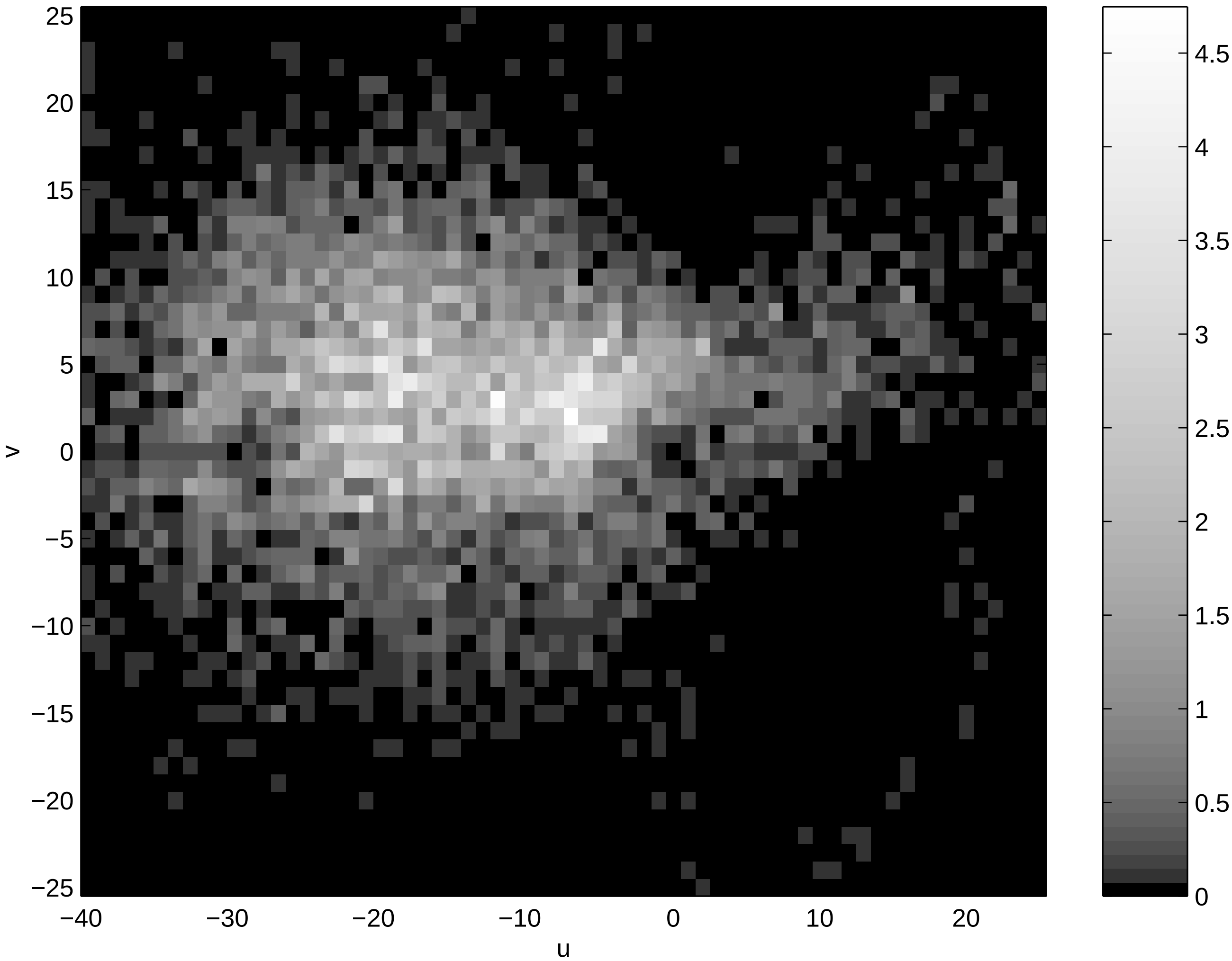
2. Using the same data set that you used for problem set 1 ('rcm00001.str' from the course website), map the joint probability density function for u and v velocities and the covariance of u and v. Are u and v uncorrelated? How would you rotate the pdf to make u and v be uncorrelated? Is there any relationship between the dominant variability and the direction of mean flow?

solution:

The covariance matrix is:

$$C = \begin{bmatrix} 162.6 & 0.33 \\ 0.33 & 38.09 \end{bmatrix} \qquad (4)$$

The mean of u is $\langle u \rangle = -15.48$ and the mean of v is $\langle v \rangle = 3.2741$. Thus the mean flow is oriented $\text{atan2}(\langle v \rangle, \langle u \rangle) = 168^\circ$ clockwise relative to east.



The correlation coefficient for u and v is 0.0041, which is quite low. (Given that $N = 10207$, the correlation between u and v is not statistically significant at a 5% level.) Nonetheless, we can rotate the joint pdf. In this case, we seek the angle θ, where:

$$\theta = \frac{1}{2}\text{atan2}(2\langle u'v'\rangle, \langle u'^2\rangle - \langle v'^2\rangle) \tag{5}$$

The dominant variability is not aligned with the direction of mean flow, and there is no a priori reason to expect them to be in the same direction, although in some physical examples, this may occur.

3. (Problem 6 from the notes on fundamentals of statistics). The sum of two normally distributed variables is also normally distributed. Thus the low moments of the sum should be related as in equation (30) of the notes. Let $Z = X + Y$ where X and Y are independent normally distributed variables with $\langle X\rangle = \langle Y\rangle = 0$. What is the third moment of Z in terms of the statistics of X and Y? Show that the fourth moment of Z obeys equation (30).

solution: Since X and Y are independent, we can simply integrate over their separate pdfs to determine the behavior of Z. If we wante dto know the pdf of Z we could compute it efficiently using characteristic functions, as we did in class.

$$\begin{aligned}
\mu_3 &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)^3\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)dx\,dy \\
&= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^3+3x^2y+3xy^2+y^3)\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)dx\,dy \\
&= \langle X^3\rangle + 3\langle X^2\rangle\langle Y\rangle + 3\langle Y^2\rangle\langle X\rangle + \langle Y^3\rangle = 0. \qquad (6)
\end{aligned}$$

and

$$\begin{aligned}
\mu_4 &= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y)^4\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)dx\,dy \\
&= \frac{1}{2\pi\sigma^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^4+4x^3y+6x^2y^2+4xy^3+y^4)\exp\left(-\frac{x^2}{2\sigma^2}\right)\exp\left(-\frac{y^2}{2\sigma^2}\right)dx\,dy \\
&= \langle X^4\rangle + 4\langle X^3Y\rangle + 6\langle X^2Y^2\rangle + \langle XY^3\rangle + \langle Y^4\rangle \\
&= 3\langle X^2\rangle\langle X^2\rangle + 6\langle X^2\rangle\langle Y^2\rangle + 3\langle Y^2\rangle\langle Y^2\rangle \\
&= 3\langle X^2\rangle(\langle X^2\rangle + \langle Y^2\rangle) + 3\langle Y^2\rangle(\langle X^2\rangle + \langle Y^2\rangle) \\
&= 3\langle Z^2\rangle^2 \qquad (7)
\end{aligned}$$

where here we've used the fact that $\langle X^4\rangle = 3\langle X^2\rangle$ in a Gaussian distribution.

Equation 30, applied to $\langle Z^4\rangle$ says that if Z is normally distributed, then

$$\langle Z^4\rangle = 3\langle Z^2\rangle\langle Z^2\rangle \qquad (8)$$

and this is the result that we have found.

4. Standard random number generating packages on most computer systems will produce random numbers with a uniform distribution between 0 and 1. That is:

$$F_1(x)dx = \begin{cases} dx & \text{for } 0 < x < 1 \\ 0. & \text{otherwise} \end{cases} \qquad (9)$$

(In Matlab and Fortran, uniform distributions are generated by the "rand" function.)

In contrast geophysical noise rarely has a uniform distribution. In order to carry out a Monte Carlo simulation, suppose that you require noise with the symmetrical exponential distribution that you considered in Problem set 1:

$$F_2(x)dx = \frac{1}{\sigma\sqrt{2}}\exp\left[\frac{-|x|\sqrt{2}}{\sigma}\right] \qquad (10)$$

with $\sigma = 1$.

a. Verify that the pdf $F_2(x)$ is normalized correctly.

b. Derive the appropriate algorithm to generate noise that is distributed as specified by $F_2(x)$.

c. Use your algorithm to convert uniform random numbers from $F_1(x)$ so that they have the distribution $F_2(x)$. Plot the pdf of the original uniform distribution and your record with the new distribution.

solution: a. Simply integrate the pdf:

$$\begin{aligned}\int_{-\infty}^{\infty} F_2(r)\,dr &= \frac{\sqrt{2}}{\sigma}\int_0^{\infty} \exp\left[\frac{-r\sqrt{2}}{\sigma}\right] dr \\ &= -\frac{\sqrt{2}}{\sigma}\frac{\sigma}{\sqrt{2}} \exp\left[\frac{-r\sqrt{2}}{\sigma}\right]\Bigg|_0^{\infty} \\ &= 1. \end{aligned} \tag{11}$$

This is the expected result, so the pdf is normalized correctly.

b. We require that:

$$F_1(x)\,dx = F_2(y)\,dy \tag{12}$$

over any appropriately chosen interval. Since F_2 has an absolute value, we need to split the solution into two parts. For $y < 0$ and $x < 1/2$,

$$\int_0^{x_1} dx = \int_{-\infty}^{y_1} \frac{1}{\sigma\sqrt{2}} \exp\left[\frac{y\sqrt{2}}{\sigma}\right].$$

Integrating, this produces,

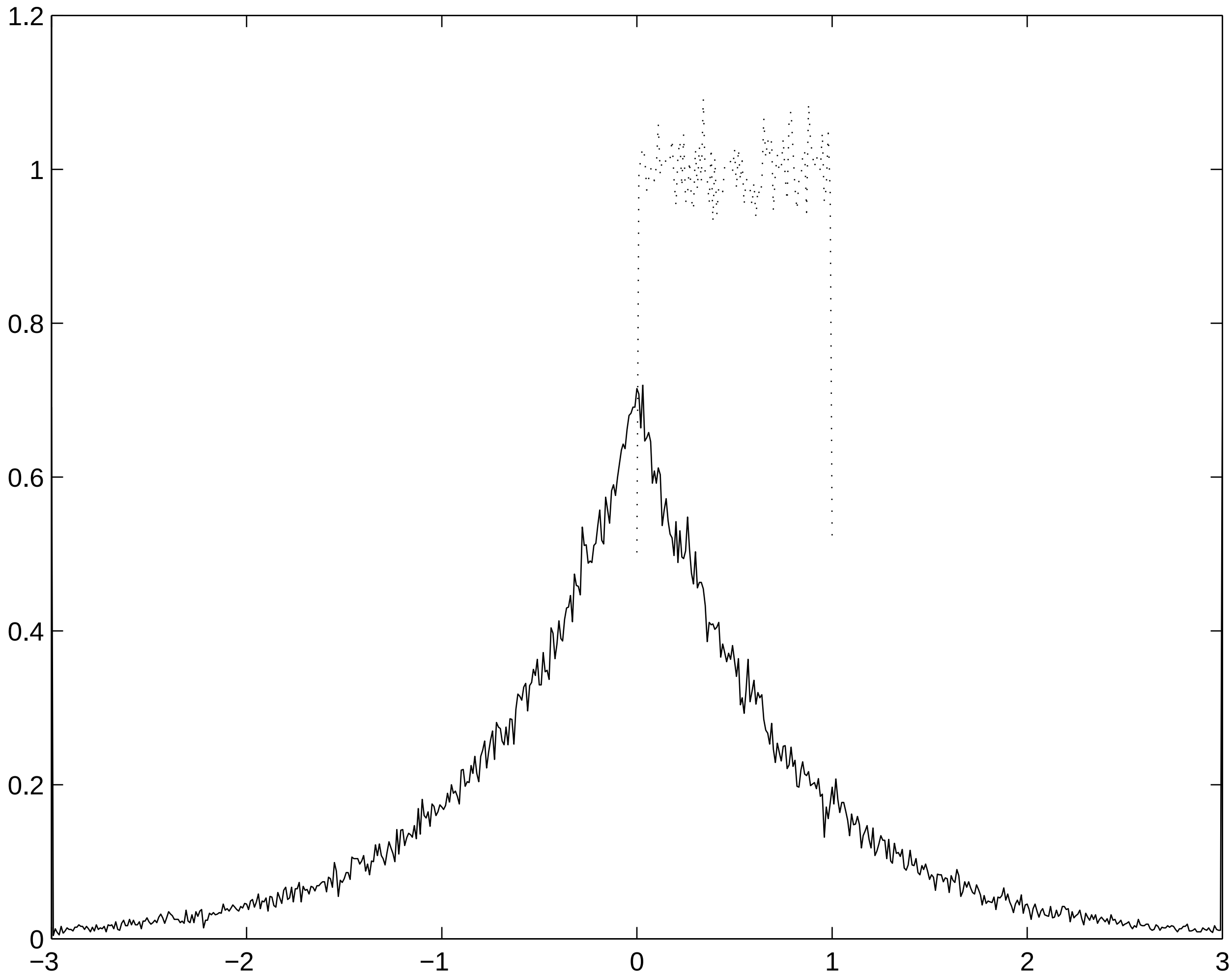
$$\begin{aligned} x|_0^{x_1} &= \frac{1}{2}\exp\left(\frac{y\sqrt{2}}{\sigma}\right)\Bigg|_{-\infty}^{y_1} = \\ x_1 &= \frac{1}{2}\exp\left(\frac{y_1\sqrt{2}}{\sigma}\right) \end{aligned}$$

so that

$$y = \frac{\sigma}{\sqrt{2}}\ln(2x).$$

For $y > 0$ and $x > 1/2$ the solution for y changes slightly:

$$\int_{1/2}^{x_2} dx = \int_{-\infty}^{y_2} \frac{1}{\sigma\sqrt{2}} \exp\left[\frac{-y\sqrt{2}}{\sigma}\right].$$



Integrating, this produces,

$$x\Big|_{1/2}^{x_2} = -\frac{1}{2}\exp\left(-\frac{y\sqrt{2}}{\sigma}\right)\Big|_0^{y_2} =$$

$$x_2 - \frac{1}{2} = \frac{1}{2}\left[\exp\left(\frac{y_2\sqrt{2}}{\sigma}\right) - 1\right]$$

so that

$$y = -\frac{\sigma}{\sqrt{2}}\ln(2-2x).$$

To summarize the algorithm, first derive a set of uniformly distributed random variables, x. Then convert them to have an exponential distribution:

$$y = \begin{cases} \frac{\sigma}{\sqrt{2}}\ln(2x) & \text{for } x \leq \frac{1}{2} \\ -\frac{\sigma}{\sqrt{2}}\ln(2-2x) & \text{for } x > \frac{1}{2} \end{cases} \tag{13}$$