Problem Set 2: MAE 127 (Solutions)

1. Use the randn function in Matlab to produce a large set ($O(10^6)$) of normally distributed random numbers. Plot the pdf of these numbers. Compute the first four moments of the empirical pdf. What are the skewness and kurtosis?

Here's code to carry out this calculation.

```
% generate random data
%  use randn to make normally distributed data
x=randn(1.e6,1);

% make the pdf
[a,b]=hist(x,-5:.1:5);
plot(b,a/.1/sum(a)); xlabel('x'); ylabel('pdf')

% compute moments using Matlab moment function.
[moment(x,1) moment(x,2) moment(x,3) moment(x,4)]

% alternatively compute moments using mean or sum function.
[mean(x) sum((x-mean(x)).^2)/(length(x)-1) ...
 sum((x-mean(x)).^3)/(length(x)-1) sum((x-mean(x)).^4)/(length(x)-1)]

% compute skewness and kurtosis using Matlab functions
[skewness(x) kurtosis(x)]

% compute skewness and kurtosis using mean, variance, sum
[sum((x-mean(x)).^3)/var(x)^1.5 sum((x-mean(x)).^4)/var(x)^2]
```

The resulting pdf is shown in Figure 1.

The first moment (the mean) is approximately zero, the second moment (the variance) is one, the third moment is approximately zero, and the fourth moment should be about 3.

Skewness is zero; kurtosis is 3. These results agree with expectations for a normal distribution.

2. Look at the wind data from the buoy data file that we used last week. Plot the wind velocity components and wind speed as pdfs. Since the data set provides wind speed and direction, you'll first you'll have to compute the eastward and northward wind velocities using the wind speed multiplied by the sine and cosine of the wind direction. Keep in mind that the Matlab sine and cosine functions use angles in radians rather than degrees. How would you describe the pdfs? Are they Gaussian? What are the first four moments of the data distributions?

Here's code to compute the empirical pdf of the wind data.

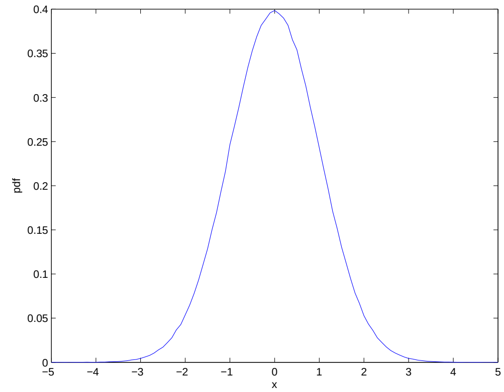

Figure 1: *Probability density function for* 10^6 *normally distributed random data points, plotted here as a line plot, using bins of 0.1 units width.*

```
% load data and compute quantities
load buoy.mat
index=find(wind_speed~=99 & wind_direction~=999);
u=-wind_speed(index).*sin(wind_direction(index)*pi/180);
v=-wind_speed(index).*cos(wind_direction(index)*pi/180);
%  wind_direction is the direction from which the wind comes,
%  measured in degrees, clockwise relative to North.  Thus
%  we need a minus sign to determine the direction towards which
%  the wind is going.

% plot pdf for eastward wind
[a,b]=hist(u,-15:.1:15);
subplot(1,3,1); plot(b,a/.1/sum(a));
xlabel('eastward wind velocity (m/s)');
ylabel('probability density')

% plot pdf for northward wind
[a,b]=hist(v,-15:.1:15);
subplot(1,3,2); plot(b,a/.1/sum(a));
xlabel('northward wind velocity (m/s)');
ylabel('probability density')

% plot pdf for wind speed
index=find(wind_speed~=99);
[a,b]=hist(wind_speed(index),0:.1:17);
subplot(1,3,3); plot(b,a/.1/sum(a));
xlabel('wind speed (m/s)');
ylabel('probability density')
```

```
% compute moments using moment function
[mean(u) moment(u,2) moment(u,3) moment(u,4) skewness(u) kurtosis(u)]
[mean(v) moment(v,2) moment(v,3) moment(v,4) skewness(v) kurtosis(v)]
[mean(wind_speed(index)) moment(wind_speed(index),2) ...
  moment(wind_speed(index),3) moment(wind_speed(index),4) ...
  skewness(wind_speed(index)) kurtosis(wind_speed(index))]

% alternatively, compute using sum function
N=length(u)-1;
[mean(u) sum((u-mean(u)).^2)/N sum((u-mean(u)).^3)/N sum((u-mean(u)).^4)/N ...
 sum((u-mean(u)).^3)/N/(sum((u-mean(u)).^2)/N)^1.5 ...
 sum((u-mean(u)).^4)/N/(sum((u-mean(u)).^2)/N)^2]

N=length(v)-1;
[mean(v) sum((v-mean(v)).^2)/N sum((v-mean(v)).^3)/N sum((v-mean(v)).^4)/N ...
 sum((v-mean(v)).^3)/N/(sum((v-mean(v)).^2)/N)^1.5 ...
 sum((v-mean(v)).^4)/N/(sum((v-mean(v)).^2)/N)^2]

s=wind_speed(index);
N=length(s)-1;
[mean(s) sum((s-mean(s)).^2)/N sum((s-mean(s)).^3)/N sum((s-mean(s)).^4)/N ...
 sum((s-mean(s)).^3)/N/(sum((s-mean(s)).^2)/N)^1.5 ...
 sum((s-mean(s)).^4)/N/(sum((s-mean(s)).^2)/N)^2]
```

The resulting wind pdfs are shown in Figure 2. The eastward and northward wind velocities

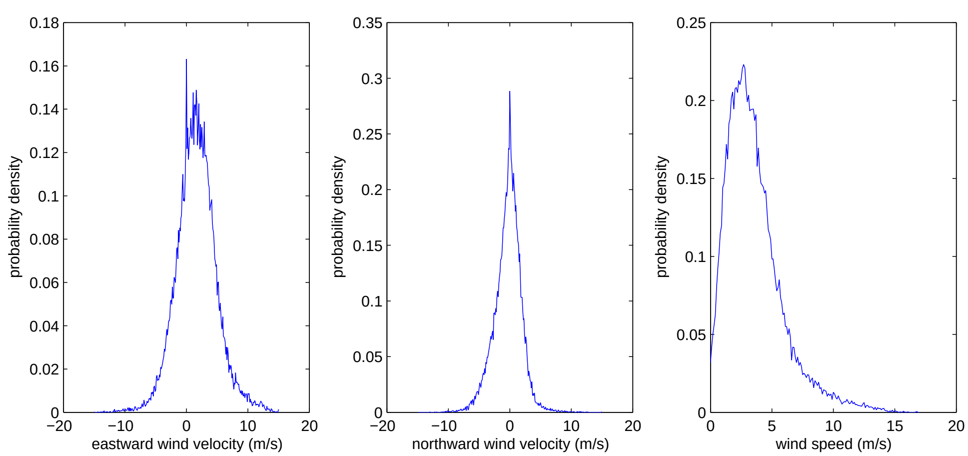

Figure 2: *Probability density functions for (left) eastward wind, (center) northward wind, and (right) wind speed from buoy measurements collected in the Santa Monica basin between 2000 and 2004.*

are roughly centered around zero and roughly symmetric though in both cases slightly skewed to positive velocities. The distributions are less rounded than would be expected for a normal distribution. The wind speed distribution contains only positive values and resembles a

"Rayleigh distribution". For all three pdfs, the kurtosis exceeds 3, so the data are not strictly Gaussian.

Moments will differ slightly depending how you calculate them, since the definition that we used in class says to divide by $N-1$, but the Matlab function simply divides by N. Moments, skewness, and kurtosis are as follows:

	μ_1	μ_2	μ_3	μ_4	skewness	kurtosis
eastward velocity (m/s)	1.62	10.66	4.61	474.5	0.1	4.2
northward velocity (m/s)	-0.33	4.85	-2.34	114.4	-0.2	4.9
wind speed (m/s)	3.58	5.41	17.24	164.3	1.4	5.6

3. Consider a pdf of the form:

$$P(x)\,dx = \begin{cases} (1+x)\,dx & \text{for } -1 < x \leq 0 \\ (1-x)\,dx & \text{for } 0 \leq x < 1 \\ 0. & \text{otherwise} \end{cases}$$

What are the mean, standard deviation, and kurtosis of data drawn from this distribution? (Hint: This problem is probably most easily done by hand, without recourse to Matlab.)

To compute the mean, we integrate $xP(x)\,dx$:

$$\begin{aligned} \langle x \rangle &= \int_{-\infty}^{\infty} xP(x)\,dx \\ &= \int_{-1}^{0} x(1+x)\,dx + \int_{0}^{1} x(1-x)\,dx \\ &= \left(\frac{x^2}{2} + \frac{x^3}{3}\right)\Big|_{-1}^{0} + \left(\frac{x^2}{2} - \frac{x^3}{3}\right)\Big|_{0}^{1} \\ &= -\frac{1}{2} + \frac{1}{3} + \frac{1}{2} - \frac{1}{3} = 0. \end{aligned}$$

Since the mean is zero, this will slightly simplify calculation of the standard deviation and kurtosis. The variance is:

$$\begin{aligned} \text{var}(x) &= \langle (x-\overline{x})^2 \rangle = \int_{-\infty}^{\infty} (x-\overline{x})^2 P(x)\,dx \\ &= \int_{-1}^{0} x^2(1+x)\,dx + \int_{0}^{1} x^2(1-x)\,dx \\ &= \left(\frac{x^3}{3} + \frac{x^4}{4}\right)\Big|_{-1}^{0} + \left(\frac{x^3}{3} - \frac{x^4}{4}\right)\Big|_{0}^{1} \\ &= \frac{1}{3} - \frac{1}{4} + \frac{1}{3} - \frac{1}{4} = \frac{1}{6} \end{aligned}$$

The standard deviation is the square root of the variance, so $\sigma_x = 1/\sqrt{6}$. To compute kurtosis, we need the fourth moment:

$$\langle (x-\overline{x})^4 \rangle = \int_{-\infty}^{\infty} (x-\overline{x})^4 P(x)\,dx$$

$$
\begin{aligned}
&= \int_{-1}^{0} x^4(1+x)\,dx + \int_{0}^{1} x^4(1-x)\,dx \\
&= \left.\left(\frac{x^5}{5}+\frac{x^6}{6}\right)\right|_{-1}^{0} + \left.\left(\frac{x^5}{5}-\frac{x^6}{6}\right)\right|_{0}^{1} \\
&= \frac{1}{5}-\frac{1}{6}+\frac{1}{5}-\frac{1}{6} = \frac{1}{15}
\end{aligned}
$$

The kurtosis is $\mu_4/\mu_2^2 = 36/15 = 12/5 = 2.4$