SIO 223A, Geophysical Data Analysis
Solutions to Problem Set 1

1.1 *Consider the model for the magnetic field reversals described in Chapter 1 of the notes: we consider time broken up into blocks Δ long, and in each block assume that there is a probability p of a reversal, independently of whatever happened before. What is the probability of getting a period $N\Delta$ with only one reversal at the end? What is the probability of getting no reversal between the times $N_1\Delta$ and $N_2\Delta$? What is the probability if N_1 is 1 and N_2 goes to infinity? Evaluate the probability of an unreversed interval longer than T for the case p/Δ constant, with Δ going to 0. You should be able to work these out from first principles; none of these answers should be given as unevaluated series.*

The probability of a reversal in each block is p, so the probability of no reversal is $1-p$. To get no reversals N times, then one, we take the product of these independent probabilities, the no-reversal ones n times, then by p for the block with a reversal:

$$(1-p)^N p$$

Because the model does not have anything about absolute time in it, the probability of no reversals for $N_1\Delta < t < N_2\Delta$ is the same as no reversals for $0 < t < (N_2 - N_1)\Delta$:

$$(1-p)^{N_2-N_1}$$

and if N_2 goes to ∞ while $N_1 = 1$ this approaches zero.
Approximately but adequately, a time T is equivalent to $N = T/\Delta$ reversal intervals, so the probability of an unreversed interval of at least T in length is $(1-p)^{T/\Delta}$. Now take $c = \frac{p}{\Delta}$ to be constant; then the probability is $p = \Delta c$, so we can rewrite the expression above as

$$(1-\Delta c)^{T/\Delta} = \big(1 - \frac{cT}{N}\big)^N$$

Now let $\Delta \to 0$ then $N \to \infty$. Now, by the binomial theorem,

$$\big(1+\frac{x}{N}\big)^N = \sum_{m=0}^{N} \binom{N}{m} \big(\frac{x}{N}\big)^m$$

since

$$\binom{N}{m} = \frac{N!}{(N-m)!m!}$$

which for large N approaches $N^m/m!$, we can write the sum as

$$\sum_{m=0}^{N} \frac{x^m}{m!} \to e^x$$

which gives the final result, that the probability is

$$e^{-cT}$$

This is what is called the **exponential distribution**, and the probability model is known as the Poisson process.

1.2. *Suppose we have a sine wave that is a function of time, t,*

$$x(t) = A\cos(t)$$

and create a random variable by sampling this at random times. What is the pdf of this random variable? Hint: what is the distribution of the argument t? Given this, transform it to give the distribution of the final random variable.
We have $x(t) = A\cos(t)$, and t is taken to be random. Since t functions as a phase, this means that we expect one phase as much as another, so the rv T is uniformly distributed over $[0, 2\pi]$; and since we are looking at a cosine function, we can just as well make T uniformly distributed over $[0, \pi)$. Let $Z = A\cos(T)$; the cdf of Z is

$$P(Z \leq z) = P(A\cos T \leq z) = P(T \leq \arccos(z/A))$$

but we know that T has the cdf

$$P(T \leq t) = t/\pi$$

and so, putting $\arccos(z/A)$ in place of t, the cdf of Z is

$$\Phi(z) = P(Z \leq z) = \frac{\arccos(z/A)}{\pi}$$

and differentiating this to get the pdf gives

$$\phi(z) = \frac{1}{\pi\sqrt{A^2 - z^2}}$$

for $|z| \leq A$; which is integrable but unbounded (a square-root singularity at each end).

1.3 *The Cauchy cumulative distribution function is*

$$\Phi(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x), \qquad -\infty < x < \infty$$

(a) Show that this is a cdf. (b) Find the density function. (c) Find x such that $P(X > x) = 0.1$.
We have a function

$$\Phi(x) = 1/2 + \frac{1}{\pi}\arctan(x)$$

This satisfies $\Phi(-\infty) = 0$ and $\Phi(\infty) = 1$. The derivative is

$$\phi(x) = \frac{1}{\pi(1+x^2)}$$

which is everywhere greater than zero, so Φ is a cumulative distribution function and ϕ is a pdf: the Cauchy pdf. Finally, for $\Phi(x) = 0.1$ we find $x = -3.078$.

1.4 *Show that if A and B are independent, then*

$$P(A \cup B) = P(A) + P(B) - P(A)P(B)$$

In general $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ by the definition of conditional probability which is $P(A \cup B) = P(A/B)P(B)$; but since independence implies $P(A/B) = P(A)$, we can combine these to give $P(A \cup B) = P(A) + P(B) - P(A)P(B)$.

1.5 *Consider the pdf of the random variable X distributed as*

$$X \sim 0.75N(0,1) + 0.25N(1,.3)$$

where $N(m,s)$ is the Normal (Gaussian) distribution, with mean m and standard deviation s, for which the pdf is

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}$$

Plot the pdf, find the first four moments, the median, the interquartile distance, the mode, and the mean error. You may use whatever mix of analytical methods, numerical methods, and looking up values in tables you like, but be sure to explain whatever method you use in adequate detail. If you use a canned routine, assume I have never heard of it and explain what it does.

We have a random variable X with the pdf

$$X \sim 0.75N(0,1) + 0.25N(1,0.3)$$

which is to say a pdf

$$\phi(x) = 0.75\phi_N(x) + \frac{0.25}{c}\phi_N(x-l,c)$$

with $l = 1$ and $c = 0.3$; ϕ_N is the standard normal distribution

$$\phi_N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

We start by getting the moments for this distribution; that is, for $N(0,1)$; we use a $\tilde{}$ to denote these moments:

$$\tilde{\mu_r} = \int_{-\infty}^{\infty} x^r \phi_N(x) dx$$

To get this, we use integral 3.461.2 in Gradshtein and Ryzhik or your favorite online integrating tool:

$$\int_{-\infty}^{\infty} x^n e^{-px^2} dx = \frac{1 \times 3 \times \cdots (2n+1)}{(2p)^n} \sqrt{\frac{\pi}{p}}$$

which gives us

$$\tilde{\mu_0} = 1 \qquad \tilde{\mu_1} = 0 \qquad \tilde{\mu_2} = 1 \qquad \tilde{\mu_3} = 0 \qquad \tilde{\mu_4} = 3$$

only the last of which is new compared with what we knew before.
Now consider the rth moment about some value p; we need to look at this because we have to compute moments about 0 and about the mean. We call this

$$\mu_r(p) = \int_{-\infty}^{\infty} (x-p)^r \frac{1}{c} \phi(\frac{x-l}{c}) dx = \int_{-\infty}^{\infty} (cu + l - p)^r \phi(u) du$$

We can identify these moments in terms of the definitions in the notes as $\mu_r' = \mu_r(0)$ and $\mu_r = \mu_r(\mu_1)$.
To compute each moment, we use the linearity of the integral. The zeroth moment is

$$0.75\tilde{\mu}_0 + 0.25\tilde{\mu}_0 = 1$$

which is what it should be for a pdf.
The first moment is, for each part

$$\int_{-\infty}^{\infty} (cu+l-p)\phi(u)du = \int_{-\infty}^{\infty} cu\phi(u)du + \int_{-\infty}^{\infty} (l-u)\phi(u)du = c\tilde{\mu}_1 + (l-p)\tilde{\mu}_0 = l-p \quad (2)$$

which we apply to the two distributions separately, the first having $l = 0$ and the second $l = 1$. The first moment about zero ($p = 0$) that is

$$\mu_1(0) = 0.75 \times 0 + 0.25 \times 1 = 0.25.$$

Of course the first moment around 0.25 is $0.75 \times 0.25 - 0.25 \times 0.75 = 0$. Using the same procedure as above in (2) the second moment is given by

$$c^2\tilde{\mu}_2 + 2c(l-p)\tilde{\mu}_1 + (l-p)^2\tilde{\mu}_0 = c^2 + (l-p)^2$$

and so the second moment about zero ($\mu_2(0)$) is

$$0.75 \times 1 + 0.25 \times (0.3^2 + 1) = 1.0225$$

and the second moment about 0.25 is

$$40.75 \times (1 + (0.25)^2) + 0.25 \times (0.3^2 + (0.75)^2) = 0.960$$

.
The third moment is given by

$$c^3\tilde{\mu} - 3 + 3c^2(l-p)\tilde{\mu}_2 + 3c(l-p)^2\tilde{\mu}_1 + (l-p)^3\tilde{\mu}_0 = 3c^2(l-p) + (l-p)^3$$

For the moment about zero ($p = 0$) this is zero for the first distribution, and $3 \times 0.3^2 + 1 = 1.27$ for the second, so $\mu_3' = 0.25 \times 1.27 = 0.3175$. For $p = 0.25$, this is $3 \times (-0.25) + (-0.25)^3 = -0.765$ for the first distribution and $3 \times 0.3^2(0.75) + (0.75)^3$ for the second; so

$$\mu_3 = -0.418$$

Finally, the fourth moment for each distribution is given by

$$c^4\tilde{\mu}_4 + 4c^3(l-p)\tilde{\mu}_3 + 6c^2(l-p)^2\tilde{\mu}_2 + 4c(l-p)^3\tilde{\mu}_1 + (l-p)^4 = 3c^4 + 6c^2(l-p)^2 2 + (l-p)^4$$

and so

$$\mu_4(0) = 2.6411$$

and

$$\mu_4(.25) = 2.6953$$

For the median and interquartile distance we need to use the expression for the cdf, which has to be found numerically using an error-function routine or a built-in function; the results are that $\Phi(x) = 0.75$ for $x = 0.988$ and that $\Phi(x) = 0.25$ for $x = -0.430$, so the interquartile range is 1.418. The median, where $\Phi(x) = 0.5$, is 0.409. Direct tabulation of the function gives the mode as 0.950; numerical integration (simple quadrature) gives the mean deviation as 0.848. Numerical integration also gives moments that check the exact values given above against algebraic blunders.