Name: ______________________________

1. Suppose you have two independent measurements $\cos\theta$ and $\sin\theta$, which you call x and y. Assuming the two measurements have Gaussian uncertainties with the same σ, write down the likelihood of the measurement and find the maximum likelihood estimate $\hat{\theta}$.

2. Consider a Gaussian distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$. Given $N > 1$ independent measurements x_i drawn from this distribution, the sample variance s^2 can be calculated using the expression

$$s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2.$$

   (a) Given that $(N-1)s^2/\sigma^2$ follows a χ^2 distribution with $N-1$ degrees of freedom, show that s^2 is an unbiased estimator of the variance σ^2, and that its variance – i.e., $\mathrm{var}\,(s^2)$ – is $2/(N-1)$.

   (b) Find the expectation values and variances of other estimates

$$s_k^2 = \frac{1}{N+k}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

   where k is an integer. Show that although the $k = 1$ value s_1^2 is a biased estimator of σ^2, its mean square deviation from unity (i.e., the MSE $d_k^2 = \mathrm{var}\,(s_k^2) + b^2$, where b is the bias) is smaller than for any other estimator s_k^2.

3. Investigate the extent to which small samples of a distribution do or do not provide good estimates of the population variance.

   (a) Generate $10,000$ samples of 4 measurements per sample, with each measurement drawn from a Gaussian distribution of zero mean and unit variance. Obtain an estimate s^2 of the variance of each sample and histogram the distribution of the $10,000$ values of s^2. Find the mean and variance of the s^2 distribution.

   (b) Repeat the calculation above using the estimator

$$s_k^2 = \frac{1}{N+k}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

   for $k = 0$ and $k = 1$.

   (c) Are the means and variances of the distributions of s_k for $k = -1, 0, 1$ what you expect?

4. The Hubble Law gives a linear relationship between the recessional velocity of a galaxy v and its distance from Earth d,

$$v = H_0 d,$$

where H_0 is Hubble's constant. Astronomers measure recessional velocities and use the Hubble Law to infer the distance to receding objects. H_0 must be determined experimentally, so uncertainties in H_0 lead to systematic uncertainties in d.

Suppose a galaxy has a measured recessional velocity $v_m = (100 \pm 5) \times 10^3$ km s^{-1}, i.e., the uncertainty is parameterized as a Gaussian with $\sigma = 5 \times 10^3$ km s^{-1}. Plot the posterior PDF for the distance to the galaxy for the four cases listed below.

(*Hints*: assume a uniform prior distribution for d in all cases. If needed, use numerical integration to plot the posterior PDF.)

(a) $H_0 = 75$ km s^{-1} Mpc^{-1} exactly, an estimate first made in the 1950s.

(b) $H_0 = 72 \pm 8$ km s^{-1} Mpc^{-1}, an estimate made 10 years ago using data from the Hubble Space Telescope. (Assume a Gaussian uncertainty in H_0.)

(c) H_0 uniformly distributed between 50 and 90 km s^{-1} Mpc^{-1}, the state of measurements prior to the mid-1990s.

(d) H_0 distributed between 50 and 90 km s^{-1} Mpc^{-1} using the Jeffreys prior. (*Jeffreys prior*: $p(x) = \frac{1}{x}\frac{1}{\log\frac{B}{A}}$ for $A \leq x \leq B$)