1. (a) The MLE is unbiased, so the risk is just the variance. And the variance of $\bar{X}$ in a binomial problem is $R(\theta, \delta_{\text{mle}}) = \theta(1 - \theta)/n$.
(b) For a uniform prior on $\theta$, the Bayes estimator is $\delta_{\text{Bayes}} = (n\bar{X} + 1)/(n + 2)$. Using the bias–variance decomposition of the risk, we get

$$R(\theta, \delta_{\text{Bayes}}) = \frac{n\theta(1 - \theta) + (1 - 2\theta)^2}{(n + 2)^2}.$$  
(c) Plot of the risk functions for $n = 10$ are shown in Figure 1 below. Here we see that, for an interval around $\theta = 0.5$, the Bayes estimator has small risk than the MLE. Outside this interval, however, the MLE is better.

Figure 1: Risk functions for Exercise 1(c). Black is Bayes; gray is MLE.

2. (a) For the MLE $\delta(X) = \bar{X}$, the risk function $R(\theta, \delta)$ is the variance, which is a constant $n^{-1}$.
(b) Under a standard normal prior, the Bayes rule (posterior mean) is $\delta_{\text{Π}}(X) = n\bar{X}/(n + 1)$. This is a linear function of $\bar{X}$, so the risk (or MSE) can be computed easily using the bias–variance decomposition. In this case,

$$R(\theta, \delta_{\text{Π}}) = \frac{(n + \theta^2)}{(n + 1)^2}.$$  
(c) Clearly, $R(0, \delta_{\text{Π}}) < R(0, \delta)$, and continuity implies that the Bayes rule has smaller risk than the MLE in a neighborhood of $\theta = 0$. To find this interval, set the two risk functions equal and solve for $\theta$: it’s a quadratic equation, so there will be two solutions, one negative and one positive. In particular

$$R(\theta, \delta) = R(\theta, \delta_{\text{Π}}) \iff \frac{n + \theta^2}{(n + 1)^2} - \frac{1}{n} = 0,$$
and the solutions are
\[ \theta = \pm \left[ \frac{(n+1)^2}{n} - n \right]^{1/2}; \]
in the interval contained by these two values, the Bayes rule has smaller risk. As you can see, the interval is symmetric around zero and, as \( n \to \infty \), it collapses to a single point. This latter point means that, when sample size is large, the risk difference between the Bayes rule and the MLE is small, which we would expect given that the MLE is asymptotically efficient.

3. Suppose the goal is to test \( H_0 : \theta \in \Theta_0 \) versus \( H_1 : \theta \notin \Theta_0 \). For decision rule (test function) \( \delta \), where \( \delta(x) = k \) means accept \( H_k \), \( k = 0, 1 \), the 0–1 loss function looks like
\[
L(\theta, \delta(x)) = I_{\{\delta(x) = 0, \theta \in \Theta_0\}} + I_{\{\delta(x) = 1, \theta \notin \Theta_0\}}.
\]
For the risk function, we consider \( \theta \in \Theta_0 \) and \( \theta \notin \Theta_0 \) cases separately. That is
\[
R(\theta, \delta) = \begin{cases} P_{\theta}\{\delta(X) = 1\} & \text{if } \theta \in \Theta_0, \\ P_{\theta}\{\delta(X) = 0\} & \text{if } \theta \notin \Theta_0. \end{cases}
\]

That is, the risk is either the Type I or Type II error probability of the test, depending on whether the particular \( \theta \) corresponds to \( H_0 \) or \( H_1 \).

5. Let \( X = (X_1, \ldots, X_n)^\top \) be a \( n \)-variate Gaussian random vector with mean \( \theta 1 = (\theta, \ldots, \theta)^\top \) and covariance matrix identity. It is easy to check that \( (X_1, \bar{X})^\top = AX \), where \( A \) is the \( 2 \times n \) matrix
\[
A = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & n^{-1} & \cdots & n^{-1} \end{pmatrix}.
\]
Since \( AX \) is a linear function, we know that \( AX \) is also normal, with mean \( (\theta, \theta)^\top \) and covariance matrix
\[
AA^\top = \begin{pmatrix} 1 & n^{-1} \\ n^{-1} & n^{-1} \end{pmatrix}.
\]
Using the general formula for multivariate normal conditional distributions\(^1\), the conditional distribution for \( X_1 \), given \( \bar{X} = t \), is normal with mean
\[
E(X_1 | \bar{X} = t) = \theta + n^{-1} n (t - \theta) = t,
\]
and variance
\[
V(X_1 | \bar{X} = t) = 1 - n^{-1} n n^{-1} = (n - 1)/n.
\]

6. (a) The MLE is unbiased, so the risk function is the variance, \( n^{-1} \). In the normal case, the risk does not depend on \( \theta \), so the Bayes risk is also \( n^{-1} \).

\(^1\)e.g., [http://en.wikipedia.org/wiki/Multivariate_normal_distribution#Conditional_distributions](http://en.wikipedia.org/wiki/Multivariate_normal_distribution#Conditional_distributions)
(b) The Bayes rule is the posterior mean, which is \( n \bar{X} / (n+1) \). Using the bias--variance decomposition, the risk of the Bayes rule is

\[ (n + \theta^2) / (n+1)^2. \]

If we take expectation of this with respect to \( \theta \sim \mathcal{N}(0,1) \), then we get the Bayes risk of the Bayes rule: \( (n+1)^{-1} \).

(c) By definition, the Bayes rule has smaller Bayes risk than any other estimator. In this case, we can clearly see this: \( n-1 > (n+1)^{-1} \).

9. (a) From the bias--variance decomposition, the risk of the posterior mean, \( \delta(x) = \frac{a+b+x}{a+b+n} \), is a quadratic function \( R(\theta, \delta) = A\theta^2 + B\theta + C \). On some simple but tedious algebra, we find that

\[
A = \frac{(a+b)^2 - n}{(a+b+n)^2}, \quad B = \frac{n-2a(a+b)}{(a+b+n)^2}, \quad C = \frac{a^2}{(a+b+n)^2}.
\]

(b) Set \( A = B = 0 \) and solve for \( a, b \):

\[
\begin{cases}
(a+b)^2 - n = 0 \\
n - 2a(a+b) = 0
\end{cases} \iff \begin{cases}
a^2 + 2ab + b^2 - n = 0 \\
-2a^2 - 2ab + n = 0
\end{cases}
\]

Solving for \( b \) in the second equation, gives \( b = (n - 2a^2)/2a \). Plugging this in to the first equation and solving for \( a \) gives \( a = \frac{1}{2}\sqrt{n} \). Substituting this back to the expression for \( b \) gives \( b = \frac{1}{2}\sqrt{n} \).

(c) The risk function for the minimax rule—the posterior mean \( \delta_\Pi \) of \( \theta \) based on a beta prior \( \Pi \) with \( a = b = \frac{1}{2}\sqrt{n} \)—is

\[ R(\theta, \delta_\Pi) = \frac{n}{4(\sqrt{n} + n)^2}, \quad \text{a constant in } \theta \text{ (by construction)}. \]

For the MLE \( \delta(x) = x/n \), the risk function is \( R(\theta, \delta) = \theta(1-\theta)/n \). Figure 2 displays these two risk functions for several values of \( n \). As \( n \) increases, the interval where the minimax rule outperforms the MLE shrinks; also, in the interval where the minimax rule is better than the MLE, the excess risk of the MLE is shrinking with \( n \). So, although the minimax rule has smaller risk than the MLE for some \( \theta \) values, for large \( n \), the MLE is probably preferred.

10. (a) Let \( \delta \) be admissible and have constant risk, say, \( r \). By admissibility, for any other \( \delta' \), there is a point \( \hat{\theta} \), depending on \( \delta' \), such that \( R(\theta, \delta') \geq r \). This means that \( \sup_\theta R(\theta, \delta') \geq r \). Therefore, \( \delta \) is minimax. (The constant risk assumption is needed to ensure that \( \sup_\theta R(\theta, \delta) \) is attained at the point \( \hat{\theta} \).)

(b) By Example 7 in the notes, we know that \( \bar{X} \) is admissible. [The example in the notes considers \( X \sim \mathcal{N}(\theta, 1) \). In this case, \( \bar{X} \sim \mathcal{N}(\theta, n^{-1}) \), but it is easy to see that the variance plays no role in the admissibility proof.] The risk is also constant, so by part (a), we conclude that \( \bar{X} \) is minimax.
(c) It is clear from the definition of the loss function that the MLE \( \delta(X) = X/n \) has constant risk. So, to get minimaxity from part (a), it remains to prove that \( \delta \) is admissible. The strategy is to find a prior \( \Pi \) so that, for the given loss function, \( \delta(X) = X/n \) is the Bayes rule. Consider \( \Pi \) as a \( \text{Unif}(0,1) \) prior. Then the posterior distribution is \( \text{Beta}(X+1,n-X+1) \), and the posterior risk is (proportional to)

\[
\int_0^1 L(\theta,a)\theta^X(1-\theta)^{n-X} d\theta = \int_0^1 (\theta-a)^2\theta^{X-1}(1-\theta)^{n-X-1} d\theta.
\]

The posterior risk for the given loss turns out to be the same as the posterior risk under squared error loss with the posterior is \( \text{Beta}(X,n-X) \). Therefore, the \( \delta(X) \) to minimize this is the mean of \( \text{Beta}(X,n-X) \), which is \( \delta(X) = X/n \), the MLE. So, for the given loss, the MLE is the Bayes rule under a proper \( \text{Unif}(0,1) \); therefore admissibility follows from Theorem 3 in the notes.

11. It is clear that, for any \( \delta \) and any prior \( \Pi \) for \( \theta \), \( \sup_\theta R(\theta,\delta) \geq r(\Pi,\delta) \). Moreover, by definition of the Bayes rule, \( r(\Pi,\delta) \geq r(\Pi,\delta_\Pi) \), where \( \delta_\Pi \) is the corresponding Bayes rule, in this case, the posterior mean. Therefore,

\[
\sup_\theta R(\theta,\delta) \geq r(\Pi,\delta) \geq r(\Pi,\delta_\Pi),
\]
so, if we can find Bayes rules with arbitrarily large Bayes risks, then the risk function 
\( R(\theta, \delta) \) must be unbounded. Consider a conjugate prior \( \Pi_{a,b} = \text{Gamma}(a, b) \), so that the posterior mean is 
\[
\delta_{a,b}(x) = \frac{b \times x + ab}{b + 1}.
\]
The corresponding risk function is
\[
R(\theta, \delta_{a,b}) = \left( \frac{b}{b + 1} \right)^2 \theta + \left( \frac{ba - \theta}{b + 1} \right)^2,
\]
and the Bayes risk is
\[
r(\Pi_{a,b}, \delta_{a,b}) = \frac{ab^3 + ab^2}{(b + 1)^2} = \frac{ab^2}{b + 1}.
\]
Thus, we can make the Bayes risk arbitrarily large by taking large enough \( b \); therefore, it must be that \( R(\theta, \delta) \) is unbounded.

12. Consider the estimator
\[
\delta'(x) = \begin{cases} 
-1 & \text{if } x < -1, \\
x & \text{if } -1 \leq x \leq 1, \\
1 & \text{if } x > 1.
\end{cases}
\]
Truncation makes sense because, if \( |X| \) is large, then it’s most likely that the real \( \theta \) is sitting on the boundary of \([-1, 1]\). The claim is that the risk function of \( \delta' \) is uniformly no larger than the risk function for \( \delta(x) = x \), the MLE. The latter risk function is constant equal to 1 in this case, so the comparison is easy. The former risk function can be found by splitting up the sample space \((-\infty, \infty)\) into three pieces, namely, \((-\infty, -1), [-1, 1], \) and \((1, \infty)\), and doing the integration separately on each. If \( \varphi \) is the standard normal density, then the risk function looks like
\[
R(\theta, \delta') = (\theta - 1)^2 \mathbb{P}_\theta(X > 1) + (\theta + 1)^2 \mathbb{P}_\theta(X < -1) + \int_{-1}^{1} (x - \theta)^2 \varphi(x - \theta) \, dx.
\]
The latter integral can be found analytically, either by direct integration or by using mean and variance formulae for the truncated normal.² Here I will use simple numerical integration, and plot the function to verify that it’s everywhere smaller than 1. Figure 3 shows the result—it’s clearly below 1 for all \( \theta \in [-1, 1] \). (You can also look at the plot of the risk function for the Bayes/minimax rule in the notes; this one is also less than 1 for all \( \theta \), proving that the MLE is inadmissible.)

15. Part (a) only. For data \( X_1, \ldots, X_n \overset{iid}{\sim} \text{Unif}(\theta - 1, \theta + 1) \), the likelihood function looks like
\[
L(\theta) = 2^{-n} I_{X_n - 1, X_{(1)} + 1}(\theta),
\]
where \( X_{(1)} \) and \( X_{(n)} \) are the sample minimum and maximum, respectively. That is, the likelihood function is positive only for \( \theta \) between \( X_{(1)} - 1 \) and \( X_{(1)} + 1 \). In this case, the Pitman estimator is
\[
\hat{\theta}_{\text{pit}} = \frac{\int_{-\infty}^{\infty} \theta L(\theta) \, d\theta}{\int_{-\infty}^{\infty} L(\theta) \, d\theta} = \frac{\int_{X_{(n)} - 1}^{X_{(n)} + 1} \theta \, d\theta}{\int_{X_{(n)} - 1}^{X_{(n)} + 1} d\theta},
\]
³http://en.wikipedia.org/wiki/Truncated_normal_distribution
The denominator is $2 - (X_{(n)} - X_{(1)})$ and the numerator is $\frac{1}{2}[(X_{(1)} + 1)^2 - (X_{(n)} - 1)^2]$; therefore, the Pitman estimator is

$$\hat{\theta}_{\text{pit}} = \frac{1}{2}[(X_{(1)} + 1)^2 - (X_{(n)} - 1)^2] = \frac{X_{(1)} + X_{(n)}}{2 - (X_{(n)} - X_{(1)})}.$$  

This is the sample mid-point.