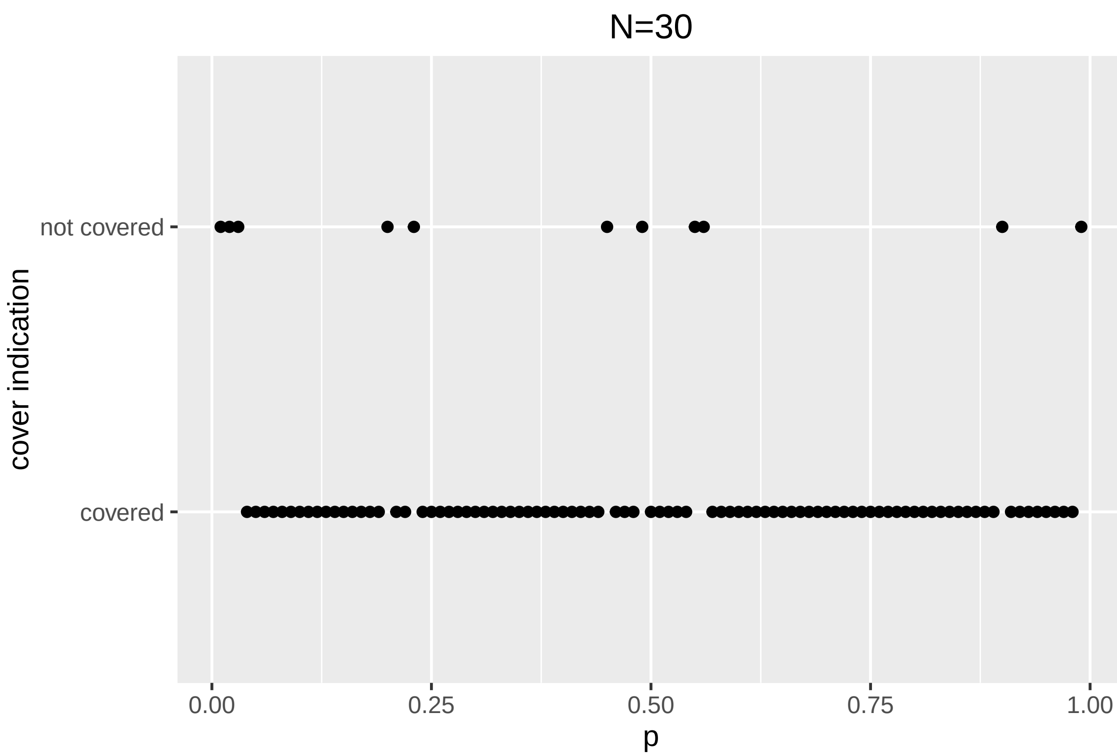
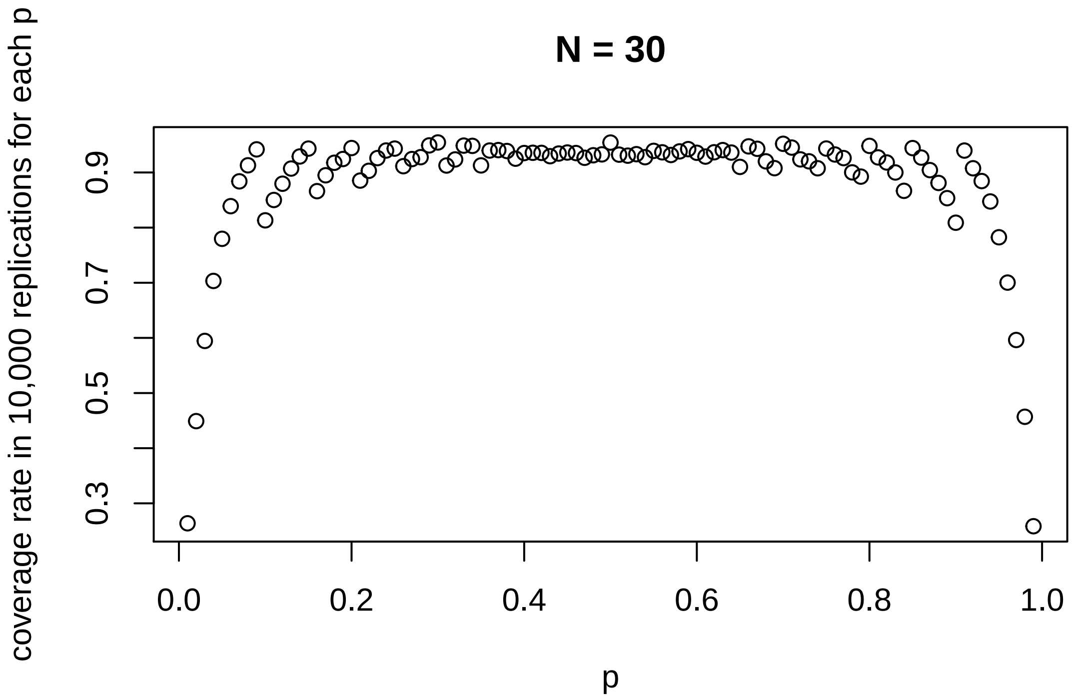
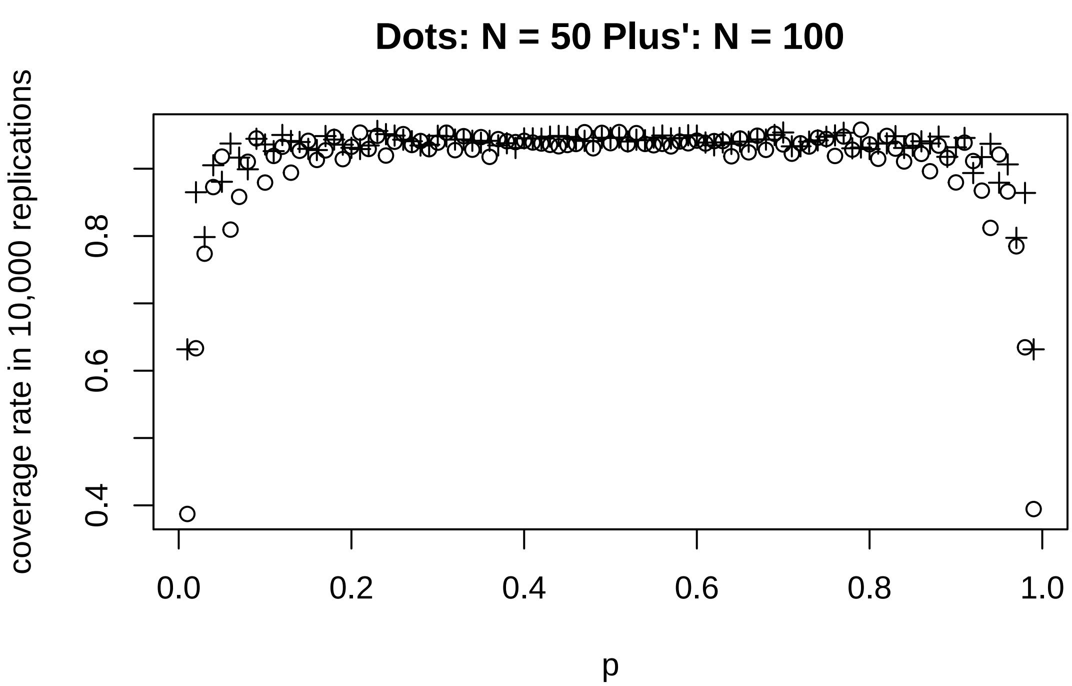
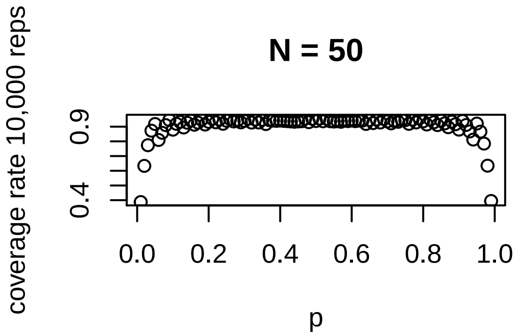
1. We see that for values for p not close to 0 or 1, that the confidence level is maintained well for both sample sizes 50 and 100 in 10,000 replications. For very small and very large proportions, it is more likely to get all failures or all successes (respectively), making for no estimate of the variation in the proportion of successes. Then, the proportion, p, is contained within fewer 10,000 confidence intervals, making the confidence level not maintained as well. For the larger sample size, 100, the confidence level is more maintained on the boundaries of p than for the smaller sample size (50). I am not surprised because for the Wald Interval estimate for p to be valid (meaning maintained specified confidence level), the expected number of successes, np, needs to be large enough to have at least 10 successes and 10 failures, which does not hold on the boundaries with the sample sizes 50 and 100. For example, 0.1*100 = 1.

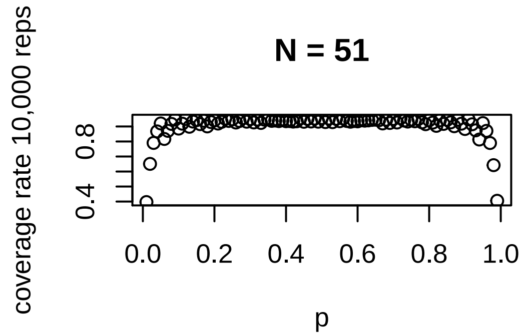


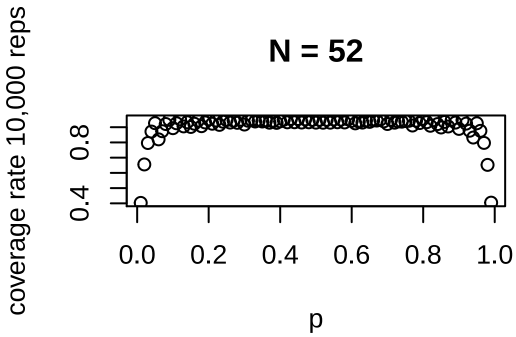
I read too quickly over the set notation, but my code did run for sample sizes between 50 and 100, so I thought I would include them as well.

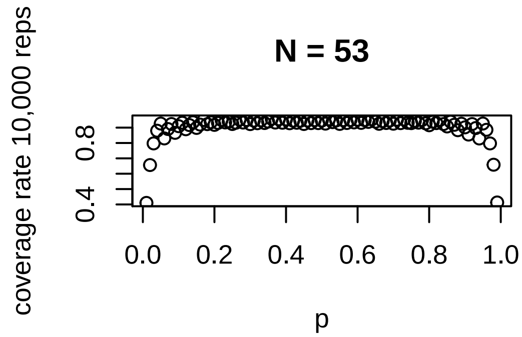

N = 30
coverage rate in 10,000 replications for each p
0.3
0.5
0.7
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

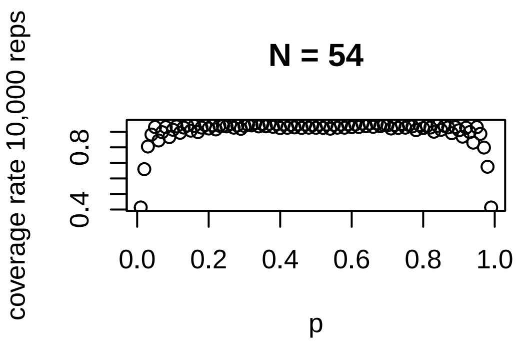

Dots: N = 50 Plus': N = 100
coverage rate in 10,000 replications
0.4
0.6
0.8
0.0
0.2
0.4
0.6
0.8
1.0
p

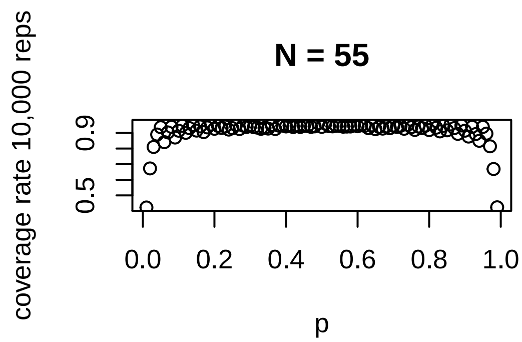

N = 50
coverage rate 10,000 reps
0.4
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

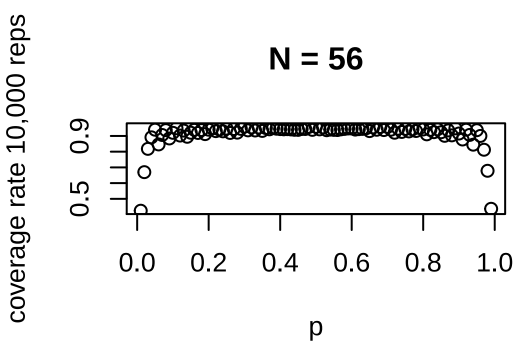

N = 51
coverage rate 10,000 reps
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
p

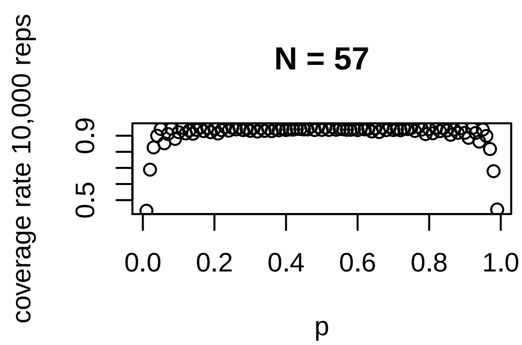

N = 52
coverage rate 10,000 reps
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
p

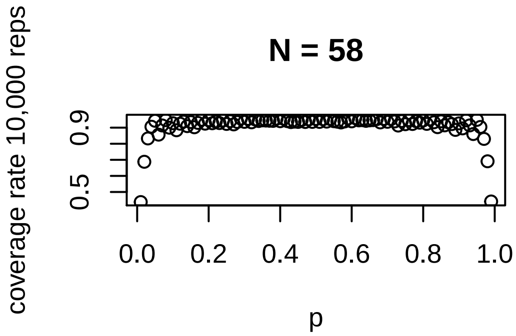

N = 53
coverage rate 10,000 reps
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
p

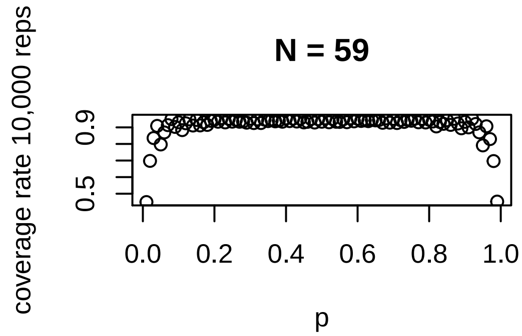

N = 54
coverage rate 10,000 reps
0.4
0.8
0.0 0.2 0.4 0.6 0.8 1.0
p

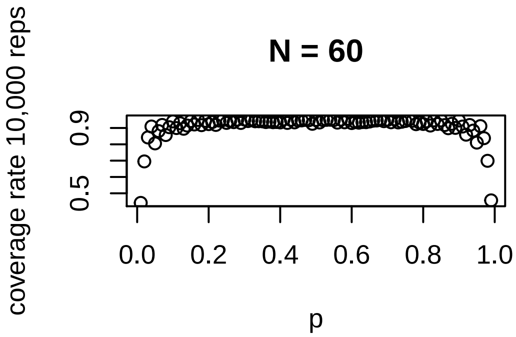

N = 55
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

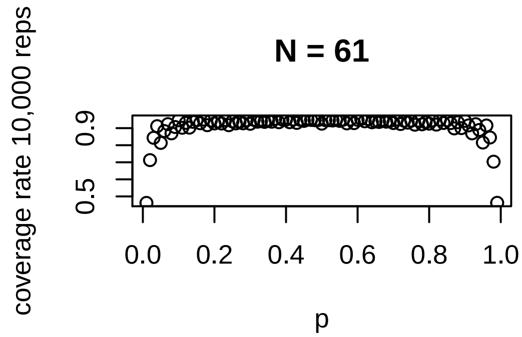

N = 56
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

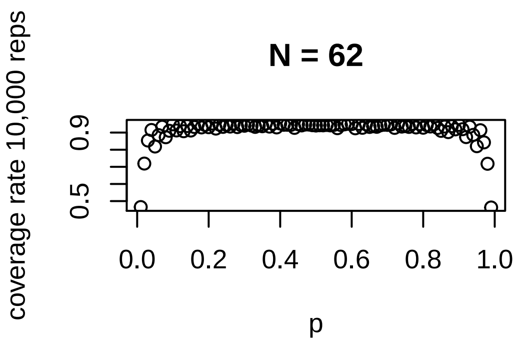

N = 57
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

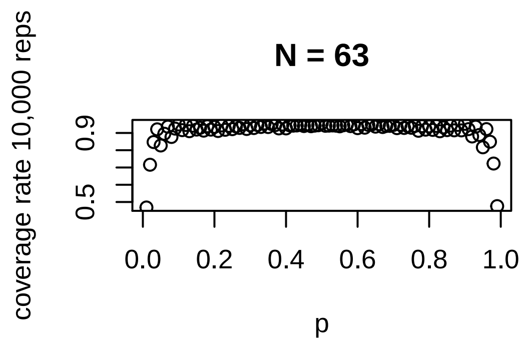

N = 58
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

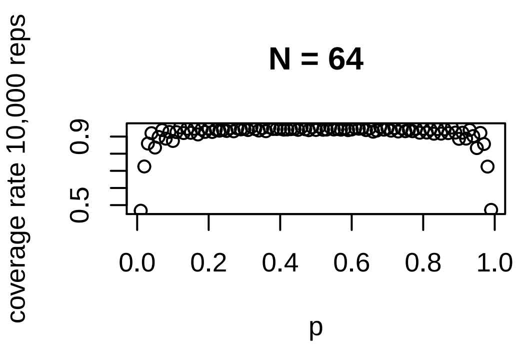

N = 59
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

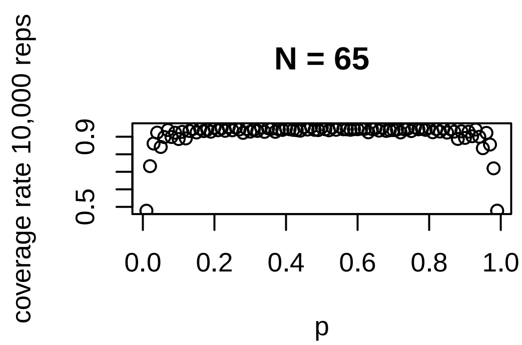

N = 60
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

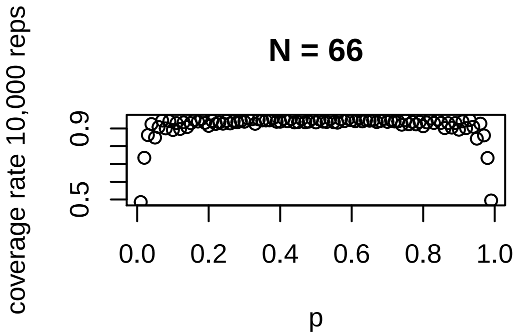

N = 61
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

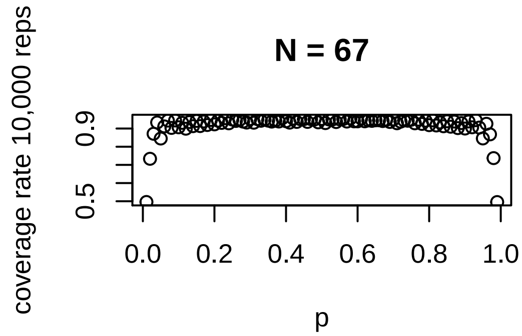

N = 62
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

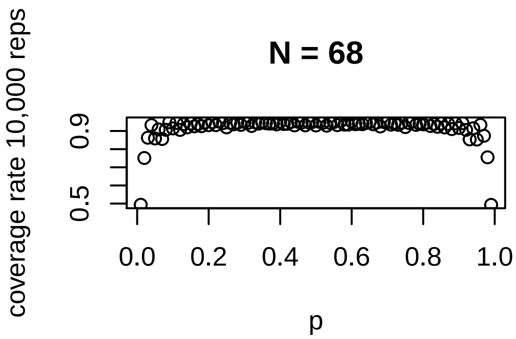

N = 63
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

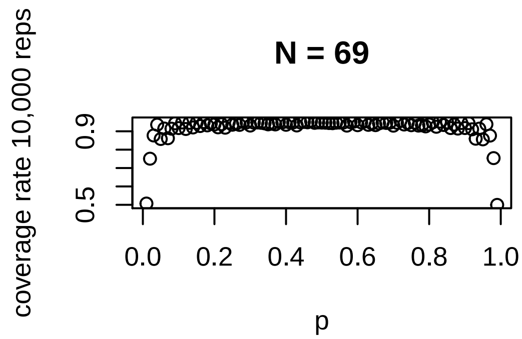

N = 64
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

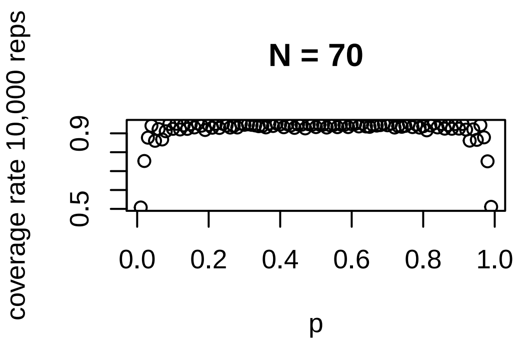

N = 65
coverage rate 10,000 reps
0.5
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

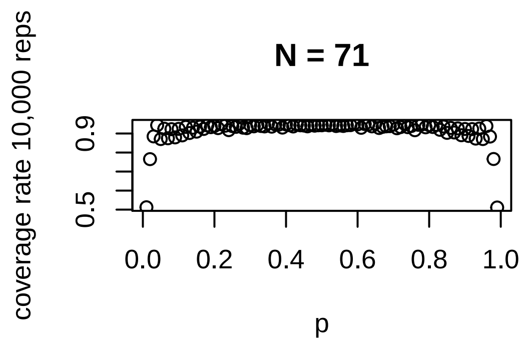

N = 66
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

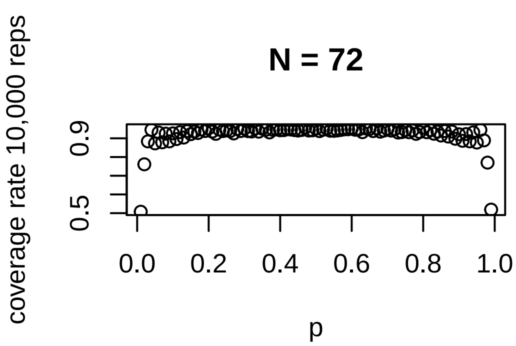

N = 67
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

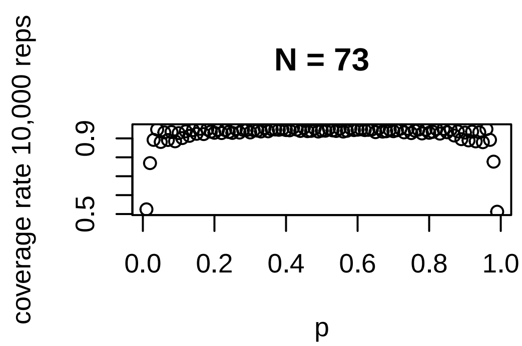

N = 68
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

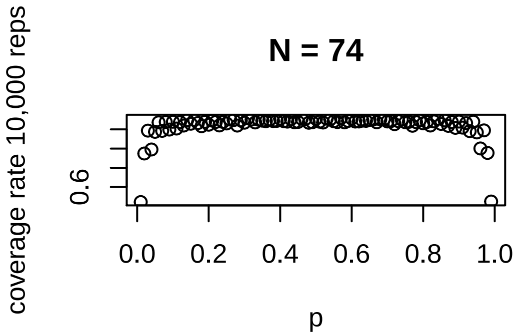

N = 69
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

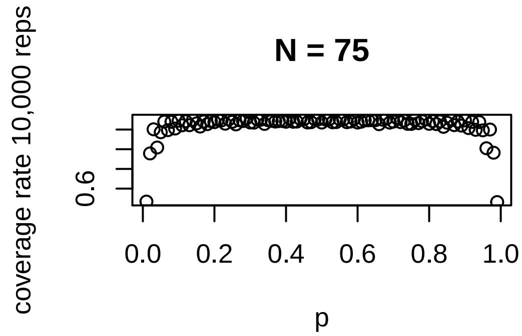

N = 70
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

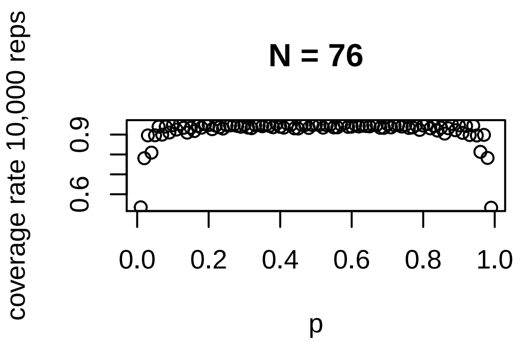

N = 71
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

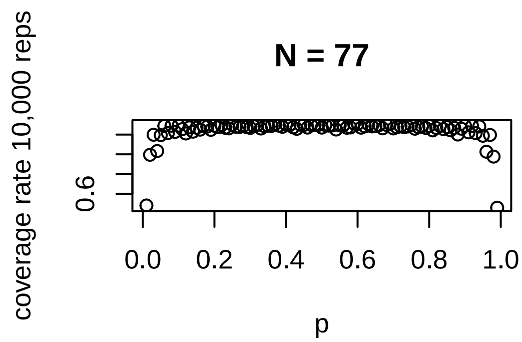

N = 72
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

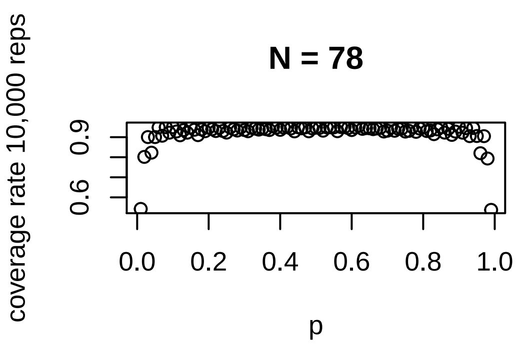

N = 73
coverage rate 10,000 reps
0.5
0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

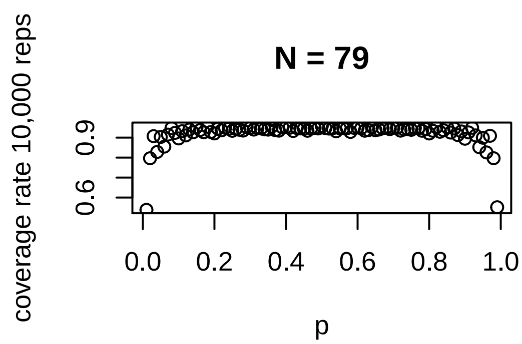

N = 74
coverage rate 10,000 reps
0.6
0.0 0.2 0.4 0.6 0.8 1.0
p

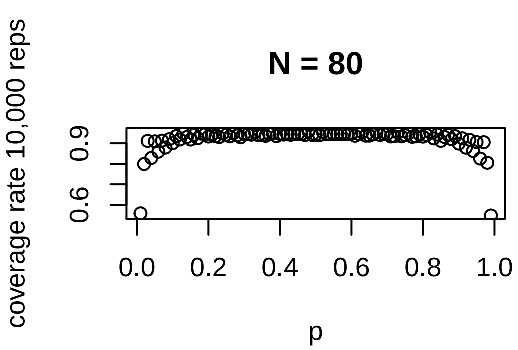

N = 75
coverage rate 10,000 reps
0.6
0.0 0.2 0.4 0.6 0.8 1.0
p

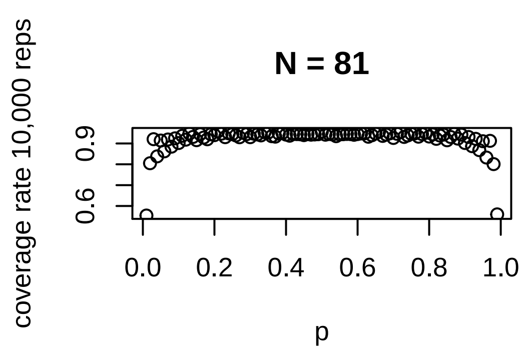

N = 76
coverage rate 10,000 reps
0.6 0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

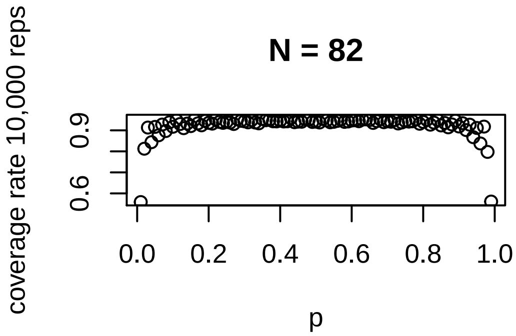

N = 77
coverage rate 10,000 reps
0.6
0.0 0.2 0.4 0.6 0.8 1.0
p

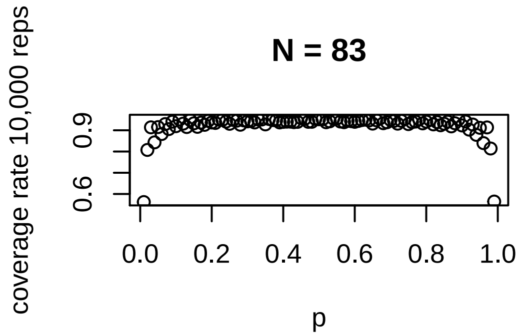

N = 78
coverage rate 10,000 reps
0.6 0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

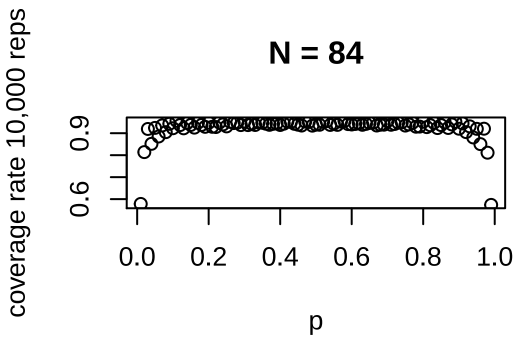

N = 79
coverage rate 10,000 reps
0.6 0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

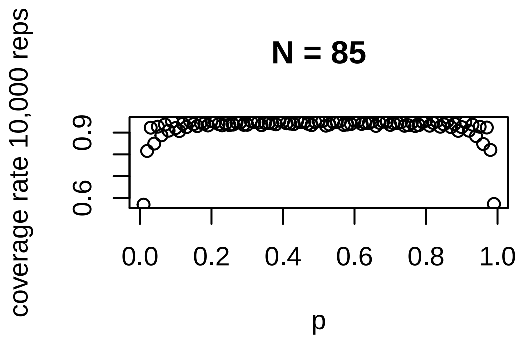

N = 80
coverage rate 10,000 reps
0.6 0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

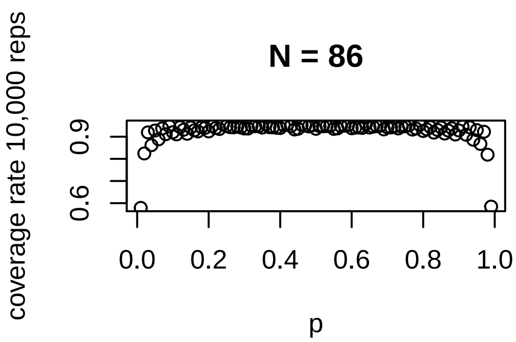

N = 81
coverage rate 10,000 reps
0.6 0.9
0.0 0.2 0.4 0.6 0.8 1.0
p

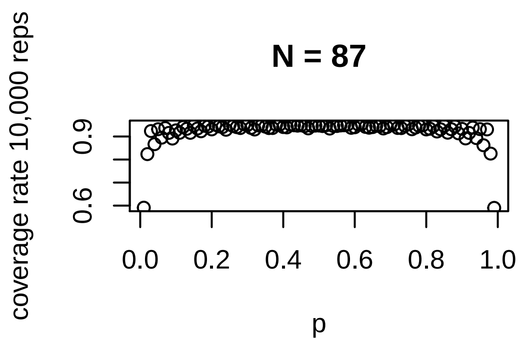

N = 82
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

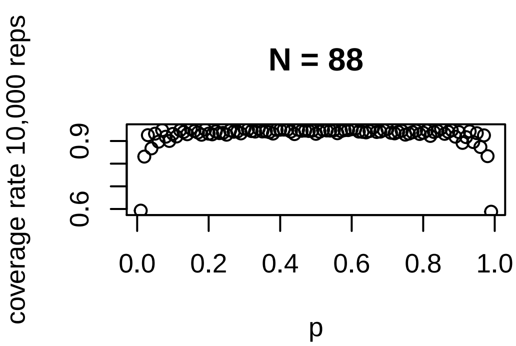

N = 83
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

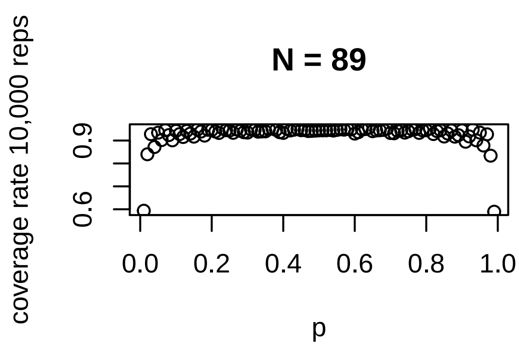

N = 84
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

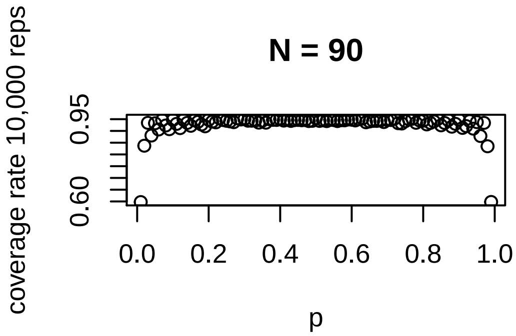

N = 85
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

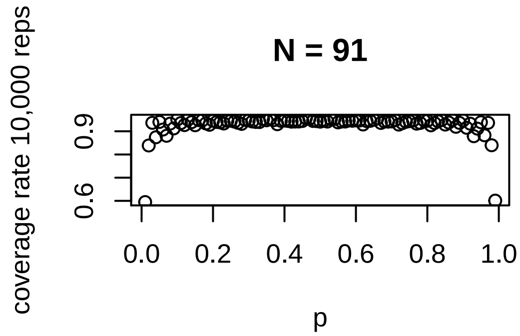

N = 86
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

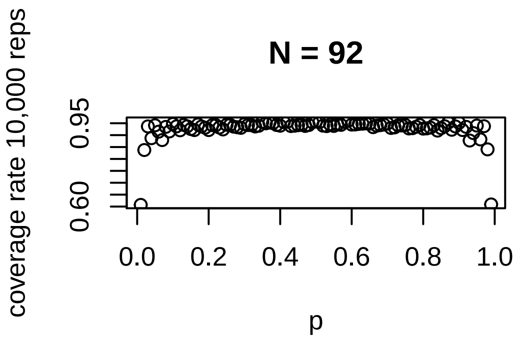

N = 87
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

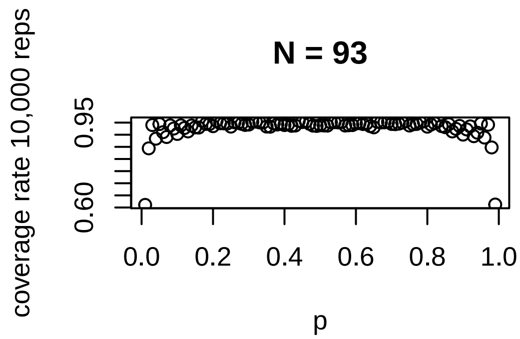

N = 88
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

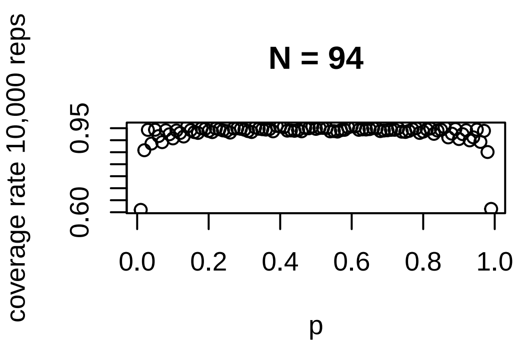

N = 89
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

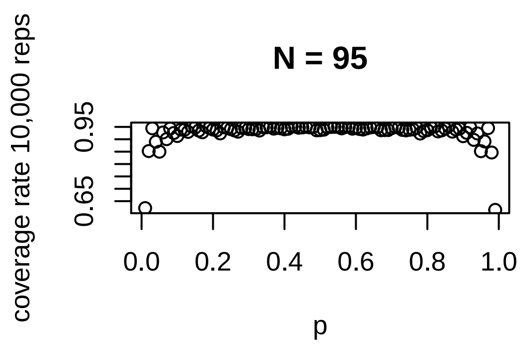

N = 90
coverage rate 10,000 reps
0.60
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p

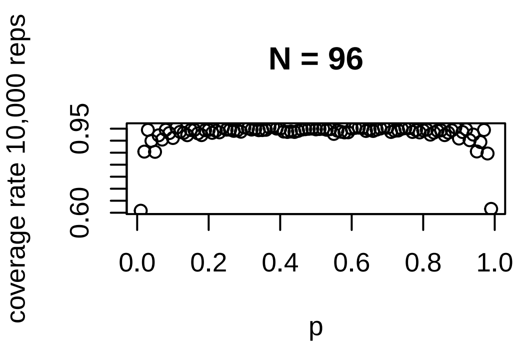

N = 91
coverage rate 10,000 reps
0.6
0.9
0.0
0.2
0.4
0.6
0.8
1.0
p

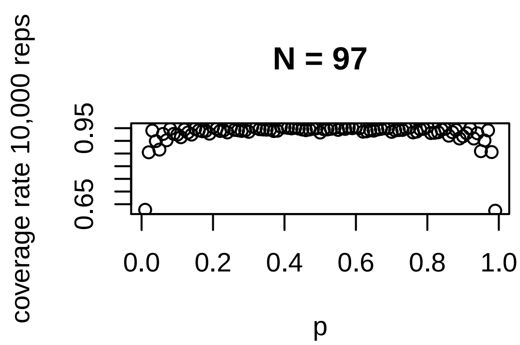

N = 92
coverage rate 10,000 reps
0.60
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p

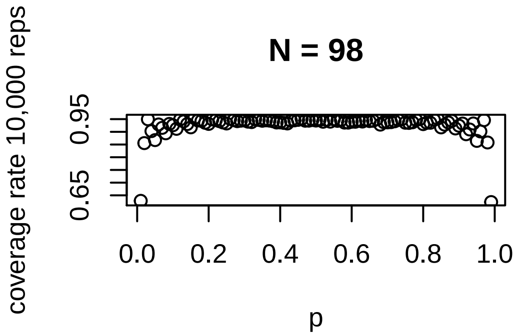

N = 93
coverage rate 10,000 reps
0.60
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p

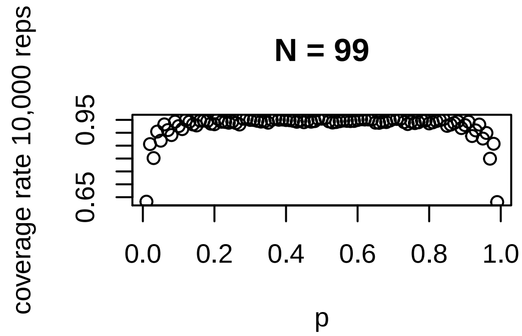

N = 94
coverage rate 10,000 reps
0.60
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p

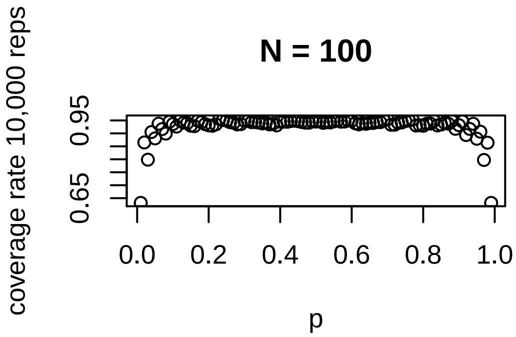

N = 95
coverage rate 10,000 reps
0.65
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p


N = 96
coverage rate 10,000 reps
0.60
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p


N = 97
coverage rate 10,000 reps
0.65
0.95
0.0
0.2
0.4
0.6
0.8
1.0
p







2. X $\sim$ BIN(N, θ)

   $\theta \sim$ UNIF(0,1)

   $\theta|$ Y $\sim$ BETA(1+Σy_i, 1 + N - Σy_i)

   $\Lambda = log(\frac{\theta}{1-\theta})$

   (a) p(θ) = 1 $\cdot$I$(\theta)_{[0,1]}$

   $\theta = \frac{e^{\Lambda}}{1+e^{\Lambda}}$

   $\frac{d\theta}{d\Lambda} = \frac{(1+e^{\Lambda})e^{\Lambda} - e^{2\Lambda}}{(1+e^{\Lambda})^2}$

   $= \frac{e^{\Lambda}}{(1+e^{\Lambda})^2}$

   $\theta = 0 \Rightarrow \Lambda =$ -infinite $\theta = 1 \Rightarrow \Lambda =$ infinite

   p(Λ) = $\frac{e^{\Lambda}}{(1+e^{\Lambda})^2}$ I$(\Lambda)_{[-infinite, infinite]}$

   $\Lambda \sim$ LOGISTIC(0,1)

   (b) $\theta = \frac{e^{\Lambda}}{1+e^{\Lambda}}$

$\Lambda \sim$ UNIF(0,1)

$\Lambda = \log(\frac{\theta}{1-\theta})$

$\frac{d\Lambda}{d\theta} = \theta^{-1}(1-\theta)^{-1}$

$\Lambda = 0 \Rightarrow \theta = \frac{1}{2}$ $\Lambda = 1 \Rightarrow \theta = \frac{e}{1+e}$

p$(\theta) = \theta^{-1}(1-\theta)^{-1}$ I$(\theta)_{[\frac{1}{2},\frac{e}{1+e}]}$

The pdf of θ resembles the kernal of the improper BETA(0,0), however, in the BETA distribution, the parameters (α and β) are strictly greater than 0. p(θ) is improper because $I(\theta)_{[\frac{1}{2},\frac{e}{1+e}]}$, and not $I(\theta)_{[0,1]}$ if we are assuming $\theta \sim$ BETA(α, β).

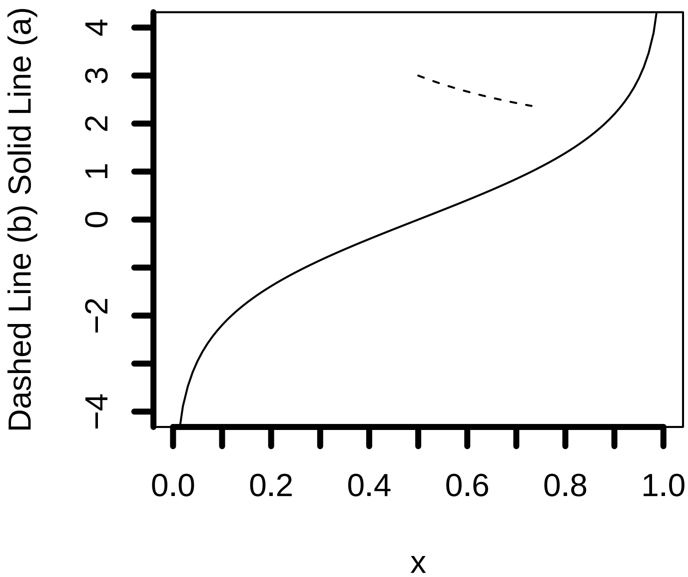
(c) The BETA distribution has domain (0,1) and so is often used as a pdf to model the prior distribution of the probability of "success" in a binary outcome variable that follows a BINOM pdf.

In both (a) and (b), we began with the same uninformative prior about a parameter, and then we did a transformation of that parameter. In (a) and (b) we actually did inverse transformations of the parameter. In both cases, we ended with an informative prior, whether a valid and proper pdf (a) or not (b), which seems counter intuitive since we can go from a pdf that should not contain information about a parameter, that when transformed does provide information about it.

In class we stated that it does not matter whether a prior is proper or improper, as long as the posterior is proper. However, in (b), the posterior will end up $\propto$ BETA(Y-1,N-Y-1), which restricts Y to being at least 1 and (N-Y) to being at least 1. So I am not sure about whether it would even be correct to use the transformation and posterior found in (b).

I think the point of this is that first, even though we start with an uninformative prior, we can transform it to get an informative prior. The nature of the informative prior depends on the transformation done, which can be subjective. It seems that if we start with an uninformative prior for a parameter, a transformation of that parameter should also be uninformative.

Below I've plotted the logistic curve for x values from -10,10 and the pdf found in (b) on it's domain.



3. Y $\sim$ POIS(θ)

   $\theta \sim$ GAM(α, β)

   p($\theta|Y$) $\propto \int_\theta \frac{\theta^{\Sigma y_i} \dot{e}^{-\theta}}{\Pi y_i!} \cdot \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^\alpha} d\theta$

   $\propto \int_\theta \theta^{\Sigma y_i + \alpha - 1} \cdot e^{-\theta(1+\frac{1}{\beta})}$

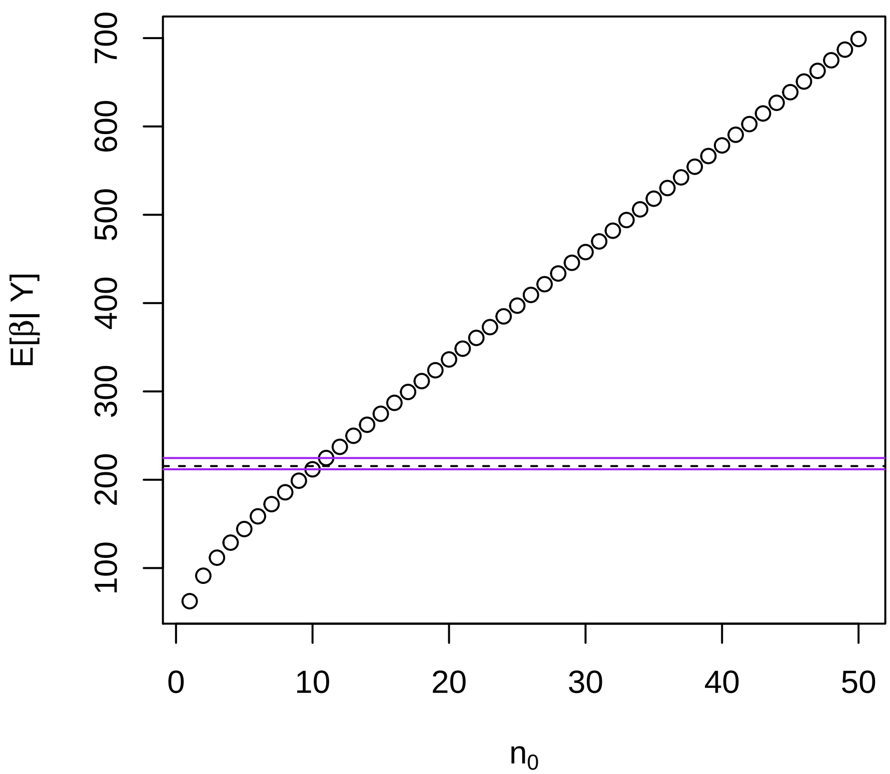
   $\theta|Y \sim$ GAM($\Sigma y_i + \alpha$, $\frac{1}{1+\frac{1}{\beta}}$)

   (a) Both posterior distributions are from the class of gamma distributions.

   $\theta_A|Y \sim$ GAM(237,$\frac{1}{1.1}$)

   $\theta_B|Y \sim$ GAM(125,$\frac{1}{2}$)

	Posterior Shape	Posterior Scale	Posterior Mean	Posterior Variance	Lower Bound	Upper Bound
A	237.00	0.91	215.45	195.87	188.90	243.74
B	125.00	0.50	62.50	31.25	52.02	73.92



(b)

The solid lines represent the posterior expections of θ_B and the dashed line represents the posetior expectation of θ_A . The E[θ_A| y] was 215.45. For the posterior of, θ_B to be be close to that of θ_A, n_o needs to be between 10 and 11, which lead to posterior expectations of 211.8182 and 224.5833 . Note that the priors on θA and θ_B are the same if n_o is 10, but different samples were observed for the two parameters, making the posterior expectations differ slightly.

4. Fisher's Information is $I(\theta)$. Binomial (which means Bernoulli as well) distributions are members of the exponential family, and so the regularity conditions hold.

   Where $\underset{\sim}{y} = \Sigma x_i$,

   $I(\theta) = -nE[(\frac{d^2}{d\theta^2} log(f(x|\theta)))]$

   $= -E[(\frac{d^2}{d\theta^2} log(f(\underset{\sim}{y}|\theta)))]$

   When $\underset{\sim}{y}|\theta \sim$ BIN(N,θ)

   $x|\theta \sim$ BERN(θ)

$I(\theta) = -nE[(\frac{d^2}{d\theta^2} log(\theta^x \cdot (1-\theta)^{1-x})]$

$= -nE[(\frac{d^2}{d\theta^2}[xlog(\theta) + (1-x)log(1-\theta)]]$

$= -nE[(\frac{d}{d\theta})[\frac{x}{\theta} - (1-\theta)^{-1} + \frac{x}{1-\theta}]]$

$= -nE[\frac{-x}{\theta^2} - (1-\theta)^{-2} + \frac{x}{(1-\theta)^2}]$

$= -n[-\theta^{-1} - (1-\theta)^{-2} + \frac{\theta}{(1-\theta)^2}]$

$= -n[-\theta^{-1} - (1-\theta)^{-1}]$

$= n[\theta^{-1} + (1-\theta)^{-1}]$

$= n\frac{[(1-\theta)+\theta]}{\theta(1-\theta)}$

$= \frac{n}{\theta(1-\theta)}$

Jeffrey's Prior $p_j(\theta)$ is then $(\frac{n}{\theta(1-\theta)})^{\frac{1}{2}}$