

STAT 221: Problem Set 2  
Kevin Eskici and Willy Xiao  
Due: Oct 7, 2014

Task 1: Examine the Posterior

Posterior is:  $f(\mu, \sigma^2, \log \vec{\theta} | Y)$

$$\begin{aligned} f(\mu, \sigma^2, \log \vec{\theta} | Y) &\propto f(Y | \log \vec{\theta}) * f(\log \vec{\theta} | \mu, \sigma^2) * f(\mu, \sigma^2) \\ &\propto \prod_{j=1}^J \left[ \left( \prod_{i=1}^N e^{-w_j e^{\log \theta_j}} (w_j e^{\log \theta_j})^{Y_{ji}} \right) * \frac{1}{\sigma} e^{-\frac{(\log \theta_j - \mu)^2}{2\sigma^2}} * \frac{1}{\sigma^2} \right] \end{aligned}$$

The Posterior of  $\log \theta$  conditional on all the other parameters is the same as the previous equation, except we can drop the prior on  $\mu$  and  $\sigma^2$  because they're given:

$$f(\log \theta | Y, \mu, \sigma^2) \propto \prod_{j=1}^J \left[ \left( \prod_{i=1}^N e^{-w_j e^{\log \theta_j}} (w_j e^{\log \theta_j})^{Y_{ji}} \right) * \frac{1}{\sigma} e^{-\frac{(\log \theta_j - \mu)^2}{2\sigma^2}} \right]$$

To check the shape of the function, we can look at the second derivative of the log-posterior:

$$\begin{aligned} \text{Let } \log p &= \log f(\log \theta | Y, \mu, \sigma^2) \\ &= \sum_{j=1}^J \left( \sum_{i=1}^N \left( -w_j e^{\log \theta_j} + Y_{ji} * (\log w_j + \log e^{\log \theta_j}) \right) - \log \sigma - \frac{(\log \theta_j - \mu)^2}{2\sigma^2} \right) \\ \frac{\partial \log p}{\partial \log \theta_j} &= \sum_{j=1}^J \left( \sum_{i=1}^N \left( -w_j e^{\log \theta_j} + Y_{ji} \log \theta_j \right) - \frac{\log \theta_j - \mu}{\sigma^2} \right) \\ \frac{\partial^2 \log p}{\partial \log \theta_j^2} &= \sum_{j=1}^J \left( \sum_{i=1}^N \left( -w_j e^{\log \theta_j} + Y_{ji} \right) - \frac{1}{\sigma^2} \right) \end{aligned}$$

Because the second derivative of the log-posterior is monotonically decreasing with respect to  $\log \vec{\theta}$ , that means our function is unimodal (ie there's a single peak).

Task 2: Write functions to simulate data from the model

Check `keskici_wxiao_ps2_functions.R`

Task 3: Evaluate coverage for a simple case

First: Estimate the amount of simulations we can do:

- a) The time it took to run one simulation on my Macbook Air was roughly 23 seconds. Accounting for startup costs (copying to multiple machines, however Odyssey decides to manage nodes etc.), we roughly said that each MCMC simulation would take 25 seconds.

- b) Calculating, we have

$$(60 \text{ seconds}) \times (60 \text{ minutes}) \times (12 \text{ nodes}) = 43200 \text{ seconds of runtime.}$$

Then

$$\frac{43200}{(4 \text{ parameters})} / (25 \text{ seconds per simulation}) = 432 \text{ simulations per parameter.}$$

We ultimately decided to go with 360 simulations per parameter to guarantee that we don't go over the hard one-hour time limit. Also 360 is a number that is divisible by 12 and easily modeled on our local machines to test (e.g. running 36 simulations rather than 360), which makes writing the .slurm job a little bit easier.

Second: Decide how many theta's to draw and how many Y's to draw:

Ultimately we weren't sure how best to do this. The example on the pset had more theta draws than Y draws, so we decided to do the same.

$$\text{theta.draws} = 30$$

$$\text{Y.draws} = 12$$

Third: RUN! :)

Unfortunately, the timing was wildly off. It didn't take into account that we had to calculate the coverage after each theta draw which means we had to recalculate how many times simulations we could do! We tried  $12 \times 15$  (ie half of what it was before, and it still went too long),  $12 \times 12$  was close. Ultimately we ran it for:

$$\text{theta.draws} = 9$$

$$\text{Y.draws} = 15$$

Task 4: Evaluate coverage with exposure weights

We ran the same simulation as Task 3, except we added in the exposure weights. Results discussion below, code in relevant file.

Task 5: Evaluate coverage with model misspecification:

We ran the same simulation as Task 4, substituting in rASL instead of rnorm as our function to generate  $\log \theta_j$ . For running the MCMC simulation, we left the parameters mu and sigmasq to be their defaults 0 and 1 respectively.

Task 6: Interpretation of results:

We noticed that the coverage percentages in task 3 seemed to peak around the true value of  $\mu$  for both the 68% and 95% confidence intervals. Additionally, we see in figures 1 and 2 that the coverage percentages were far lower than 68/95 for low values of  $\theta$  than they are in figures 3 and 4 (though in figures 1 and 2 the coverage percentages rapidly increase as we approach the true parameter value before having a slight dip). From this we can infer that when we have higher values of  $\mu$ , the frequentist coverage properties hold better. This can be explained by the prior proportionally having less weight in the posterior when our values of  $\mu$  are higher. It is easiest to see this effect by comparing figures 3 and 4 because they were generated with the same values of  $\sigma^2$  with different  $\mu$ s. It is also interesting to note that while there was an increase in  $\mu$  in figure 2 compared to figure 1, the increase in variance seems to have partially offset it leaving the 68% coverages pretty similar. In the 95% CI case the increased variance seemed to have less of an impact and the coverages were better for low values of  $\log(\theta_j)$ .

In task 4 we added exposure weights to the mix. Comparing the figures for task 4 (figures 5,6,7, and 8) to their corresponding figures for task 3 (figures 1,2,3, and 4 respectively), we see that the addition of exposure weights to the model made the coverage peaks at the true values of  $\mu$  more pronounced. Again we saw that this peak coverage was not much higher than the rest of the curve for cases where our  $\mu$ s were higher (comparing figures 5 and 6 to figures 7 and 8). We also see that the coverage curves for the  $\log(w_j)$  are almost horizontal at 68/95, and are flatter for the cases with higher  $\mu$ s as seen in figures 7 and 8 as opposed to figures 5 and 6 where they seem to curve a bit before converging to a line at 68/95 after a certain threshold.

In task 5 we evaluated coverage for a misspecified model. Given this, it is no surprise that the coverage rates were the furthest from 68/95 for this model. Comparing figures 10 and 11 we see that adding the  $.7X$  term instead of subtracting it gives us much better coverage, which again seems like a result of there being a higher mean (and the model isn't AS off (though it is still misspecified) in that case). Interestingly it seems the weights have a bit more of an effect for the parameters shown in figure 11 than they do in figure 10. Comparing figures 9 and 12 we also see that increasing our error term (and thus also the variance) doesn't have as much of an effect on coverage in our misspecified model, likely because without the exponential term our prior still has a large weight in both cases.

Overall we see that the more weight our prior is given in our posterior distribution relative to the likelihood, the less frequentist coverage properties hold.

Appendix:

Figure 1: Task 3 coverage plots for  $\log(\theta_j)$ 's :  $\mu = 1.6, \sigma^2 = 0.7^2$

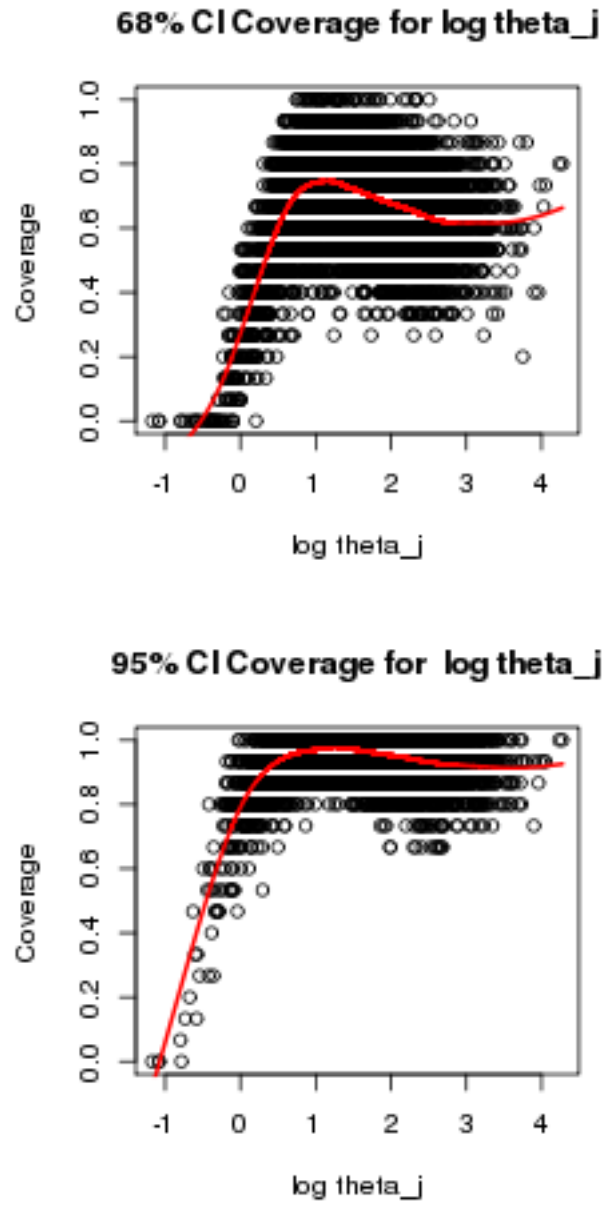


Figure 2: Task 3 coverage plots for  $\log(\theta_j)$ 's :  $\mu = 2.5, \sigma^2 = 1.3^2$

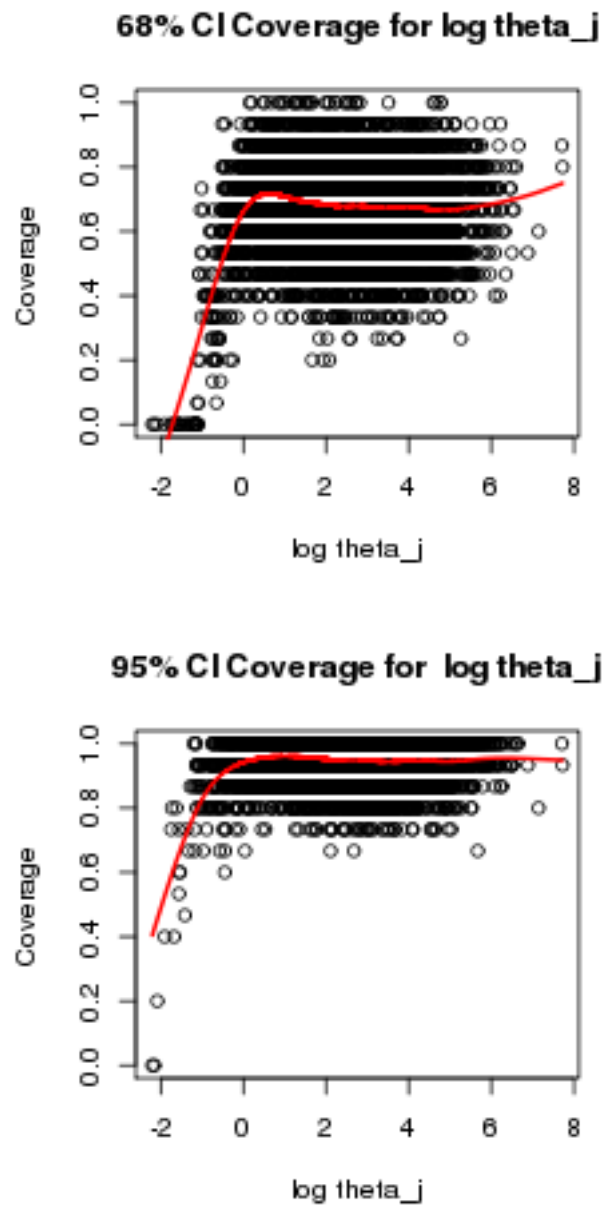


Figure 3: Task 3 coverage plots for  $\log(\theta_j)$ 's :  $\mu = 5.2, \sigma^2 = 1.3^2$

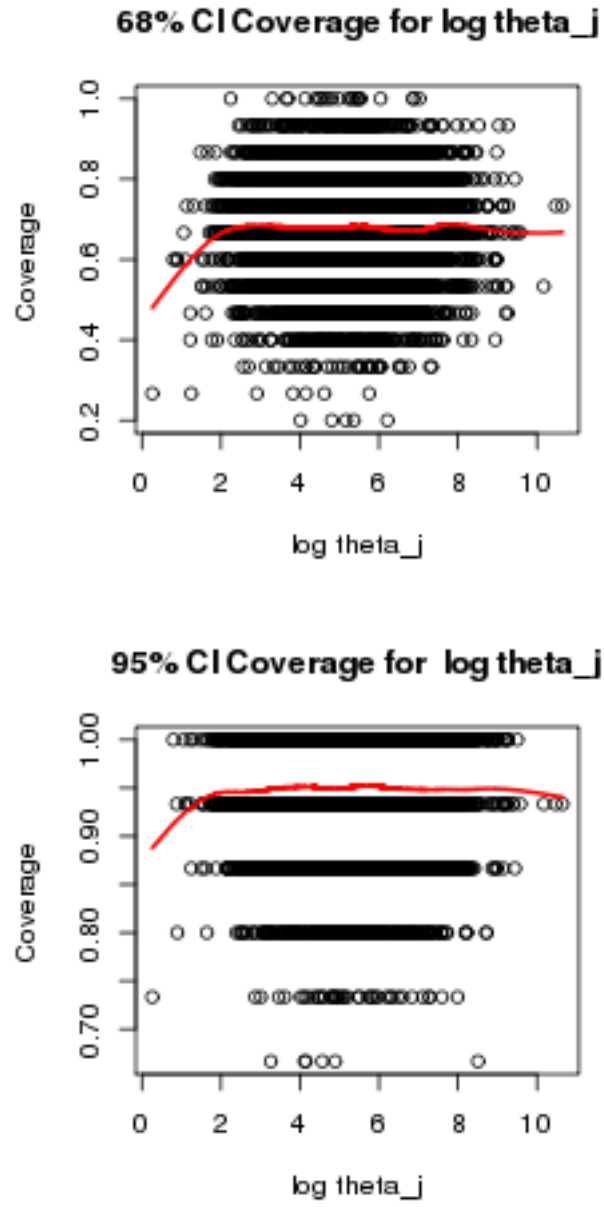


Figure 4: Task 3 coverage plots for  $\log(\theta_j)$ 's :  $\mu = 4.9$ ,  $\sigma^2 = 1.6^2$

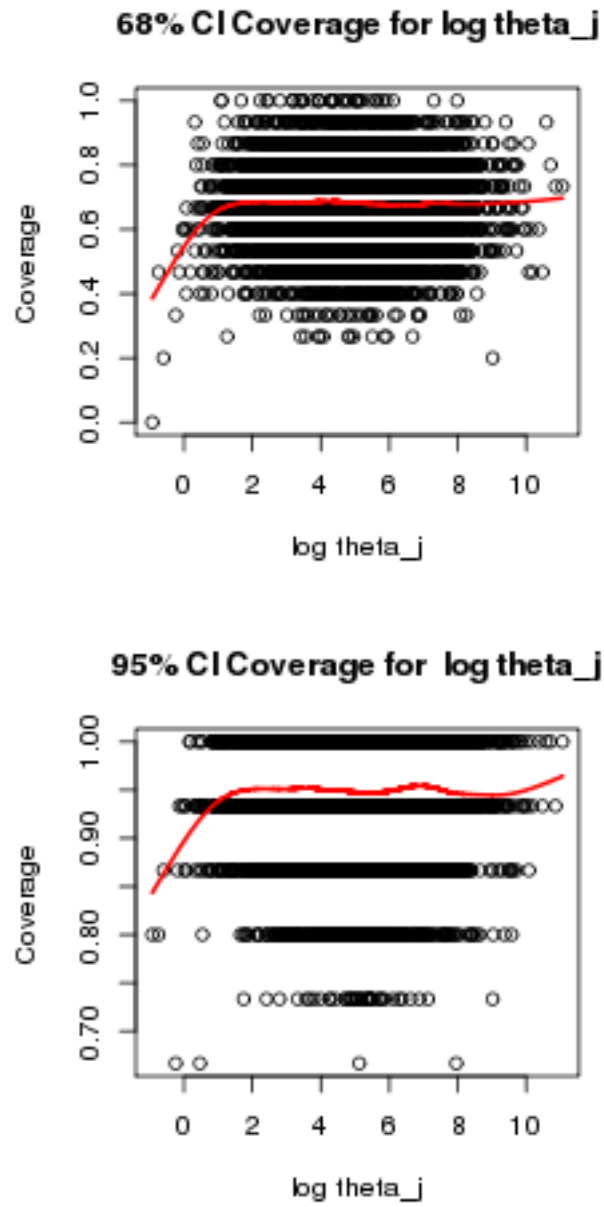


Figure 5: Task 4 coverage plots for  $\mu = 1.6$ ,  $\sigma^2 = 0.7^2$

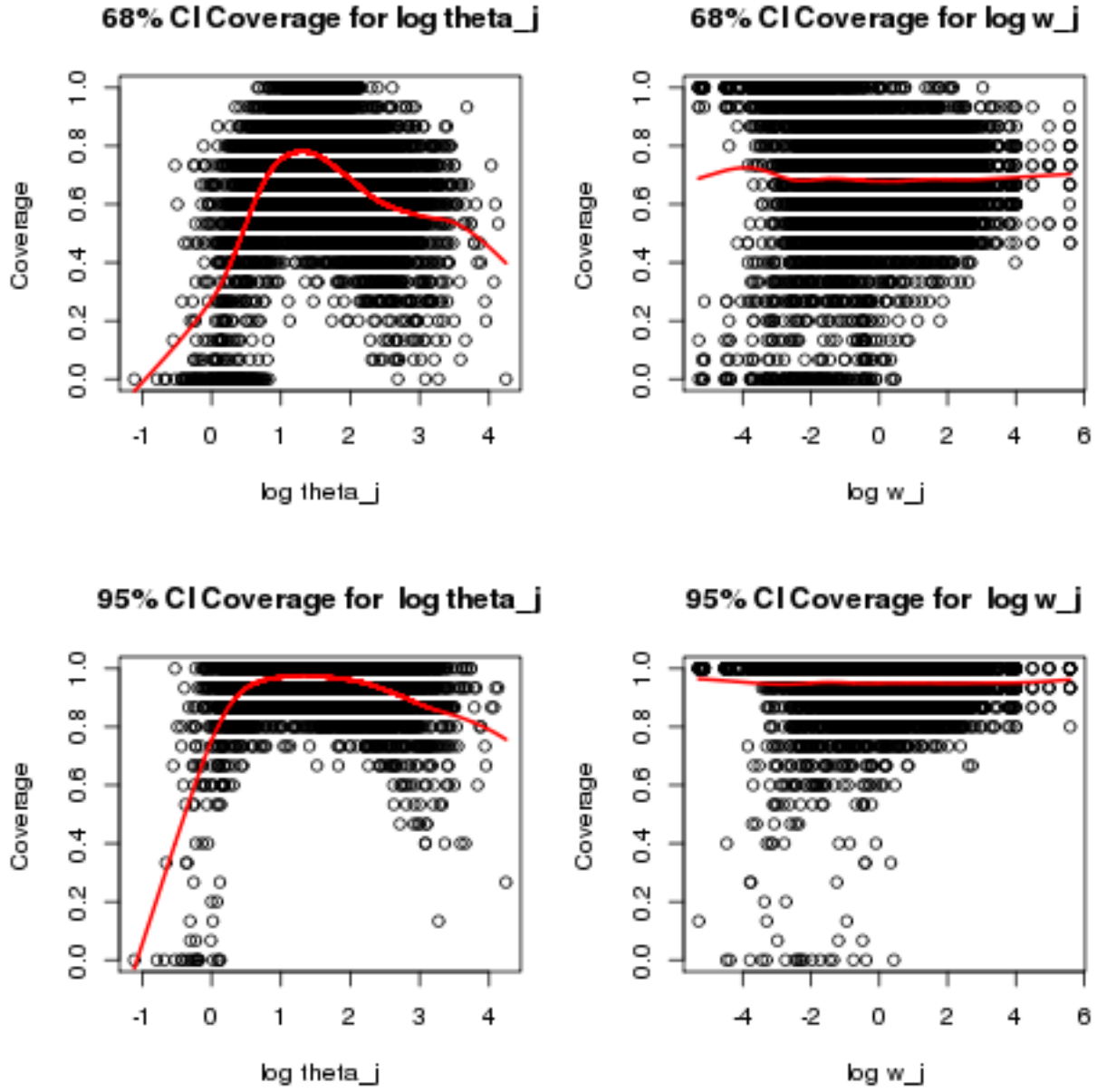




Figure 6: Task 4 coverage plots for  $\mu = 2.5$ ,  $\sigma^2 = 1.3^2$

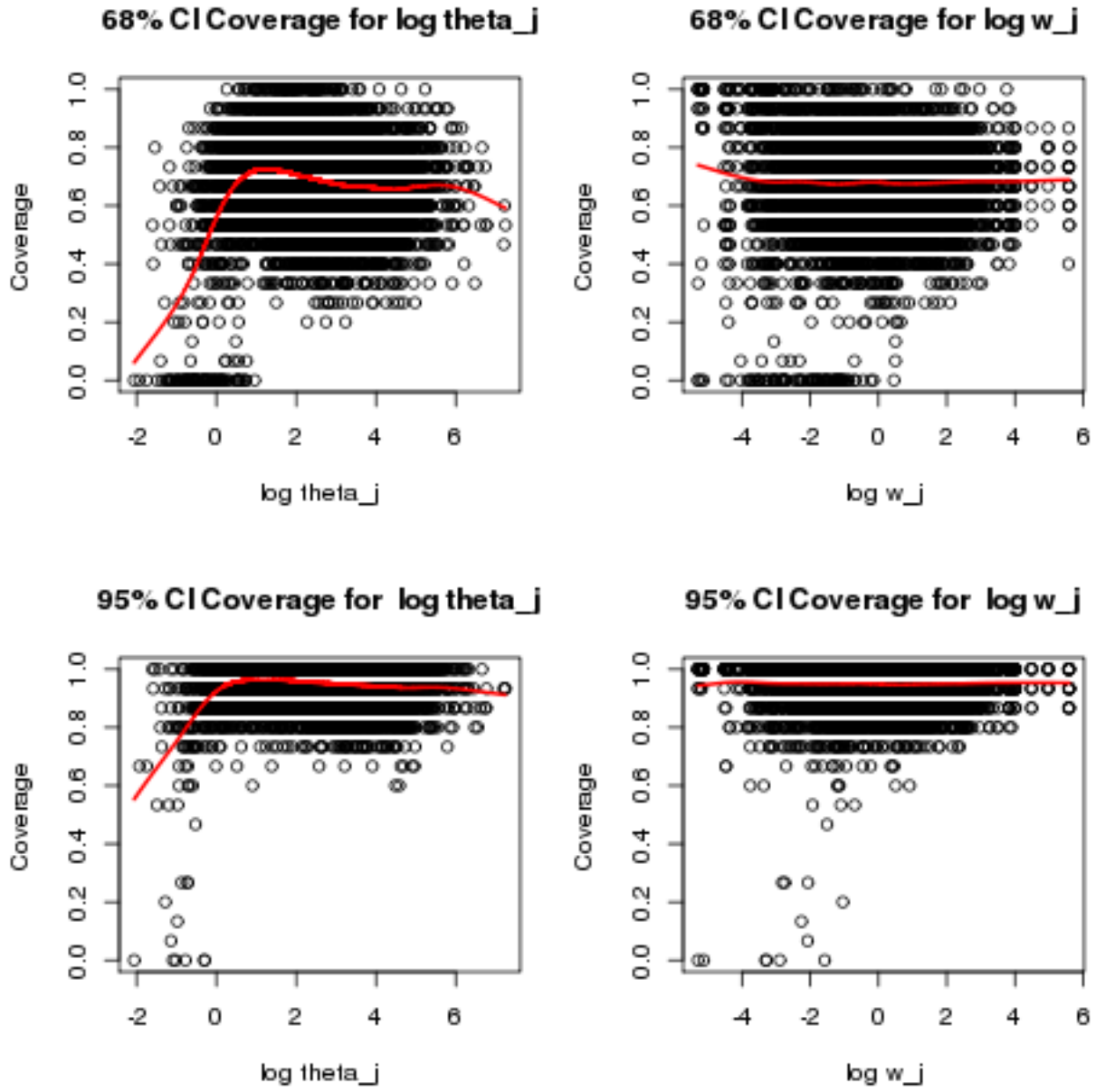


Figure 7: Task 4 coverage plots for  $\mu = 5.2$ ,  $\sigma^2 = 1.3^2$

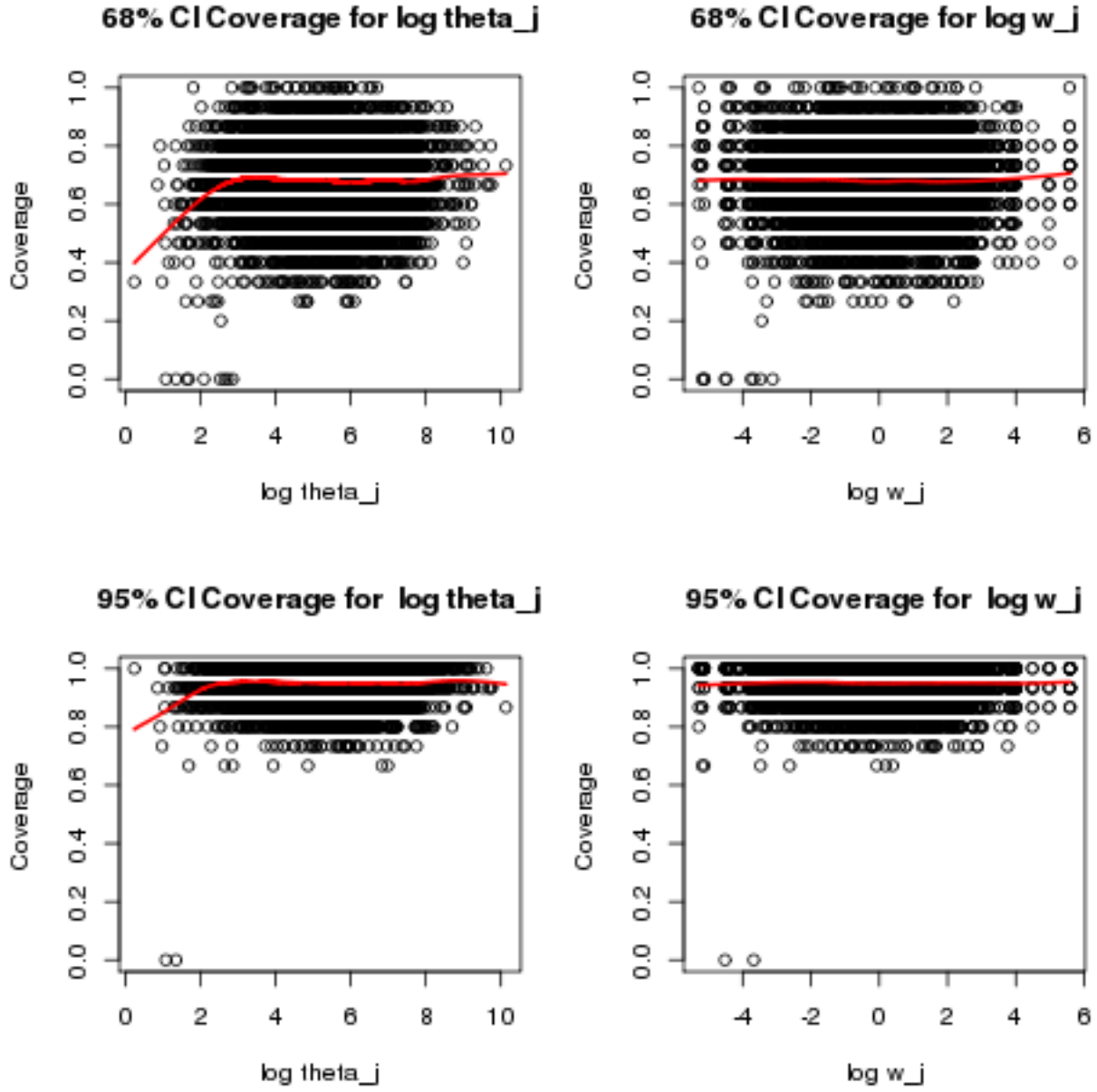


Figure 8: Task 4 coverage plots for  $\mu = 4.9$ ,  $\sigma^2 = 1.6^2$

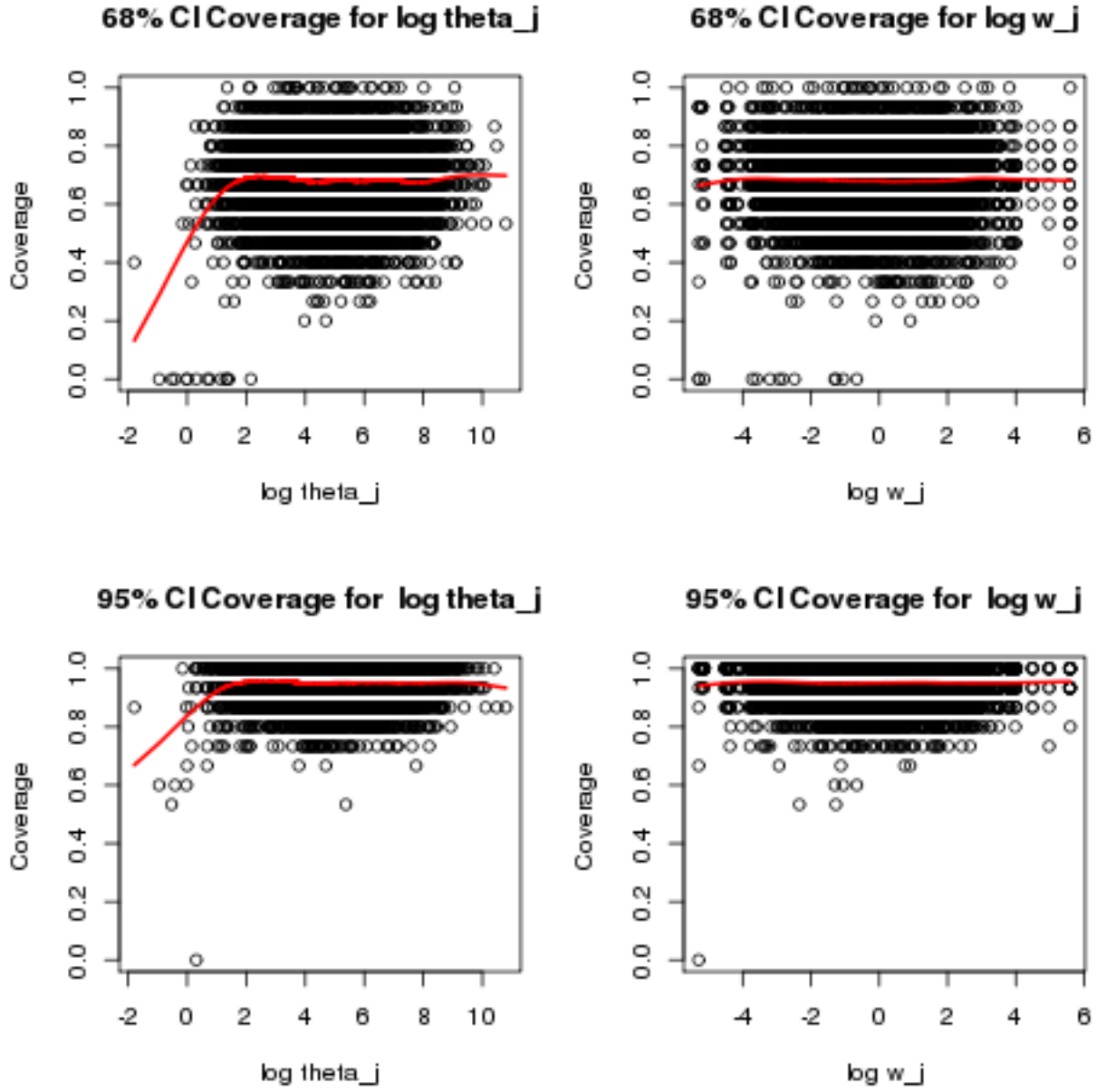


Figure 9: Task 5 coverage plots for  $\log(\theta_j)$ 's :  $x_0 = 1.6$ ,  $m = 0$ ,  $b = 1.3$

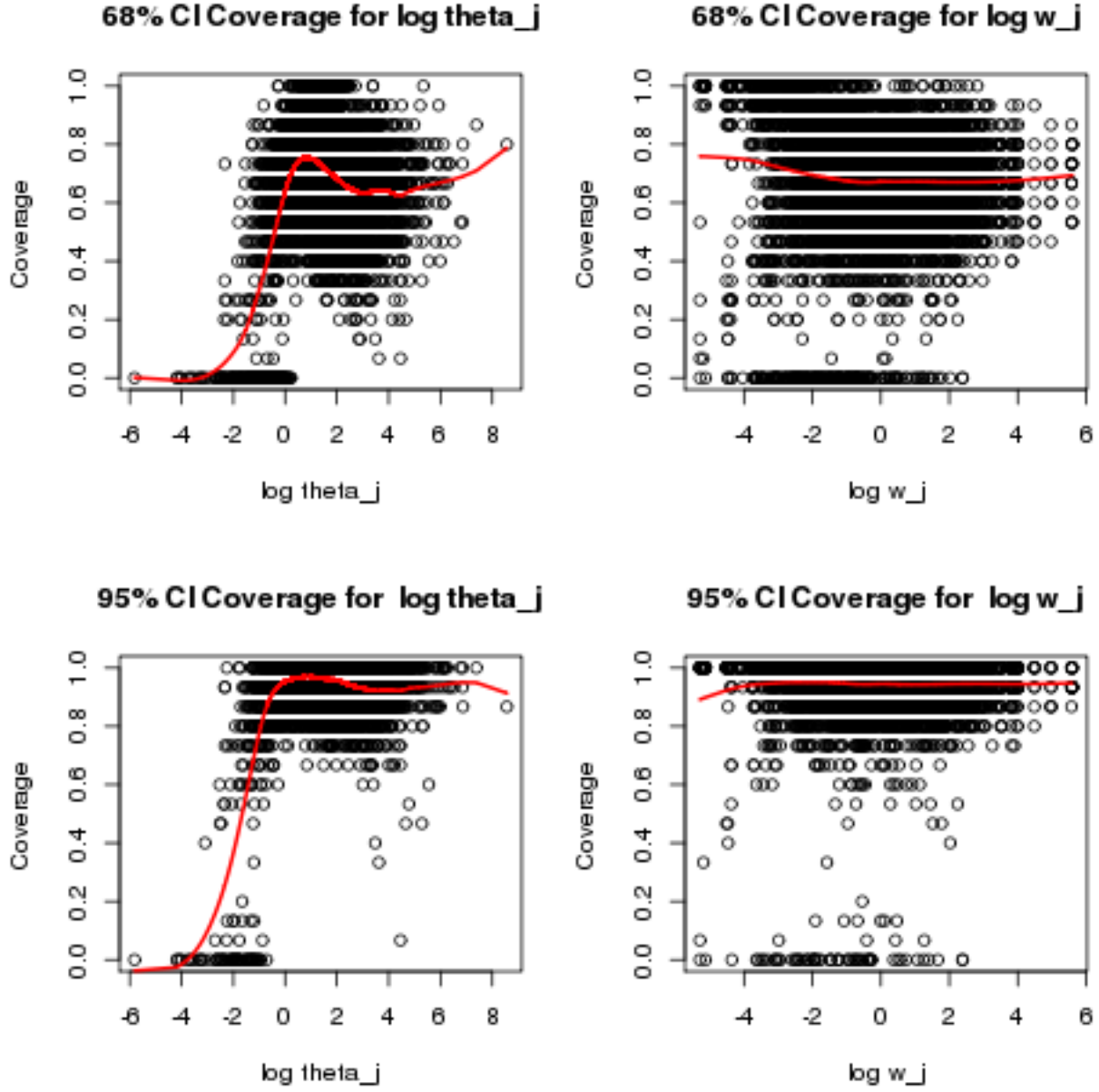


Figure 10: Task 5 coverage plots for  $\log(\theta_j)$ 's :  $x_0 = 1.6, m = -0.7, b = 1.3$

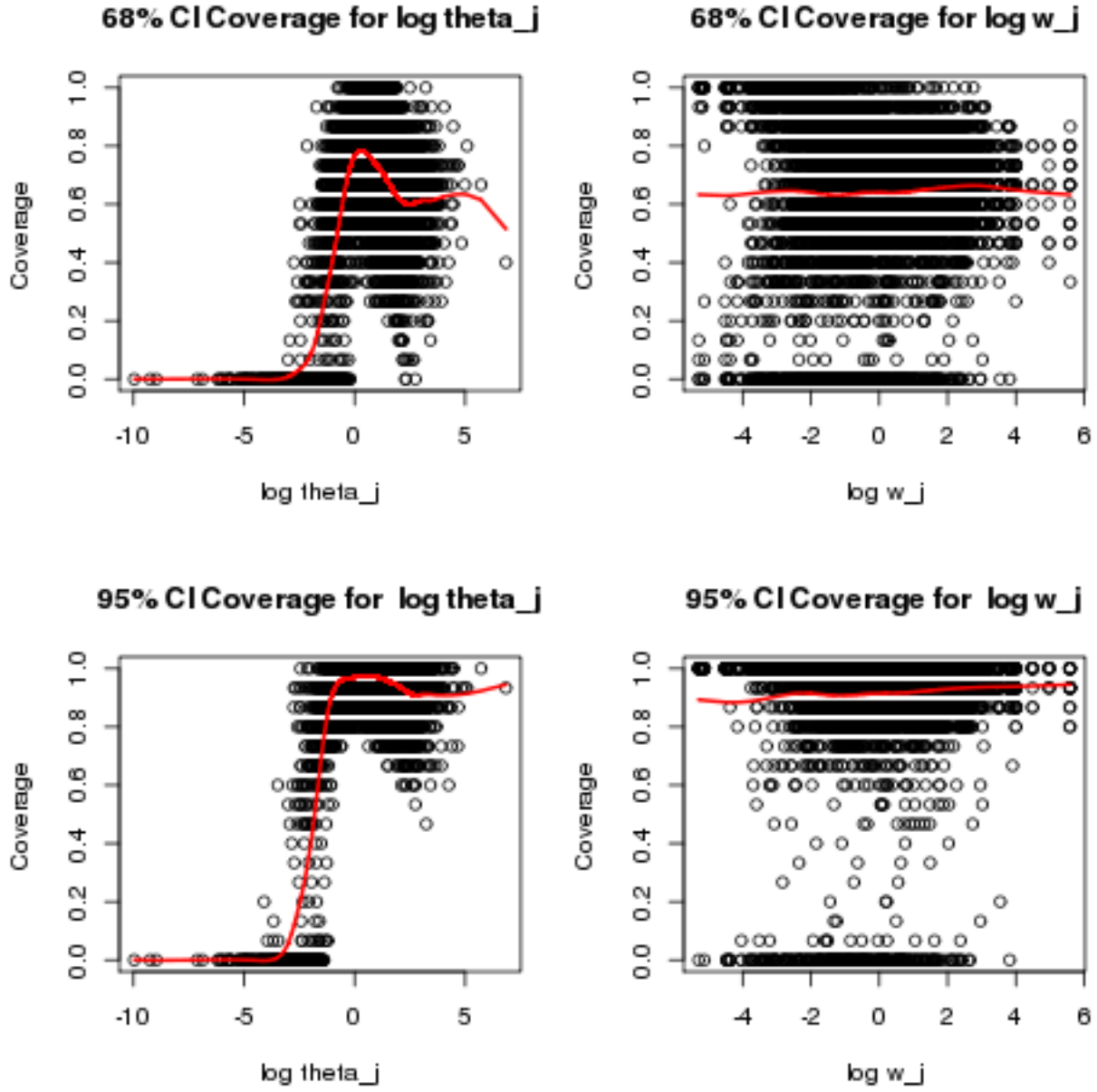


Figure 11: Task 5 coverage plots for  $\log(\theta_j)$ 's :  $x_0 = 1.6, m = -.7, b = 1.3$

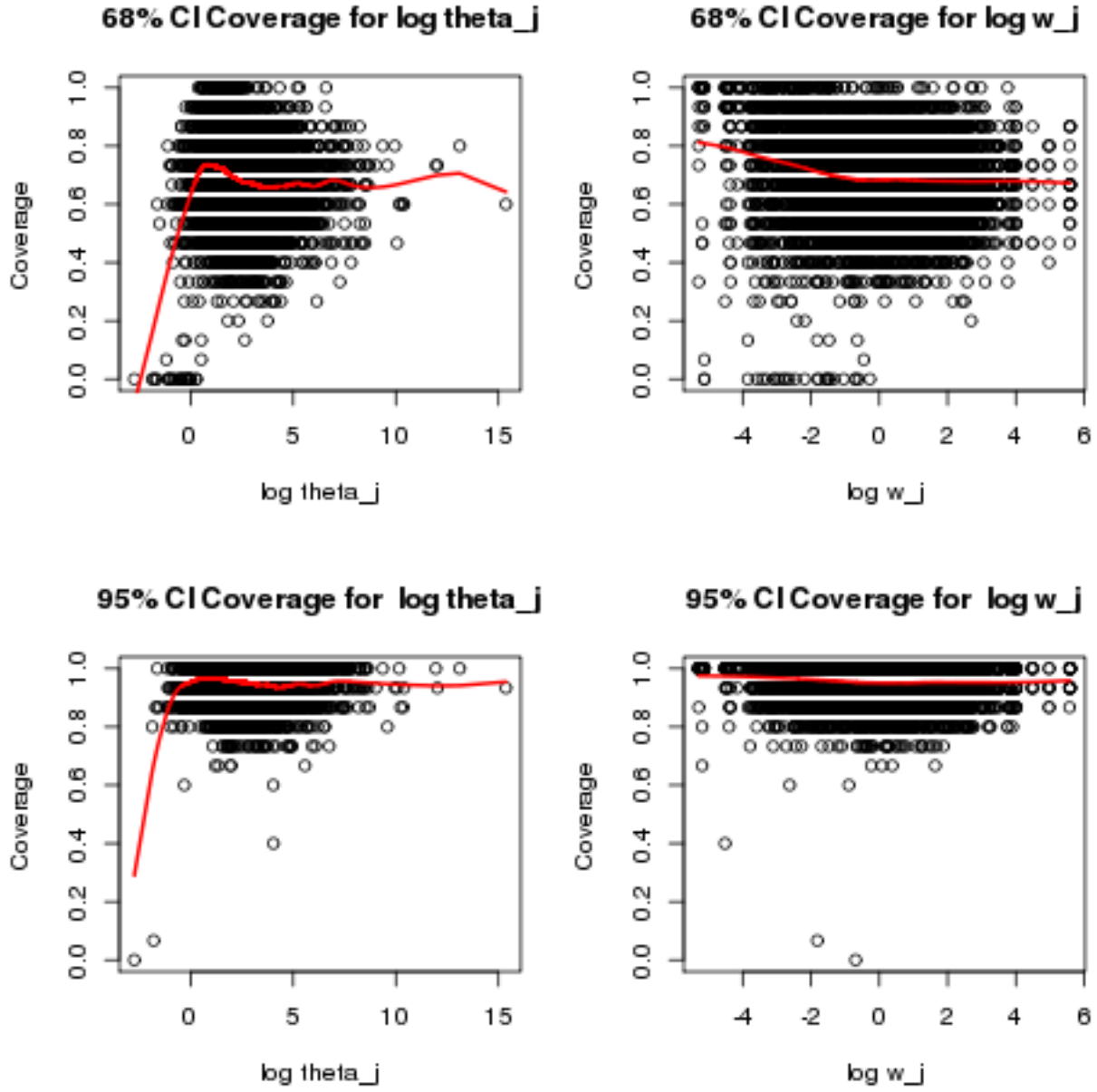


Figure 12: Task 5 coverage plots for  $x_0 = 1.6$ ,  $m = 0$ ,  $b = 2.6$

