Bayesian Analysis of Crack-Seal Veins

Daniel Pliego

March, 2025

The first model I will fit is a model with an exponential likelihood of the form:

$$\lambda_i \sim Uniform(0,1)$$
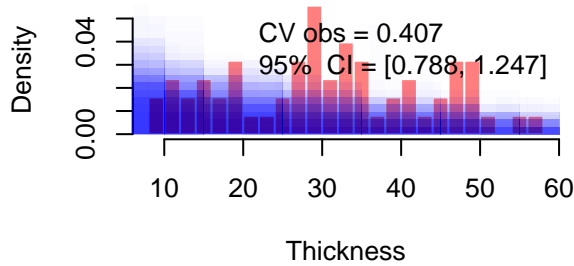$$X_n \mid \lambda_i \sim Exp(\lambda_i)$$

Where X_n is the thickness observed thickness and λ_i is the parameter for each sample i

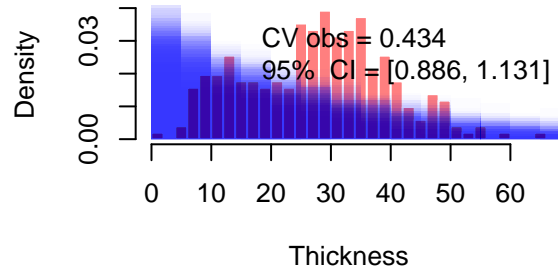This will make the stan model generate lambda's for each one of the samples

I will use CoV as a measure of the how the exponential distribution fits the data as it is the same measurement used on the paper and we should expect similar results. Along with this, the plots should give us a general idea of how the exponential distribution fits the data.

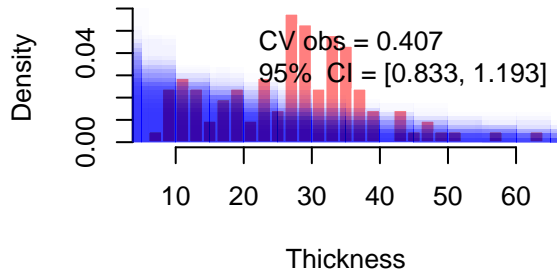I will later do posterior predictive checks to ensure the model is well specified.
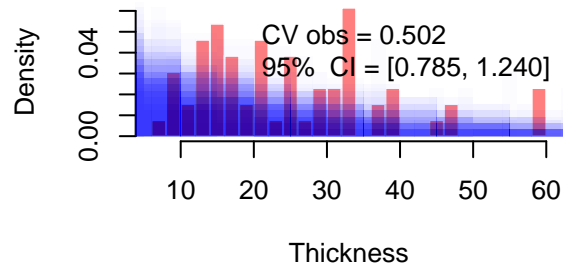
Posterior Predictive Check Sample 1

CV obs = 0.407
95% CI = [0.788, 1.247]

Posterior Predictive Check Sample 2

CV obs = 0.434
95% CI = [0.886, 1.131]

Posterior Predictive Check Sample 3

CV obs = 0.407
95% CI = [0.833, 1.193]

Posterior Predictive Check Sample 4

CV obs = 0.502
95% CI = [0.785, 1.240]

As expected in all of the samples the observed COV is far away from the 95% interval. We can also see that the posterior distribution with an exponential likelihood for the thickness does not fit the data very well. This is expected and is what was show on the paper.

The model is simple enough that I wont provide trace plots or rank plots in order to save space. I had a look at them and they looked good as expected.
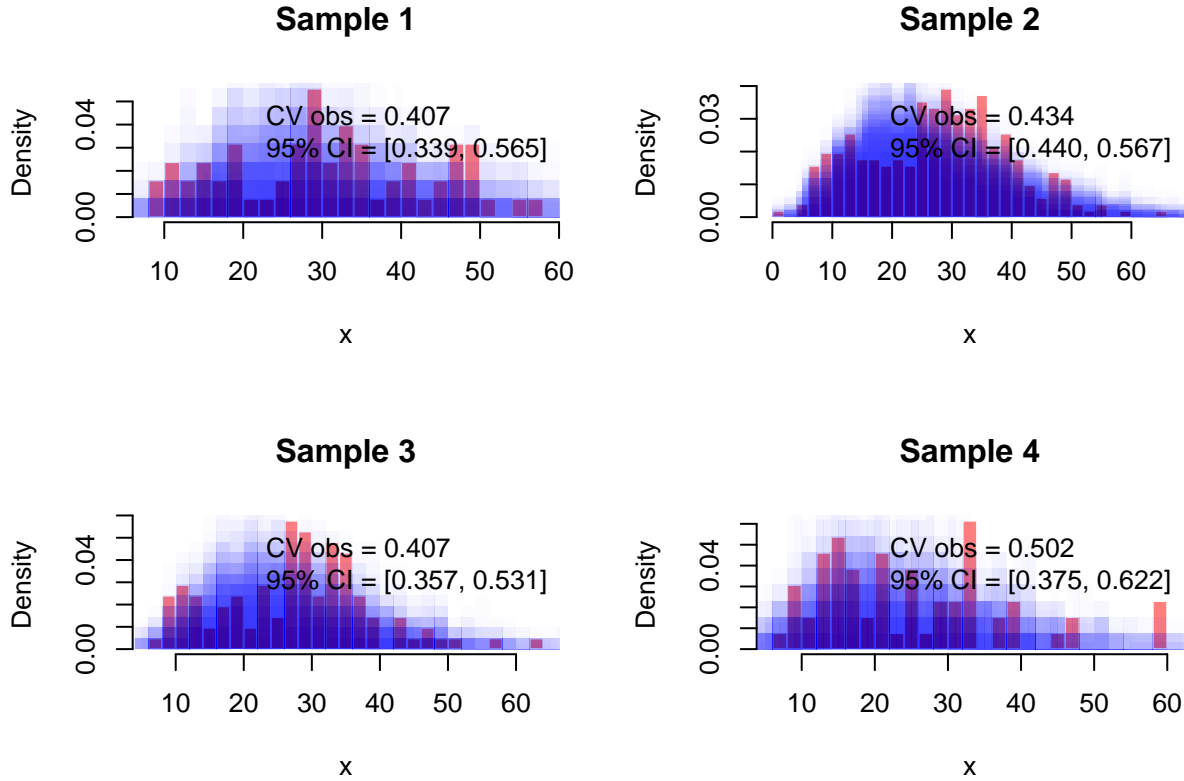
Model 2, Gamma model

Now that we have seen that the exponential model does not fit the data properly due to the assumption of equal variance and mean. I will implement a Gamma model to relax that assumption.

I will also implement a hierarchical model to help inform the prior.

$$\mu_\alpha \sim Exp(1)$$
$$\theta_\alpha \sim Exp(1)$$
$$\mu_\beta \sim Exp(1)$$
$$\theta_\beta \sim Exp(1)$$
$$\alpha_i \mid \mu_\alpha, \theta_\alpha \sim Gamma(\mu_\alpha, \theta_\alpha)$$
$$\beta_i \mid \mu_\beta, \theta_\beta \sim Gamma(\mu_\beta, \theta_\beta)$$
$$X_n \mid \alpha_i, \beta_i \sim Gamma(\alpha_i, \beta_i)$$

I will provide posterior predictive checks, however due to the hierarchical model, I will perform mixed predictive replication for hierarchical models as specified in the stan documentation. This are posterior predictive checks with the difference that the hyper parameters remain fixed

Looking at the plots, the hierarchical gamma model appears to fit the data much better, additionally the CoV measurement is now inside the 95% CI provided by the posterior distribution in all samples but sample 2.

Due to the hierarchical model, to compute posterior predictive I would need to fit n stan models, if I have time this should be ran and checked, I dont have the time to run it currently. I will provide trace plots and rank plots.
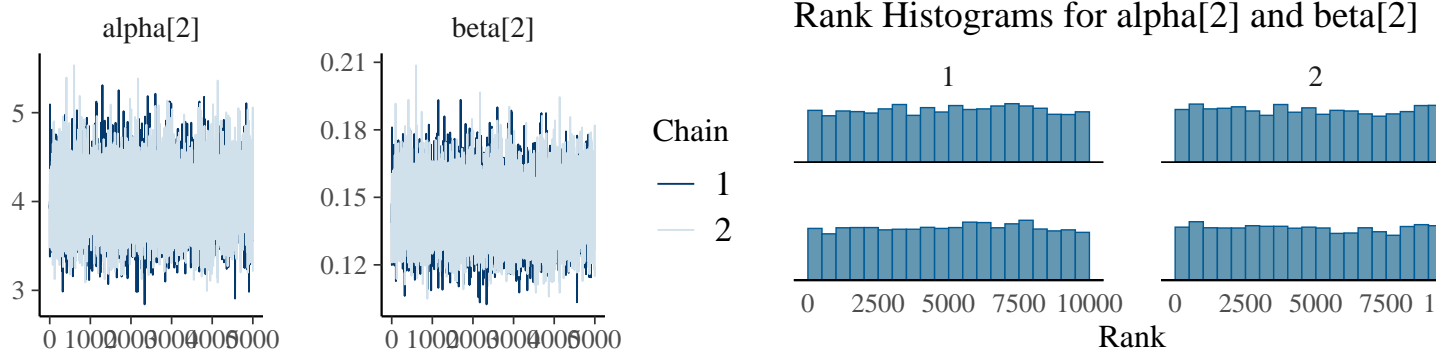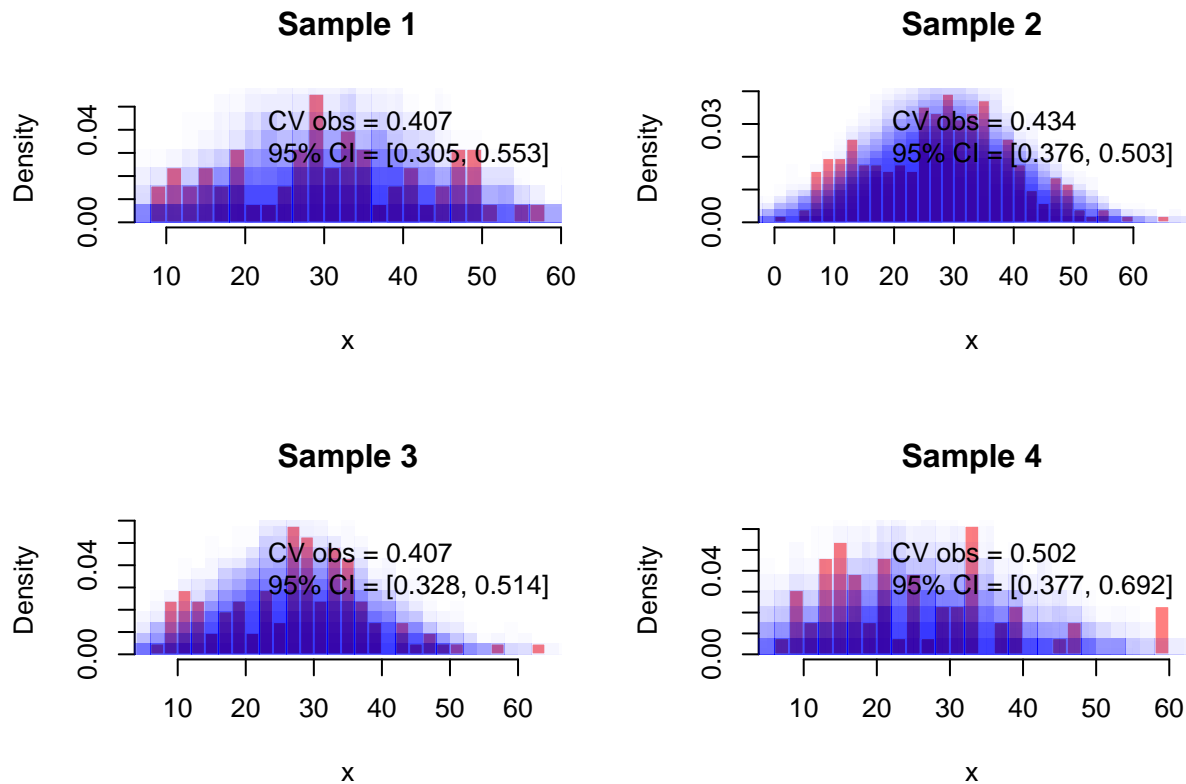

Everything looks good. I only made trace plots from the sample 2 parameters in order to have a clear view at them. I inspected all the trace plots and they all look well as expected. Both chains on the trace plot appear to behave similarly and the rank plots appear uniform as desired.

Model 3, Normal model

Finally I want to test a normal distribution model. The normal distribution does not have a support on the positive real numbers as the gamma and exponential distribution do. However the thickness values are on average a few standard deviations away from 0. Which makes it so that fitting a normal distribution should have little to no predictions which are negative. I will also use a hierarchical model to help inform the prior.

model:

$$\alpha_\mu \sim Exp(1)$$
$$\alpha_\theta \sim Exp(1)$$
$$\beta_\mu \sim Exp(1)$$
$$\beta_\theta \sim Exp(1)$$
$$\mu_i \mid \alpha_\mu, \beta_\mu \sim Gamma(\alpha_\mu, \beta_\mu)$$
$$\theta_i \mid \alpha_\theta, \beta_\theta \sim Gamma(\alpha_\theta, \beta_\theta)$$
$$X_n \mid \mu_i, \theta_i \sim Normal(\mu_i, \theta_i)$$

Sample 1

CV obs = 0.407
95% CI = [0.305, 0.553]

Sample 2

CV obs = 0.434
95% CI = [0.376, 0.503]

Sample 3

CV obs = 0.407
95% CI = [0.328, 0.514]

Sample 4

CV obs = 0.502
95% CI = [0.377, 0.692]

From looking at the plots and the CoV, I think the normal model performs better than the gamma distribution. the fited distribution seems to follow close the observed data and the CoV intervals are narrower with all of the observed values being inside them.

I will make trace plots and ranks plot for the normal model

Rank Histograms for mu[2] and theta[2]
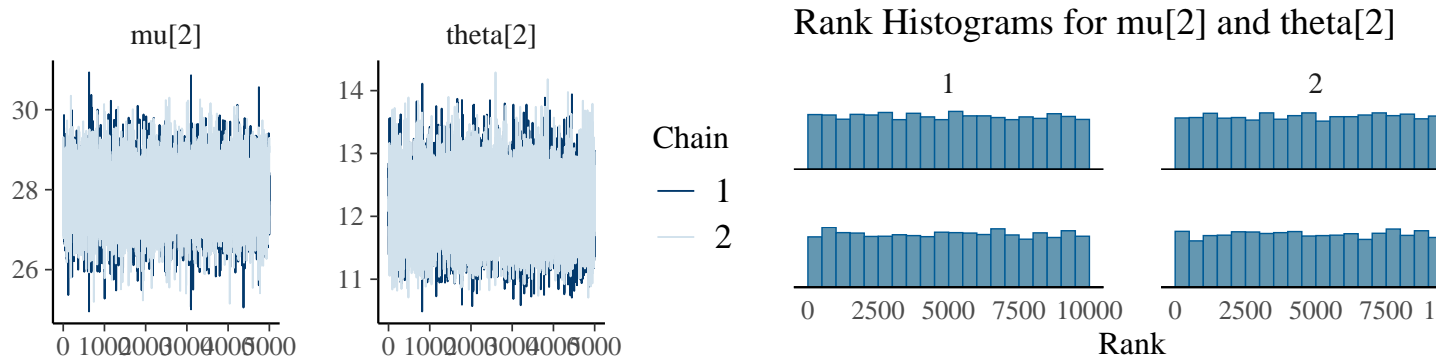
They are all good. the normal likelihood seems to perform the better out of the three distributions even though it has an incorrect support compared to the other 2 distributions. Because of this, I will use a normal distribution to fit an AR(1) model

AR(1) model using a normal dstrbution