

CS7DS3 Applied Statistical Modelling

Assignment 2

Harinath Babu (24340502)

Q1

Given that the time T can be modeled using a exponential distribution as,

$$p(t|\theta) = \theta \exp(-\theta t)$$

This can be rewritten as,

$$p(t|\theta) = 1.\theta \exp(-\theta t)$$

Which resembles the form of the standard exponential family with,

$$g(\theta) = \theta$$

Normalising constant; $h(t) = 1$

Natural parameter; $\phi(\theta) = -\theta$

Sufficient statistic; $s(t) = t$

Q2

i)

- No, the choice to model θ in a log-linear setting as an exponential function of the covariates is not surprising in this case.
- As θ is the rate parameter and must be $\theta > 0$, using the exponential function naturally ensures that this constraint is satisfied.
- Modeling θ in this way allows the expected time to failure to be expressed as an exponential function of the covariates, which gives us an relatively interpretable relationship, as the covariates have a multiplicative effect on the expected time.
- The covariates remain additive on the log-rate scale, which also simplifies interpretation with respect to the effects of each covariate on the failure rate.

ii)

Given,

$$E[T] = \frac{1}{\theta}$$

and the rate θ can be expressed in the form

$$\theta = \exp(-\beta_0 - \beta_1 x_1 - \beta_2 x_2)$$

The expectation can be rewritten as,

$$E[T] = \frac{1}{\exp(-(\beta_0 + \beta_1 x_1 + \beta_2 x_2))}$$

$$E[T] = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

Q3

i)

- The closer the value of **Rhat** is to 1, the better the MCMC performance. In this case the diagnostic value for all the coefficients is 1, which is the ideal value.
- Relative to the post-warmup sample size (4000), both the Bulk ESS and the Tail ESS are quite high which implies efficient sampling in both the central and tail regions of the posterior distribution.
- These factors indicate that the MCMC performance was satisfactory.

ii)

- β_0 can be taken as the expected log-time to failure when both the other covariates are 0 (i.e) for a *Type A component* ($x_1 = 0$) of *Grade 0* ($x_2 = 0$).
- Even though a grade of 0 is outside the scale of x_2 , since grade was treated as a continuous variable, a grade of 0 is valid for interpretation.
- This expected time to failure in this case can be calculated directly as,

$$E[T] = \exp(\beta_0) = \exp(0.96)$$

$$E[T] = 2.61$$

- So, a Type A component ($x_1 = 0$) of Grade 0 ($x_2 = 0$) has a expected time to failure of 2.61 units as calculated using β_0 , which can thus be quantified as the log-time to failure of a component of this specification.

iii)

- From the MCMC output, the parameter that corresponds to the type of component **b_typeB** (β_1) is only used when the component is type B and thus its coefficient can be taken to represent the effect of the component time on the expected time to failure.
- The expected time to failure,

$$E[T] = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$$

- The expected time is monotonic with respect to each of the parameters; a higher value of any coefficients indicates a higher time to failure and lower values indicates lower time to failure.
- The estimated value of the parameter **b_typeB** is -0.50 with a 95% confidence interval of [-0.88, -0.11], which does not include 0 and is statistically significant.
- Thus it can be stated with high confidence that the component type affects the expected time to failure, with type A components taking a longer time to fail on average.

iv)

- For assessing the time to failure of components, the type of component seems more important than the grade.
 - The coefficient for x_1 (type) is estimated at -0.50 with a 95% confidence interval of [-0.88, -0.11] which does not include 0. Thus we can say the parameter has a significant effect on the result with low uncertainty.
 - Whereas the coefficient for x_2 (grade) is estimated at -0.13 with a 95% confidence interval of [-0.27, 0.01] which has a relatively higher density around 0 suggesting higher uncertainty and does not provide enough evidence to conclude that its effect is statistically significant.
 - The larger magnitude of the coefficient of the type also implies a higher impact on the expected time to failure.
 - The higher magnitude of the estimate and the relatively lower posterior uncertainty imply that the type of the component is more important than the grade when assessing the time to failure of the components.
-