STAT40180 — Stochastic Models

Brendan Murphy

Week 2

Multiple Models

Multiple Models

- There are multiple plausible models for the Kiama blowhole data.
- For example:
    - Exponential
    - Weibull
    - Gamma
    - Log-normal
- We have already seen how to fit the first two models.
- The methods for the second two are similar.

Time-to-Event: Kiama Blowhole

- The gamma distribution also generalizes the exponential

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp\left(-\frac{x}{\beta}\right), \text{ where } \alpha, \beta > 0.$$

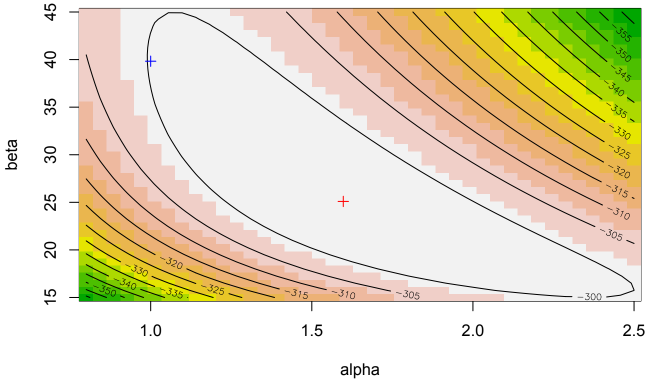
- If $(\alpha, \beta) = (1, 1/\theta)$ then the gamma model is the same as an exponential distribution.
- However, when $\alpha \neq 1$ it is has a different shape.
- Also,

$$\mathbb{E}(X) = \alpha\beta \text{ and } \mathbb{V}\text{ar}(X) = \alpha\beta^2$$

So, method of moments is straightforward for this model.

Maximum likelihood: Likelihood

- We can produce a contour plot of the likelihood function to see how it varies with the value of $\theta = (\alpha, \beta)$.



- The maximum is marked (red) and the exponential fit is also marked (blue).

Maximum Likelihood: Code

- The code for doing the maximum likelihood estimation.

```
loglik<-function(theta,x)
{
alpha<-theta[1]
beta<-theta[2]
sum(dgamma(x,shape=alpha,scale=beta,log=TRUE))
}

alpha0<-1
beta0<-mean(x)
theta0<-c(alpha0,beta0)
fit<-optim(par=theta0,loglik,method="BFGS",x=x,control=list(fnscale=-1),hessian=TRUE)
```

- We can see that

$$\hat{\theta} = (\hat{\alpha}, \hat{\beta}) = (1.60, 25.1)$$

and approximate 95% confidence intervals are:

$$\alpha : 1.60 \pm 0.52 \text{ and } \beta : 25.1 \pm 9.6.$$

Time-to-Event: Kiama Blowhole

- The log-normal distribution is another potential model

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2}\left(\frac{\log x - \mu}{\sigma}\right)^2\right], \text{ where } \sigma > 0.$$

- This is equivalent to saying that $\log X \sim \text{Normal}(\mu, \sigma^2)$
- Also,

$$\mathbb{E}(X) = \exp(\mu + \sigma^2/2) \text{ and } \mathbb{V}\text{ar}(X) = \exp(2\mu + \sigma^2)\exp(\sigma^2 - 1)$$

In principle, method of moments can be done.

Maximum Likelihood: Code

- The code for doing the maximum likelihood estimation.

```
loglik<-function(theta,x)
{
mu<-theta[1]
sigma<-theta[2]
sum(dlnorm(x,meanlog=mu,sdlog=sigma,log=TRUE))
}

mu0<-mean(log(x))
sigma0<-sd(log(x))
theta0<-c(mu0,sigma0)
fit<-optim(par=theta0,loglik,method="BFGS",x=x,control=list(fnscale=-1),hessian=TRUE)
```

- We can see that

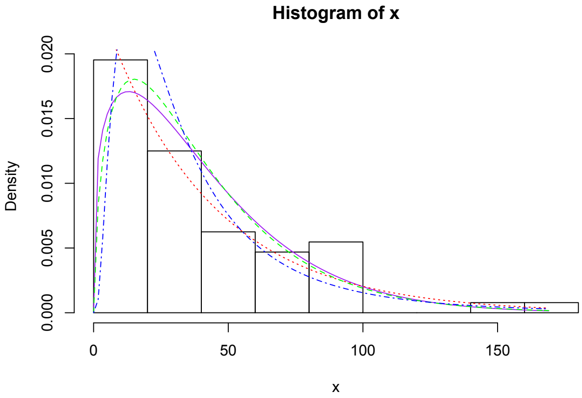
$$\hat{\theta} = (\hat{\mu}, \hat{\sigma}) = (3.35, 0.84)$$

and approximate 95% confidence intervals are:

$$\alpha : 3.35 \pm 0.11 \text{ and } \beta : 0.84 \pm 0.07.$$

Model Fit

- We can informally compare all of the model fits to the data histogram.



- Exponential=red, Weibull=purple, gamma=green,log-Normal=blue

Model Fit

- We can compare the models using the log-likelihood and the number of parameters.

Model	Log-likelihood	Parameters
Exponential	-299.8	1
Weibull	-296.9	2
Gamma	-295.9	2
Log-normal	-293.9	2

- We need to balance the quality of fit (log-likelihood) and model complexity (parameters).

Model Fit: Information Criteria

- Information criteria balance quality of fit ($\hat{\ell}$) and the number of parameters (p).

$$AIC = 2\hat{\ell} - 2p \text{ (Akaike Information Criterion)}$$

$$BIC = 2\hat{\ell} - log(n)p \text{ (Bayesian Information Criterion)}$$

Model	Log-likelihood	Parameters	AIC	BIC
Exponential	-299.8	1	-601.6	-603.8
Weibull	-296.9	2	-597.8	-602.1
Gamma	-295.9	2	-595.8	-600.1
Log-normal	-293.9	2	-591.8	-596.1

- The log-normal model has the highest AIC and BIC values[1].

[1] Some people define *AIC* and *BIC* in an equivalent but different manner