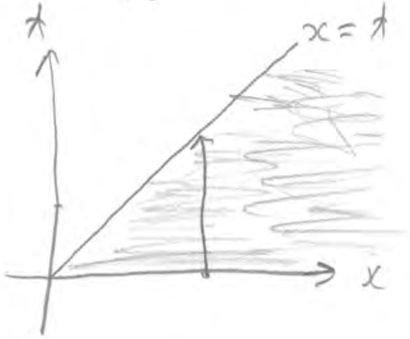
# Sample Questions: Survival and Hazard Functions

STA312 Spring 2019. Copyright information is at the end of the last page.

For all these questions, $T$ is a continuous random variable with $P(T > 0) = 1$, density $f_T(t)$ and cumulatve distribution function $F(t) = P(T \leq t)$.

1. The survival function is $S(t) = P(T > t)$. Prove $E(T) = \displaystyle\int_0^\infty S(t)\,dt.$

$$\int_0^\infty \left( \int_t^\infty f_T(x)\,dx \right) dt$$



$$\int_0^\infty \int_0^x f_T(x) \quad dt \; dx = \int_0^\infty f_T(x) \left( \int_0^x dt \right) dx$$

$$= \int_0^\infty x \, f_T(x)\, dx = E(T) \checkmark$$

2. The hazard function is defined by $h(t) = \lim\limits_{\Delta \to 0} \dfrac{P(t < T < t + \Delta | T > t)}{\Delta}$.

Prove $h(t) = \frac{f(t)}{S(t)}$.

where $\Delta > 0$

$$\lim_{\Delta \to 0} \frac{P(t < T < t + \Delta \cap T > t)}{P(T > t)\ \Delta}$$

$$= \lim_{\Delta \to 0} \frac{P(t < T < t + \Delta)}{S(t)\ \Delta} = \frac{1}{S(t)} \lim_{\Delta \to 0} \frac{F(t+\Delta) - F(t)}{\Delta}$$

$$= \frac{f(t)}{S(t)}$$ ▨

3. Prove $S(t) = e^{-\int_0^t h(x)\,dx}$.

$$\boxed{\int_0^t h(x)\,dx} = \int_0^t \frac{f(x)}{S(x)}\,dx = -\int_0^t \frac{-f(x)}{1-F(x)}\,dx$$

$u = 1 - F(x)$

$du = -f(x)\,dx$

| $x$ | $u = 1-F(x)$ |
|---|---|
| $t$ | $1-F(t) = S(t)$ |
| $0$ | $1-F(0) = 1$ |

$$= -\int_1^{S(t)} \frac{1}{u}\,du = -\log u \Big|_1^{S(t)}$$

$$= (-1)\big(\log S(t) - \log(1)\big)$$

$$= \boxed{-\log(S(t))} \Rightarrow \log S(t) = -\int_0^t h(x)\,dx$$

$$\Longrightarrow \boxed{S(t) = e^{-\int_0^t h(x)\,dx}}$$

4. Let $T \sim \exp(\lambda)$. Find the hazard function $h(t)$ for $t > 0$.

$F_T(t) = 1 - e^{-\lambda t}$ so $S(t) = e^{-\lambda t}$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}}$$

$$= \boxed{\lambda}$$

5. Let $T$ have the Pareto density $f(t|\theta) = \begin{cases} \frac{\theta}{t^{\theta+1}} & \text{for } t \geq 1 \\ 0 & \text{for } t < 1 \end{cases}$

(a) Find the hazard function $h(t)$ for $t > 1$.

$$S(t) = \int_t^\infty \theta x^{-\theta-1} dx$$

$$= \cancel{\theta} \frac{x^{-\theta}}{-\cancel{\theta}} \Big|_t^\infty = (-1)\left(\underbrace{\lim_{x \to \infty} \frac{1}{x^\theta}}_{=0} - \frac{1}{t^\theta}\right)$$

$$= \frac{1}{t^\theta}$$

$$h(t) = \frac{f(t)}{S(t)} = \frac{\frac{\theta}{t^{\theta+1}}}{\frac{1}{t^\theta}} = \frac{\theta}{t} \quad \downarrow$$

$$h(t) = \frac{\theta}{t}$$

(b) Earlier, we found the MLE $\widehat{\theta}_n = \frac{n}{\sum_{i=1}^{n} \log t_i}$, and $\widehat{v}_n = \frac{\widehat{\theta}_n^2}{n}$.

i. Give $\widehat{h(t)}$, the maximum likelihood estimate of the hazard function evaluated at a particular time $t > 1$. Your answer is a formula involving $t$ and $\widehat{\theta}_n$.

$$\widehat{h(t)} = \frac{\hat{\theta}_n}{t}$$

ii. We want a confidence interval for $h(t)$, the hazard function evaluated at a particular time $t > 1$. Give formulas for the lower and upper 95% confidence limits. Show your work.

95% CI for $\theta$ ~ $\hat{\theta}_n \pm 1.96\sqrt{\hat{v}_n}$, so

$$0.95 \approx P\left(\hat{\theta}_n - 1.96\sqrt{\hat{\theta}_n^2/n} < \theta < \hat{\theta}_n + 1.96\sqrt{\hat{\theta}_n^2/n}\right)$$

$$= P\left(\hat{\theta}_n - 1.96\frac{\hat{\theta}_n}{\sqrt{n}} < \theta < \hat{\theta}_n + 1.96\frac{\bar{\theta}}{\sqrt{n}}\right)$$

$$= P\left(\underbrace{\frac{\hat{\theta}_n}{t} - 1.96\frac{\hat{\theta}_n}{t\sqrt{n}}}_{\text{lower 95\% Confidence limit}} < \underbrace{\frac{\theta}{t}}_{h(t)} < \underbrace{\frac{\hat{\theta}_n}{t} + 1.96\frac{\bar{\theta}}{t\sqrt{n}}}_{\text{upper 95\% confidence limit}}\right)$$

$$h(t) = \frac{f(t)}{S(t)}$$

6. Let $T$ have a gamma distribution with parameters $\alpha > 0$ and $\lambda > 0$.

   (a) What is the hazard function?

$$h(t) = \frac{\cancel{\frac{\lambda^\alpha}{\Gamma(\alpha)}} e^{-\lambda t} t^{\alpha-1}}{\int_t^\infty \cancel{\frac{\lambda^\alpha}{\Gamma(\alpha)}} e^{-\lambda x} x^{\alpha-1} \, dx}$$

   (b) Using R, plot the hazard function for several values of $\alpha$ and $\lambda$. How do the parameter values influence the shape of the hazard function?

---

This assignment was prepared by Jerry Brunner, Department of Mathematical and Computational Sciences, University of Toronto. The LaTeX source code is available from the course website:

`http://www.utstat.toronto.edu/~brunner/oldclass/312s19`

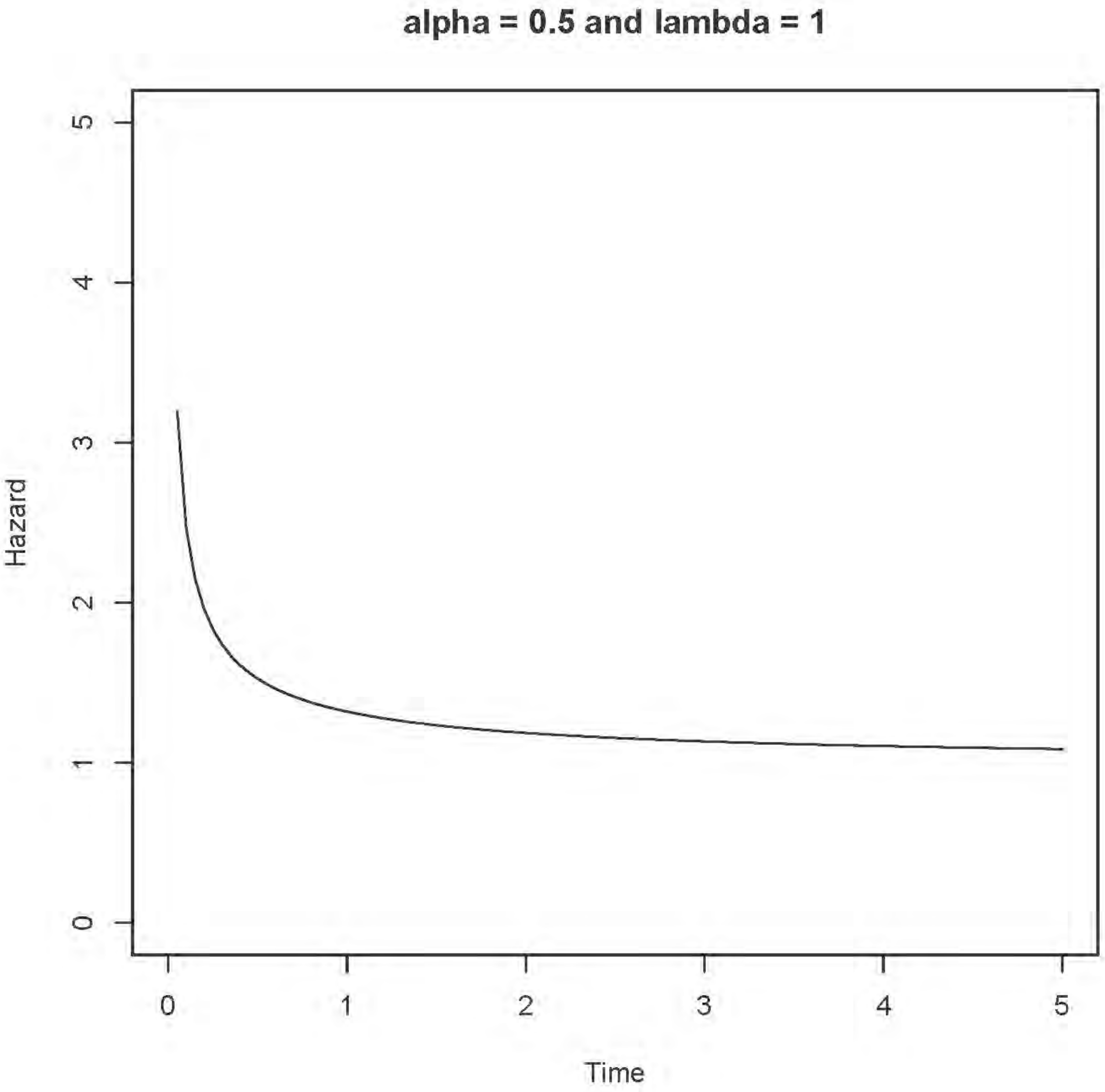
# Exploring the hazard function of the gamma distribution[*]

```
> # Hazard function of Gamma(alpha,lambda) distribution: : h(t) = f(t)/S(t)
>
> # You might think the following would work, but watch.
> # Remember that a gamma distribution with alpha=1 is exponential.
> # The hazard function of an exponential distribution is a constant lambda.
>
> Time = 1:10
>
> alpha=1; lambda=1
> dgamma(Time,shape=alpha,rate=lambda) / (1-pgamma(Time,shape=alpha,rate=lambda))
 [1] 1 1 1 1 1 1 1 1 1 1
>
> alpha=1; lambda=2
> dgamma(Time,shape=alpha,rate=lambda) / (1-pgamma(Time,shape=alpha,rate=lambda))
 [1] 2 2 2 2 2 2 2 2 2 2
>
> alpha=1; lambda=3
> dgamma(Time,shape=alpha,rate=lambda) / (1-pgamma(Time,shape=alpha,rate=lambda))
 [1] 3.000000 3.000000 3.000000 3.000000 3.000000 3.000000 3.000000 2.999999
 [9] 3.000051 2.999501
>
> alpha=1; lambda=4
> dgamma(Time,shape=alpha,rate=lambda) / (1-pgamma(Time,shape=alpha,rate=lambda))
 [1] 4.000000 4.000000 4.000000 4.000000 4.000000 3.999999 3.999960 4.002409
 [9] 4.178481      Inf
>
> alpha=1; lambda=10
> dgamma(Time,shape=alpha,rate=lambda) / (1-pgamma(Time,shape=alpha,rate=lambda))
 [1] 10.000000 10.000000  9.998336       Inf       Inf       Inf       Inf
 [8]       Inf       Inf       Inf
>
> # Use logs to get around division by a number close to zero
> alpha=1; lambda=10
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> exp(logh)
 [1] 10 10 10 10 10 10 10 10 10 10
```

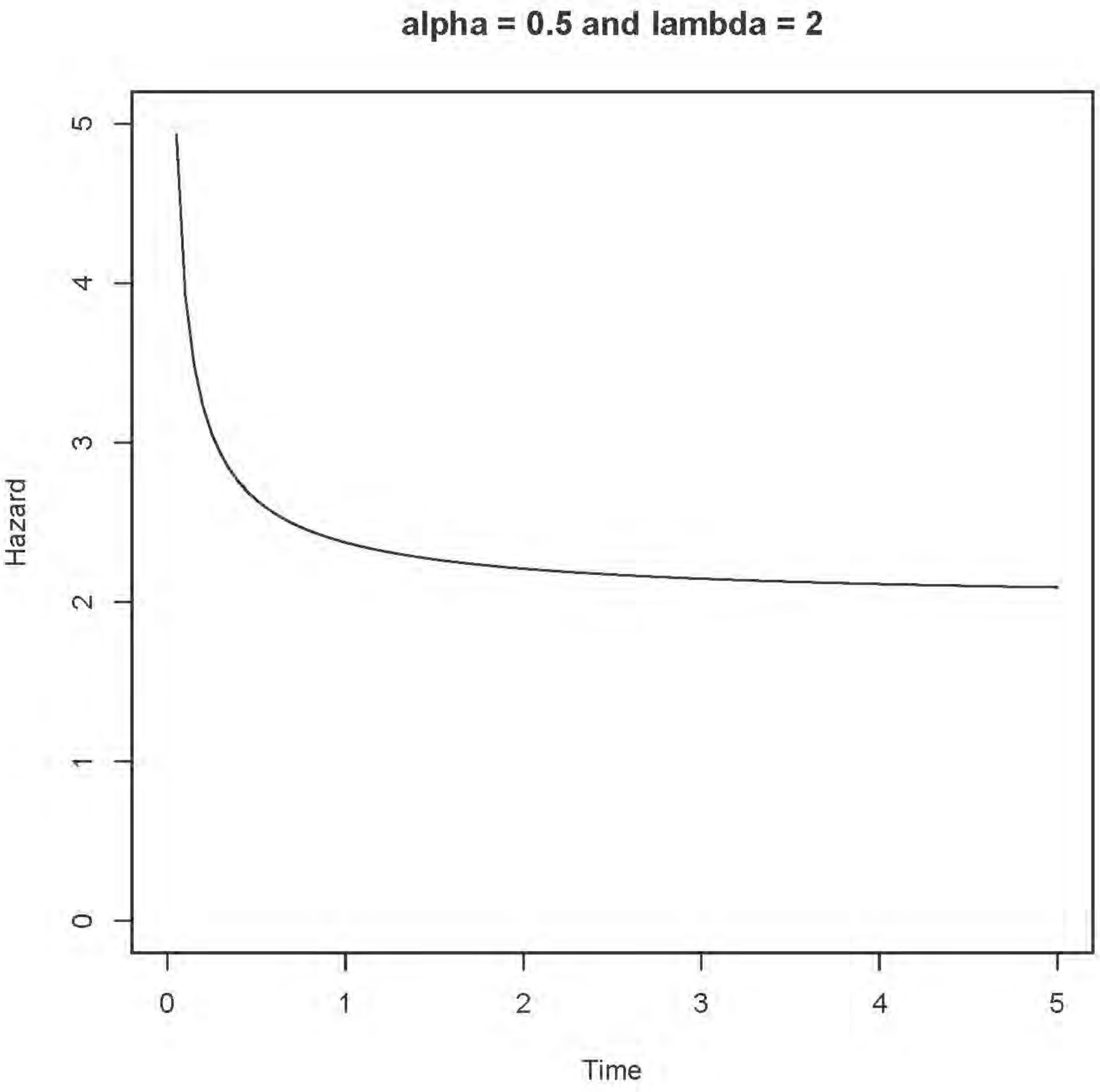
---

[*] This document is free and open source. See last page for copyright information.

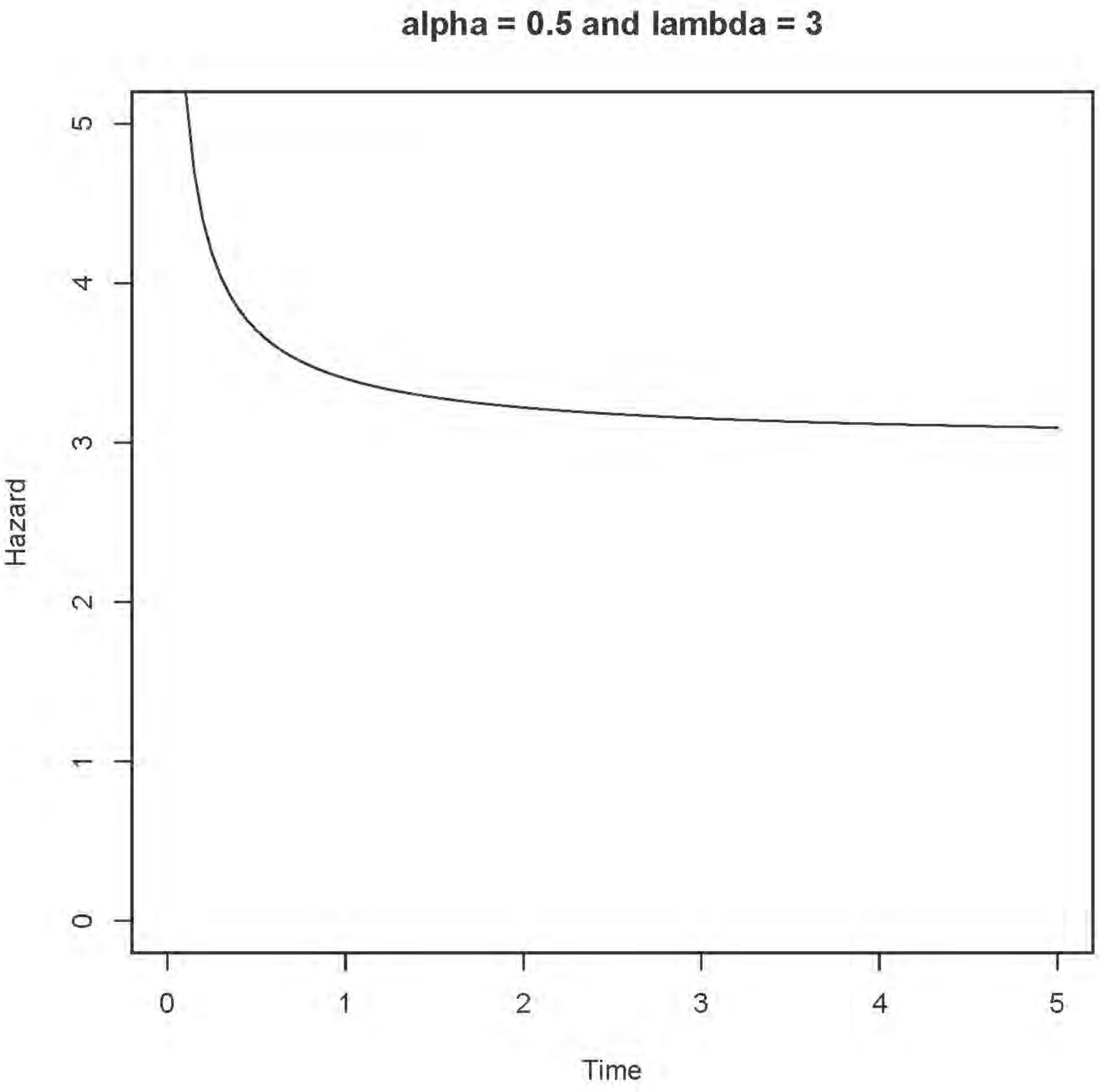
```
> # Now plot
> alpha=1/2;lambda=1
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
> # Notice explicit limits on y to keep all plots on the same scale.
```



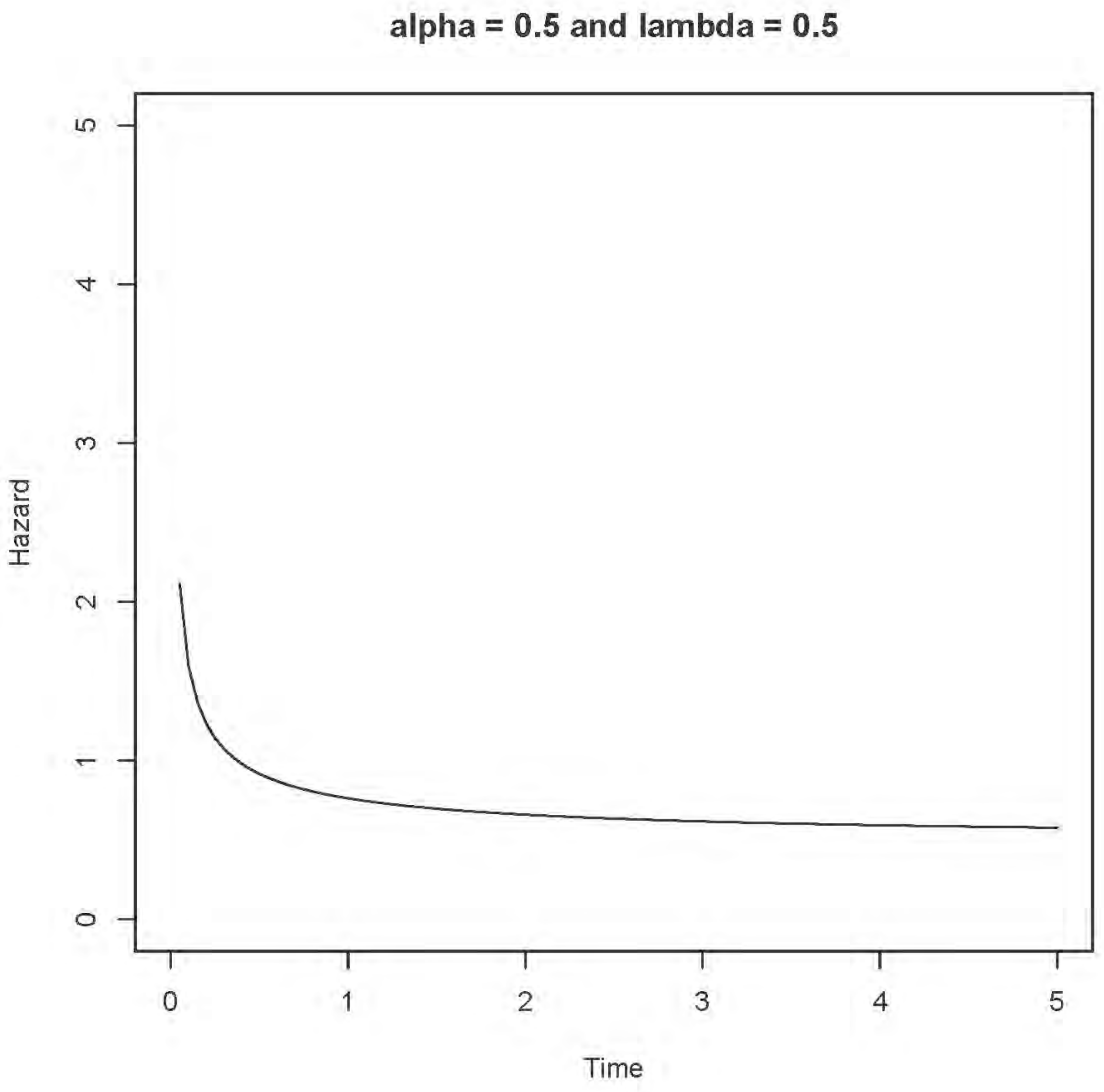
```
> alpha=1/2;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



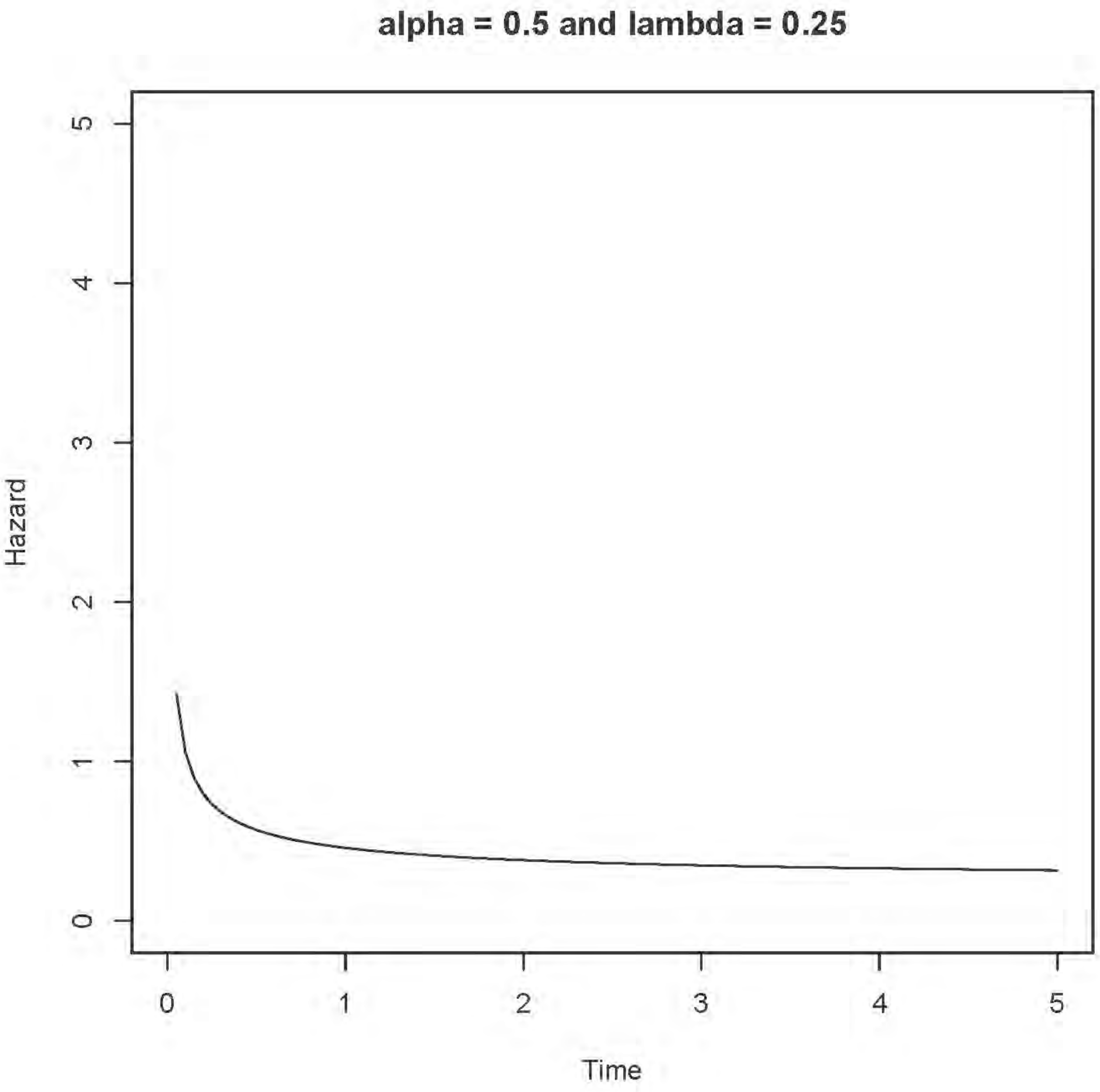
```
> alpha=1/2;lambda=3
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



```
> # Try lambda small
> alpha=1/2;lambda=1/2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```

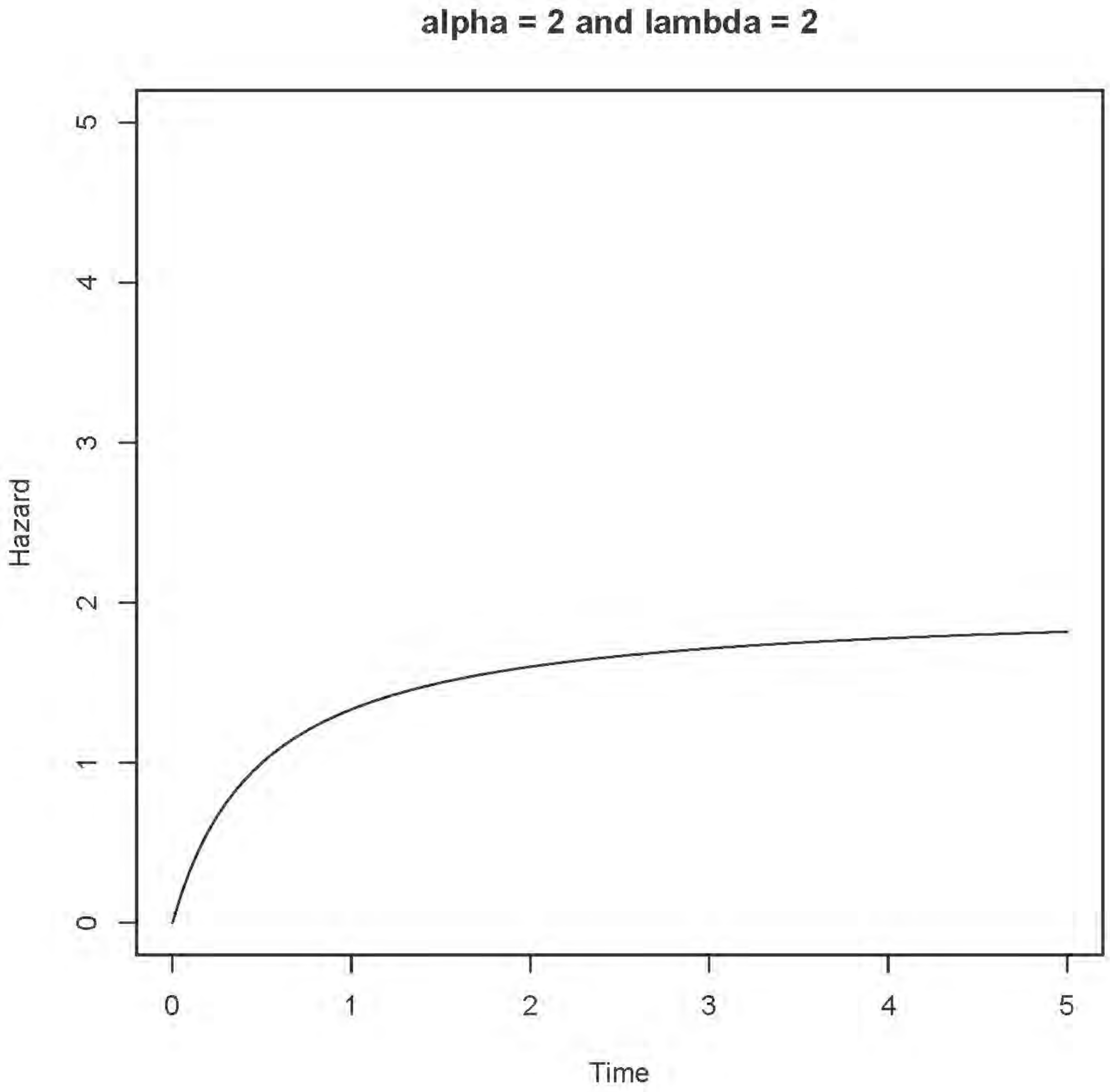


```
> alpha=1/2;lambda=1/4
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```

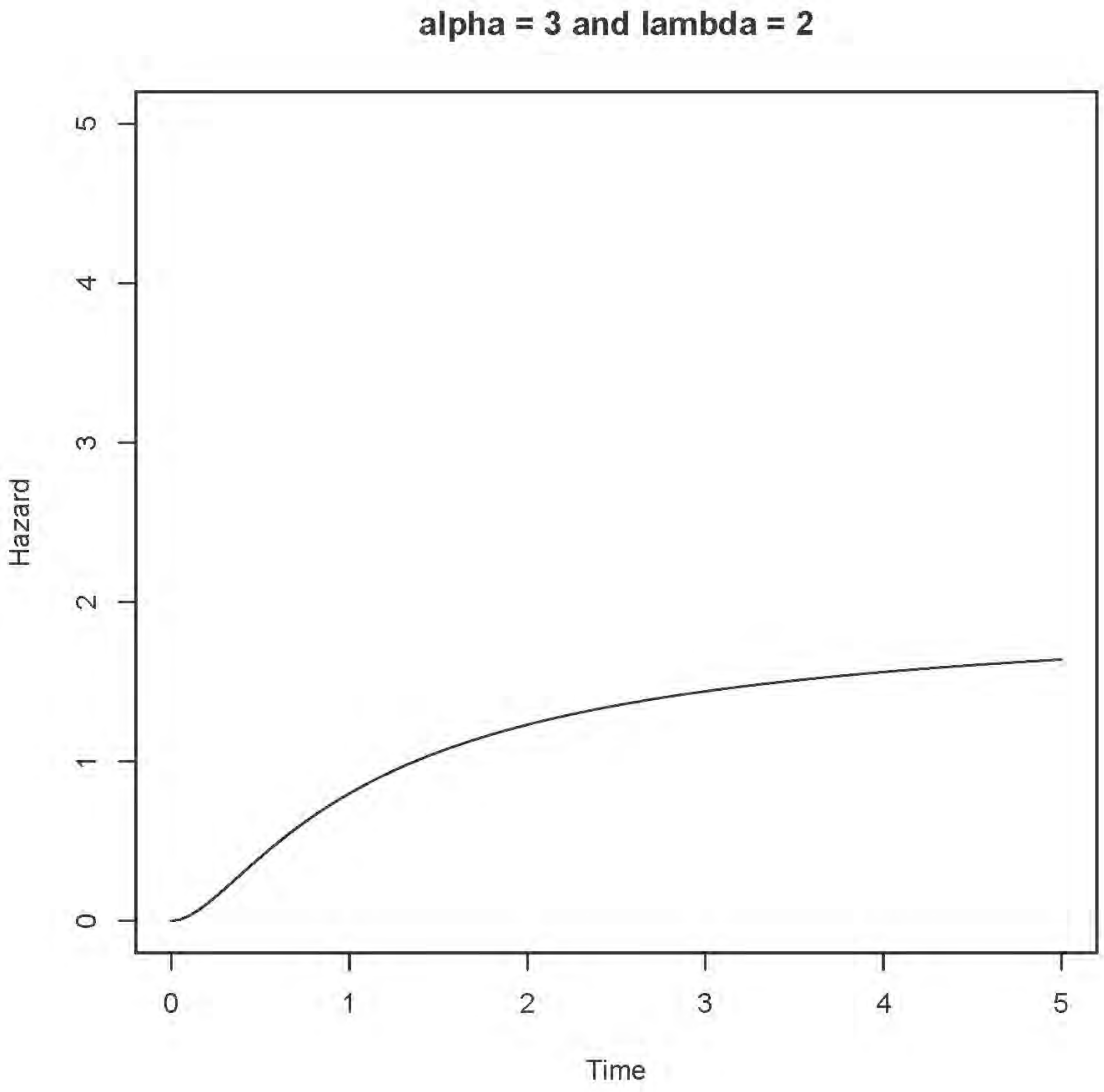


```
>
> # Lambda appears to control the overall level of the function, with larger
> # numbers higher. Investigate alpha, initially holding lambda fixed at 2.
```

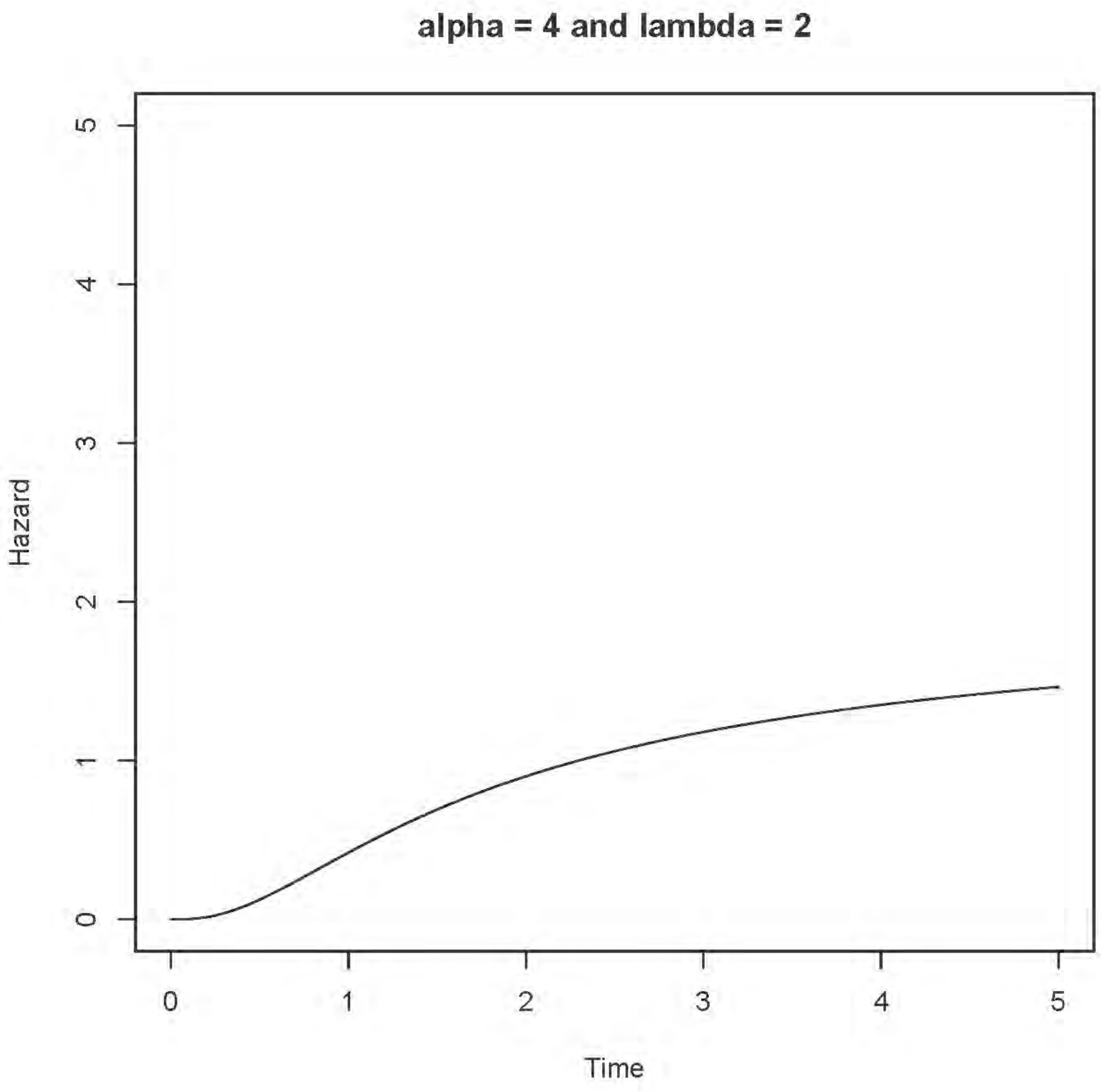
```
> # Investigate alpha, initially holding lambda fixed at 2
>
> alpha=2;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



```
> alpha=3;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```

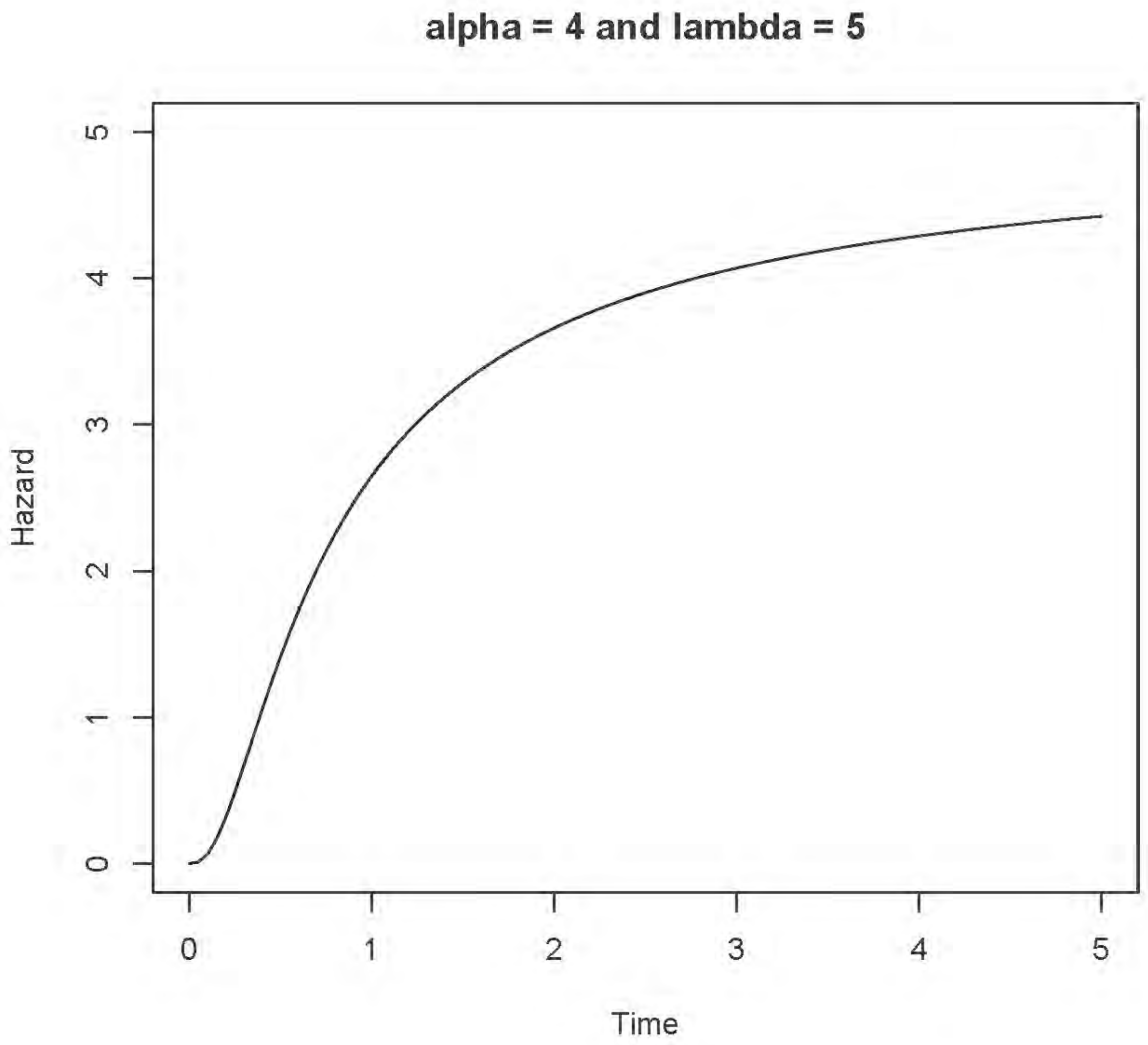


```
> alpha=4;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
>
```

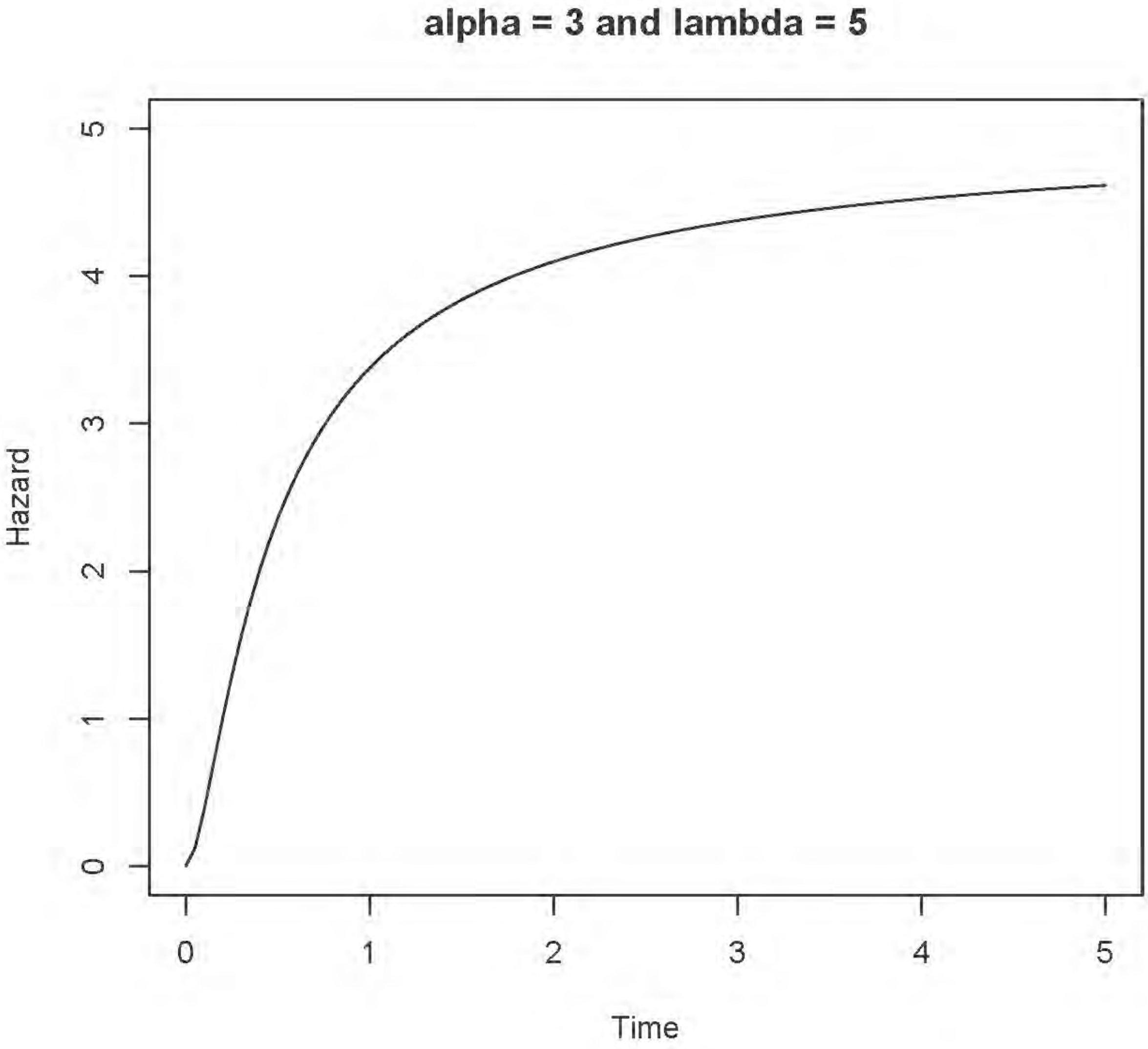


```
> # Increase lambda to 5
```

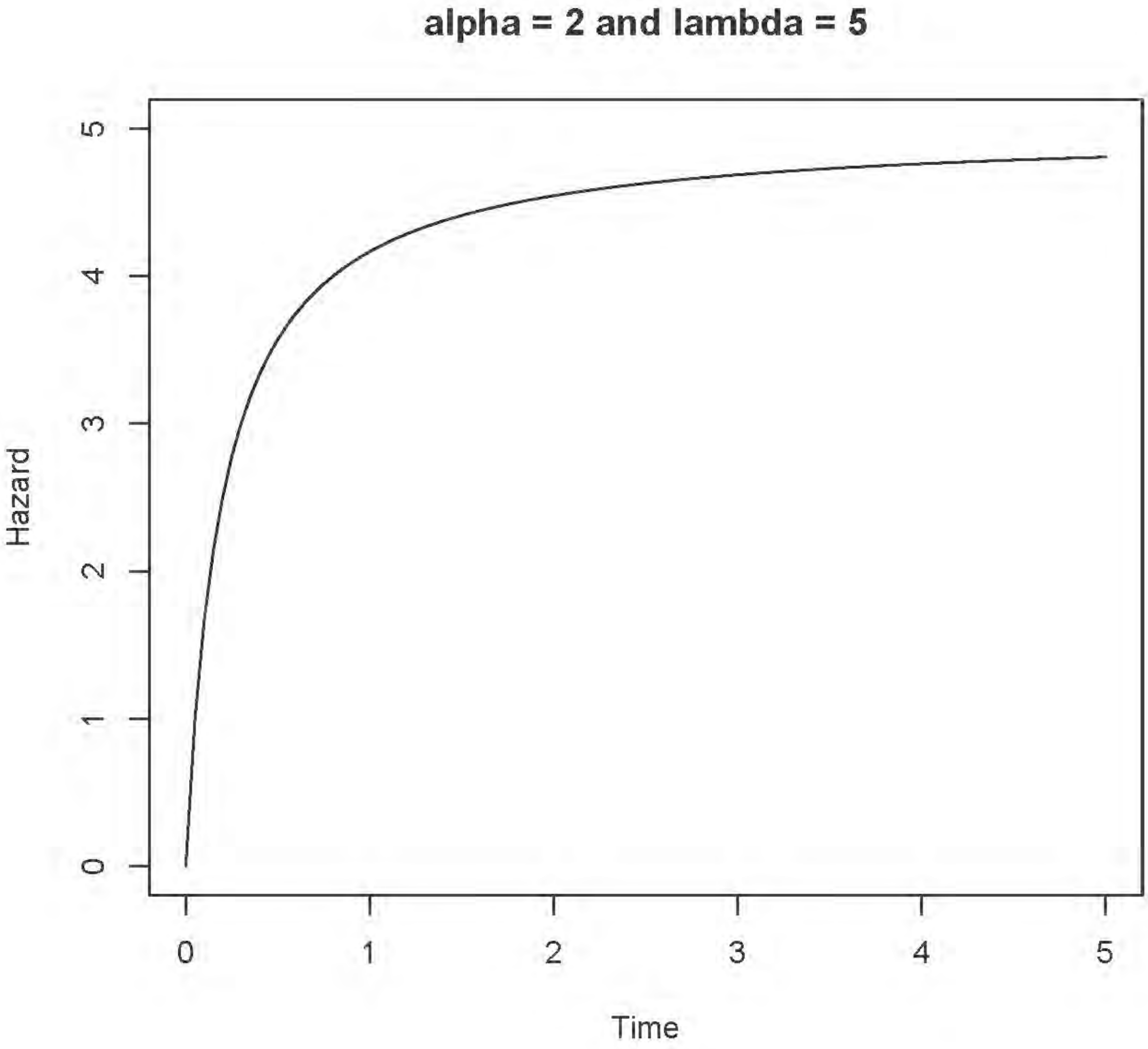
```
> # Increase lambda to 5
>
> alpha=4;lambda=5
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



```
> alpha=3;lambda=5
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```

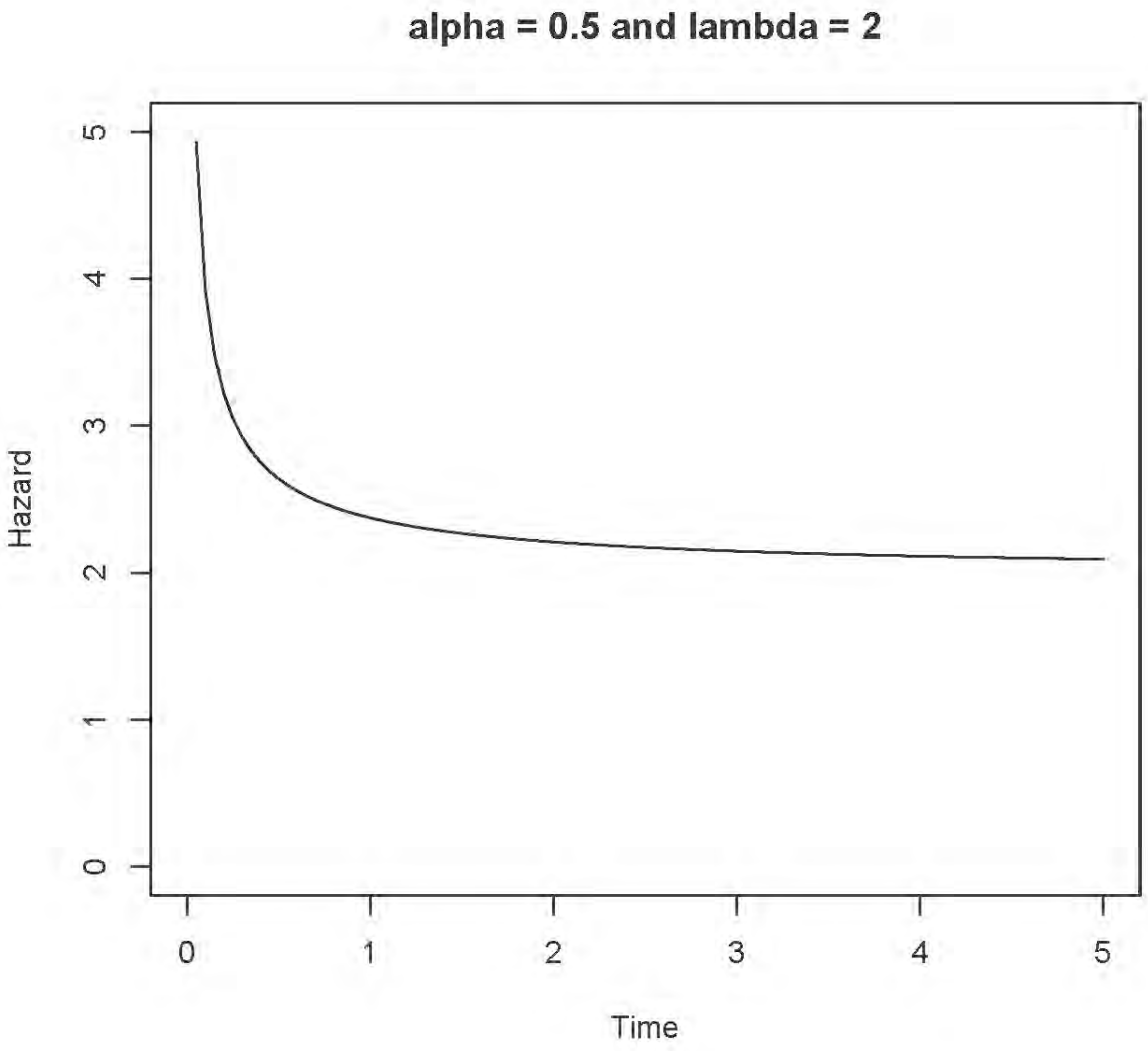


```
> alpha=2;lambda=5
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```

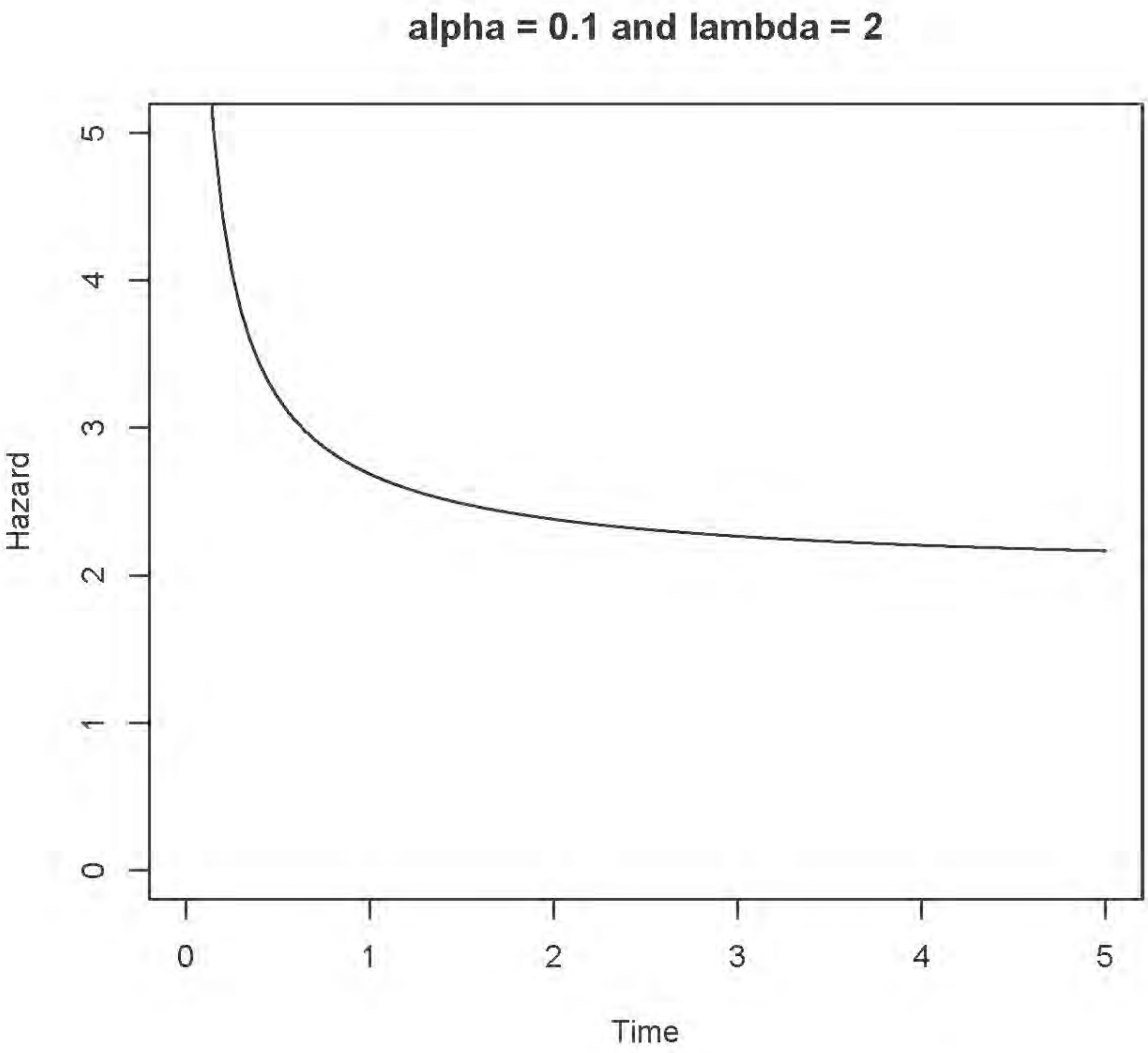


```
> # So far, we know that
>     # If alpha=1, h(t) is constant at lambda
>     # If alpha > 1, h(t) is increasing and bounded above by lambda
>     # If alpha < 1, h(t) is decreasing and bounded below by lambda
>     # alpha values close to one may make the big change occur closer to zero.
```

```
> # Now alpha < 1 again
> alpha=1/2;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



```
> alpha=1/10;lambda=2
> Time = seq(from=0,to=5,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
```



```
> # Is the hazard function asymptotic to lambda?
>
> alpha=1/10;lambda=2
> Time = seq(from=0,to=50,length=101)
> logh = dgamma(Time,shape=alpha,rate=lambda,log=TRUE) -
+     pgamma(Time,shape=alpha,rate=lambda, lower.tail=FALSE,log.p=TRUE)
> Hazard = exp(logh)
> tstring = paste('alpha =',alpha,'and lambda =',lambda)
> plot(Time,Hazard,type='l',main=tstring,ylim=c(0,5)) # That's a lower case ell
>
> # Add the line y = lambda = 2
> xx = c(0,50); yy =c(2,2); lines(xx,yy,lty=2)
```