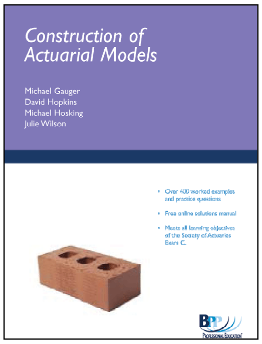


# Construction of Actuarial Models

**By J. Wilson, M. Hosking, D. Hopkins, and M. Gauger**

**Published by BPP Professional Education**

## *Solutions to practice questions – Chapter 1*

### *Solution 1.1*

An estimate is a number, which is calculated using some sample data.

An estimator is a random variable. So its value depends on the outcome of some experiment and it has a statistical distribution.

---

### *Solution 1.2*

The likelihood function is:

$$\begin{aligned} L &= \prod_{i=1}^{n} f(x_i) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \\ &= C\sigma^{-n} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \end{aligned}$$

where $C$ is a constant. Because there are now 2 unknown parameters, we have to differentiate with respect to each parameter and solve 2 simultaneous equations.
Taking logs, we obtain:

$$\ln L = \ln C - n\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

If we make the substitution $\nu = \sigma^2$, then this becomes:

$$\ln L = \ln C - \frac{n}{2}\ln\nu - \frac{1}{2\nu}\sum_{i=1}^{n}(x_i-\mu)^2$$

Differentiating with respect to $\mu$ and $\nu$:

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\nu}\sum_{i=1}^{n}(x_i-\mu) = \frac{1}{\nu}\left(\sum_{i=1}^{n}x_i - n\mu\right)$$

$$\frac{\partial \ln L}{\partial \nu} = -\frac{n}{2\nu} + \frac{1}{2\nu^2}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{2\nu}\left[-n + \frac{1}{\nu}\sum_{i=1}^{n}(x_i-\mu)^2\right]$$

Setting these equal to 0 gives:

$$\sum_{i=1}^{n} x_i - n\hat{\mu} = 0 \Rightarrow \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

and:

$$-n + \frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}\left(x_i - \hat{\mu}\right)^2 = 0 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

So what we have shown here is that the MLEs for $\mu$ and $\sigma^2$ in the normal distribution are the sample mean and the sample variance (calculated using a denominator of $n$). These results seem intuitively reasonable.

---

## *Solution 1.3*

The estimator is:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - 2\bar{X}\sum_{i=1}^{n} X_i + n\bar{X}^2\right] = \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right]$$

$$= \frac{1}{n}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right]$$

and its expected value is:

$$E\left(\tilde{\sigma}^2\right) = E\left[\frac{1}{n}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right)\right] = \frac{1}{n}\left[\sum_{i=1}^{n} E\left(X_i^2\right) - nE\left(\bar{X}^2\right)\right]$$

Since $X_i \sim N\left(\mu, \sigma^2\right)$ for $i = 1, 2, \ldots, n$, it follows that $\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ and:

$$E\left(X_i^2\right) = \text{var}\left(X_i\right) + \left[E\left(X_i\right)\right]^2 = \sigma^2 + \mu^2 \quad \text{for } i = 1, 2, \ldots, n$$

$$E\left(\bar{X}^2\right) = \text{var}\left(\bar{X}\right) + \left[E\left(\bar{X}\right)\right]^2 = \frac{\sigma^2}{n} + \mu^2$$

Substituting these into the expression for $E\left(\tilde{\sigma}^2\right)$, we obtain:

$$E\left(\tilde{\sigma}^2\right) = \frac{1}{n}\left[\sum_{i=1}^{n}\left(\sigma^2 + \mu^2\right) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right]$$

$$= \frac{1}{n}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right]$$

$$= \frac{n-1}{n}\sigma^2$$

Since $E\left(\tilde{\sigma}^2\right) \neq \sigma^2$, $\tilde{\sigma}^2$ is a biased estimator of $\sigma^2$.

 

## *Solution 1.4*

We need the mean and variance of $\bar{X}$. Using the standard results, we have:

$$E(\bar{X}) = \mu \qquad \text{and:} \qquad \text{var}(\bar{X}) = \sigma^2 / n$$

We can see from these that both the conditions for consistency are satisfied, and so $\bar{X}$ is a consistent estimator for $\mu$.

*In fact it is also true that* $S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$ *is a consistent estimator for* $\sigma^2$. *You might like to check to see whether you can prove that this is also true.*

---

## *Solution 1.5*

The MSE measures the amount of "squared deviation" of the estimator from the parameter. If this squared deviation is small, then the estimator is fairly close to the true value of the parameter. The smaller the MSE, the closer the estimator is on average, whatever the true value of the parameter.

You might have thought that there would have been other measures of this closeness that might be superior, for example $E(\theta - \hat{\theta})$ or $E(|\theta - \hat{\theta}|)$. However, there are problems with both of these apparently simpler possibilities. The first can be minimized by any unbiased estimator, and so is not sufficiently distinctive. The second has the problems of calculation that are associated with modulus functions. So the definition of mean squared error given here is preferable as a measure of "goodness-of-fit" to the true parameter value.

---

## *Solution 1.6*

From Example 1.1 we have:

$$\frac{d^2 \ln L}{d\theta^2} = \frac{400}{\theta^2} - \frac{2}{\theta^3}\sum_{j=1}^{10} T_j$$

This has expected value:

$$\begin{aligned}
E\left(\frac{d^2 \ln L}{d\theta^2}\right) &= \frac{400}{\theta^2} - \frac{2}{\theta^3}\sum_{j=1}^{10} E\left(T_j\right) \\
&= \frac{400}{\theta^2} - \frac{2}{\theta^3}\sum_{j=1}^{10} 40\,\theta \qquad \text{since } T_j \sim Gamma(40, \theta) \\
&= \frac{400}{\theta^2} - \frac{800}{\theta^2} \\
&= -\frac{400}{\theta^2}
\end{aligned}$$

So the asymptotic variance of $\tilde{\theta}$ is $\frac{\theta^2}{400}$. As we have estimated the value of $\theta$ to be 1.8825, we estimate the asymptotic variance to be $\frac{1.8825^2}{400} = 0.00886$.

## *Solution 1.7*

From the *Tables*, we have:

$$E(X) = e^{\mu + \frac{1}{2}\sigma^2}$$

and:

$$E\left(X^2\right) = e^{2\mu + 2\sigma^2}$$

So:

$$\begin{aligned}\text{var}(X) &= E\left(X^2\right) - \left[E(X)\right]^2 \\ &= e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2} \\ &= e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)\end{aligned}$$

Equating the theoretical mean to the sample mean and the theoretical variance to the sample variance, we obtain the equations:

$$e^{\mu+\frac{1}{2}\sigma^2} = 600$$

$$e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right) = 1,600$$

Squaring the first of these and substituting into the second gives:

$$600^2\left(e^{\sigma^2} - 1\right) = 1,600$$

So:

$$\sigma^2 = \ln\left(\frac{1,600}{600^2} + 1\right) = 0.00443$$

and:

$$\mu = \ln 600 - \frac{1}{2}\sigma^2 = 6.39471$$

Note. The *Tables* refer to $\mu$ and $\sigma$ as the parameters. Here we have called $\mu$ and $\sigma^2$ the parameters. No difference is intended.

---

## *Solution 1.8*

From the Tables:

$$E(X^{-1}) = \theta^{-1}\Gamma(2) = \frac{1}{\theta}$$

(Note that for an integer $n$, the gamma function is $\Gamma(n) = (n-1)!$)

(Note also that the formula is valid for $k < 1$ for this distribution.)

Taking the sample values from the Example 1.7, we have:

$$\sum x_i^{-1} = 0.266917$$

Equating these, we find that $\hat{\theta} = 56.1972$.

## *Solution 1.9*

The median lies between the 4th and 5th of these values.
The 4th value, or $400/9 = 44.44$ th percentile is 547.

The 5th value, or $500/9 = 55.56$ th percentile is 609.

Interpolating gives:

$$\pi_{50} \approx 547 + \frac{(50-400/9)}{(500/9-400/9)} \times (609-547) = 578$$

Alternatively, for the median you could just take the average of the 4th and 5th sample values.

---

## *Solution 1.10*

To find the percentile matching estimate of $\theta$, we equate the theoretical median to the sample median:

$$\begin{aligned} F_X(1,045) = 0.5 &\Rightarrow 1 - e^{-1,045/\theta} = 0.5 \\ &\Rightarrow e^{-1,045/\theta} = 0.5 \\ &\Rightarrow -\frac{1,045}{\theta} = \ln 0.5 \\ &\Rightarrow \theta = -\frac{1,045}{\ln 0.5} = 1,507.616 \end{aligned}$$

---

## *Solution 1.11*

From the *Tables*, the distribution function of the inverse Weibull distribution is:

$$F(x) = e^{-(\theta/x)^\tau}$$

Equating the theoretical median to the sample median, and the theoretical 90th percentile to the 90th percentile of the sample:

$$e^{-(\theta/1,000)^\tau} = 0.5$$

$$e^{-(\theta/2,500)^\tau} = 0.9$$

Taking logs:

$$\left(\frac{\theta}{1,000}\right)^\tau = -\ln 0.5$$

$$\left(\frac{\theta}{2,500}\right)^\tau = -\ln 0.9$$

So:

$$\left(\frac{\theta}{1,000}\right)^\tau \Bigg/ \left(\frac{\theta}{2,500}\right)^\tau = 2.5^\tau = \frac{\ln 0.5}{\ln 0.9}$$

and taking logs again:

$$\tau \ln 2.5 = \ln\left(\frac{\ln 0.5}{\ln 0.9}\right) \Rightarrow \tau = 2.05596$$

We can also find the estimate for $\theta$ by substituting back if necessary.

## *Solution 1.12*

Using the notation given in the text, we have $n=100$, $n_1=58$, $n_2=32$, $n_3=10$, $a_1=0$, $b_1=a_2=100$, $b_2=a_3=200$, $b_3=500$.

So the first raw moment is:

$$E(X_e)=\int_0^{100} x\frac{0.58}{100}dx+\int_{100}^{200} x\frac{0.32}{100}dx+\int_{200}^{500} x\frac{0.1}{300}dx$$

$$=0.0058\frac{100^2}{2}+0.0032\frac{200^2-100^2}{2}+\frac{0.1}{300}\frac{500^2-200^2}{2}=112$$

Similarly, the second raw moment is:

$$E(X_e^2)=\int_0^{100} x^2\frac{0.58}{100}dx+\int_{100}^{200} x^2\frac{0.32}{100}dx+\int_{200}^{500} x^2\frac{0.1}{300}dx$$

$$=0.0058\frac{100^3}{3}+0.0032\frac{200^3-100^3}{3}+\frac{0.1}{300}\frac{500^3-200^3}{3}=22,400$$

---

## *Solution 1.13*

Let $X_i$ denote the amount of the $i$th claim. The posterior distribution of $\theta$ is given by the conditional probabilities $\Pr\left(\theta=1,000\,|\sum_{i=1}^5 X_i=6,258\right)$, $\Pr\left(\theta=1,200\,|\sum_{i=1}^5 X_i=6,258\right)$ and $\Pr\left(\theta=1,500\,|\sum_{i=1}^5 X_i=6,258\right)$.

The first of these is given by:

$$\Pr\left(\theta=1,000\left|\sum_{i=1}^5 X_i=6,258\right.\right)=\frac{f\left(\sum_{i=1}^5 X_i=6,258\left|\theta=1,000\right.\right)\Pr(\theta=1,000)}{f\left(\sum_{i=1}^5 X_i=6,258\right)}$$

However, since the $X_i$'s are independent $Exp(\theta)$ random variables, $\sum_{i=1}^5 X_i \sim Gamma(5,\theta)$. (This is a well-known result and is easily proved using moment generating functions.)

So:

$$f\left(\sum_{i=1}^5 X_i=6,258\left|\theta\right.\right)=\frac{(6,258/\theta)^5 e^{-6,258/\theta}}{6,258\,\Gamma(5)}=K\theta^{-5}e^{-6,258/\theta}$$

where $K=\dfrac{6,258^4}{\Gamma(5)}$ (a constant), and:

$$\Pr\left(\theta=1,000\left|\sum_{i=1}^5 X_i=6,258\right.\right)=\frac{K(1,000)^{-5}e^{-6,258/1,000}\Pr(\theta=1,000)}{f\left(\sum_{i=1}^5 X_i=6,258\right)}$$

$$=\frac{K(1,000)^{-5}e^{-6,258/1,000}\times\frac{1}{3}}{f\left(\sum_{i=1}^5 X_i=6,258\right)}$$

Similarly:

$$\Pr\left(\theta = 1,200 \left| \sum_{i=1}^{5} X_i = 6,258 \right.\right) = \frac{K(1,200)^{-5}\ e^{-6,258/1,200} \times \frac{1}{3}}{f\left(\sum_{i=1}^{5} X_i = 6,258\right)}$$

and:

$$\Pr\left(\theta = 1,500 \left| \sum_{i=1}^{5} X_i = 6,258 \right.\right) = \frac{K(1,500)^{-5}\ e^{-6,258/1,500} \times \frac{1}{3}}{f\left(\sum_{i=1}^{5} X_i = 6,258\right)}$$

The probability in the denominator is:

$$\begin{aligned} f\left(\sum_{i=1}^{5} X_i = 6,258\right) &= K\Big[(1,000)^{-5}\ e^{-6,258/1,000} \times \tfrac{1}{3} + (1,200)^{-5}\ e^{-6,258/1,200} \times \tfrac{1}{3} \\ &\qquad + (1,500)^{5}\ e^{-6,258/1,500} \times \tfrac{1}{3}\Big] \\ &= K\Big[6.385574 \times 10^{-19} + 7.2799231 \times 10^{-19} \\ &\qquad + 6.7693329 \times 10^{-19}\Big] \\ &= 2.0432830 \times 10^{-18} \times K \end{aligned}$$

So:

$$\Pr\left(\theta = 1,000 \left| \sum_{i=1}^{5} X_i = 6,258 \right.\right) = \frac{6.383574 \times 10^{-19} \times K}{2.0432830 \times 10^{-18} \times K} = 0.31$$

$$\Pr\left(\theta = 1,200 \left| \sum_{i=1}^{5} X_i = 6,258 \right.\right) = \frac{7.2799231 \times 10^{-19} \times K}{2.0432830 \times 10^{-18} \times K} = 0.36$$

$$\Pr\left(\theta = 1,500 \left| \sum_{i=1}^{5} X_i = 6,258 \right.\right) = \frac{6.7693329 \times 10^{-19} \times K}{2.0432830 \times 10^{-18} \times K} = 0.33$$

---

## *Solution 1.14*

$D$ can be modeled as a Binomial random variable since:

- we have a fixed number of "trials", *ie* 500 lives under observation
- each trial has two possible outcomes, *ie* each life under observation will either die during the year or survive to the end of the year
- the lives are assumed to be independent
- each life is assumed to have the same probability of dying during the year as every other life.

## Solution 1.15

The PDF of the prior distribution of $\lambda$ is:

$$f_{prior}(\lambda) = \frac{(2{,}800/\lambda)^2\, e^{-2{,}800/\lambda}}{\lambda\,\Gamma(2)} \qquad \lambda > 0$$

The likelihood function is given by:

$$L(\lambda) = \prod_{i=1}^{5} f_X(x_i) = \prod_{i=1}^{5} \frac{1}{\lambda} e^{-x_i/\lambda} = \frac{1}{\lambda^5} e^{-\sum x_i/\lambda} = \frac{1}{\lambda^5} e^{-7{,}504/\lambda}$$

So the PDF of the posterior distribution of $\lambda$ is:

$$f_{post}(\lambda) = C\,\frac{1}{\lambda^8}\, e^{-10{,}304/\lambda} \qquad \lambda > 0$$

where $C$ is a constant. Hence the posterior distribution of $\lambda$ is inverse gamma with parameters $\alpha = 7$ and $\theta = 10{,}304$.

---

## Solution 1.16

The Bayesian estimate of $q_{60}$ under all-or-nothing loss is the mode of the posterior distribution of $q_{60}$. The mode is the value of $q$ that maximizes the posterior PDF. We have:

$$f_{post}(q) = C\, q^{15} (1-q)^{1{,}495}$$

So:

$$\begin{aligned} f'_{post}(q) &= 15C\, q^{14}(1-q)^{1{,}495} - 1{,}495C\, q^{15}(1-q)^{1{,}494} \\ &= 5C\, q^{14}(1-q)^{1{,}494}\left[3(1-q) - 299q\right] \end{aligned}$$

Setting this equal to 0:

$$3(1-q) - 299q = 0 \Rightarrow 302q = 3 \Rightarrow q = 0.00993$$

---

## Solution 1.17

The PDF of the prior distribution of $\lambda$ is:

$$f_{prior}(\lambda) = \frac{(\lambda/2)^8\, e^{-\lambda/2}}{\lambda\,\Gamma(8)} \qquad \lambda > 0$$

Now let $N_j$ denote the number of claims in year $j$. Since $N_j \sim Poi(\lambda)$, the likelihood function is given by:

$$L(\lambda) = \Pr(N_1 = 12)\Pr(N_2 = 12) = \frac{e^{-\lambda}\lambda^{12}}{12!}\,\frac{e^{-\lambda}\lambda^{12}}{12!} = \frac{e^{-2\lambda}\lambda^{24}}{(12!)^2}$$

The PDF of the posterior distribution of $\lambda$ is therefore:

$$f_{post}(\lambda) = C\,\lambda^{31}\, e^{-5\lambda/2} \qquad \lambda > 0$$

where $C$ is a constant. We recognize this as the PDF of a $Gamma(32, 0.4)$ distribution. The mean of this distribution is $32 \times 0.4 = 12.8$, so the required Bayesian estimate of $\lambda$ is 12.8.

 

## *Solution 1.18*

*The hazard function was defined in Course M.*

Let $Y = X \mid 40 < X < 80$. Then:

$$h_Y(50) = \frac{f_X(50)}{S_X(50) - S_X(80)}$$

Since $X \sim Pareto(1.5, 90)$:

$$f_X(x) = \frac{1.5(90)^{1.5}}{(x+90)^{2.5}} \qquad \text{and} \qquad S_X(x) = \left(\frac{90}{x+90}\right)^{1.5}$$

So:

$$h_Y(50) = \frac{\dfrac{1.5(90)^{1.5}}{(140)^{2.5}}}{\left(\dfrac{90}{140}\right)^{1.5} - \left(\dfrac{90}{170}\right)^{1.5}} = 0.0424$$

---

## *Solution 1.19*

(i) Let $T_i$ denote the lifetime random variable for the $i$ th life to die and $T_j$ denote the lifetime random variable for the $j$ th life to be censored. The likelihood function is:

$$\begin{aligned} L &= \prod_{i \in D} f_{T_i}(t_i) \prod_{j \in C} \Pr(T_j > 10) \\ &= \prod_{i \in D} \frac{1}{\theta} e^{-t_i/\theta} \prod_{j \in C} e^{-10/\theta} \\ &= \frac{1}{\theta^4} \exp\left(-\sum_{i \in D} t_i / \theta\right) (e^{-10/\theta})^3 \\ &= \frac{1}{\theta^4} e^{-20.16/\theta} e^{-30/\theta} \\ &= \frac{1}{\theta^4} e^{-50.16/\theta} \end{aligned}$$

(ii) The log-likelihood is:

$$\log L = -4 \log \theta - \frac{50.16}{\theta}$$

Differentiating with respect to $\theta$ gives:

$$\frac{d \log L}{d\theta} = -\frac{4}{\theta} + \frac{50.16}{\theta^2}$$

Setting the derivative equal to 0 and rearranging, we then obtain:

$$\theta = \frac{50.16}{4} = 12.54$$

Finally, since:

$$\frac{d^2 \log L}{d\theta^2} = \frac{4}{\theta^2} - \frac{100.32}{\theta^3} = -0.0254 \text{ when } \theta = 12.54$$

$\log L$ has a maximum turning point at $\theta = 12.54$. Thus 12.54 is the maximum likelihood estimate of $\theta$.

*A quicker way to establish this result is to note that the log-likelihood is the PDF of an inverse gamma random variable with parameters $\alpha = 3$ and $\theta = 50.16$. This is maximized at the mode.*

*The formula for the mode of the inverse gamma distribution is given in the Tables as:*

$$\frac{\theta}{\alpha+1} = \frac{50.16}{4} = 12.54$$

(iii) The hazard function of $Exp(\theta)$ is:

$$h(t) = \frac{1}{\theta} \text{ for } t > 0$$

Using the maximum likelihood estimate of $\theta$ obtained in (ii), the hazard function becomes:

$$h(t) = \frac{f(t)}{S(t)} = \frac{1}{12.54} = 0.079745 \text{ for } t > 0$$

So the survival function is:

$$S(t) = e^{-t/12.54} \text{ for } t > 0$$

---

## *Solution 1.20*

(i) Let $X_i$ denote the $i$ th loss random variable. The observed values of the last six loss amounts are:
1250, 1535, 1490, 1604, 2205, 2090
Since only losses in excess of \$1000 are observed, the likelihood function is given by:

$$L = \prod_{i=1}^{6} \frac{f_{X_i}(x_i)}{\Pr(X_i > 1000)}$$

Now, since $X_i$ has a Pareto distribution with parameters $\alpha$ and $\theta$, we have:

$$f_{X_i}(x_i) = \frac{\alpha\theta^\alpha}{(x_i+\theta)^{\alpha+1}}$$

and:

$$\Pr(X_i > 1000) = \left(\frac{\theta}{1000+\theta}\right)^\alpha$$

So:

$$L(\theta) = \prod_{i=1}^{6}\left(\frac{\alpha\theta^\alpha}{(x_i+\theta)^{\alpha+1}}\right)\left(\frac{\theta}{1000+\theta}\right)^{-\alpha}$$

$$= \prod_{i=1}^{6}\frac{\alpha(1000+\theta)^\alpha}{(x_i+\theta)^{\alpha+1}}$$

$$= \frac{\alpha^6(1000+\theta)^{6\alpha}}{\left[(1250+\theta)(1535+\theta)(1490+\theta)(1604+\theta)(2205+\theta)(2090+\theta)\right]^{\alpha+1}}$$

(ii) The last two values in the list represent censored observations. So the likelihood function becomes:

$$L = \prod_{i=1}^{4} \frac{f_{X_i}(x_i)}{\Pr(X_i > 1000)}\left[\frac{\Pr(X_i > 2000)}{\Pr(X_i > 1000)}\right]^2$$

Now:

$$\Pr(X_i > 2000) = \left(\frac{\theta}{2000+\theta}\right)^\alpha$$

So the likelihood is:

$$L=\left(\frac{1000+\theta}{2000+\theta}\right)^{2\alpha}\frac{\alpha^4(1000+\theta)^{4\alpha}}{\left[(1250+\theta)(1535+\theta)(1490+\theta)(1604+\theta)\right]^{\alpha+1}}$$