ETH Zürich, D-MATH
HS 2021
Prof. Dr. Mario V. Wüthrich

Non-Life Insurance: Mathematics and Statistics

Solution sheet 12

Solution 12.1 (Inhomogeneous) Credibility Estimators for Claim Counts

First, we note that

$$\mu_0 \stackrel{\text{def}}{=} \mathbb{E}[\mu(\Theta_i)] = \mathbb{E}[\Theta_i \lambda_0] = \lambda_0 = 0.088.$$

Then, we define

$$X_{i,1} = \frac{N_{i,1}}{v_{i,1}},$$

for all $i \in \{1, \ldots, 5\}$. We have

$$\mathbb{E}[X_{i,1}|\Theta_i] = \frac{1}{v_{i,1}} \mathbb{E}[N_{i,1}|\Theta_i] = \frac{1}{v_{i,1}} \mu(\Theta_i)\, v_{i,1} = \mu(\Theta_i)$$

and

$$\text{Var}(X_{i,1}|\Theta_i) = \frac{1}{v_{i,1}^2} \text{Var}(N_{i,1}|\Theta_i) = \frac{1}{v_{i,1}^2} \mu(\Theta_i)\, v_{i,1} = \frac{\mu(\Theta_i)}{v_{i,1}} = \frac{\sigma^2(\Theta_i)}{v_{i,1}},$$

with

$$\sigma^2(\Theta_i) = \mu(\Theta_i) = \Theta_i \lambda_0,$$

for all $i \in \{1, \ldots, 5\}$. Moreover, since

$$\mathbb{E}[\mu(\Theta_i)^2] = \text{Var}(\mu(\Theta_i)) + \mathbb{E}[\mu(\Theta_i)]^2 = \tau^2 + \lambda_0^2 < \infty$$

and

$$\mathbb{E}[X_{i,1}^2|\Theta_i] = \text{Var}(X_{i,1}|\Theta_i) + \mathbb{E}[X_{i,1}|\Theta_i]^2 = \frac{\mu(\Theta_i)}{v_{i,1}} + \mu(\Theta_i)^2,$$

we get

$$\mathbb{E}[X_{i,1}^2] = \mathbb{E}[\mathbb{E}[X_{i,1}^2|\Theta_i]] = \mathbb{E}\left[\frac{\mu(\Theta_i)}{v_{i,1}} + \mu(\Theta_i)^2\right] = \frac{\lambda_0}{v_{i,1}} + \tau^2 + \lambda_0^2 < \infty,$$

for all $i \in \{1, \ldots, 5\}$. In particular, Model Assumptions 8.12 of the lecture notes (version of December 17, 2020) for the Bühlmann-Straub model are satisfied. The (expected) volatility σ^2 within the regions defined in formula (8.4) of the lecture notes (version of December 17, 2020) is given by

$$\sigma^2 = \mathbb{E}[\sigma^2(\Theta_i)] = \mathbb{E}[\mu(\Theta_i)] = \lambda_0 = 0.088.$$

(a) Let $i \in \{1, \ldots, 5\}$. Then, according to Theorem 8.16 of the lecture notes (version of December 17, 2020), the inhomogeneous credibility estimator is given by

$$\widehat{\widehat{\mu(\Theta_i)}} = \alpha_{i,T}\, \widehat{X}_{i,1:T} + (1 - \alpha_{i,T})\, \mu_0,$$

with credibility weight $\alpha_{i,T}$ and observation based estimator $\widehat{X}_{i,1:T}$

$$\alpha_{i,T} = \frac{v_{i,1}}{v_{i,1} + \frac{\sigma^2}{\tau^2}} \qquad \text{and} \qquad \widehat{X}_{i,1:T} = \frac{1}{v_{i,1}} v_{i,1} X_{i,1} = X_{i,1}.$$

Hence, we get

$$\widehat{\widehat{\mu(\Theta_i)}} = \frac{v_{i,1}}{v_{i,1} + \frac{\sigma^2}{\tau^2}} X_{i,1} + \frac{\frac{\sigma^2}{\tau^2}}{v_{i,1} + \frac{\sigma^2}{\tau^2}} \mu_0 = \frac{v_{i,1}}{v_{i,1} + \frac{0.088}{0.00024}} X_{i,1} + \frac{\frac{0.088}{0.00024}}{v_{i,1} + \frac{0.088}{0.00024}} 0.088.$$

The results for the five regions are summarized in the following table:

	region 1	region 2	region 3	region 4	region 5
$\alpha_{i,T}$	99.3%	96.5%	99.7%	99.0%	92.0%
$\widehat{X}_{i,1:T}$	7.8%	7.8%	7.4%	9.8%	7.5%
$\widehat{\widehat{\mu(\Theta_i)}}$	7.8%	7.9%	7.4%	9.8%	7.6%

Table 1: Estimated credibility weights $\alpha_{i,T}$, observation based estimates $\widehat{X}_{i,1:T}$ and inhomogeneous credibility estimates $\widehat{\widehat{\mu(\Theta_i)}}$ in regions $i = 1, \ldots, 5$.

Note that since the credibility coefficient $\kappa = \sigma^2/\tau^2 \approx 367$ is rather small compared to the volumes $v_{1,1}, \ldots, v_{5,1}$, the credibility weights $\alpha_{1,T}, \ldots, \alpha_{5,T}$ are fairly high. Moreover, the observation based estimates are almost the same for the regions $1, 2, 3$ and 5, only $\widehat{X}_{4,1:T}$ is roughly 2% higher. As a result, only for the smallest two credibility weights $\alpha_{2,T}$ and $\alpha_{5,T}$ we see a slight upwards deviation of the corresponding inhomogeneous credibility estimates $\widehat{\widehat{\mu(\Theta_2)}}$ and $\widehat{\widehat{\mu(\Theta_5)}}$ from the observation based estimates $\widehat{X}_{2,1:T}$ and $\widehat{X}_{5,1:T}$ towards $\mu_0 = 8.8\%$. If we decreased the volatility τ^2 between the risk classes, the credibility coefficient $\kappa = \sigma^2/\tau^2$ would increase and, thus, the credibility weights $\alpha_{1,T}, \ldots, \alpha_{5,T}$ would decrease. Consequently, the credibility estimates would move stronger towards $\mu_0 = 8.8\%$.

(b) Since the number of policies grows 5% in each region, next year's numbers of policies $v_{1,2}, \ldots, v_{5,2}$ are given by

	region 1	region 2	region 3	region 4	region 5
$v_{i,2}$	52'564	10'642	127'376	36'797	4'402

Table 2: Next year's numbers of policies in regions $i = 1, \ldots, 5$.

Similarly to part (a), we define

$$X_{i,2} = \frac{N_{i,2}}{v_{i,2}},$$

for all $i \in \{1, \ldots, 5\}$. According to the exercise sheet, next year's numbers of claims stay within the Bühlmann-Straub model framework assumed for this year's numbers of claims. Thus, according to formula (8.16) of the lecture notes (version of December 17, 2020), the mean square error of prediction is given by, for all $i \in \{1, \ldots, 5\}$,

$$\mathbb{E}\left[\left(\frac{N_{i,2}}{v_{i,2}} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \mathbb{E}\left[\left(X_{i,2} - \widehat{\widehat{\mu(\Theta_i)}}\right)^2\right] = \frac{\sigma^2}{v_{i,2}} + (1 - \alpha_{i,T})\,\tau^2. \tag{1}$$

We get the following root mean square errors of prediction for the five regions:

	region 1	region 2	region 3	region 4	region 5
$\sqrt{\text{mean square errors of prediction}}$	0.185%	0.408%	0.119%	0.221%	0.627%
in % of the credibility estimates	2.4%	5.2%	1.6%	2.2%	8.3%

Table 3: Root mean square errors of prediction in regions $i = 1, \ldots, 5$.

Note that we get the highest root mean square errors of prediction for regions 2 and 5, i.e. exactly for those regions for which we also have the lowest volumes and, consequently, the lowest credibility weights. Of course, this is due to formula (1).

Solution 12.2 (Homogeneous) Credibility Estimators for Claim Sizes

We define

$$X_{i,t} \;=\; \frac{Y_{i,t}}{v_{i,t}},$$

for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. Then, we have

$$\mathbb{E}[X_{i,t}|\Theta_i] \;=\; \frac{1}{v_{i,t}}\,\mathbb{E}[Y_{i,t}|\Theta_i] \;=\; \frac{1}{v_{i,t}}\,\frac{\mu(\Theta_i)cv_{i,t}}{c} \;=\; \mu(\Theta_i)$$

and

$$\operatorname{Var}(X_{i,t}|\Theta_i) \;=\; \frac{1}{v_{i,t}^2}\operatorname{Var}(Y_{i,t}|\Theta_i) \;=\; \frac{1}{v_{i,t}^2}\,\frac{\mu(\Theta_i)cv_{i,t}}{c^2} \;=\; \frac{\mu(\Theta_i)}{cv_{i,t}} \;=\; \frac{\sigma^2(\Theta_i)}{v_{i,t}},$$

with

$$\sigma^2(\Theta_i) \;=\; \frac{\mu(\Theta_i)}{c} \;=\; \frac{\Theta_i}{c},$$

for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. Moreover, using that

$$\mathbb{E}[X_{i,t}^2|\Theta_i] \;=\; \operatorname{Var}(X_{i,t}|\Theta_i) + \mathbb{E}[X_{i,t}|\Theta_i]^2 \;=\; \frac{\mu(\Theta_i)}{cv_{i,t}} + \mu(\Theta_i)^2 \;=\; \frac{\Theta_i}{cv_{i,t}} + \Theta_i^2,$$

we get

$$\mathbb{E}[X_{i,t}^2] \;=\; \mathbb{E}[\mathbb{E}[X_{i,t}^2|\Theta_i]] \;=\; \mathbb{E}\left[\frac{\Theta_i}{cv_{i,t}} + \Theta_i^2\right] \;<\; \infty$$

by assumption, for all $i \in \{1,2,3,4\}$ and $t \in \{1,2\}$. In particular, Model Assumptions 8.12 of the lecture notes (version of December 17, 2020) for the Bühlmann-Straub model are satisfied.

(a) First, following Theorem 8.16 of the lecture notes (version of December 17, 2020), we define the observation based estimator $\widehat{X}_{i,1:T}$ as

$$\widehat{X}_{i,1:T} \;=\; \frac{1}{\sum_{t=1}^T v_{i,t}} \sum_{t=1}^T v_{i,t}X_{i,t} \;=\; \frac{v_{i,1}X_{i,1} + v_{i,2}X_{i,2}}{v_{i,1}+v_{i,2}} \;=\; \frac{Y_{i,1}+Y_{i,2}}{v_{i,1}+v_{i,2}},$$

for all $i \in \{1,2,3,4\}$. Then, we need to estimate the structural parameters $\sigma^2 = \mathbb{E}[\sigma^2(\Theta_1)]$ and $\tau^2 = \operatorname{Var}(\mu(\Theta_1))$. According to formula (8.14) of the lecture notes (version of December 17, 2020), σ^2 can be estimated by

$$\widehat{\sigma}_T^2 \;=\; \frac{1}{I}\sum_{i=1}^I \frac{1}{T-1}\sum_{t=1}^T v_{i,t}\left(X_{i,t} - \widehat{X}_{i,1:T}\right)^2 \;\approx\; 1.3 \cdot 10^{10}.$$

In order to estimate τ^2, we define first the weighted sample mean $\bar{X}$ over all observations by

$$\bar{X} \;=\; \frac{1}{\sum_{i=1}^I\sum_{t=1}^T v_{i,t}} \sum_{i=1}^I\sum_{t=1}^T v_{i,t}X_{i,t} \;=\; \frac{\sum_{i=1}^I Y_{i,1}+Y_{i,2}}{\sum_{i=1}^I v_{i,1}+v_{i,2}} \;\approx\; 7\text{'}004.$$

Then, following the lecture notes, we define $\widehat{v}_T^2$, c_w and $\widehat{t}_T^2$ as

$$\widehat{v}_T^2 \;=\; \frac{I}{I-1}\sum_{i=1}^I \frac{v_{i,1}+v_{i,2}}{\sum_{j=1}^I v_{j,1}+v_{j,2}}\left(\widehat{X}_{i,1:T} - \bar{X}\right)^2 \;\approx\; 9.3\cdot 10^7,$$

$$c_w \;=\; \frac{I-1}{I}\left[\sum_{i=1}^I \frac{v_{i,1}+v_{i,2}}{\sum_{j=1}^I v_{j,1}+v_{j,2}}\left(1 - \frac{v_{i,1}+v_{i,2}}{\sum_{j=1}^I v_{j,1}+v_{j,2}}\right)\right]^{-1} \;\approx\; 1.425$$

and

$$\widehat{t}_T^2 \;=\; c_w \left(\widehat{v}_T^2 - \frac{I\,\widehat{\sigma}_T^2}{\sum_{i=1}^I v_{i,1} + v_{i,2}} \right) \;\approx\; 1.25 \cdot 10^8.$$

Then, using formula (8.15) of the lecture notes (version of December 17, 2020), τ^2 is estimated by

$$\widehat{\tau}_T^2 \;=\; \max\left\{\widehat{t}_T^2, 0\right\} \;=\; \widehat{t}_T^2 \;\approx\; 1.25 \cdot 10^8.$$

Now let $i \in \{1,2,3,4\}$. According to Theorem 8.16 of the lecture notes (version of December 17, 2020), the homogeneous credibility estimator is given by

$$\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}} \;=\; \alpha_{i,T}\,\widehat{X}_{i,1:T} + (1-\alpha_{i,T})\,\widehat{\mu}_T,$$

with credibility weight $\alpha_{i,T}$ and estimate $\widehat{\mu}_T$

$$\alpha_{i,T} \;=\; \frac{v_{i,1}+v_{i,2}}{v_{i,1}+v_{i,2}+\widehat{\sigma}_T^2/\widehat{\tau}_T^2} \qquad \text{and} \qquad \widehat{\mu}_T \;=\; \frac{1}{\sum_{i=1}^I \alpha_{i,T}} \sum_{i=1}^I \alpha_{i,T}\widehat{X}_{i,1:T} \;\approx\; 14\text{'}538.$$

The results for the four risk classes are summarized in the following table:

	risk class 1	risk class 2	risk class 3	risk class 4
$\alpha_{i,T}$	95.4%	98.4%	82.5%	89.6%
$\widehat{X}_{i,1:T}$	10'493	1'907	18'375	29'197
$\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}$	10'677	2'107	17'702	27'665

Table 4: Estimated credibility weights $\alpha_{i,T}$, observation based estimates $\widehat{X}_{i,1:T}$ and homogeneous credibility estimates $\widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}$ in risk classes $i = 1,2,3,4$.

Looking at the credibility weights $\alpha_{1,T}, \alpha_{2,T}, \alpha_{3,T}$ and $\alpha_{4,T}$, we see that the estimated credibility coefficient $\hat{\kappa} = \hat{\sigma}_T^2/\hat{\tau}_T^2 \approx 104$ has the biggest impact on risk classes 3 and 4, where we have less volumes compared to risk classes 1 and 2. As a result, the smoothing of the observation based estimates $\widehat{X}_{1,1:T}, \widehat{X}_{2,1:T}, \widehat{X}_{3,1:T}$ and $\widehat{X}_{4,1:T}$ towards $\hat{\mu}_T \approx 14\text{'}538$ is strongest for risk classes 3 and 4.

(b) Since the number of claims grows 5% in each risk class, next year's numbers of claims $v_{1,3}, \ldots, v_{4,3}$ are given by

	risk class 1	risk class 2	risk class 3	risk class 4
$v_{i,3}$	1'167	3'468	262	479

Table 5: Next year's numbers of claims in risk classes $i = 1,2,3,4$.

Similarly to part (a), we define

$$X_{i,3} \;=\; \frac{Y_{i,3}}{v_{i,3}},$$

for all $i \in \{1,2,3,4\}$. According to the exercise sheet, next year's total claim sizes stay within the Bühlmann-Straub model framework assumed for the previous year's total claim sizes. Thus, according to formula (8.17) of the lecture notes (version of December 17, 2020), the mean square error of prediction can be estimated by, for all $i \in \{1,2,3,4\}$,

$$\widehat{\mathbb{E}}\left[\left(\frac{Y_{i,3}}{v_{i,3}} - \widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}\right)^2\right] = \widehat{\mathbb{E}}\left[\left(X_{i,3} - \widehat{\widehat{\mu(\Theta_i)}}^{\text{hom}}\right)^2\right] = \frac{\widehat{\sigma}_T^2}{v_{i,3}} + (1-\alpha_{i,T})\widehat{\tau}_T^2\left(1 + \frac{1-\alpha_{i,T}}{\sum_{i=1}^I \alpha_{i,T}}\right). \tag{2}$$

We get the following estimated root mean square errors of prediction for the four risk classes:

	risk class 1	risk class 2	risk class 3	risk class 4
$\sqrt{\text{mean square errors of prediction}}$	4'108	2'392	8'508	6'360
in % of the credibility estimates	38.5%	113.5%	48.1%	23.0%

Table 6: Estimated root mean square errors of prediction in risk classes $i = 1,2,3,4$.

According to formula (2), the smaller the volumes in a particular risk class, the bigger the corresponding estimated root mean square error of prediction. Moreover, note that these estimated root mean square errors of prediction are rather high compared to the credibility estimates, which indicates a high variability within the individual risk classes.

Solution 12.3 Degenerate MLE and the Poisson-Gamma Model

(a) We observe that $N_t = 0$ for all $t = 1,\ldots,T$. In this case, the log-likelihood function $\ell_{\boldsymbol{N}}(\lambda)$ of the data $\boldsymbol{N} = (N_1,\ldots,N_T)$ for the unknown parameter $\lambda > 0$ is given by

$$\ell_{\boldsymbol{N}}(\lambda) = \sum_{t=1}^T \log\left(\exp\{-\lambda v_t\}\frac{(\lambda v_t)^{N_t}}{N_t!}\right) = \sum_{t=1}^T \log\left(\exp\{-\lambda v_t\}\right) = -\lambda\sum_{t=1}^T v_t.$$

As the volumes $v_1,\ldots,v_T$ are positive, we see that $\ell_{\boldsymbol{N}}(\lambda)$ increases as λ decreases, i.e. here we are in the situation of a degenerate Poisson model with MLE $\widehat{\lambda}_T = 0$. If we used this degenerate model for premium calculations, we would get a pure risk premium of 0, as we do not expect any claims. Of course, a model with zero pure risk premium does not make any sense, i.e. we need to circumvent this degenerate case. This can be done for example with the Poisson-gamma model considered in part (b).

(b) (i) The prior estimator λ_0 of the unknown parameter Λ is given by

$$\lambda_0 = \mathbb{E}[\Lambda] = \frac{\gamma}{c} = \frac{1}{50}.$$

According to Theorem 8.2 of the lecture notes (version of December 17, 2020), we have

$$\Lambda|\boldsymbol{N} \sim \Gamma\left(\gamma + \sum_{t=1}^T N_t, c + \sum_{t=1}^T v_t\right),$$

where we write $\boldsymbol{N} = (N_1,\ldots,N_T)$. Therefore, the posterior estimator $\widehat{\lambda}_T^{\text{post}}$ of the unknown parameter Λ is given by

$$\widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{N}] = \frac{\gamma + \sum_{t=1}^T N_t}{c + \sum_{t=1}^T v_t} = \frac{1+0}{50+50} = \frac{1}{100}.$$

(ii) According to Corollary 8.4 of the lecture notes (version of December 17, 2020), we can write

$$\widehat{\lambda}_T^{\text{post}} = \alpha_T \, \widehat{\lambda}_T + (1 - \alpha_T)\, \lambda_0$$

by setting

$$\alpha_T = \frac{\sum_{t=1}^T v_t}{c + \sum_{t=1}^T v_t} \in (0,1).$$

In our case we get

$$\alpha_T = \frac{50}{50+50} = \frac{1}{2}.$$

Indeed, we check

$$\alpha_T \, \widehat{\lambda}_T + (1-\alpha_T)\, \lambda_0 = \frac{1}{2} \cdot 0 + \left(1 - \frac{1}{2}\right) \cdot \frac{1}{50} = \frac{1}{100} = \widehat{\lambda}_T^{\text{post}}.$$

(iii) Similarly as in item (i), for the posterior estimator $\widehat{\lambda}_{T+1}^{\text{post}}$, conditionally given data $(N_1, v_1), \ldots, (N_{T+1}, v_{T+1})$, we get

$$\widehat{\lambda}_{T+1}^{\text{post}} = \frac{\gamma + \sum_{t=1}^{T+1} N_t}{c + \sum_{t=1}^{T+1} v_t} = \frac{1+1}{50+60} = \frac{2}{110}.$$

According to Corollary 8.6 of the lecture notes (version of December 17, 2020), we can write

$$\widehat{\lambda}_{T+1}^{\text{post}} = \beta_{T+1} \frac{N_{T+1}}{v_{T+1}} + (1 - \beta_{T+1})\, \widehat{\lambda}_T^{\text{post}}$$

by setting

$$\beta_{T+1} = \frac{v_{T+1}}{c + \sum_{t=1}^{T+1} v_t} \in (0,1).$$

In our case we get

$$\beta_{T+1} = \frac{10}{50+60} = \frac{1}{11}.$$

Indeed, we check

$$\beta_{T+1} \frac{N_{T+1}}{v_{T+1}} + (1-\beta_{T+1})\, \widehat{\lambda}_T^{\text{post}} = \frac{1}{11}\frac{1}{10} + \left(1 - \frac{1}{11}\right) \frac{1}{100} = \frac{1}{110} + \frac{1}{110} = \frac{2}{110} = \widehat{\lambda}_{T+1}^{\text{post}}.$$

(c) Note that, by definition, a Poisson random variable requires a positive frequency parameter. In case of a frequency parameter which is equal to 0, we are in the degenerate Poisson model, see also part (a). However, if $\Lambda \sim \mathcal{N}(\mu, \sigma^2)$, then the probability that Λ is negative is given by

$$\mathbb{P}[\Lambda < 0] = \mathbb{P}\left[\frac{\Lambda - \mu}{\sigma} < -\frac{\mu}{\sigma}\right] = \Phi\left(-\frac{\mu}{\sigma}\right) > 0,$$

where Φ denotes the distribution function of a standard Gaussian distribution. As there is a positive probability that the frequency parameter Λ is negative, we conclude that a Poisson-normal model is not well-defined and, thus, not reasonable.

Solution 12.4 Pareto-Gamma Model

(a) Let $f_{\boldsymbol{Y}|\Lambda}$ denote the density of $\boldsymbol{Y}|\Lambda$ and f_Λ the density of Λ. Then, we have

$$\begin{aligned} f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=\alpha) &= \prod_{t=1}^T \frac{\alpha}{\theta}\left(\frac{y_t}{\theta}\right)^{-(\alpha+1)} \cdot 1_{\{y_t\geq\theta\}} \\ &= \alpha^T\theta^{-T}\left(\prod_{t=1}^T \frac{y_t}{\theta}\right)^{-\alpha}\left(\prod_{t=1}^T \frac{y_t}{\theta}\right)^{-1} \cdot 1_{\{y_t\geq\theta\}} \end{aligned}$$

and

$$f_\Lambda(\alpha) = \frac{c^\gamma}{\Gamma(\gamma)}\alpha^{\gamma-1}\exp\{-c\alpha\}\cdot 1_{\{\alpha>0\}}.$$

Let $f_{\Lambda|\boldsymbol{Y}}$ denote the density of $\Lambda|\boldsymbol{Y}$. Then, for all $\alpha>0$ and $y_1,\ldots,y_T\geq\theta$, we have

$$\begin{aligned} f_{\Lambda|\boldsymbol{Y}}(\alpha|Y_1=y_1,\ldots,Y_T=y_T) &= \frac{f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=\alpha)\,f_\Lambda(\alpha)}{\int_0^\infty f_{\boldsymbol{Y}|\Lambda}(y_1,\ldots,y_T|\Lambda=x)\,f_\Lambda(x)\,dx} \\ &\propto \alpha^T\left(\prod_{t=1}^T \frac{y_t}{\theta}\right)^{-\alpha}\alpha^{\gamma-1}\exp\{-c\alpha\} \\ &= \alpha^{\gamma+T-1}\exp\left\{-\alpha\sum_{t=1}^T\log\frac{y_t}{\theta}\right\}\exp\{-c\alpha\} \\ &= \alpha^{\gamma+T-1}\exp\left\{-\alpha\left(\sum_{t=1}^T\log\frac{y_t}{\theta}+c\right)\right\}. \end{aligned}$$

We conclude that

$$\Lambda|\boldsymbol{Y} \sim \Gamma\left(\gamma+T, c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right).$$

(b) First, we observe that

$$\lambda_0 = \mathbb{E}[\Lambda] = \frac{\gamma}{c} \qquad \text{and} \qquad \widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{Y}] = \frac{\gamma+T}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}.$$

Then, we can write

$$\begin{aligned} \widehat{\lambda}_T^{\text{post}} &= \frac{\gamma+T}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}} = \frac{\sum_{t=1}^T\log\frac{Y_t}{\theta}}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{T}{\sum_{t=1}^T\log\frac{Y_t}{\theta}} + \frac{c}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\frac{\gamma}{c} \\ &= \alpha_T\,\widehat{\lambda}_T + (1-\alpha_T)\,\lambda_0, \end{aligned}$$

with

$$\alpha_T = \frac{\sum_{t=1}^T\log\frac{Y_t}{\theta}}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}.$$

(c) For the (conditional mean square error) uncertainty of the posterior estimator $\widehat{\lambda}_T^{\text{post}} = \mathbb{E}[\Lambda|\boldsymbol{Y}]$ we have

$$\begin{aligned} \mathbb{E}\left[\left(\Lambda-\widehat{\lambda}_T^{\text{post}}\right)^2\middle|\boldsymbol{Y}\right] &= \mathbb{E}\left[(\Lambda-\mathbb{E}[\Lambda|\boldsymbol{Y}])^2\middle|\boldsymbol{Y}\right] = \operatorname{Var}(\Lambda|\boldsymbol{Y}) = \frac{\gamma+T}{\left(c+\sum_{t=1}^T\log\frac{Y_t}{\theta}\right)^2} \\ &= \frac{1}{c+\sum_{t=1}^T\log\frac{Y_t}{\theta}}\,\widehat{\lambda}_T^{\text{post}} = (1-\alpha_T)\,\frac{1}{c}\,\widehat{\lambda}_T^{\text{post}}. \end{aligned}$$

(d) Analogously to $\widehat{\lambda}_T^{\text{post}}$, the posterior estimator $\widehat{\lambda}_{T-1}^{\text{post}}$ in the sub-model where we only have observed $(Y_1, \ldots, Y_{T-1})$ is given by

$$\widehat{\lambda}_{T-1}^{\text{post}} = \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}.$$

Thus, we can write

$$\begin{aligned}\widehat{\lambda}_T^{\text{post}} &= \frac{\gamma + T}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} = \frac{\log \frac{Y_T}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \frac{1}{\log \frac{Y_T}{\theta}} + \frac{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}} \frac{\gamma + T - 1}{c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}} \\ &= \beta_T \frac{1}{\log \frac{Y_T}{\theta}} + (1 - \beta_T) \widehat{\lambda}_{T-1}^{\text{post}},\end{aligned}$$

with

$$\beta_T = \frac{\log \frac{Y_T}{\theta}}{c + \sum_{t=1}^{T} \log \frac{Y_t}{\theta}}.$$

Remark: Suppose we want to use the observations $Y_1, \ldots, Y_{T-1}$ in order to estimate Y_T in a Bayesian sense. Then, we have

$$\begin{aligned}\mathbb{E}[Y_T | Y_1, \ldots, Y_{T-1}] &= \mathbb{E}\left[\mathbb{E}\left[Y_T | Y_1, \ldots, Y_{T-1}, \Lambda\right] | Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.} \\ &= \mathbb{E}\left[\mathbb{E}\left[Y_T | \Lambda\right] | Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.},\end{aligned}$$

where in the second equality we used that, conditionally given Λ, $Y_1, \ldots, Y_T$ are independent. Now, by assumption,

$$Y_T | \Lambda \sim \text{Pareto}(\theta, \Lambda).$$

In particular, $\mathbb{E}[Y_T|\Lambda] < \infty$ if and only if $\Lambda > 1$. However, according to part (a) (for $T - 1$ instead of T observations), we have

$$\Lambda | (Y_1, \ldots, Y_{T-1}) \sim \Gamma\left(\gamma + T - 1, c + \sum_{t=1}^{T-1} \log \frac{Y_t}{\theta}\right).$$

Since the range of a gamma distribution is the whole positive real line, this implies that

$$0 < \mathbb{P}\left[\Lambda \leq 1 | Y_1, \ldots, Y_{T-1}\right] = \mathbb{P}\left[\mathbb{E}\left[Y_T | \Lambda\right] = \infty | Y_1, \ldots, Y_{T-1}\right] \quad \text{a.s.}$$

We conclude that

$$\mathbb{E}[Y_T | Y_1, \ldots, Y_{T-1}] = \infty \quad \text{a.s.}$$