1. We just need to write out all the terms and simplify. We have

\[ \text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2] = \mathbb{E}[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2, \]

\[ B(\hat{\theta})^2 = (\mathbb{E}[\hat{\theta}] - \theta)^2 = (\mathbb{E}[\hat{\theta}])^2 - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 \]

and

\[ \text{Var}(\hat{\theta}) = \mathbb{E}[(\mathbb{E}[\hat{\theta}] - \hat{\theta})^2] = \mathbb{E}[(\mathbb{E}[\hat{\theta}])^2 - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2] \]

\[ = (\mathbb{E}[\hat{\theta}])^2 - 2(\mathbb{E}[\hat{\theta}])^2 + \mathbb{E}[\hat{\theta}^2] = - (\mathbb{E}[\hat{\theta}])^2 + \mathbb{E}[\hat{\theta}^2], \]

which indeed yields

\[ \text{MSE}(\hat{\theta}) = \mathbb{E}[\hat{\theta}^2] - 2\theta\mathbb{E}[\hat{\theta}] + \theta^2 = B(\hat{\theta})^2 + \text{Var}(\hat{\theta}). \]

2. a) Note that by definition of a random sample, for every \( i = 1, \ldots, n \) the rv \( X_i \) has the same distribution as \( X \) (and thus also has pdf \( f \)) and hence we can compute

\[ \mathbb{E}[X_i] = \int_{\mathbb{R}} x f(x|\theta) \, dx = \int_0^\infty x \frac{1}{\theta} e^{-x/\theta} \, dx \overset{\text{ibp}}{=} -xe^{-x/\theta} \bigg|_0^\infty - \int_0^\infty e^{-x/\theta} \, dx = 0 + \theta = \theta. \]

(ibp stands for integration by parts). Using this we get

\[ \mathbb{E}[\hat{\theta}_1] = \mathbb{E}[X_1] = \theta, \]

\[ \mathbb{E}[\hat{\theta}_2] = \frac{1}{2} \mathbb{E}[X_1] + \frac{1}{2} \mathbb{E}[X_2] = \theta/2 + \theta/2 = \theta, \]

\[ \mathbb{E}[\hat{\theta}_3] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \frac{1}{n} n\theta = \theta. \]

As each estimator has mean equal to \( \theta \) they are indeed all unbiased. For computing their MSE’s we may use the result from Lemma 3.3.6 (see also Exercise 1 above) with the understanding that they all have bias equal to 0 to compute

\[ \text{MSE}(\hat{\theta}_1) = \text{Var}(\hat{\theta}_1) = \text{Var}(X_1) = \theta^2 \]

(for the last step, just use the pdf and repeated integration by parts to compute the second moment of \( X \)),

\[ \text{MSE}(\hat{\theta}_2) = \text{Var}(\hat{\theta}_2) = \text{Var}((X_1 + 2X_2)/3) = \frac{1}{9} (\text{Var}(X_1) + 4\text{Var}(X_2)) = \frac{1}{9} (\theta^2 + 4\theta^2) = \frac{5}{9} \theta^2, \]

\[ \text{MSE}(\hat{\theta}_3) = \text{Var}(\hat{\theta}_3) = \text{Var}((X_1 + 2X_2)/3) = \frac{1}{9} (\text{Var}(X_1) + 4\text{Var}(X_2)) = \frac{1}{9} (\theta^2 + 4\theta^2) = \frac{5}{9} \theta^2, \]
\[
\text{MSE}(\hat{\theta}_3) = \text{Var}(\hat{\theta}_3) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \theta^2 = \frac{\theta^2}{n}.
\]

Hence, only \( \hat{\theta}_3 \) has the property that its MSE vanishes as \( n \to \infty \), for the others the MSE stays constant. So it would be better to use \( \hat{\theta}_3 \) as it performs (a lot) better than the other two. (Note that \( \hat{\theta}_3 \) is also the only one that makes use of all the available information in the random sample, the others use only a bit of it).

b) As we have only one unknown parameter, for the method of moments we need only one equation to solve \( \theta \) from, namely the equation

\[
\mathbb{E}[X] = \frac{1}{n} \sum_{i=1}^{n} X_i.
\]

Now, in part a) we already computed \( \mathbb{E}[X] = \theta \). As \( X \) has the same distribution as the \( X_i \)'s, we also have that \( \mathbb{E}[X] = \theta \). Hence the equation we need to solve becomes

\[
\theta = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

and of course we don’t have to any solving: the equation immediately yields \( \hat{\theta}_{\text{mom}} = (1/n) \sum_{i=1}^{n} X_i \). Hence indeed \( \hat{\theta}_{\text{mom}} = \hat{\theta}_3 \).

3. a) Recall that \( X \) takes values in \( \{0, 1, \ldots \} \) with pmf \( p(k|\lambda) = \lambda^k e^{-\lambda} / k! \) for \( k = 0, 1, \ldots \). Hence the likelihood function of the observed sample is given by

\[
L(\lambda) = \prod_{i=1}^{n} p(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{e^{-n\lambda \bar{x}}}{\bar{x}^n},
\]

where \( \bar{x} = (1/n) \sum_{i=1}^{n} X_i \) is the sample mean and \( \bar{x} := \prod_{i=1}^{n} x_i! \). Hence the log likelihood \( l \) is given by

\[
l(\lambda) = \log(L(\lambda)) = \log \left( \frac{e^{-n\lambda \bar{x}}}{\bar{x}^n} \right) = -n\lambda + n\bar{x} \log(\lambda) - \log(\bar{x})
\]

and we can compute

\[
l'(\lambda) = -n + n\bar{x} / \lambda = 0 \iff \lambda = \bar{x}.
\]

Checking the boundary conditions we see that \( l(\lambda) \to -\infty \) as \( \theta \) tends to 0 or \( \infty \), and hence we conclude that the zero of \( l' \) is indeed the global maximum of \( l \), hence we arrive at \( \hat{\lambda}_{\text{mle}} = \bar{x} \).

b) Recall that by definition \( \hat{\lambda}_{\text{mle}} \) is consistent if it converges to \( \lambda \) in probability as \( n \) (the length of a random sample) tends to \( \infty \). (We consider the estimator as a function of a random sample \( X_1, \ldots, X_n \) now). Rather than proving this directly we make use of
Lemma 3.3.13 from the lecture notes: it is enough to show that \( \text{MSE}(\hat{\lambda}_{\text{mle}}) \) tends to 0 as \( n \to \infty \). In order to show this, by Lemma 3.3.6 (see also Exercise 1 above) we have

\[
\text{MSE}(\hat{\lambda}_{\text{mle}}) = B(\hat{\lambda}_{\text{mle}})^2 + \text{Var}(\hat{\lambda}_{\text{mle}}). \tag{1}
\]

Now

\[
\mathbb{E}[\hat{\lambda}_{\text{mle}}] = \mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^{n} X_i \right] = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} \lambda = \lambda,
\]

hence \( B(\hat{\lambda}_{\text{mle}}) = \mathbb{E}[\hat{\lambda}_{\text{mle}}] - \lambda = 0 \). Furthermore

\[
\text{Var}(\hat{\lambda}_{\text{mle}}) = \text{Var} \left( \frac{1}{n} \sum_{i=1}^{n} X_i \right) = \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \lambda = \frac{\lambda}{n}.
\]

Hence, plugging these back in (1) yields \( \text{MSE}(\hat{\lambda}_{\text{mle}}) = \lambda/n \), so indeed \( \text{MSE}(\hat{\lambda}_{\text{mle}}) \to 0 \) as \( n \to \infty \) and we are done.

4. a) The insurer has to pay the claim in full by itself if the claim amount is not more than the retention level \( K = 10 \), that is

\[
\mathbb{P}(X \leq K) = \mathbb{P}(X \leq 10) = \int_{0}^{10} \alpha (1 + x)^{-\alpha-1} \, dx = -(1 + x)^{-\alpha} \bigg|_{x=10}^{x=0} = 1 - 11^{-\alpha}.
\]

b) Your records will contain 3 exactly observed claims (the ones less than \( K \)) and 4 claims of the amount \( K \) (where the original claim is larger than \( K \)). The likelihood function \( L \) for such a sample is hence given by:

\[
L(\alpha) = \mathbb{P}(X > K)^m \prod_{i=1}^{n} f_X(x_i | \alpha),
\]

where we know \( K = 10, m = 4, n = 3 \) and \( x_1 = 7, x_2 = 3 \) and \( x_3 = 9 \). We can compute

\[
\mathbb{P}(X > K) = \int_{K}^{\infty} f_X(x) \, dx = \int_{K}^{\infty} \frac{\alpha}{(1 + x)^{\alpha+1}} \, dx = -(1 + x)^{-\alpha} \bigg|_{x=\infty}^{x=K} = (1 + K)^{-\alpha}
\]

and hence

\[
L(\alpha) = (1 + K)^{-am} \alpha^n \prod_{i=1}^{n} (1 + x_i)^{-\alpha-1}. \tag{2}
\]

To maximise this function we may as usual consider the log likelihood rather:

\[
l(\alpha) = \log L(\alpha) = -am \log(1 + K) + n \log \alpha - (\alpha + 1) \sum_{i=1}^{n} \log(1 + x_i)
\]
and compute

\[ l'(\alpha) = -m \log(1 + K) + \frac{n}{\alpha} - \sum_{i=1}^{n} \log(1 + x_i) = 0 \iff \alpha = \frac{m \log(1 + K) + \sum_{i=1}^{n} \log(1 + x_i)}{n}. \]

We see that \( l(\alpha) \to -\infty \) as \( \alpha \downarrow 0 \) or \( \alpha \to \infty \), hence it follows that the above zero of \( l' \) indeed coincides with the mle, i.e. that

\[ \hat{\alpha}_{mle} = \frac{m \log(1 + K) + \sum_{i=1}^{n} \log(1 + x_i)}{n}. \]

Plugging the numbers we have, i.e. \( K = 10, \ m = 4, \ n = 3 \) and \( x_1 = 7, \ x_2 = 3 \) and \( x_3 = 9 \), this can be computed to be 0.20 (rounded to two decimals).

c) First we compute the cdf \( F_{\hat{Z}} \). You can either apply Lemma 3.3.22 from the lecture notes or compute it directly. For the latter, realise that \( \hat{Z} \) only takes a value if \( X > K \), and this value equals \( X - K \). Hence we find for \( z > 0 \):

\[ F_{\hat{Z}}(z) = \mathbb{P}(\hat{Z} \leq z) = \mathbb{P}(X \leq K + z \mid X > K) = \frac{\mathbb{P}(X \in (K, K + z])}{\mathbb{P}(X > K)} \]

where the final equality uses the definition of conditional probabilities. Using that

\[ \mathbb{P}(X \in (K, K + z]) = \int_{K}^{K+z} f_X(x) \, dx = -(1 + K + z)^{-\alpha} + (1 + K)^{-\alpha} \]

and the expression for \( \mathbb{P}(X > K) \) we already computed in a) it follows that

\[ F_{\hat{Z}}(z) = \frac{-(1 + K + z)^{-\alpha} + (1 + K)^{-\alpha}}{(1 + K)^{-\alpha}} = 1 - \frac{(1 + K + z)^{-\alpha}}{(1 + K)^{-\alpha}} = 1 - (1 + K)^{\alpha}(1 + K + z)^{-\alpha}. \]

Differentiating \( F_{\hat{Z}} \) to get the pdf \( f_{\hat{Z}} \) gives the required result.

d) For estimating \( \alpha \) by the method of moments given samples \( z_1, \ldots, z_n \) we need to solve \( \alpha \) from the equation \( \mathbb{E}[Z] = \bar{z} \) (only one equation needed as there is only one unknown), where \( \bar{z} = (1/n) \sum_{i=1}^{n} z_i \). So let us first compute \( \mathbb{E}[Z] \) under the assumption \( \alpha > 1 \). We get

\[ \mathbb{E}[Z] = \int_{0}^{\infty} z \alpha (1 + K)^{\alpha} (1 + K + z)^{-\alpha-1} \, dz \]

\[ = (1 + K)^{\alpha} \left( -z(1 + K + z)^{-\alpha}|_{z=0}^{z=\infty} + \int_{0}^{\infty} (1 + K + z)^{-\alpha} \, dz \right) \]

\[ = (1 + K)^{\alpha} \left( 0 + \frac{(1 + K)^{1-\alpha}}{\alpha - 1} \right) = \frac{1 + K}{\alpha - 1}, \]

where the third equation uses integration by parts. Note that it is crucial to have \( \alpha > 1 \).

Note furthermore that \( \alpha \mapsto \mathbb{E}[Z] \) is a bijection from \( (1, \infty) \) to \( \mathbb{R}_{>0} \), hence for any \( \bar{z} \in \mathbb{R}_{>0} \)
there is indeed a unique solution to the equation $\mathbb{E}[Z] = \bar{z}$. Plugging in the numbers that we have, namely $K = 10, n = 4$ and $z_1 = 15 - 10 = 5, z_2 = 11 - 10 = 1, z_3 = 21 - 10 = 11$ and $z_4 = 13 - 10 = 3$ we find the equation $11/((\alpha - 1) = (1/4) \cdot 20 = 5$ and hence we conclude that $\hat{\alpha}^{\text{mom}} = 16/5$.

5. a) We have

$$\mathbb{E}[Y] = \mathbb{E}[\min\{X, K\}] \overset{(1)}{=} \int_{\mathbb{R}} \min\{x, K\} f(x) \, dx \overset{(2)}{=} \int_{0}^{\infty} \min\{x, K\} f(x) \, dx$$

$$= \int_{0}^{K} x f(x) \, dx + \int_{K}^{\infty} K f(x) \, dx = \int_{0}^{\infty} x f(x) \, dx - \int_{K}^{\infty} x f(x) \, dx + \int_{K}^{\infty} K f(x) \, dx \overset{(3)}{=} \mathbb{E}[X] - \int_{K}^{\infty} (x - K) f(x) \, dx \overset{(4)}{=} \mathbb{E}[X] - \int_{K}^{\infty} u f(u + K) \, du,$$

where:

(1): here we use $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x) f(x) \, dx$ with $g(x) = \min\{x, K\}$,

(2): here we use that $X$ takes non-negative values, hence $f(x) = 0$ for $x < 0$,

(3): in the previous equation we have written things in such a way that we have the term $\int_{\mathbb{R}} f(x) \, dx = \mathbb{E}[X]$ present,

(4): here we use the substitution $u = x - K$.

b) Once the reinsurance contract is in force, the insurance company only has to pay the amount $Y = \min\{X, K\}$. Hence the goal of the insurance company, to reduce $\mathbb{E}[X]$ by at least 1, means that they should set $K$ such that $\mathbb{E}[Y] \leq \mathbb{E}[X] - 1$. Using the result from part a), this means that $K$ should be set such that

$$\int_{0}^{\infty} x f(x + K) \, dx \geq 1. \quad (3)$$

Now it is a matter of computing the integral in the lhs:

$$\int_{0}^{\infty} x f(x + K) \, dx = \int_{0}^{\infty} x(x + K)e^{-(x+K)} \, dx = e^{-K}\int_{0}^{\infty} x^2e^{-x} \, dx + e^{-K}K \int_{0}^{\infty} xe^{-x} \, dx = e^{-K/2} + e^{-K}K1 = e^{-K/(K+2)},$$

where the integrals are computed using integration by parts. Plugging this back in (3) indeed yields the result $e^{-K/(K+2)} \geq 1$.

6. a) Since we have $n$ values $x_i \in (0, K)$ and $m$ values $K$, the likelihood function $L$ is given by (cf. equation (3.20) in the notes)

$$L(c) = \mathbb{P}(X > K)^m \prod_{i=1}^{n} f(x_i | c),$$
where
\[ P(X > K) = \int_{K}^{\infty} f(x) \, dx = \int_{K}^{\infty} 2cx e^{-cx^2} \, dx = -e^{-cx^2} \bigg|_{x=\infty}^{x=K} = e^{-cK^2}. \]

In order to find \( \hat{c}_{\text{mle}} \), consider the log likelihood \( l \) given by
\[
l(c) = \log L(c) = m \log P(X > K) + \sum_{i=1}^{n} \log f(x_i|c) = -mcK^2 + \sum_{i=1}^{n} \log(2cx_i) - c \sum_{i=1}^{n} x_i^2
\]

hence
\[
l'(c) = -mK^2 + n - \sum_{i=1}^{n} x_i^2 = 0 \iff c = \frac{mK^2 + \sum_{i=1}^{n} x_i^2}{n}.
\]

Since \( l(c) \to -\infty \) as \( c \downarrow 0 \) and \( c \to \infty \) it follows that \( l \) attains its maximum in this value of \( c \), hence we conclude that
\[
\hat{c}_{\text{mle}} = \frac{mK^2 + \sum_{i=1}^{n} x_i^2}{n}.
\]

b) First Lemma 3.3.22 yields that the cdf \( F \) of \( \hat{Z} \) is given by
\[
F(z) = \frac{F_X(z + K) - F_X(K)}{1 - F_X(K)},
\]
where \( F_X \) is the cdf of \( X \). We can easily compute
\[
F_X(x) = \mathbb{P}(X \leq x) = \int_{0}^{x} 2cy e^{-cy^2} \, dy = -e^{-cy^2} \bigg|_{y=x}^{y=0} = 1 - e^{-cx^2}
\]

and hence we get
\[
F(z) = \frac{1 - e^{-c(z+K)^2} - (1 - e^{-cK^2})}{1 - (1 - e^{-cK^2})} = \frac{e^{-cK^2} - e^{-c(z+K)^2}}{e^{-cK^2}} = 1 - e^{-c(z+K)^2 + cK^2} = 1 - e^{-cz(z+2K)}.
\]

We can now obtain the pdf by differentiating \( F \) and this indeed yields the formula given.

7. Recall Chebychev’s inequality, i.e. for any rv \( Y \) with mean \( \mu \) and standard deviation \( \sigma := \sqrt{\text{Var}(Y)} \) and for any \( a > 0 \) we have
\[
\mathbb{P}(|Y - \mu| \geq a\sigma) \leq \frac{1}{a^2}.
\]

Now, suppose that we have \( \hat{\theta} \) such that \( \text{MSE}(\hat{\theta}) \to 0 \) as \( n \to \infty \). Recall that we may write
\[
\text{MSE}(\hat{\theta}) = B(\hat{\theta})^2 + \text{Var}(\hat{\theta}) = (\mathbb{E}[\hat{\theta}] - \theta)^2 + \text{Var}(\hat{\theta}),
\]
and since both terms on the rhs are positive it follows from \( \text{MSE}(\hat{\theta}) \to 0 \) that both

\[
\text{Var}(\hat{\theta}) \to 0 \quad \text{as } n \to \infty
\]  

and

\[
\mathbb{E}[\hat{\theta}] \to \theta \quad \text{as } n \to \infty.
\]

Take any \( \varepsilon > 0 \). Choose \( n_0 \) such that for all \( n \geq n_0 \) we have \( |\mathbb{E}[\hat{\theta}] - \theta| < \varepsilon / 2 \) (which is possible by (6)) then we find for all \( n \geq n_0 \):

\[
\mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) = \mathbb{P}(|\hat{\theta} - \mathbb{E}[\hat{\theta}] + \mathbb{E}[\hat{\theta}] - \theta| > \varepsilon) \leq \mathbb{P}(|\hat{\theta} - \mathbb{E}[\hat{\theta}]| + |\mathbb{E}[\hat{\theta}] - \theta| > \varepsilon) \overset{\text{choice of } n_0}{\leq} \mathbb{P}(|\hat{\theta} - \mathbb{E}[\hat{\theta}]| > \varepsilon/2) \leq \frac{4}{\varepsilon^2} \text{Var}(\hat{\theta})
\]

and indeed from (5) we see that the ultimate rhs tends to 0 as \( n \to \infty \), hence the proof is done.

An alternative proof is as follows. For any \( \varepsilon > 0 \) we can write:

\[
\text{MSE}(\hat{\theta}) = \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}\left[1_{|\hat{\theta} - \theta| \leq \varepsilon}(\hat{\theta} - \theta)^2\right] + \mathbb{E}\left[1_{|\hat{\theta} - \theta| > \varepsilon}(\hat{\theta} - \theta)^2\right].
\]

Now both terms on the rhs are positive, hence if \( \text{MSE}(\hat{\theta}) \to 0 \) it must be the case that both these terms tend to 0 as well. But for the second term we have

\[
\mathbb{E}\left[1_{|\hat{\theta} - \theta| > \varepsilon}(\hat{\theta} - \theta)^2\right] \geq \mathbb{E}\left[1_{|\hat{\theta} - \theta| > \varepsilon}\varepsilon^2\right] = \varepsilon^2 \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon).
\]

If we now let \( n \to \infty \) then the lhs tends to 0, hence also the rhs has to tend to 0. This implies that \( \mathbb{P}(|\hat{\theta} - \theta| > \varepsilon) \to 0 \) and hence we are done.