UNIVERSITY OF OSLO

Faculty of Mathematics and Natural Sciences

Examination in: STK4540 — Non-Life Insurance Mathematics

Day of examination: Tuesday, December 16th 2014

Examination hours: 09.00 – 13.00

This problem set consists of 6 pages.

Appendices: None

Permitted aids: Approved calculator

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Problem 1

In this problem (part d) and e)) you need to know the following result: Gaussian variables deviate less than 2 standard deviations from their mean with probability close to 95 %. This leads to the *Wald* criterion of pronouncing a parameter θ different from zero if $|\hat{\theta}/\hat{\sigma}| > 2$.

a) How is risk premium (pure premium) defined? Why is it important in general insurance? How do you estimate its components?

b) What model is often used for claim frequency? How can you check that the model is suitable? If the model is not suitable which other model can be chosen instead? What is the main conceptual difference between the two models?

Table 1 on page 2 shows estimates of claim intensity and claim size for an automobile insurance portfolio.

c) How do you interpret that the driving limit coefficients increase from left to right? Is this reasonable?

How much does the claim intensity increase in percent from 8 000 km in driving limit to 16 000 km in driving limit? Relate this increase to the increase in driving limit and speculate on the reason.

d) Apply the Wald criterion to the regression coefficients for claim frequency in Table 1 and examine whether some of them are different from zero.

Do the same exercise for the coefficients for claim size.

Do you draw the same conclusion for claim frequency and claim size?

(Continued on page 2.)

	Intercept	Age	
		≤ 26	> 26
Freq.	-2.43(0.08)	0(0)	-0.55(0.07)
Size	8.33(0.07)	0(0)	-0.36(0.07)

	Distance limit on policy (1000 km)					
	8	12	16	20	25-30	No limit
Freq.	0(0)	0.17(0.04)	0.28(0.04)	0.50(0.04)	0.62(0.05)	0.82(0.08)
Size	0(0)	0.02(0.04)	0.03(0.04)	0.09(0.04)	0.11(0.05)	0.11(0.08)

	Geographical regions with traffic density from *high to low*					
	Region 1	Region 2	Region 3	Region 4	Region 5	Region 6
Freq.	0(0)	-0.19(0.04)	-0.24(0.06)	-0.29(0.04)	-0.39(0.05)	-0.36(0.04)
Size	0(0)	-0.10(0.04)	-0.03(0.05)	-0.07(0.04)	-0.02(0.05)	0.06(0.08)

Table 1: *Estimated coefficients of claim intensity and claim size for automobile data (standard deviation in parenthesis).*

e) Imagine you have many variable candidates and you only want a few in your final model. All variable candidates are statistically significant. How can the Wald criterion be used to select which variables to include in your final model?

Problem 2

Assume policy risks $X_1, ..., X_J$ are stochastically independent with mean and variance for the portfolio payout $\mathcal{X}$ being

$$E(\mathcal{X}) = \pi_1 + ... + \pi_J \qquad \text{and} \qquad \text{var}(\mathcal{X}) = \sigma_1^2 + ... + \sigma_J^2 \tag{1}$$

where $\pi_j = E(X_j)$ and $\sigma_j = \text{sd}(X_j)$.

a) Use average portfolio expectation and average portfolio standard deviation to explain the concept of diversification.

Introduce $\xi(\mathbf{x}) = E(Y|\mathbf{x})$ and $\sigma(\mathbf{x}) = \text{sd}(Y|\mathbf{x})$. Consider the identities

$$E(Y) = E\{\xi(\mathbf{X})\} = E\{E(Y|\mathbf{X})\} \tag{2}$$

and

$$\text{var}(Y) = \text{var}\{\xi(\mathbf{X})\} + E\{\sigma(\mathbf{X})\}, \tag{3}$$

referred to as the rule of double expectation and double variance.

Let $\mathcal{X} = Z_1 + ... + Z_\mathcal{N}$ be the portfolio payout and assume that $\mathcal{N}, Z_1, Z_2, ...$ are stochastically independent. Let $E(Z_i) = \xi_Z$ and $\text{sd}(Z_i) = \sigma_Z$. Elementary rules for random sums imply

$$E(\mathcal{X}|\mathcal{N}) = \mathcal{N}\xi_Z \qquad \text{and} \qquad \text{var}(\mathcal{X}|\mathcal{N}) = \mathcal{N}\sigma_Z^2$$

(Continued on page 3.)

so that $Y = \mathcal{X}$ and $\mathbf{X} = \mathcal{N}$ in (2) and (3) yield

$$E(\mathcal{X}) = E(\mathcal{N})\xi_Z \qquad \text{and} \qquad var(\mathcal{X}) = E(\mathcal{N})\sigma_Z^2 + \text{var}(\mathcal{N})\xi_Z^2. \tag{4}$$

If $\mathcal{N}$ is Poisson distributed so that $E(\mathcal{N})$ =var$(\mathcal{N}) = J\mu T$, then

$$E(\mathcal{X}) = J\mu T\xi_Z \qquad \text{and} \qquad \text{var}(\mathcal{X}) = J\mu T(\sigma_Z^2 + \xi_Z^2). \tag{5}$$

Suppose each X_j depends on some underlying ω_j. Suppose all pairs (ω_j, X_j) follow the same distribution and let $X_1, ..., X_J$ be conditionally independent given $\omega_1, ..., \omega_J$.

The sampling regimes may be two different ones:

$$\omega_1 = \cdots = \omega_J = \omega \qquad \text{and} \qquad \omega_1, ..., \omega_J \text{ independent.}$$

On the left, ω is a common background factor affecting the entire portfolio whereas on the right, ω_j is attached to each X_j individually.

Consider the case where ω is common background. We are assuming $\xi(\omega) = E(X_j|\omega)$ and $\sigma(\omega) = \text{sd }(X_j|\omega)$, the same for all j.

b) Calculate *the coefficient of variation*, $\frac{\text{sd}(\mathcal{X})}{E(\mathcal{X})}$, and comment on the implications for diversification. (*Hint: Find $E(\mathcal{X}|\omega)$ and $var(\mathcal{X}|\omega)$. Then, use the double rules to find $var(\mathcal{X})$ and $E(\mathcal{X})$*)

Assume now that each X_j is attached to a separate and independently drawn ω_j.

c) Calculate *the coefficient of variation*, $\frac{\text{sd}(\mathcal{X})}{E(\mathcal{X})}$ in this case, and comment on the implications for diversification. (*Hint: Find the mean and variance of each X_j by inserting $J = 1$ in the formulas for $E(\mathcal{X})$ and $sd(\mathcal{X})$ from part b). Then add these over all j for the mean and variance of $\mathcal{X}$*).

The preceding argument enables understanding of how random intensities $\mu_1, ..., \mu_J$ in general insurance influence portfolio claim numbers $\mathcal{N} = N_1 + ... + N_J$ and payoffs $\mathcal{X} = Z_1 + ... + Z_{\mathcal{N}}$. Consider the sampling regimes

$$\mu_1 = \cdots = \mu_J = \mu \qquad \text{and} \qquad \mu_1, ..., \mu_J \text{ independent.}$$

where μ on the left affects all policy holders jointly whereas there is one μ_j for each individual on the right. Claim numbers are in either case conditionally independent given $\mu_1, ..., \mu_J$. To understand how the uncertainty of $\mathcal{N}$ is linked to common or individual sampling we may invoke the expressions for $E(\mathcal{X})$ and sd$(\mathcal{X})$ derived in part b) and c). The functions $\xi(\omega)$ and $\sigma(\omega)$ are now $\xi(\mu) = \mu T$ and $\sigma(\mu) = \sqrt{\mu T}$ assuming the usual Poisson model. This implies $E(\mathcal{N}) = JT\xi_\mu$.

d) Calculate sd$(\mathcal{N})$ when μ is common and when μ is individual.

How these results are passed on to $\mathcal{X}$ may be examined by inserting the expression for $E(\mathcal{N})$ and sd$(\mathcal{N})$ into (5). This leads to $E(\mathcal{X}) = JT\xi_\mu\xi_Z$.

Calculate sd$(\mathcal{X})$ when μ is common and when μ is individual.

(Continued on page 4.)

Problem 3

The *over-threshold* model is important when the extreme right tail of a claim size distribution is analyzed. The over-threshold model is the distribution of $Z_b = Z - b$ given $Z > b$. Its density function is

$$f_b(z) = \frac{f(b+z)}{1-F(b)}, \; z > 0.$$

The density function for the *Pareto* model is

$$f(z) = \frac{\alpha/\beta}{(1+z/\beta)^{1+\alpha}}, \; z > 0.$$

Over-threshold distributions for the *Pareto* model is also Pareto. These over-threshold distributions preserve the Pareto model and its shape. The mean (if it exists) is known as the *the mean excess* function and becomes

$$E(Z_b|Z > b) = \frac{\beta + b}{\alpha - 1} = \xi + \frac{b}{\alpha - 1} \text{ (requires } \alpha > 1) \tag{6}$$

where $\xi = E(Z)$.
The *sample mean excess function* is defined as

$$e_n(b) = \frac{\sum_{i=1}^{n}(X_i - b)\mathcal{I}(X_i > b)}{\sum_{i=1}^{n}\mathcal{I}(X_i > b)}, \tag{7}$$

where $\mathcal{I}(X_i > b)$ is the indicator function such that $\mathcal{I}(X_i > b) = 1$ if $(X_i > b)$ and 0 otherwise. To view this function the *sample mean excess plot* is constructed

$$\{(X_{i,n}, e_n(X_{i,n})) \; : \; 2 \le i \le n\}, \tag{8}$$

where $X_{i,n}$ denotes the i-th order statistic.

a) If the data is Pareto over a high treshold what shape would this plot have? (*Hint: how does $E(Z_b|Z > b)$ in equation (6) above change when b changes?*)

b) What does the result of Pickands say about over-threshold distributions in general?

c) Why can it be difficult to use Pickands' result in practice?

Let $Y_b = (Z - b)/\beta_b$ given $Z > b$, where $\beta_b > 0$, is some parameter.

It can be shown that $\Pr(Y_b > y) = \bar{F}(b+y\beta_b)/\bar{F}(b)$ where $\bar{F}(z) = \Pr(Z \ge z)$. l'Hôpital's rule can also be used to argue that the limit of $\Pr(Y_b > y)$ as $b \to \infty$ coincides with the limit of

$$L_b(y) = \frac{f(b+y\beta_b)}{f(b)}(1 + y\frac{d\beta_b}{db}) \tag{9}$$

where $f(z)$ is the density function of Z.

(Continued on page 5.)

d) Insert $f(z) = cz^{\theta-1}/(1+z/\beta)^{\alpha+\theta}$ into the function $L_b(y)$ in (9) and show that

$$L_b(y) = (1 + \frac{y\beta_b}{b})^{\theta-1}(1 + \frac{y\beta_b}{\beta+b})^{-(\alpha+\theta)}(1 + y\frac{d\beta_b}{db}). \tag{10}$$

e) If $\beta_b = \beta + b$, show that $L_b(y) \to (1+y)^{-\alpha}$ as $b \to \infty$, and over-threshold distributions of extended Pareto variables belong to the Pareto family when b is large.

Problem 4

Solvency is an important task in general insurance. Solvency is the financial control of liabilities under nearly worst-case scnearios. The target is the *reserve*. Portfolio liabilities can estimated using simple approximation. The portfolio loss $\mathcal{X}$ for independent risks becomes Gaussian as the number of policies $J \to \infty$. This is a consequence of the central limit theorem. In this problem (partb), c) and d)) you are asked to develop some programs. You may write sketches of programs (pseudo code) or R code.

a) Assume that $E(\mathcal{X}) = 121$ million NOK, $\mathrm{sd}(\mathcal{X}) = 20$ million NOK and that $\phi_\epsilon = 1.645$ is the upper 5% percentile of the standard normal distribution. What is the reserve estimated by simple approximation in this case?

A *mixed distribution* is often used in extreme value modelling. In this model non-parametric sampling is used for all claims that do not exceed a threshold b.

Suppose the original claims $z_1, ..., z_n$ are ranked in ascending order as $z_{(1)} \leq \cdots \leq z_{(n)}$ so that observations from $z_{(n_1)}$ and smaller are below the threshold. Choose some small probability p and let $n_1 = n(1-p)$ and $b = z_{(n_1)}$. Then take

$$Z_{\leq b} = \hat{Z} \text{ and } Z_{>b} = z_{(n_1)} + \text{ Pareto}(\alpha, \beta), \tag{11}$$

where $\hat{Z}$ is the empirical distribution over $z_{(1)}, ..., z_{(n_1)}$, i.e.,

$$\Pr(\hat{Z} = z_{(i)}) = \frac{1}{n_1}, \; i = 1, ..., n. \tag{12}$$

b) Write a program that carrys out m simulations of the claim size under this model.

(*Hint: To sample from the Pareto distribution use the fact that the cumulative distribution of a Pareto variable is* $F(x) = 1 - \frac{1}{(1+x/\beta)^\alpha}$. *Then use the inversion principle* ($X = F^{-1}(U)$) *where* $U \sim Uniform[0,1]$ *to find out how* X *should be sampled in the tail. You can assume that* α *and* β *are estimated from likelihood estimation (you do not have to specify this in the code).*)

(Continued on page 6.)

c) Write a program that simulates the portfolio liabilities. For the number of claims you can assume a Poisson distribution with parameter λ. For the claim size use the code from part b).

d) Assume that the insurer buys an $a \times b$ contract drawn up on single events.

Modify the program in part c) so that this reinsurance contract is reflected in the portfolio liability.

THE END