Advanced Methods of Non-Life Insurance

Prof. dr. Tim Verdonck

FORMULARY

PART A: FOUNDATIONS OF LOSS DATA ANALYTICS

Introduction non-life insurance

- **Expectation**
  - **Definition:** Assume rv $X \sim F$ and $h : \mathbb{R} \to \mathbb{R}$ a sufficiently nice measurable function. The expected value of $h(X)$ is

$$E[h(X)] = \int_{\mathbb{R}} h(x)dF(x) = \begin{cases} \sum_{k \in \mathcal{A}} h(x)f(x) & X \text{ discrete} \\ \\ \int_{-\infty}^{\infty} h(x)f(x)dx & X \text{ continuous} \end{cases}$$

    provided the right side converges absolutely.

  - **Mean**, expectation or first moment of $X \sim F$

$$\mu_X = E[X] = \int_{\mathbb{R}} xdF(x)$$

  - **Moment of order** k : measure of variation around 0

$$\alpha_k = E(X^k) = \begin{cases} \sum_x x^k f(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} x^k f(x)dx & X \text{ continuous} \end{cases}$$

    If $E[X^k]$ exists ($E[|X|^k] < \infty$), all lower order moments $E[X^l]$, $l \leq k$ exist.

  - **Central moment order** k: measure of variation around μ

$$\mu_k = E\left[(X-\mu)^k\right] = \begin{cases} \sum_x (x-\mu)^k f(x) & X \text{ discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^k f(x)dx & X \text{ continuous} \end{cases}$$

  - **Variance**: second central moment

$$\sigma_X^2 = \text{var}(X) = E\left[(X-\mu)^2\right]$$

    and its square root σ is called the standard deviation.

- **Coefficient of variation** is ratio of σ to μ

$$\text{Vco}(X) = \frac{\sigma_X}{\mu_X}$$

- **Skewness**: if positive (negative), then long right (left) tail

$$\zeta_X = \frac{\mu_3}{\sigma^3} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = E[Z^3].$$

where $Z = \frac{X-\mu}{\sigma}$ is the so-called standardized rv.

- **(Excess) Kurtosis**: measure of peakedness (or thickness in the tails)

$$\gamma_X = \frac{\mu_4}{\sigma_4} = E[Z^4] - 3.$$

-3 sets $\gamma_X = 0$ for the normal distribution

- **Weak law of large numbers (LLN):** Suppose:
  - Probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - $Y_1, \ldots, Y_n$ uncorrelated and identically distributed
  - Finite mean $\mu = E(Y_1)$

  then the weak law of large numbers is:

$$\forall \epsilon > 0 \lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i - \mu\right| \geq \epsilon\right) = 0 \quad \text{or} \quad \overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \xrightarrow{P} \mu.$$

- **Central limit theorem (CLT):** Suppose:
  - Probability space $(\Omega, \mathcal{F}, \mathbb{P})$
  - $Y_1, \ldots, Y_n$ independent and identically distributed (iid)
  - Finite mean $\mu = E(Y_1)$ and finite variance $\sigma^2 = \text{var}(Y_1)$

  then the central limit theorem provides the asymptotic limit distribution (convergence in distribution) as $n \to \infty$:

$$\frac{\sum_{i=1}^{n} Y_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} Z \sim \mathcal{N}(0,1) \quad \text{or} \quad P\left[\left(\frac{\sum_{i=1}^{n} Y_i - n\mu}{\sqrt{n}\sigma}\right) \leq x\right] \xrightarrow{D} \Phi(x).$$

  Practically, sums of rv's can often be approximated by those from normal distribution

$$\sum_{i=1}^{n} Y_i \approx N\left(n\mu, n\sigma^2\right) \quad \text{or} \quad \overline{Y} \approx N\left(\mu, \frac{1}{n}\sigma^2\right)$$

- **Moment generating function (mgf) of** X
  - **Definition:**
$$M(r) = M_X(r) = E(e^{rX}) = \int_{\mathbb{R}} e^{rx} dF(x)$$
  - Alternative: **characteristic function:**
$$\varphi_X(r) = E[e^{irX}].$$
  - **Lemma:** Choose $X \sim F$ and assume that there exists $r_0 > 0$ such that $\forall r \in (-r_0, r_0)$: $M_X(r) < \infty$, then $M_X(r)$ has power series expansion for $r \in (-r_0, r_0)$ with
$$M_X(r) = \sum_{k \geq 0} \frac{r^k}{k!} E[X^k]$$
  - **Property:** The derivatives at the origin are given by
$$\frac{d^k}{dr^k} M_X(r)|_{r=0} = E(X^r) < \infty$$
- **Cumulant generating function (cgf)**
  - **Definition:**
$$\log M_X(r) = \log E[e^{rX}]$$
  - **Lemma:** Assume that M_X is finite on $(-r_0, r_0)$ with $r_0 > 0$. Then $\log M_X(.)$ is a convex function on $(-r_0, r_0)$.
- **Survival function**
$$\overline{F}(x) = 1 - F(X) = \Pr[X > x]$$
- **Property:** Assume that $X \sim F$ is non-negative, $\mathbb{P}$-a.s. and has finite first moment. Then
$$E(X) = \int_0^\infty x dF(x) = \int_0^\infty [1 - F(x)] dx = \int_0^\infty \mathbb{P}[X > x] dx = \int_0^\infty \overline{F}(x) dx$$
Moreover, the second moment $E[X^2]$ (when it exists) is
$$E[X^2] = \int_0^{+\infty} x^2 f_X(x) dx = \int_0^{+\infty} 2x\overline{F}(x) dx$$
- **Tower property or double expectation theorem:** For any sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$ on our probability space $(\Omega, \mathcal{F}, \Pr)$ we have for any integrable rv $X \sim F$
$$E[X] = E[E[X|\mathcal{G}]]$$
In particular, if X and Y are rv's on $(\Omega, \mathcal{F}, \mathbb{P})$ we have
$$E[X] = E[E[X|Y]]$$
where $E[E[X|Y]]$ is an abbreviation for $E[X|\sigma(Y)]$ with $\sigma(Y) \subset \mathcal{F}$ denoting the σ-algebra generated by the random variable Y.
Assume that X is square integrable then the tower proper implies
$$\text{var}(X) = E[\text{var}(X|\mathcal{G})] + \text{var}(E[X|\mathcal{G}]).$$

- **Inverse of F or p-quantile of $X \sim F$:** Let F be right-continuous and non-decreasing. The **generalized inverse** of F for $p \in (0,1)$ is then

$$F^{\leftarrow}(p) = inf\{x; F(x) \geq p\}$$

where $inf(\emptyset) = \infty$.

Properties:

1. $F^{\leftarrow}(p)$ is non-decreasing and left-continuous.
2. F is continuous iff $F^{\leftarrow}(p)$ is strictly increasing.
3. F is strictly increasing iff $F^{\leftarrow}(p)$ is continuous.
4. (If F is right-continuous then) $F(x) \geq p$ iff $F^{\leftarrow}(p) \leq x$.
5. $F^{\leftarrow}(F(x)) \leq x$.
6. $F(F^{\leftarrow}(p)) \geq p$.
7. If F is strictly increasing, then $F^{\leftarrow}(F(x)) = x$.
8. If F is continuous, then $F(F^{\leftarrow}(p)) = p$.

- **Popular distributions**

  - Normal distribution: $X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu \in \mathbb{R}, \sigma^2 > 0$

$$\begin{aligned} f_X(x) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad x \in \mathbb{R} \\ M_X(r) &= e^{r\mu + r^2\sigma^2/2} < \infty \qquad r \in \mathbb{R} \\ E[X] &= \mu \\ \text{var}[X] &= \sigma^2 \end{aligned}$$

  - Exponential distribution: $X \sim \text{Exp}(\lambda)$ with $\lambda > 0$

$$\begin{aligned} f_X(x) &= \lambda e^{-\lambda x} \qquad x \in [0, +\infty[\\ M_X(r) &= \frac{\lambda}{\lambda - r} \qquad r < \lambda \\ E[X] &= \frac{1}{\lambda} \\ \text{var}[X] &= \frac{1}{\lambda^2} \end{aligned}$$

  - Geometric distribution: $X \sim \text{Geo}(p)$ with $0 < p \leq 1$

$$\begin{aligned} f_X(k) &= (1-p)^k p \qquad k \in \{0, 1, 2, 3, \ldots\} \\ M_X(r) &= \frac{p}{1-(1-p)e^r} \\ E[X] &= \frac{1-p}{p} \\ \text{var}[X] &= \frac{1-p}{p^2} \end{aligned}$$

Introduction aggregate loss modelling

- **Compound distribution**
  - **Model assumptions**
    The total claim amount S is given by the following compound distribution

$$S = Y_1 + \ldots + Y_N = \sum_{i=1}^{N} Y_i$$

    with 3 standard assumptions

    1. N is a discrete rv which only takes values in $\mathcal{A} \subset \mathbb{N}_0$
    2. $Y_1, Y_2, \ldots \overset{iid}{\sim} G$ with $G(0) = 0$.
    3. N and $(Y_1, Y_2, \ldots)$ are independent.

  - **Basic recognition features**
    Assume S has a compound distribution. We have (whenever they exist)

$$\begin{aligned} E[S] &= E[N]E[Y_1] \\ \mathrm{var}(S) &= \mathrm{var}(N)E[Y_1]^2 + E[N]\,\mathrm{var}(Y_1) \\ \mathrm{Vco}(S) &= \sqrt{\mathrm{Vco}(N)^2 + \frac{1}{E[N]}\mathrm{Vco}(Y_1)^2} \\ M_S(r) &= M_N\left(\log(M_{Y_1(r)})\right) \qquad \text{for } r \in \mathbb{R}. \end{aligned}$$

  - If assumptions above hold, then distribution function of S can be written as

$$\begin{aligned} F_S(x) &= \Pr[S \le x] = \sum_{k\in\mathcal{A}} \Pr\left[\sum_{i=1}^{N} Y_i \le x \middle| N = k\right] \Pr\left[N = k\right] \\ &= \sum_{k\in\mathcal{A}} \Pr\left[\sum_{i=1}^{k} Y_i \le x\right] \Pr\left[N = k\right] = \sum_{k\in\mathcal{A}} G^{*k}(x) \Pr\left[N = k\right] \end{aligned}$$

    G^{*k} denotes the k-th convolution of the distribution function G.
    In particular, we have for $Y_1, Y_2 \overset{iid}{\sim} G$

$$\begin{aligned} G^{*2}(x) &= \Pr[Y_1 + Y_2 \le x] = \int G(x-y)dG(y) \\ G^{*k}(x) &= \int G^{*(k-1)}(x)dG(y) \end{aligned}$$

Modelling loss frequency

- **Binomial distribution**
  - We choose fixed volume $v \in \mathbb{N}$ and fixed default probability $p \in (0,1)$. We say N has a binomial distribution $N \sim \mathrm{Binom}(v, p)$ if

$$p_k = \Pr(N = k) = \binom{v}{k} p^k (1-p)^{v-k} \qquad \forall k \in \{0, \ldots, v\} = \mathcal{A}$$

- Assume $N \sim \text{Binom}(v, p)$ for fixed $v \in \mathbb{N}$ and $p \in (0, 1)$

$$\begin{aligned} E[N] &= vp \\ \text{var}(N) &= vp(1-p) \\ \text{Vco}(N) &= \sqrt{\frac{1-p}{vp}} \\ M_N(r) &= (pe^r + (1-p))^v \qquad \forall r \in \mathbb{R} \end{aligned}$$

- **Second characterisation of binomial distribution:**
  Assume that $N \sim \text{Binom}(v, p)$ with given $v \in \mathbb{N}$ and $p \in (0, 1)$.
  Choose $X_1, \ldots, X_v \overset{iid}{\sim} \text{Bernouilli}(p)$. Then we have

$$N \overset{(d)}{=} \sum_{i=1}^{v} X_i$$

- **Poisson distribution**
  - We choose fixed volume $v > 0$ and fixed expected claims frequency $\lambda > 0$. N has Poisson distribution $N \sim Poi(\lambda v)$, if

$$p_k = \Pr[N = k] = e^{-\lambda v} \frac{(\lambda v)^k}{k!} \qquad \forall k \in \mathcal{A} = \mathbb{N}$$

  - Assume $N \sim Poi(\lambda v)$ for fixed $\lambda, v > 0$. Then

$$\begin{aligned} E[N] &= \lambda v \\ \text{var}(N) &= \lambda v \\ \text{Vco}(N) &= \sqrt{\frac{1}{\lambda v}} \\ M_N(r) &= e^{\lambda v(e^r - 1)} \qquad \forall r \in \mathbb{R} \end{aligned}$$

- **Mixed poisson distribution**
  - Assume $\Lambda \sim H$ with $H(0) = 0, E[\Lambda] = \lambda$ and $\text{var}(\Lambda) > 0$.
  - Conditionally, given Λ, $N \sim Poi(\Lambda v)$ for fixed volume $v > 0$.
  - If N satisfies this definition, then we have $E[N] < \text{var}(N)$.
- **Gamma distribution**
  $X \sim \Gamma(\gamma, c)$ with shape parameter $\gamma > 0$ and scale parameter $c > 0$ if X is non-negative, absolutely continuous rv with density

$$f(x) = \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-cx}$$

  with Gamma function $\Gamma(.)$ is defined as

$$\Gamma(\gamma) = \int_0^\infty x^{\gamma-1} e^{-x} dx \qquad (\gamma > 0)$$

  - $\Gamma(\gamma + 1) = \gamma\Gamma(\gamma)$

- $\Gamma(1) = \Gamma(2) = 1 \qquad \Gamma(\frac{1}{2}) = \sqrt(\pi)$
- $\Gamma(n) = (n-1)!$

- **Negative-binomial distribution**
  - **Definition:** $X \sim \text{NegBin}(\lambda v, \gamma)$ with volume $v > 0$, expected claims frequency $\lambda > 0$ and dispersion parameter $\gamma > 0$ if
    - $\Theta \sim \Gamma(\gamma, \gamma)$
    - Conditionally, given Θ, $N \sim Poi(\Theta \lambda v)$
  - **Second definition:** Negative-binomial distribution satisfies for $k \in \mathcal{A} = \mathbb{N}_0$

$$p_k = \Pr[N = k] = \binom{k + \gamma - 1}{k} (1-p)^\gamma p^k$$

    where we choose $p = (\lambda v)/(\gamma + \lambda v) \in (0, 1)$
  - **Proposition:** Assume $N \sim \text{NegBin}(\lambda v, \gamma)$ for fixed $\lambda, v, \gamma > 0$. Then

$$\begin{aligned} E[N] &= \lambda v \\ \text{var}(N) &= \lambda v (1 + \lambda v/\gamma) > \lambda v \\ \text{Vco}(N) &= \sqrt{\frac{1}{\lambda v}} \sqrt{1 + \lambda v/\gamma} > \gamma^{-1/2} > 0 \\ M_N(r) &= \left(\frac{1-p}{1-pe^r}\right)^\gamma \qquad \forall r < -\log(p) \end{aligned}$$

    where $p = (\lambda v)/(\gamma + \lambda v) \in (0, 1)$.

- **$(a, b, 0)$ class**
  - N belongs to $(a, b, 0)$ class (or is a Panjer distribution) if there exist constants $a, b \in \mathbb{R}$ such that for all $k = 1, 2, 3, \ldots$ we have the recursion

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right)$$

  - **Lemma:** Assume N has a non-degenerate Panjer distribution or belongs to $(a, b, 0)$ class. N is either binomially, Poisson or negative-binomially distributed.

- **$(a, b, 1)$ class**
  A count distribution belongs to $(a, b, 1)$ class if there exist constants $a, b \in \mathbb{R}$ such that for all $k = 2, 3, \ldots$ the probabilities p_k satisfy

$$p_k = p_{k-1}\left(a + \frac{b}{k}\right)$$

- **Zero truncation or modification**
  - Consider p_k^0 to be a probability for this member of $(a, b, 0)$.
  - Let p_k^M be the corresponding probability for a member of $(a, b, 1)$ where M stands for *modified*.
  - Pick a new probability of a zero claim, p_0^M, and define $c = \frac{1-p_0^M}{1-p_0^0}$.

- We then calculate zero modified distribution as $p_k^M = c p_k^0$.
- Note that $\sum_{k=0}^{\infty} p_k^M = 1$!

Assume that $p_0^M = 0$, so that probability of $N = 0$ is zero (*truncated at zero*). Then we get zero truncated probabilities (where we use T instead of M now):

$$p_k^T = \begin{cases} 0 & k = 0 \\ \frac{1}{1-p_0^0} p_k^0 & k \geq 1 \end{cases}$$

Modelling loss severity

- **Empirical distribution function**

$$\hat{G}_n(y) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{Y_i \leq y\}}$$

- **Loss size index function and empirical version**

$$\mathcal{I}(G(y)) = \frac{\int_0^y z dG(z)}{\int_0^\infty z dG(z)} \quad \text{and} \quad \hat{\mathcal{I}}_n(\alpha) = \frac{\sum_{n=1}^{\lfloor n\alpha \rfloor} Y_{(i)}}{\sum_{i=1}^{n} Y_i} \quad \text{for } \alpha \in [0,1]$$

- **Tail analysis**
  Assume that G has **infinite support** and that $\overline{G} = 1 - G$ is survival function.
  - $\overline{G} \in \mathcal{R}_{-\alpha}$: $\overline{G} = 1 - G$ is **regularly varying at infinity** with (tail) index $\alpha > 0$ if

$$\lim_{x \to \infty} \frac{\overline{G}(xt)}{\overline{G}(x)} = \lim_{x \to \infty} \frac{1 - G(xt)}{1 - G(x)} = t^{-\alpha} \qquad \forall t > 0$$

  - If the above holds true for $\alpha = 0$ then $\overline{G} \in \mathcal{R}_0$.
  - If the above holds true for $\alpha = \infty$ then $\overline{G}$ is **rapidly varying at infinity**: $\overline{G} \in \mathcal{R}_{-\infty}$.
  - **(empirical) mean excess plot**

$$u \mapsto e(u) = E[Y_i - u | Y > u] \quad \text{and} \quad u \mapsto \hat{e}_n(u) = \frac{\sum_{i=1}^{n} (Y_i - u) 1_{\{Y_i > u\}}}{\sum_{i=1}^{n} 1_{\{Y_i > u\}}}$$

  - **(empirical) log-log plot**

$$y \mapsto (\log y, \log(1 - G(y))) \quad \text{and} \quad y \mapsto \left(\log y, \log(1 - \hat{G}_n(y)\right)$$

- **Parametric claim size distributions** We use following notation for rv $Y \sim G$:
  - g: **density** of Y for G absolutely continuous
  - $M_Y(r)$: **moment generating function** of Y in $r \in \mathbb{R}$, where it exists
  - μ_Y: **expected value** of Y, if it exists
  - σ_Y^2: **variance** of Y, if it exists
  - $\mathrm{Vco}(Y)$: **coefficient of variation** of Y, if it exists

– ζ_Y: **skewness** of Y, if it exists

– $\overline{G} = 1 - G$: **survival function** of Y, i.e. $\overline{G}(y) = \Pr[Y > y]$

For the analysis of G also following quantities are of interest

– $E[Y 1_{\{u_1 < Y \le u_2\}}]$: **expected value of Y within layer** $(u_1, u_2]$

– $I\left(G(y)\right) = E[Y 1_{\{Y \le y\}}]/\mu_Y$: **loss size index function** for level y

– $e(u) = E[Y - u | Y > u]$: **mean excess function** of Y above u

- **Gamma distribution**

  – $Y \sim \Gamma(\gamma, c)$: Gamma distribution with shape parameter $\gamma > 0$ and scale parameter $c > 0$

$$g(y) = \frac{c^\gamma}{\Gamma(\gamma)} y^{\gamma-1} e^{-cy} \qquad \text{for } y \ge 0$$

  – No closed form solution for distribution function G

$$G(y) = \int_0^y \frac{c^\gamma}{\Gamma(\gamma)} x^{\gamma-1} e^{-cx} dx = \frac{1}{\Gamma(\gamma)} \int_0^{cy} z^{\gamma-1} e^{-z} dz = \mathcal{G}(\gamma, cy) \qquad y \ge 0$$

  where $\mathcal{G}(.,.)$ is incomplete gamma function.

  – Family of gamma distributions is closed towards multiplication with positive constant, that is, for $\rho > 0$ we have

$$\rho Y \sim \Gamma(\gamma, c/\rho)$$

  – For mfg and moments we have

$$\begin{aligned} M_Y(r) &= \left(\frac{c}{c-r}\right)^\gamma \qquad \text{for } r < c \\ \mu_Y &= \frac{\gamma}{c} \\ \sigma_Y^2 &= \frac{\gamma}{c^2} \\ \mathrm{Vco}(Y) &= \gamma^{-1/2} \\ \zeta_Y &= 2\gamma^{-1/2} > 0 \end{aligned}$$

  – For $0 \le u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned} E[Y 1_{\{u_1 < Y \le u_2\}}] &= \frac{\gamma}{c} \left[\mathcal{G}(\gamma+1, cu_2) - \mathcal{G}(\gamma+1, cu_1)\right] \\ I(G(y)) &= \mathcal{G}(\gamma+1, cy) \\ e(u) &= \frac{\gamma}{c} \left(\frac{1 - \mathcal{G}(\gamma+1, cu)}{1 - \mathcal{G}(\gamma, cu)}\right) - u \end{aligned}$$

  – Gamma distribution does not have a regularly varying tail at infinity. In fact, $\overline{\mathrm{G}}(y) = 1 - G(y)$ decays roughly as e^{-cy} to 0 as $y \to \infty$ because e^{-cy} gives asymptotic lower bound and $e^{-(c-\epsilon)y}$ as an asymptotic upper bound for any $\epsilon > 0$ on $\overline{\mathrm{G}}(y)$.

  – Method of moment estimators are given by

$$\hat{c}^{MM} = \frac{\hat{\mu}_n}{\hat{\sigma}_n^2} \quad \text{and} \quad \hat{\gamma}^{MM} = \frac{\hat{\mu}_n^2}{\hat{\sigma}_n^2}$$

- For MLE we have log-likelihood function, set $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$

$$\ell_{\boldsymbol{Y}}(\gamma, c) = \sum_{i=1}^{n} \gamma \log c - \log \Gamma(\gamma) + (\gamma - 1) \log Y_i - cY_i$$

  Then MLE $\hat{\gamma}^{MLE}$ of γ is solution

$$\log \gamma - \log \hat{\mu}_n - \frac{\Gamma'(\gamma)}{\Gamma(\gamma)} + \frac{1}{n} \sum_{i=1}^{n} \log Y_i = 0$$

  This is solved numerically and MLE for c is then given by

$$\hat{c}^{MLE} = \frac{\hat{\gamma}^{MLE}}{\hat{\mu}_n}$$

- **Weibull distribution**
  - $Y \sim \text{Weibull}(\tau, c)$ Weibull distributed with shape parameter $\tau > 0$ and scale parameter $c > 0$

$$g(y) = (c\tau)(cy)^{\tau-1} e^{-(cy)^\tau}$$

  - Survival function does not have regularly varying tail at infinity, but decay of

$$G(y) = 1 - e^{-(cy)^\tau} \qquad \text{for } y \geq 0$$

    is slower than in gamma case for $\tau < 1$. In fact $\overline{\mathrm{G}}(y) = 1 - G(y)$ decays as $e^{-(cy)^\tau}$ to 0 for $y \to \infty$.
  - Family of Weibull distributions is closed towards multiplication with $\rho > 0$

$$\rho Y \sim \text{Weibull}(\tau, c/\rho)$$

  - Mgf does not exist for $\tau < 1$ and $r > 0$ and moments are

$$\begin{aligned} \mu_Y(r) &= \frac{\Gamma(1+1/\tau)}{c} \\ \sigma_Y^2 &= \frac{\Gamma(1+2/\tau)}{c^2} - \mu_Y^2 \\ \zeta_Y &= \frac{1}{\sigma_Y^3}\left[\frac{\Gamma(1+3/\tau)}{c^3} - 3\mu_Y\sigma_Y^2 - \mu_Y^3\right] \end{aligned}$$

  - For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned} E[Y 1_{u_1 < Y \leq u_2}] &= \frac{\Gamma(1+1/\tau)}{c}\left[\mathcal{G}(1+1/\tau, (cu_2)^\tau) - \mathcal{G}(1+1/\tau, (cu_1)^\tau)\right] \\ I(G(y)) &= \mathcal{G}(1+1/\tau, (cy)^\tau) \\ e(u) &= \frac{\Gamma(1+1/\tau)}{c}\left(\frac{1 - \mathcal{G}(1+1/tau, (cu)^\tau)}{e^{-(cu)^\tau}}\right) - u \end{aligned}$$

  - Generating Weibull random numbers by observing that we have identity $Y \stackrel{(d)}{=} \frac{1}{c} Z^{1/\tau}$ with $Z \sim \exp(1) \stackrel{(d)}{=} \Gamma(1,1)$: `rgamma(n,shape=1,rate=1)`

– Method of moment estimators are given by

$$\begin{aligned}\hat{c}^{MM} &= \frac{\Gamma(1+1/\hat{\tau}^M M)}{\hat{\mu}_n} \\ \frac{\hat{\sigma}_n^2}{\hat{\mu}_n^2}+1 &= \frac{1+2/\hat{\tau}^{MM}}{\Gamma(1+1/\hat{\tau}^M M)^2}\end{aligned}$$

which needs to be solved numerically in `R`

– For MLE we need to solve system of equations

$$\begin{aligned}c &= \left(\frac{1}{n}\sum_{i=1}^n Y_i^\tau\right)^{-1/\tau} \\ \tau\frac{1}{n}\sum_{i=1}^n \log(cY_i)((cY_i)^\tau-1) &= 1\end{aligned}$$

- **Log-normal distribution**

– $Y \sim \mathrm{LN}(\mu, \sigma^2)$ log-normal distributed with mean parameter $\mu \in \mathbb{R}$ and standard deviation parameter $\sigma > 0$

$$\begin{aligned}g(y) &= \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{y}e^{-\frac{1}{2}\frac{(\log y-\mu)^2}{\sigma^2}} \qquad \text{for } y \geq 0 \\ G(y) &= \Phi\left(\frac{\log y-\mu}{\sigma}\right)\end{aligned}$$

with $\Phi(.)$ denoting standard Gaussian distribution function.

– Family of log-normal distributions is closed towards multiplication with positive constant, that is, for $\rho > 0$ we have

$$\rho Y \sim \log n(\mu+\log\rho, \sigma^2)$$

– Mgf does not exist for $r > 0$ and we have following moments

$$\begin{aligned}\mu_Y &= e^{\mu+\sigma^2/2} \\ \sigma_Y^2 &= e^{2\mu+\sigma^2}(e^{\sigma^2}-1) \\ \mathrm{Vco}(Y) &= \left(e^{\sigma^2}-1\right)^{1/2} \\ \zeta_Y &= \left(e^{\sigma^2}+2\right)\left(e^{\sigma^2}-1\right)^{1/2}\end{aligned}$$

– For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned}E[Y 1_{u_1<Y\leq u_2}] &= \mu_Y\left[\Phi\left(\frac{\log u_2-(\mu+\sigma^2)}{\sigma}\right)-\Phi\left(\frac{\log u_1-(\mu+\sigma^2)}{\sigma}\right)\right] \\ I(G(y)) &= \Phi\left(\frac{\log y-(\mu+\sigma^2)}{\sigma}\right) \\ e(u) &= \mu_Y\left(\frac{1-\Phi\left(\frac{\log u-(\mu+\sigma^2)}{\sigma}\right)}{1-\Phi\left(\frac{\log u-\mu}{\sigma}\right)}\right)-u\end{aligned}$$

- Log-normal distribution does not have regularly varying survival function at infinity.
- Generating log-normal random numbers
  - Choose standard Gaussian numbers $Z \sim \Phi$
  - Set $Y = e^{\mu+\sigma Z}$
- Method of moment estimators are given by

$$\begin{aligned}\hat{\sigma}^{MM} &= \left[\log\left(\frac{\hat{\sigma}_n^2}{\hat{\mu}_n^2}+1\right)\right]^{1/2} \\ \hat{\mu}^{MM} &= \log \hat{mu}_n - (\hat{\sigma}^{MM})^2/2\end{aligned}$$

- MLE is given by

$$\begin{aligned}\hat{\mu}^{MLE} &= \frac{1}{n}\sum_{i=1}^{n}\log Y_i \\ (\hat{\sigma}^{MLE})^2 &= \frac{1}{n}\sum_{i=1}^{n}\sum_{i=1}^{n}\left(\log Y_i - \hat{\mu}^{MLE}\right)^2\end{aligned}$$

- **Log-gamma distribution**
  - Log-gamma distribution is more heavy tailed than log-normal distribution.
  - It is obtained by assuming $\log Y \sim \Gamma(\gamma, c)$ for positive parameters γ and c

$$\begin{aligned}g(y) &= \frac{c^\gamma}{\Gamma(\gamma)}(\log y)^{\gamma-1}y^{-(c+1)} \qquad \text{for } y \geq 1 \\ G(y) &= \mathcal{G}(\gamma, c\log y)\end{aligned}$$

  - Mgf does not exist for $r > 0$ and for moments we have

$$\begin{aligned}\mu_Y &= \left(\frac{c}{c-1}\right)^\gamma \qquad \text{for } c > 1 \\ \sigma_Y^2 &= \left(\frac{c}{c-2}\right)^\gamma - \mu_Y^2 \qquad \text{for } c > 2 \\ \zeta_Y &= \frac{1}{\sigma_Y^3}\left[\left(\frac{c}{c-3}\right)^\gamma - 3\mu_Y\sigma_Y^2 - \mu_Y^3\right] \qquad \text{for } c > 3\end{aligned}$$

  - For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned}E[Y1_{u_1<Y\leq u_2}] &= \left(\frac{c}{c-1}\right)^\gamma [\mathcal{G}(\gamma,(c-1)\log u_2) - \mathcal{G}(\gamma,(c-1)\log u_1)] \\ I(G(y)) &= \mathcal{G}(\gamma,(c-1)\log y) \\ e(u) &= \left(\frac{c}{c-1}\right)^\gamma\left(\frac{1-\mathcal{G}(\gamma,(c-1)\log u)}{1-\mathcal{G}(\gamma, c\log u)}\right) - u\end{aligned}$$

  - Log-gamma has regularly varying survival function at infinity with $c > 0$
  - Method of moment estimators are given by (solved numerically)

$$\begin{aligned}\hat{\gamma}^{MM} &= \frac{\log\hat{\mu}_n}{\log\frac{\hat{c}^{MM}}{\hat{c}^{MM}-1}} \\ \frac{\log(\hat{\sigma}_n^2+\hat{\mu}_n^2)}{\log\hat{\mu}_n} &= \frac{\log\hat{c}^{MM}-\log(\hat{c}^{MM}-2)}{\log\hat{c}^{MM}-\log(\hat{c}^{MM}-1)}\end{aligned}$$

– MLE is obtained analogously to MLE for gamma observations by simply replacing $_i$ by $\log Y_i$

- **Pareto distribution**
  - $Y \sim \text{Pareto}(\theta, \alpha)$ with threshold $\theta > 0$ and tail index $\alpha > 0$

$$\begin{aligned} g(y) &= \frac{\alpha}{\theta}\left(\frac{y}{\theta}\right)^{-(\alpha+1)} \qquad \text{for } y \geq \theta \\ G(y) &= 1 - \left(\frac{y}{\theta}\right)^{-\alpha} \end{aligned}$$

  - Claims above threshold θ are assumed to have regularly varying tails with $\alpha > 0$.
  - We have closedness towards multiplication with a positive constant $\rho > 0$

$$\rho Y \sim \text{Pareto}(\theta\rho, \alpha)$$

  - Mgf does not exist for $r > 0$ and for moments we have

$$\begin{aligned} \mu_Y &= \theta\frac{\alpha}{\alpha-1} \qquad \text{for } \alpha > 1 \\ \sigma_Y^2 &= \theta^2\frac{\alpha}{(\alpha-1)^2(\alpha-2)} \qquad \text{for } \alpha > 2 \\ \zeta_Y &= \frac{2(1+\alpha)}{\alpha-3}\left(\frac{\alpha-2}{\alpha}\right)^{1/2} \qquad \text{for } \alpha > 3 \end{aligned}$$

  - For $0 \leq u_1 < u_2$ and $u, y > 0$ we obtain

$$\begin{aligned} E[Y 1_{u_1 < Y \leq u_2}] &= \theta\frac{\alpha}{\alpha-1}\left[\left(\frac{u_1}{\theta}\right)^{-\alpha+1} - \left(\frac{u_2}{\theta}\right)^{-\alpha+1}\right] \\ I(G(y)) &= 1 - \left(\frac{y}{\theta}\right)^{-\alpha+1} \\ e(u) &= \frac{1}{\alpha-1}u \end{aligned}$$

  - As soon as we only study tails of distributions we should use MLEs for parameter estimation (MM is not sufficiently robust against outliers). Since threshold θ has natural meaning we only need to estimate α

$$\hat{\alpha}^{MLE} = \left(\frac{1}{n}\sum_{i=1}^{n}\log Y_i - \log\theta\right)^{-1}$$

  - **Lemma:** Assume $Y_1, \ldots, Y_n \overset{iid}{\sim} \text{Pareto}(\theta, \alpha)$, then

$$\begin{aligned} E[\hat{\alpha}^{MLE}] &= \frac{n}{n-1}\alpha \\ \text{var}\left(\hat{\alpha}^{MLE}\right) &= \frac{n^2}{(n-1)^2(n-2)}\alpha^2 \end{aligned}$$

  - We order claims accordingly to $Y_{(1)} \leq Y_{(2)} \leq \ldots \leq Y_{(n)}$ and define **Hill estimator**

$$\hat{\alpha}_{k,n}^{H} = \left(\frac{1}{n-k+1}\sum_{i=k}^{n}\log Y_{(i)} - \log Y_{(k)}\right)^{-1} \qquad \text{for } k < n$$

- Hill estimator is based on rationale that Pareto distribution is closed towards increasing thresholds, .i.e. for $Y \sim \text{Pareto}(\theta_0, \gamma)$ and $\theta_1 > \theta_0$ we have

$$\Pr[Y > y | Y \geq \theta_1] = \frac{\left(\frac{y}{\theta_0}\right)^{-\alpha}}{\left(\frac{\theta_1}{\theta_0}\right)^{-\alpha}} = \left(\frac{y}{\theta_1}\right)^{-\alpha} \quad \text{for } y \geq \theta_1$$

- Therefore if data comes from Pareto distribution we should observe stability in $\hat{\alpha}^H_{k,n}$ for changing k.

- **Creating new distributions**
  - **Multiplication by a constant**
    Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$. Consider transformation $\boldsymbol{Y = cX}$ with $c > 0$. Then

$$F_Y(y) = F_X\left(\frac{y}{c}\right) \quad \text{and} \quad f_Y(y) = \frac{1}{c} f_X\left(\frac{y}{c}\right)$$

  - **Raising to a power**
    Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$.
    Let $\boldsymbol{Y = X^{\tau}}$. Then, if $\boldsymbol{\tau > 0}$,

$$F_Y(y) = F_X(y^{1/\tau}) \quad \text{and} \quad f_Y(y) = \frac{1}{\tau} y^{\frac{1}{\tau}-1} f_X(y^{1/\tau})$$

    while, if $\boldsymbol{\tau < 0}$,

$$F_Y(y) = 1 - F_X(y^{1/\tau}) \quad \text{and} \quad f_Y(y) = \left|\frac{1}{\tau}\right| y^{\frac{1}{\tau}-1} f_X(y^{1/\tau})$$

  - **Exponentiation**
    Let X be a continuous r.v. with pdf $f_X(x)$ and cdf $F_X(x)$ with $f_X(x) > 0$ for all real x.
    Let $\boldsymbol{Y = \exp(X)}$. Then, for $y > 0$,

$$F_Y(y) = F_X(\ln y), \quad \text{and} \quad f_Y(y) = \frac{1}{y} f_X(\ln y)$$

- **Probability integral transformation (PIT)**
  We consider rv X with cdf F, where F is strictly increasing on some interval I, $F = 0$ to the left of I and $F = 1$ to the right of I. F^{-1} is well defined for $x \in I$.

  1. Let $Y = F(X)$, then Y has a uniform distribution on $[0, 1]$.
  2. Let U be uniform on $[0, 1]$ and let $Z = F^{-1}(U)$. Then the cdf of Z is F.

- **$\boldsymbol{k}$-point mixture distribution**
  - Consider rv X generated from k distinct subpopulations, where subpopulation i is modeled by the continuous distribution $f_{X_i}(x)$, then the pdf of X is given by

$$f_X(x) = \sum_{i=1}^{k} p_i f_{X_i}(x), \quad \text{with } 0 < p_i < 1, \quad \sum_{i=1}^{k} p_i = 1$$

- The cdf, k-th moment and moment generating function of the k-th point mixture are given as

$$\begin{aligned} F_X(x) &= \sum_{i=1}^{k} p_i F_{X_i}(x) \\ \mathbb{E}(X^k) &= \sum_{i=1}^{k} p_i \mathbb{E}(X_i^k) \\ M_X(r) &= \sum_{i=1}^{k} p_i M_{X_i}(r) \end{aligned}$$

- **Continuous mixture distribution**
  - Let X have conditional distribution $f_X(x|\theta)$ at a particular value of θ and let $g(\theta)$ be the pdf of the unknown rv θ. The unconditional pdf of X is

$$f_X(x) \quad = \quad \int_{-\infty}^{\infty} f_X(x|\theta) g(\theta) d\theta$$

    The pdf $g(\theta)$ is known as the prior distribution of θ (prior information or expert opinion is used in the analysis).
  - The cdf, k-moment and moment generating function of the continuous mixture are given as

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{\infty} F_X(x|\theta) g(\theta) d\theta \\ \mathbb{E}(X^k) &= \int_{-\infty}^{\infty} \mathbb{E}(X^k|\theta) g(\theta) d\theta = \mathbb{E}\left[\mathbb{E}(X^k|\theta)\right] \\ M_X(r) = \mathbb{E}(e^{rX}) &= \int_{-\infty}^{\infty} \mathbb{E}(e^{tX}|\theta) g(\theta) d\theta \end{aligned}$$

  - In particular the mean and variance of X are given by

$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}\left[\mathbb{E}(X|\theta)\right] \\ \mathrm{Var}(X) &= \mathbb{E}\left[\mathrm{Var}(X|\theta)\right] + \mathrm{Var}\left[\mathbb{E}(X|\theta)\right] \end{aligned}$$

- **Kolmogorov-Smirnov (KS) test**
  Assume an iid sequence $Y_1, Y_2, \ldots$ from unknown distribution function G and corresponding empirical distribution function $\hat{G}_n$ of finite sample size n. The non-parametric KS test investigates whether the continuous distribution function G_0 fits to given sample $Y_1, Y_2, \ldots, Y_n$.
  - **Glivenko-Cantelli theorem**: the empirical distribution function of an iid sample converges uniformly to true underlying distribution function $\mathbb{P}$-a.s. $(n \to \infty)$.
  - $H_0 : G = G_0$
  - $H_1 : G \neq G_0$
  - **KS test statistic:**

$$D_n = D_n(Y_1, \ldots, Y_n) = \|\hat{G}_n - G_0\|_\infty = \sup_y |\hat{G}_n(y) - G_0(y)|$$

- $\sqrt{n}D_n \to$ **Kolmogorov distribution $\boldsymbol{K}$** (as $n \to \infty$)

$$K(y) = 1 - 2\sum_{j=1}^{\infty}(-1)^{j+1}e^{-2j^2y^2} \qquad (y \in \mathbb{R}_+).$$

- H_0 is rejected on significance level $q \in (0,1)$ if

$$D_n > n^{-1/2}K^{\leftarrow}(1-q)$$

where $K^{\leftarrow}(1-q)$ denotes the $(1-q)$-quantile of the Kolmogorov distribution K .

q	20%	10%	5%	2%	1%
$K^{\leftarrow}(1-q)$	1.07	1.22	1.36	1.52	1.63

- **Anderson-Darling (AD) test**
  - **AD test statistic:**

$$\sup_y |\hat{G}_n(y) - G_0(y)|\sqrt{\psi(G_0(y))}$$

where $\psi : [0,1] \to \mathbb{R}_+$ is a weight function.

  - $\psi(t(1-y))^{-1}$ to investigate tails.

- **Pearson's $\boldsymbol{\chi^2}$ test**
  - Splits support of G_0 into K disjoint intervals $I_k = [c_k, c_{k+1})$ and groups data accordingly.
  - O_k counts the number of observed realisations $Y_1, \ldots, Y_n$ in I_k and E_k denotes the expected number of observations in I_k according to G_0.
  - **Test statistic of $\boldsymbol{n}$ observations:**

$$X^2_{n,K} = \sum_{k=1}^{K}\frac{(O_k - E_k)^2}{E_k}$$

  - If d parameters were estimated, then $X^2_{n,K}$ is compared to χ^2_{K-1-d} distribution.
  - Rule of thumb: $E_k > 4$

- **Information criteria**
Assume we want to compare different densities g_1 and g_2 that were fitted to $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$. The **Akaike** and **Bayesian Information Criterion** are

$$\begin{aligned} AIC^{(i)} &= -2\ell^{(i)}_{\mathbf{Y}} + 2d^{(i)} \\ BIC^{(i)} &= -2\ell^{(i)}_{\mathbf{Y}} + \log(n)d^{(i)} \end{aligned}$$

where $\ell^{(i)}_{\mathbf{Y}}$ is **log-likelihood function** of density g_i for data $\boldsymbol{Y}$ and $d^{(i)}$ denotes **number of estimated parameters** in g_i.

- **Re-insurance layers and deductibles**

– In this case the **pure risk premium for claim $Y \sim G$** is given by

$$\begin{aligned} E[(Y-d)_+] &= \int_d^\infty (y-d)dG(y) = E[Y1_{\{Y>d\}}] - dP[Y>d] \\ &= P[Y>d](E[Y|Y>d]-d) = P[Y>d]e(d) \end{aligned}$$

under the assumption that $P[Y > d] > 0$ and that the mean excess function $e(.)$ of Y exists.

– Insurance company covers $(Y \wedge M)$ and pure risk premium for this (bounded) claim is given by

$$\begin{aligned} E[Y \wedge M] &= \int_0^M ydG(y) + MP[Y>M] = E[Y1_{\{Y\le M\}}] + MP[Y>M] \\ &= E[Y] - \big(E[Y1_{\{Y>M\}}] - MP[Y>M]\big) \\ &= E[Y] - P[Y>M]e(M) = E[Y] - E[(Y-M)_+] \end{aligned}$$

– If we combine deductibles with maximal covers we obtain excess-of-loss (XL) (re-)insurance treaties. Assume we have deductible $u_1 > 0$. Insurance treaty "$u_2 XL u_1$" covers claims layer $(u_1, u_1+u_2]$ that is, this contract covers maximal excess of u_2 above priority u_1. The pure risk premium for such contracts is then given by

$$E[((Y-u_1)_+) \wedge u_2] = E[(Y-u_1)_+] - E[(Y-u_1-u_2)_+]$$

– **Theorem: leverage effect of claims inflation:** Choose a fixed deductible $d > 0$ and assume that claim at time 0 is given by Y_0. Assume that there is a deterministic inflation index $i > 0$ such that claim at time 1 can be represented by $Y_1 \overset{(d)}{=} (1+i)Y_0$. We have

$$E[(Y_1-d)_+] \ge (1+i)E[(Y_0-d)_+]$$

Aggregate loss models or compound distributions

- **Compound binomial model**

  – The total claim amount S has a **compound binomial distribution**, write

$$S \sim \mathrm{CompBinom}(v,p,G)$$

  if S has a compound distribution with $N \sim \mathrm{Binom}(v,p)$ for given $v \in \mathbb{N}$ and $p \in (0,1)$ and individual claim size distribution G.

  – **Proposition:** Assume $S \sim \mathrm{CompBinom}(v,p,G)$. We have

$$\begin{aligned} E[S] &= vpE[Y_1] \\ \mathrm{var}(S) &= vp\left(E[Y_1^2] - pE[Y_1]^2\right) \\ \mathrm{Vco}(S) &= \sqrt{\frac{1}{vp}}\sqrt{1-p+\mathrm{Vco}(Y_1)^2} \\ M_S(r) &= (pM_{Y_1}(r) + (1-p))^v \qquad r \in \mathbb{R} \end{aligned}$$

  whenever they exist.

- **Corollary - Aggregation property:** Assume that $S_1, \dots, S_n$ are independent with $S_j \sim \text{CompBinom}(v_j, p, G)$ for all $j = 1, \dots, n$. The aggregated claim has a compound binomial distribution with

$$S = \sum_{j=1}^{n} S_j \sim \text{CompBinom}\left(\sum_{j=1}^{n} v_j, p, G\right)$$

- **Compound Poisson model**
  - The total claim amount S has a **compound Poisson distribution**

$$S \sim \text{CompPoi}(\lambda v, G)$$

    if S has compound distribution with $N \sim \text{Poi}(\lambda v)$ for given $\lambda, v > 0$ and individual claim size distribution G.
  - **Proposition:** Assume $S \sim \text{CompPoi}(\lambda v, G)$. We have

$$\begin{aligned} E[S] &= \lambda v E[Y_1] \\ \text{var}(S) &= \lambda v E[Y_1^2] \\ \text{Vco}(S) &= \sqrt{\frac{1}{\lambda v}} \sqrt{1 + \text{Vco}(Y_1)^2} \\ M_S(r) &= e^{\lambda v (M_{Y_1}(r) - 1)} \qquad \text{for } r \in \mathbb{R} \end{aligned}$$

    whenever they exist.
  - **Theorem - Aggregation of compound Poisson distributions:** Assume $S_1, \dots, S_n$ are independent with $S_j \sim \text{CompPoi}(\lambda_j v_j, G_j)$ $\forall j = 1, \dots, n$. The aggregated claim has compound Poisson distribution

$$S = \sum_{j=1}^{n} S_j \sim \text{CompPoi}(\lambda v, G)$$

    with

$$v = \sum_{j=1}^{n} v_j \qquad \lambda = \sum_{j=1}^{n} \frac{v_j}{v} \lambda_j \qquad G = \sum_{j=1}^{n} \frac{\lambda_j v_j}{\lambda v} G_j.$$

- **Extension of the compound poisson model**
  - Let $(p_j^+)_{j=1,\dots,m}$ be a discrete probability distribution on finite set $\{1, \dots, m\}$. Assume $p_j^+ > 0$ for all j.
  - Assume G_j corresponding claim size distributions of the sub-portfolios with $G_j(0) = 0$.
  - Define mixture distribution

$$G(y) = \sum_{j=1}^{m} p_j^+ G_j(y) \qquad \text{for } y \in \mathbb{R}$$

    Former theorem exactly provides such a mixture distribution with $p_j^+ = \frac{\lambda_j v_j}{\lambda v}$ if we aggregate sub-portfolios.

- Define a discrete random variable I which indicates to which sub-portfolio particular claim Y belongs

$$\mathbb{P}[I=j]=p_j^+ \qquad \text{for all } j\in\{1,\dots,m\}$$

- **Definition - Extended compound poisson model:** The total claim amount $S=\sum_{i=1}^{N} Y_i$ has a compound Poisson distribution as defined before. In addition, we assume that $(Y_i,I_i)_{i\geq 1}$ are iid and independent of N with Y_i having marginal distribution function G with $G(0)=0$ and I_i having marginal distribution function given before.

- (Y_1,I_1) takes values in $\mathbb{R}_+\times\{1,\dots,m\}$ and let $A_1,\dots,A_n$ be a measurable disjoint decomposition of $\mathbb{R}_+\times\{1,\dots,m\}$, i.e.
  - $A_k\cap A_l=\emptyset$ for all $k\neq l$
  - $\cup_{i=1}^{n}A_k=\mathbb{R}_+\times\{1,\dots,m\}$

  This measurable disjoint decomposition is called admissible for (Y_1,I_1) if for all $k=1,\dots,n$

$$p^{(k)}=\Pr\left[(Y_1,I_1)\in A_k\right]>0.$$

  Note that $\sum_{k=1}^{n}p^{(k)}=1$.

- **Theorem - Disjoint decomposition property:** Assume that S fulfils extended compound Poisson model assumptions. We choose an admissible, measurable disjoint decomposition $A_1,\dots,A_n$ for (Y_1,I_1). Define for $k=1,\dots,n$ the random variables

$$S_k=\sum_{i=1}^{N}Y_i 1_{\{(Y_i,I_i)\in A_k\}}$$

  S_k are independent and $\text{CompPoi}(\lambda_k v_k,G_k)$ distributed for $k=1,\dots,n$ with

$$\lambda_k v_k=\lambda v p^{(k)}>0 \text{ and } G_k(y)=\Pr[Y_1\leq y|(Y_1,I_1)\in A_k]$$

- **Compound negative-binomial model**
  - The total claim amount S has a **compound Negative-binomial distribution**

$$S\sim\text{CompNB}(\lambda v,\gamma,G)$$

    if S has compound distribution with $N\sim\text{NegBin}(\lambda v,\gamma)$ for given $\lambda,v,\gamma>0$ and individual claim size distribution G.
  - **Proposition:** Assume $S\sim\text{CompNB}(\lambda v,\gamma,G)$. We have, whenever they exist

$$\begin{aligned}
E[S] &= \lambda v E[Y_1]\\
\text{var}(S) &= \lambda v E[Y_1^2]+(\lambda v)^2E[Y_1]^2/\gamma\\
\text{Vco}(S) &= \sqrt{\frac{1}{\lambda v}}\sqrt{1+\text{Vco}(Y_1)^2+\lambda v/\gamma}>\gamma^{-1/2}\\
M_S(r) &= \left(\frac{1-p}{1-pM_{Y_1}(r)}\right)^{\gamma} \quad \text{for } r\in\mathbb{R} \text{ such that } M_{Y_1}(r)<1/p
\end{aligned}$$

    with $p=(\lambda v)/(\gamma+\lambda v)\in(0,1)$.

Parameter estimation

- **Method of moments (specific case)**
  - Assume that there exist strictly positive volumes $v_1, \ldots, v_T$ such that the components of $\boldsymbol{F} = (N_1/v_1, \ldots, N_T/v_T)'$ are **independent** with

$$\lambda = E[N_t/v_t] \qquad \text{and} \qquad \tau_t^2 = \text{var}(N_t/v_t) \in (0, \infty)$$

  - **Lemma:** (assumption above holds) Unbiased linear (in $\boldsymbol{F}$) estimator for λ with minimal variance is given by

$$\hat{\lambda}_T^{MV} = \left(\sum_{t=1}^T \frac{1}{\tau_t^2} \right)^{-1} \sum_{t=1}^T \frac{N_t/v_t}{\tau_t^2}$$

    The variance of this estimator is given by

$$\text{var}(\hat{\lambda}_T^{MV}) = \left(\sum_{t=1}^T \frac{1}{\tau_t^2} \right)^{-1}$$

- **Maximum Likelihood Estimation (MLE) method**
  Assume that the components of $\boldsymbol{N} = (N_1, \ldots, N_T)'$ are independent with probability weights $p_k^{(t)}(\vartheta) = \Pr_\vartheta[N_t = k] = \Pr[N_t = k]$ which depend on common unknown parameter ϑ.
  - **Joint likelihood function for observation $\boldsymbol{N}$**

$$\mathcal{L}_{\boldsymbol{N}}(\vartheta) = \prod_{t=1}^T p_{N_t}^{(t)}(\vartheta)$$

  - **Joint log-likelihood function**

$$\ell_{\boldsymbol{N}}(\vartheta) = \sum_{t=1}^T \log p_{N_t}^{(t)}(\vartheta)$$

  - $\hat{\vartheta}_T^{MLE}$ for ϑ based on observation $\boldsymbol{N}$ is given by

$$\hat{\vartheta}_T^{MLE} = \underset{\vartheta}{\text{argmax}}\, \mathcal{L}_{\boldsymbol{N}}(\vartheta) = \underset{\vartheta}{\text{argmax}}\, \ell_{\boldsymbol{N}}(\vartheta)$$

  - Under suitable regularity properties and real valued parameter, $\hat{\vartheta}_T^{MLE}$ is solution of

$$\frac{\partial}{\partial \vartheta} \ell_{\boldsymbol{N}}(\vartheta) = \frac{\partial}{\partial \vartheta} \sum_{t=1}^T \log p_{N_t}^{(t)}(\vartheta) = 0$$

- **Statistical test - χ^2 goodness-of-fit test**
  - $H_0 : N_t \overset{iid}{\sim} \text{Poi}(\lambda v_t)$ for $t = 1, \ldots, T$.

- **Test statistic**

$$\chi^* = \chi^*(\boldsymbol{N}) = \sum_{t=1}^{T} \frac{(N_t/v_t - \lambda)^2}{\lambda/v_t}$$

- Aggregation and disjoint decomposition theorems imply that $N_t \sim \text{Poi}(\lambda v_t)$ can be understood as a sum of v_t iid random variables $X_i \sim \text{Poi}(\lambda)$. Hence

$$N_t \overset{(d)}{=} \sum_{i=1}^{v_t} X_i$$

with $E[X_1] = \lambda$ and $\text{var}(X_1) = \lambda$. But then CLT applies as $v_t \to \infty$

$$\tilde{Z}_t = \frac{N_t/v_t - \lambda}{\sqrt{\lambda/v_t}} = \frac{N_t - \lambda v_t}{\sqrt{\lambda v_t}} \overset{(d)}{=} \frac{\sum_{i=1}^{v_t} X_i - \lambda v_t}{\sqrt{\lambda v_t}} \xrightarrow{D} Z_t \sim \mathcal{N}(0,1)$$

- Approximate $\tilde{Z}_t$ by $Z_t \sim \mathcal{N}(0,1)$ for v_t sufficiently large.
- If $Z_1, \ldots, Z_n \overset{iid}{\sim} \mathcal{N}(0,1)$ then $\sum_{t=1}^{T} Z_t^2$ has χ_T^2 distribution. Therefore

$$\chi^* = \chi^*(\boldsymbol{N}) = \sum_{t=1}^{T} \frac{(N_t/v_t - \lambda)^2}{\lambda v_t} = \sum_{t=1}^{T} \tilde{Z}_t^2 \overset{(d)}{\approx} \sum_{t=1}^{T} Z_t^2 \sim \chi_T^2$$

- We replace unknown λ by $\hat{\lambda}_T^{MLE}$ and lose 1 df.

$$\hat{\chi}^* = \sum_{t=1}^{T} v_t \frac{\left(N_t/v_t - \hat{\lambda}_T^{MLE}\right)^2}{\hat{\lambda}_T^{MLE}} \overset{(d)}{\approx} \chi_{T-1}^2$$

Approximations for compound distributions

- **Compound distributions - distribution of S**
  - **Basic recognition features of compound distributions:** Assume S has a compound distribution. We have (whenever they exist)

$$\begin{aligned} E[S] &= E[N]E[Y_1] \\ \text{var}(S) &= \text{var}(N)E[Y_1]^2 + E[N]\,\text{var}(Y_1) \\ \text{Vco}(S) &= \sqrt{\text{Vco}(N)^2 + \frac{1}{E[N]}\text{Vco}(Y_1)^2} \\ M_S(r) &= M_N\left(\log(M_{Y_1(r)})\right) \qquad \text{for } r \in \mathbb{R}. \end{aligned}$$

  - If assumptions above hold, the **distribution function of S** can be written as

$$\begin{aligned} F_S(x) &= \mathbb{P}[S \le x] = \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{N} Y_i \le x \,\middle|\, N = k\right] \mathbb{P}[N = k] \\ &= \sum_{k \in \mathcal{A}} \mathbb{P}\left[\sum_{i=1}^{k} Y_i \le x\right] \mathbb{P}[N = k] = \sum_{k \in \mathcal{A}} G^{*k}(x)\mathbb{P}[N = k] \end{aligned}$$

G^{*k} denotes the k-th convolution of the distribution function G.
In particular, we have for $Y_1, Y_2 \overset{iid}{\sim} G$

$$\begin{aligned} G^{*2}(x) &= \mathbb{P}[Y_1 + Y_2 \leq x] = \int G(x-y)dG(y) \\ G^{*k}(x) &= \int G^{*(k-1)}(x)dG(y) \end{aligned}$$

- **Normal approximation**
  - **Theorem:** Assume $S \sim \text{CompPoi}(\lambda v, G)$ with G having a finite second moment. We have
    $$\frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}} \Rightarrow N(0,1) \text{ as } v \to \infty$$
  - **Approximation of the distribution function of S:**
    $$P[S \leq x] = P\left[\frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}} \leq \frac{x - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}\right] \approx \Phi\left(\frac{x - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}\right)$$
- **Translated gamma and log-normal approximation**
  - We choose $k \in \mathbb{R}$ and define **translated or shifted** random variables
    $$X = k + Z \qquad \text{where } Z \sim \Gamma(\gamma, c) \quad \text{or} \quad Z \sim LN(\mu, \sigma^2)$$

Translated gamma case	**Translated log-normal** case
$E[X] = k + \gamma/c$	$E[X] = k + e^{\mu+\sigma^2/2}$
$\text{var}(X) = \gamma/c^2$	$\text{var}(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right)$
$\zeta_X = 2\gamma^{-1/2}$	$\zeta_X = \left(e^{\sigma^2} + 2\right)\left(e^{\sigma^2} - 1\right)^{1/2}$

  - Assume S has a finite third moment. Then we choose
    $$X = k + Z \qquad \text{where } Z \sim \Gamma(\gamma, c) \quad \text{or} \quad Z \sim LN(\mu, \sigma^2)$$
    such that the three parameters of X fulfill
    $$E[X] = E[S] \qquad \text{var}(X) = \text{var}(S) \qquad \zeta_X = \zeta_S$$
- **Edgeworth approximation**
  Assume S is compound Poisson distributed with claim size distribution G having a positive radius of convergence $\rho_0 > 0$.
  - Normalized random variable
    $$Z = \frac{S - \lambda v E[Y_1]}{\sqrt{\lambda v E[Y_1^2]}}$$
    $\Rightarrow E[Z] = 0, \text{var}(Z) = 1$ and $\zeta_Z = \zeta_S$.
  - **Taylor expansion around origin, choose $n \geq 3$**
    $$\log M_Z(r) = \sum_{k=0}^{n} \frac{\frac{d^k}{dr^k} \log M_Z(r)|_{r=0}}{k!} r^k + o(r^n) \qquad \text{as } r \to 0$$

– Set
$$a_k = \frac{\frac{d^k}{dr^k}\log M_Z(r)|_{r=0}}{k!}$$
$\Rightarrow a_0 = \log M_Z(0) = 0, a_1 = E[Z] = 0$ and $a_2 = \mathrm{var}(Z)/2! = 1/2$.

– **Approximation**
$$M_Z(r) \approx e^{\frac{1}{2}r^2 + \sum_{k=3}^n a_k r^k} = e^{\frac{1}{2}r^2} e^{\sum_{k=3}^n a_k r^k}$$

– Using second Taylor expansion for $e^x = 1 + x + x^2/2! + \ldots$ applied to latter exponential function in last expression, **mgf of Z is approximated by**
$$M_Z(r) \approx e^{r^2/2}\left[1 + \sum_{k=3}^n a_k r^k + \frac{\left(\sum_{k=3}^n a_k r^k\right)^2}{2!} + \ldots\right]$$

– For appropriate constants $b_k \in \mathbb{R}$ we get approximation (for small r)
$$M_Z(r) \approx e^{r^2/2}\left[1 + a_3 r^3 + \sum_{k\geq 4} b_k r^k\right] \tag{1}$$

– **Lemma:** Let Φ denote the standard Gaussian distribution function and $\Phi^{(k)}$ its k-th derivative. For $k \in \mathbb{N}_0$ and $r \in \mathbb{R}$ we have
$$r^k e^{r^2/2} = (-1)^k \int_{-\infty}^{\infty} e^{rx}\Phi^{(k+1)}(x)dx$$

– Set $X \sim N(0,1)$ and rewrite approximation (1) as (using the above Lemma):
$$\begin{aligned} M_Z(r) &\approx E[e^{rX}] - a_3\int_{-\infty}^{\infty} e^{rx}\Phi^{(4)}(x)dx + \sum_{k\geq 4} b_k(-1)^k \int_{-\infty}^{\infty} e^{rx}\Phi^{(k+1)}(x)dx \\ &= \int_{-\infty}^{\infty} e^{rx}\left[\Phi'(x) - a_3\Phi^{(4)}(x) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k+1)}(x)\right]dx \end{aligned}$$
Let Z have distribution function F_Z, then the latter suggests approximation
$$dF_Z(z) \approx \left[\Phi'(z) - a_3\Phi^{(4)}(z) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k+1)}(z)\right]dz$$

– Integration provides the **Edgeworth approximation** ($x = \sqrt{\lambda v E[Y_1^2]}z + \lambda v E[Y_1]$)
$$P[S \leq x] = F_Z(z) \approx EW(z) \stackrel{def}{=} \Phi(z) - a_3\Phi^{(3)}(z) + \sum_{k\geq 4} b_k(-1)^k\Phi^{(k)}(z)$$

– The first order approximationof Φ is corrected by higher order terms involving skewness and other higher order terms reflected by a_S and b_k.

– Consider derivatives $\Phi^{(k)}$ for $k \geq 1$
$$\begin{aligned} \Phi'(z) &= \frac{1}{\sqrt{2\pi}}e^{-z^2/2} \\ \Phi^{(k)}(z) &= \frac{d^{k-1}}{dz^{k-1}}\frac{1}{\sqrt{2\pi}}e^{-z^2/2} = \mathcal{O}\left(z^{k-1}e^{-z^2/2}\right) \qquad \text{for } |z| \to \infty; k \geq 2 \end{aligned}$$

– From this it follows that
$$\lim_{z\to-\infty} EW(z) = 0 \qquad \text{and} \qquad \lim_{z\to\infty} EW(z) = 1$$

PART B: COPULAS

- A copula C is supermodular (2-increasing) if the inequality

$$C(v_1, v_2) - C(u_1, v_2) - C(v_1, u_2) + C(u_1, u_2) \geq 0$$

  is valid for any $u_1 \leq v_1$ and $u_2 \leq v_2$.

- Independence copula

$$C^{\text{ind}}(u_1, u_2) = u_1 u_2$$

- (Co-)monotonicity copula or Fréchet upper bound copula

$$C^{\text{M}}(u_1, u_2) = \min(u_1, u_2)$$

- Counter-monotonicity copula or Fréchet lower bound copula

$$C^{\text{CM}}(u_1, u_2) = \max(u_1 + u_2 - 1, 0)$$

- Gaussian copula

$$C_\rho^{\text{Gauss}}(u_1, u_2) = \int_{-\infty}^{\phi^{-1}(u_1)} \int_{-\infty}^{\phi^{-1}(u_2)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\right) ds_1 ds_2$$

- Probability density function of the Gaussian copula

$$\begin{aligned} c_\rho^{\text{Gauss}}(u_1, u_2) &= \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{\xi_1^2 - 2\rho\xi_1\xi_2 + \xi_2^2}{2(1-\rho^2)}\right) \frac{d}{du_1}\Phi^{-1}(u_1) \frac{d}{du_2}\Phi^{-1}(u_2) \\ &\overset{(*)}{=} \frac{1}{\sqrt{1-\rho^2}} \exp\left(-\frac{\xi_1^2 - 2\rho\xi_1\xi_2 + \xi_2^2}{2(1-\rho^2)}\right) \exp\left(\frac{\xi_1^2 + \xi_2^2}{2}\right) \end{aligned}$$

  where $\xi_1 = \Phi^{-1}(u_1)$ and $\xi_2 = \Phi^{-1}(u_2)$.
  $^{(*)}$ Note that $\frac{d}{du_i}\Phi^{-1}(u_i) = \frac{1}{\phi(\Phi^{-1}(u_i))} = \sqrt{2\pi}\exp(\frac{\xi_i^2}{2})$.

- Univariate t-distribution with ν degrees of freedom: $Y_1 \sim t_\nu$

$$Y_1 = \frac{X_1}{\sqrt{\xi/\nu}} = \frac{\sqrt{\nu}X_1}{\sqrt{Z_1^2 + \ldots + Z_\nu^2}}$$

  where $X_1, Z_1, \ldots Z_\nu \sim$ Normal(0,1) and $\xi \sim \chi_\nu^2$, X_1 independent on $\mathbf{Z}$

- Bivariate t-distribution with mean $\boldsymbol{\mu}$ and shape matrix $\frac{\nu}{\nu-2}\boldsymbol{\Sigma}$:

$$(Y_1, Y_2) = \boldsymbol{\mu} + \left(\frac{X_1}{\sqrt{\xi/\nu}}, \frac{X_2}{\sqrt{\xi/\nu}}\right)$$

  where $\mathbf{X} = (X_1, X_2) \sim \text{Normal}(\mathbf{0}, \boldsymbol{\Sigma})$

- Student copula

$$\begin{aligned} C^{\text{Student}}_{\rho,\nu}(u_1,u_2) &= T_{\rho,\nu}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\right) \\ &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{ds_1 ds_2}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{\nu(1-\rho^2)}\right)^{\frac{-(\nu+2)}{2}} \end{aligned}$$

  where $t_\nu \equiv$ univariate Student distribution with ν df and $T_{\rho,\nu} \equiv$ bivariate Student distribution with ν df and $0 \le \rho \le 1$.

- Form of an Archimedean copula

$$C(u_1,u_2) = \phi^{-1}\left(\phi(u_1) + \phi(u_2)\right)$$

- Archimedean copulas

Name	Generator	Bivariate copula
Frank copula	$\phi(t) = -\log\left(\frac{\exp(-\theta t)-1}{\exp(-\theta)-1}\right)$ $(-\infty < \theta < +\infty)$	$C^{\mathrm{Fr}}(u_1, u_2) = \frac{-1}{\theta}\log\left(1 + \frac{(\exp(-\theta u_1)-1)(\exp(-\theta u_2)-1)}{\exp(-\theta)-1}\right)$
Clayton copula	$\phi(t) = \frac{t^{-\theta}-1}{\theta}$ $(\theta > 0)$	$C^{\mathrm{Cl}}(u_1, u_2) = \left(u_1^{-\theta} + u_2^{-\theta} - 1\right)^{-1/\theta}$
Gumbel copula	$\phi(t) = (-\log(t))^{\theta}$ $(\theta \geq 1)$	$C^{\mathrm{Gu}}(u_1, u_2) = exp\left(-\left\{(-\log(u_1))^{\theta} + (-\log(u_2))^{\theta}\right\}^{1/\theta}\right)$

- Survival copula $\overline{C}$ associated with C

$$\overline{C}(u_1, u_2) = C(1 - u_1, 1 - u_2) + u_1 + u_2 - 1$$

- The construction of the Marhall-Olkin survival copula leads to following copula family

$$\begin{aligned} C_{\alpha_1,\alpha_2}(u_1, u_2) &= \min(u_1^{1-\alpha_1} u_2, u_1 u_2^{1-\alpha_2}) \\ &= \begin{cases} u_1^{1-\alpha_1} u_2 & u_1^{\alpha_1} \geq u_2^{\alpha_2} \\ u_2^{1-\alpha_2} u_1 & u_1^{\alpha_1} < u_2^{\alpha_2} \end{cases} \end{aligned}$$

- The co-copula C^* and dual $\tilde{C}$ are defined as

$$\begin{aligned} C^*(u_1, u_2) &= 1 - C(1 - u_1, 1 - u_2) \\ \tilde{C}(u_1, u_2) &= u_1 + u_2 - C(u_1, u_2) \end{aligned}$$

- $M_{X,Y} = M_C$ is measure of concordance between rvs X and Y (with copula C) $\Leftrightarrow$
  1. it is defined for every pairs of rvs (completeness)
  2. it is a relative (normalized) measure, i.e. $M_{X,Y} \in [-1, 1]$
  3. it is symmetric, i.e. $M_{X,Y} = M_{Y,X}$
  4. if X and Y are independent, then $M_{X,Y} = 0$
  5. $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$
  6. if $\{(X_n, Y_n)\}$ is sequence of continuous rvs with copula C_n and $\lim_{n\to+\infty} C_n(x, y) = C(x, y), \forall(x, y) \in [0, 1]^2$ then $\lim_{n\to+\infty} M_{X_n,Y_n} = M_{X,Y}$.
  7. it respects concordance order: if $C1 \prec C2$, then $M_{C_1} \leq M_{C_2}$

- Pearson correlation coefficient between X_1 and X_2

$$\rho = \frac{\mathrm{cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} = \frac{E[X_1 X_2] - E[X_1]E[X_2]}{\sigma_{X_1}\sigma_{X_2}}$$

- Kendall's tau

$$\begin{aligned} \rho_\tau(X_1, X_2) &= P\left((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) > 0\right) \\ &\quad -P\left((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)}) < 0\right) \\ &= E\left[\mathrm{sign}\left((X_1^{(1)} - X_1^{(2)})(X_2^{(1)} - X_2^{(2)})\right)\right] \\ &\Updownarrow \\ \rho_\tau(X_1, X_2) &= 4\int\int C_X(u, v)dC_X(u, v) - 1 \end{aligned}$$

- Sample Kendall's tau of a bivariate sample of size M

$$\hat{\rho}_\tau(X_1, X_2) = \begin{pmatrix} M \\ 2 \end{pmatrix}^{-1} \sum_{1 \le i \le j \le M} \text{sign}\left((X_1^{(i)} - X_1^{(j)})(X_2^{(i)} - X_2^{(j)})\right)$$

where

$$\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{k!(n-k)!}$$

- Spearman's rho

$$\begin{aligned} \rho_S(X_1, X_2) &= \text{Corr}\left(F_{X_1}(X_1), F_{X_2}(X_2)\right) \\ &\Updownarrow \\ \rho_S(X_1, X_2) &= 12 \int\int (C_X(u,v) - uv)\,du\,dv \end{aligned}$$

- Sample Spearman's rho of a bivariate sample of size M

$$\hat{\rho}_S(X_1, X_2) = \frac{12}{M(M^2-1)} \sum_{i=1}^{M} \left(\text{rank}(X_1^{(i)}) - \frac{M+1}{2}\right)\left(\text{rank}(X_2^{(i)}) - \frac{M+1}{2}\right)$$

- ρ_τ and ρ_S for Archimedean copulas

Family	Kendall's tau	Spearman's rho
Independence	0	0
Clayton	$\frac{\theta}{\theta+2}$	Complicated
Gumbel	$1 - \frac{1}{\theta}$	No closed-form
Frank	$1 - \frac{4}{\theta}(D_1(-\theta) - 1)$	$1 - \frac{12}{\theta}(D_2(-\theta) - D_1(-\theta))$

where the Debye function $D_k(.)$ is defined as

$$D_k(x) = \frac{k}{x^k}\int_0^x \frac{t^k}{e^t - 1}dt, \; k = 1, 2 \quad \text{and} \quad D_k(-x) = D_k(x) + \frac{kx}{k+1}$$

- **Theorem:** Let $\mathbf{X} = (X_1, \ldots, X_N)$ have a meta-Gaussian distribution with continuous marginal cdfs and copula $C^{\text{Gauss}}(\cdot\,;\mathbf{\Omega}), (\mathbf{\Omega})_{ij} = \rho_{ij}$, then

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi}\arcsin(\rho_{ij})$$

$$\rho_S(X_i, X_j) = \frac{6}{\pi}\arcsin\left(\frac{\rho_{ij}}{2}\right) \approx \rho_{ij}$$

- **Theorem:** Let $\mathbf{X} = (X_1, \ldots, X_N)$ have a meta-t_ν distribution with continuous marginal cdfs and copula $C^{\text{Student}}(\cdot\,;\mathbf{\Omega}, \nu)$, then

$$\rho_\tau(X_i, X_j) = \frac{2}{\pi}\arcsin(\rho_{ij})$$

- If the limit $\lambda_u \in [0,1]$ exists, the coefficient of upper tail dependence of X_1 and X_2 is defined by

$$\lambda_u = \lim_{q \to 1^-} P\left(X_2 > F_{X_2}^{-1}(q) | X_1 > F_{X_1}^{-1}(q)\right)$$

- If the limit $\lambda_\ell \in [0,1]$ exists, the coefficient of lower tail dependence of X_1 and X_2 is defined by

$$\lambda_\ell = \lim_{q \to 0^+} P\left(X_2 \le F_{X_2}^{-1}(q) | X_1 \le F_{X_1}^{-1}(q)\right)$$

PART C: GENERALIZED LINEAR MODELS (GLMs)

Tariffication

- We focus on risk adjusted premia using the compound Poisson model $S \sim \text{CompPoi}(\lambda v, G)$ and consider

$$S = \sum_{i=1}^{N} Y_i = \sum_{\ell=1}^{v} \sum_{i=1}^{N^{(\ell)}} Y_i^{(\ell)} = \sum_{\ell=1}^{v} S_\ell$$

  where $S_\ell = \sum_{i=1}^{N^{(\ell)}} Y_i^{(\ell)}$ models total claim amount of policy $\ell = 1, \ldots, v$.

- This decoupling provides independent compound Poisson distributions S_ℓ

$$S_\ell \sim \text{CompPoi}(\lambda_\ell, G_\ell)$$

  where we set

  - volume $v_\ell = 1$
  - $\lambda_\ell > 0$ is expected number of claims of policy ℓ
  - $Y_i^{(\ell)} \sim G_\ell$ describes claim size distribution of policy ℓ.

- This implies for mean value of S the following decomposition

$$E[S] = \sum_{\ell=1}^{v} E[S_\ell] = \sum_{\ell=1}^{v} \lambda_\ell E[Y_1^{(\ell)}] = \lambda E[Y_1] \sum_{\ell=1}^{v} \frac{\lambda_\ell E[Y_1^{(\ell)}]}{\lambda E[Y_1]} = \mu \sum_{\ell=1}^{v} \chi^{(\ell)}$$

  where

  - $\mu = E[S]/v = \lambda E[Y_1]$ is average claim over all policies
  - $\chi^{(\ell)} > 0$ reflects risk characteristics of policy $\ell = 1, \ldots, v$.

- **Multiplicative tariff**
  We assume that we have only $k = 2$ tariff criteria but generalisation is straightforward. We set up a multiplicative tariff structure

  - First criterion has I risk characteristics $i \in \{1, \ldots, I\}$
    Second criterion has J risk characteristics $j \in \{1, \ldots, J\}$
  - Thus we have $M = I \cdot J$ different risk classes
  - Organize data with observations having **one index per rating factor**. This is suitable for displaying data in a table (tabular form).
  - Assume policy ℓ belongs to risk class (i, j), write $\chi^{(\ell)} = \chi^{(i,j)}$. Hence

$$E[S] = \mu \sum_{i,j} v_{i,j} \chi^{(i,j)}$$

    where $v_{i,j}$ denotes number of policies belonging to risk class (i, j).

- Our aim is to set up a multiplicative tariff structure for these $K = 2$ tariff criteria, i.e. we assume
$$\chi^{(i,j)} = \chi_{1,i}\chi_{2,j}$$
where χ_{k,ℓ_k} describes specifics of criterion k if it has risk characteristics ℓ_k.

We need to find appropriate multiplicative pricing factors $\chi_{1,i}$, $i \in \{1,\dots,I\}$ and $\chi_{2,j}$, $j \in \{1,\dots,J\}$ that describe risk classes (i,j) according to multiplicative tariff structure.

- $S_{i,j}$ is total claim of risk class (i,j) and $v_{i,j}$ is corresponding volume with
$$\sum_{i,j} v_{i,j} = v \qquad \text{and} \qquad \sum_{i,j} S_{i,j} = S$$
- This implies that we need to study
$$E[S_{i,j}] = v_{i,j}\frac{E[S]}{v}\chi^{(i,j)} = v_{i,j}\mu\chi_{1,i}\chi_{2,j}$$
  * $\mu = \lambda E[Y_1]$ is average claim per policy over whole portfolio v
    $\Rightarrow E[S] = v\mu$.
  * $\chi^{(i,j)} = \chi_{1,i}\chi_{2,j}$ describes multiplicative tariff structure fro two tariff criteria.

Simple tariffication methods

- **Bailey and Simon**
  - Specify $\mu, \chi_{1,i}$ and $\chi_{2,j} > 0$ such that following expression (which describes the test statistic of the χ^2-goodness-of-fit test) is minimized:
$$X^2 = \sum_{i,j} \frac{\left(S_{i,j} - v_{i,j}\mu\chi_{1,i}\chi_{2,j}\right)^2}{v_{i,j}\mu\chi_{1,i}\chi_{2,j}}$$
    We denote the minimizers by $\hat{\mu}, \hat{\chi}_{1,i}$ and $\hat{\chi}_{2,j}$.
  - **Lemma:** The minimizers have a (systematic) positive bias:
$$\sum_{i,j} v_{i,j}\hat{\mu}\hat{\chi}_{1,i}\hat{\chi}_{2,j} \geq \sum_{i,j} S_{i,j} = S$$
- **Bailey and Jung**
  This method imposes unbiasedness of rows and columns by definition: choose $\mu, \chi_{1,i}$ and $\chi_{2,j} > 0$ such that rows i and columns j satisfy
$$\begin{aligned}\sum_{j=1}^{J} v_{i,j}\mu\chi_{1,i}\chi_{2,j} &= \sum_{j=1}^{J} S_{i,j}\\ \sum_{i=1}^{I} v_{i,j}\mu\chi_{1,i}\chi_{2,j} &= \sum_{i=1}^{I} S_{i,j}\end{aligned}$$

Gaussian approximation

We consider claims ratio in risk class (i, j) defined by

$$R_{i,j} = S_{i,j}/v_{i,j}$$

- **Expected value for this claim ratio**

$$E[R_{i,j}] = \mu\chi_{1,i}\chi_{2,j}$$

- Model that we consider:

$$X_{i,j} \stackrel{def}{=} \log R_{i,j} \sim \mathcal{N}\left(\beta_0 + \beta_{1,i} + \beta_{2,j}, \sigma^2\right)$$

  Consequently,

$$\begin{aligned} E[R_{i,j}] &= e^{\beta_0+\sigma^2/2}e^{\beta_{1,i}}e^{\beta_{2,j}} \\ \mathrm{var}(R_{i,j}) &= E[R_{i,j}]^2\left(e^{\sigma^2}-1\right) \end{aligned}$$

- The mean has a right multiplicative structure:

$$\text{Set } \mu = e^{\beta_0+\sigma^2/2}, \chi_{1,i} = e^{\beta_{1,i}} \text{ and } \chi_{2,j} = e^{\beta_{2,j}}$$

- Set $M = I \cdot J$ and define for $X_{i,j} = \log R_{i,j} = \log(S_{i,j}/v_{i,j})$ the vector

$$\boldsymbol{X} = (X_1, \ldots, X_M)' = (X_{1,1}, \ldots, X_{1,J}, \ldots, X_{I,1}, \ldots, X_{I,J})' \in \mathbb{R}^M.$$

  Index m will always refer to

$$m = m(i, j) = (i-1)J + j \in \{1, \ldots, M = I \cdot J\}$$

  We assume that $\boldsymbol{X}$ has a multivariate Gaussian distribution

$$\boldsymbol{X} \sim \mathcal{N}(Z\boldsymbol{\beta}, \Sigma)$$

  with

  - diagonal covariance matrix $\Sigma = \sigma^2 diag(w_1, \ldots, w_M)$
  - parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_{1,2}, \ldots, \beta_{1,I}, \beta_{2,2}, \ldots, \beta_{2,J})' \in \mathbb{R}^{r+1}$
  - design matrix $Z \in \mathbb{R}^{M\times(r+1)}$ with $r + 1 = I + J - 1$ (Z has full rank)

  such that for $m = m(i, j)$

$$E[X_{i,j}] = (Z\boldsymbol{\beta})_m = \beta_0 + \beta_{1,i} + \beta_{2,j}$$

  We initialize $\beta_{1,1} = \beta_{2,1} = 0$ and β_0 plays role of intercept.
  For weights w_m, one often sets $w_m = v_{i,j}^{-1}$, because then for $v_{i,j}$ large:

$$\mathrm{var}(R_{i,j}) = \mathrm{var}\left(e^{X_{ij}}\right) = E[R_{i,j}]^2\left(e^{\sigma^2/v_{i,j}}-1\right) \approx \frac{\sigma^2}{v_{i,j}}E[R_{i,j}]^2$$

- **Goodness-of-fit analysis**
  We assume homoscedasticity, i.e. identical weights $w_{i,j} = w$ and $\Sigma = \sigma^2 w\mathcal{I}$ which implies $\hat{\beta}^{MLE} = (Z'Z)^{-1}Z'\boldsymbol{X}$.

  - **Total sum of squares:**

  $$\begin{aligned} SS_{tot} &= \sum_m \left(X_m - \overline{X}\right)^2 = \sum_m \left(\hat{X}_m - \overline{X}\right)^2 + \sum_m \left(X_m - \hat{X}_m\right)^2 \\ &= SS_{reg} + SS_{err} \end{aligned}$$

  with $\overline{X} = \frac{1}{M}\sum\limits_{m=1}^{M} X_m$ and $\hat{\boldsymbol{X}} = Z\hat{\boldsymbol{\beta}}^{MLE}$

  - **Coefficient of determination $\boldsymbol{R^2}$:**

  $$R^2 = \frac{SS_{reg}}{SS_{tot}} = 1 - \frac{SS_{err}}{SS_{tot}} \in [0,1]$$

  - **Adjusted coefficient of determination $\boldsymbol{R_a^2}$:**

  $$R_a^2 = 1 - \frac{SS_{err}/(M-r-1)}{SS_{tot}/(M-1)} \in [0,1]$$

  - **The residual standard deviation $\boldsymbol{\sigma}$** is estimated by:

  $$\hat{\sigma}^2 = \frac{1}{M}\sum_m \left(X_m - \hat{X}_m\right)^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{M} = \frac{SS_{err}}{M}$$

    * Set $r = I + J - 2$, i.e. dimension of parameter $\boldsymbol{\beta}$ is $r+1$
    * $\hat{\sigma}^2$ is MLE for σ^2 and $\frac{M\hat{\sigma}^2}{\sigma^2}$ follows χ^2_{M-r-1} distribution.
  - **Unbiased variance parameter estimator:**

  $$\hat{s}^2 = \frac{M}{M-r-1}\hat{\sigma}^2$$

  - **Likelihood ratio test:**
    * We have $r + 1 = I + J - 1$ dimensional parameter vector given by

    $$\boldsymbol{\beta} = (\beta_0, \beta_{1,2}, \ldots, \beta_{1,I}, \beta_{2,2}, \ldots, \beta_{2,J})' \in \mathbb{R}^{r+1}$$

    * Note that model is invariant under permutation of parameters and components.
    * We define

    $$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_r)' \in \mathbb{R}^{r+1}$$

    so that we have ordering of components that is appropriate for next layout.
    * $H_0 : \beta_0 = \ldots = \beta_{p-1} = 0$ for given $p < r + 1$
      1. Calculate residual differences SS_{err}^{full} and $\hat{\sigma}_{full}$ in **full model** with $r+1$ dimensional parameter vector $\boldsymbol{\beta} \in \mathbb{R}^{r+1}$
      2. Calculate residual differences $SS_{err}^{H_0}$ and $\hat{\sigma}_{H_0}$ in **reduced model** $(\beta_p, \ldots, \beta_r)' \in \mathbb{R}^{r+1-p}$

* Denote the design matrix of the reduced model by Z_0; calculate the **likelihood ratio $\boldsymbol{\Lambda}$**

$$\begin{aligned}\Lambda &= \frac{\hat{f}_{H_0}(\boldsymbol{X})}{\hat{f}_{full}(\boldsymbol{X})} = \left(\frac{\hat{\sigma}_{H_0}}{\hat{\sigma}_{full}}\right)^{-M} \frac{exp\left\{-\frac{1}{2\hat{\sigma}^2_{H_0}}(\boldsymbol{X} - Z_0\hat{\boldsymbol{\beta}}^{MLE}_{H_0})'(\boldsymbol{X} - Z_0\hat{\boldsymbol{\beta}}^{MLE}_{H_0})\right\}}{exp\left\{-\frac{1}{2\hat{\sigma}^2_{full}}(\boldsymbol{X} - Z_0\hat{\boldsymbol{\beta}}^{MLE}_{full})'(\boldsymbol{X} - Z_0\hat{\boldsymbol{\beta}}^{MLE}_{full})\right\}}\\ &= \left(\frac{\frac{SS^{H_0}_{err}}{M}}{\frac{SS^{full}_{err}}{M}}\right)^{-M/2} = \left(\frac{SS^{H_0}_{err}}{SS^{full}_{err}}\right)^{-M/2} = \left(1 + \frac{SS^{H_0}_{err} - SS^{full}_{err}}{SS^{full}_{err}}\right)^{-M/2}\end{aligned}$$

  · Likelihood ratio test rejects null hypothesis for small values of Λ.

  · This is equivalent to rejection for large values of $\frac{SS^{H_0}_{err} - SS^{full}_{err}}{SS^{full}_{err}}$.

* This motivates to consider the following test statistic F

$$F = \frac{SS^{H_0}_{err} - SS^{full}_{err}}{SS^{full}_{err}} \frac{M-r-1}{p} = \frac{SS^{H_0}_{err} - SS^{full}_{err}}{p\hat{s}^2_{full}}$$

F has F distribution with $df_1 = p$ and $df_2 = M - r - 1$ degrees of freedom, hence we reject H_0 on significance level $1 - \alpha$ if

$$F > F^{\leftarrow}_{p,M-r-1}(\alpha)$$

Generalized Linear Models (GLM) and tariffication

- **Components of a GLM**
  Denote response Y_i and independent variables $X_i = (x_{i1}, \ldots, x_{ik})$ for $i = 1, \ldots, n$.

  1. **Random component**: Y_i $(1 \leq i \leq n)$ independent with density from the exponential family, i.e

  $$f(y;\theta,\phi) \;=\; \exp\left(\frac{\theta y - b(\theta)}{a(\phi)} + c(y,\phi)\right).$$

  ϕ is dispersion parameter and functions $b(.)$, $a(.)$ and $c(.,.)$ are known.

  2. **Systematic component** (linear predictor):

  $$\eta_i = \beta_1 x_{i1} + \ldots + \beta_k x_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta} \text{ with } \boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)$$

  3. **Link function**: monotone and differentiable link function $g(\mu_i) = \eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ which combines the linear predictor with the mean $\mu_i = E[Y_i]$ (link between random and systematic component).
  g is chosen to ensure that the estimated parameter lies in the admissible space of values.

- **Exponential dispersion family**
  $X \sim f_X$ belongs to exponential dispersion family if f_X is of the form

$$f_X(x;\theta,\phi) = e^{\frac{x\theta - b(\theta)}{\phi/w} + c(x,\phi,w)}$$

Write $X \sim EDF(\theta, \phi, w, b(.))$ where

- $w > 0$ is given weight
- $\phi > 0$ is dispersion parameter
- $\theta \in \Theta$ is unknown parameter of distribution
- $\Theta \subset \mathbb{R}$ is open set op possible parameters θ
- $b : \Theta \to \mathbb{R}$ is cumulant function
- $c(., ., .)$ is normalisation, not depending on θ

f_X can be a density in absolutely continuous sense, probability weights in discrete case or mixture.

- **Lemma - Moment generating function:** Choose fixed $b(.)$ and assume that $EDF(\theta, \phi, w, b(.))$ gives well-defined densities with identical supports for all θ in open set Θ. Assume that for any $\theta \in \Theta$ there exists neighbourhood of zero such that mgf $M_X(r)$ of $X \sim EDF(\theta, \phi, w, b(.))$ is finite in this neighbourhood of zero (for r). Then we have for all $\theta \in \Theta$ and r sufficiently close to zero

$$M_X(r) = e^{\frac{b(\theta + r\phi/w) - b(\theta)}{\phi/w}}$$

- **Corollary:** Same assumptions as in Lemma and in addition we assume that $b \in C^2$ in interior of Θ. Then we have

$$E[X] = b'(\theta) \qquad \text{and} \qquad \text{var}(X) = \frac{\phi}{w} b''(\theta)$$

- **Examples of link functions**

  1. **log** :
  $$g(\mu) = \log(\mu)$$
  2. **logit**:
  $$g(\mu) = \log\left\{\frac{\mu}{1-\mu}\right\}$$
  3. **probit**:
  $$g(\mu) = \Phi^{-1}(\mu)$$
  where $\Phi(.)$ is the normal cumulative distribution function.
  4. **log-log** :
  $$g(\mu) = \log(-\log\mu)$$
  5. **complementary log-log**:
  $$g(\mu) = \log\{-\log(1-\mu)\}.$$

- **Generalized linear models - goal**
  Aim is to express expected claim of risk class (i, j) as expected number of claims times average claims
  $$E[S_{i,j}] = E[N_{i,j}]E[Y_{i,j}^{(\ell)}]$$
  - $N_{i,j}$ describes number of claims in risk class (i, j)

- $Y_{i,j}^{(\ell)}$ corresponding iid claim sizes for $\ell = 1, \dots, N_{i,j}$ in risk class (i, j)

- **GLM for Poisson claims counts**
  We assume that $N_{i,j}$ are independent with $N_{i,j} \sim \mathrm{Poi}(\lambda_{i,j} v_{i,j})$ and $v_{i,j}$ counting number of policies in risk class (i, j).
  - **Proposition:** Solution to MLE problem in Poisson case is given by solution of
    $$Z' V e^{Z\boldsymbol{\beta}} = Z' V \boldsymbol{X}$$
  - Poisson case is rewritten as
    $$Z' V e^{Z\boldsymbol{\beta}^{MLE}} - Z' N = 0$$

- **GLM for gamma claim sizes**
  We denote by $n_{i,j}$ number of observations $Y_{i,j}^{\ell}$ in risk class (i, j).
  - **Proposition:** Solution to MLE problem in gamma case is given by solution of
    $$Z' V_\theta e^{Z\boldsymbol{\beta}} = Z' V_\theta \boldsymbol{X}$$

- **Variable reduction analysis**
  - Having observations $\boldsymbol{X} = (X_1, \dots, X_M)'$ with independent components, we determine MLE $\hat{\beta}^{MLE}$ for $\beta \in \mathbb{R}^{r+1}$ within exponential dispersion family with log-link function g and design matrix $Z \in \mathbb{R}^{M \times (r+1)}$.
  - This provides estimate for mean
    $$\hat{\mu}_m = b'(\hat{\theta}_m) = e^{(Z\hat{\boldsymbol{\beta}}^{MLE})_m}$$
  - We define **inverse function** $h = (b')^{-1}$ which implies that $\hat{\theta}_m = h(\hat{\mu}_m)$.
  - **Log-likelihood function** at this estimate is then
    $$\ell_{\boldsymbol{X}}(\hat{\boldsymbol{\mu}}) = \sum_m \frac{X_m h(\hat{\mu}_m) - b(h(\hat{\mu}_m))}{\phi / w_m} + c(X_m, \phi, w_m)$$
    where we assume that $\phi_m = \phi$ for all $m = 1, \dots, M$.
  - Consider model $Z\boldsymbol{\beta}$ and compare it to **saturated model** which has as many parameters as observations
    $$\ell_{\boldsymbol{X}}(\boldsymbol{X}) = \sum_m \frac{X_m h(X_m) - b(h(X_m))}{\phi / w_m} + c(X_m, \phi, w_m)$$
  - **Scaled deviance:**
    $$\begin{aligned} D^*(\boldsymbol{X}, \hat{\boldsymbol{\mu}}) &= 2(\ell_{\boldsymbol{X}}(\boldsymbol{X}) - \ell_{\boldsymbol{X}}(\hat{\boldsymbol{\mu}})) \\ &= \frac{2}{\phi} \sum_m w_m \left[X_m h(X_m) - b(h(X_m)) - X_m h(\hat{\mu}_m) + b(h(\hat{\mu}_m))\right] \end{aligned}$$
  - **Deviance statistics:**
    $$D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}) = \phi D^*(\boldsymbol{X}, \hat{\boldsymbol{\mu}}) = 2\phi(\ell_{\boldsymbol{X}}(\boldsymbol{X}) - \ell_{\boldsymbol{X}}(\hat{\boldsymbol{\mu}}))$$

- $H_0 : \beta_0 = \ldots = \beta_{p-1} = 0$ for given $p < r+1$
  1. Calculate deviance statistics $D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{full})$ in full model $\boldsymbol{\beta} \in \mathbb{R}^{r+1}$
  2. Calculate deviance statistics $D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{H_0})$ under H_0
- **Test statistic $\boldsymbol{F}$**

$$F = \frac{D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{H_0}) - D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{full})}{D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{full})} \frac{M-r-1}{p} \geq 0$$

- F is approximated by F-distribution with df given by $df_1 = p$ and $df_2 = M - r - 1$.
- **Second test statistic $\boldsymbol{X^2}$**

$$X^2 = D^*(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{H_0}) - D^*(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{full}) \geq 0$$

- X^2 is approximately χ^2-distributed with $df = p$
- For this test statistic, we need to estimate the dispersion parameter ϕ.

• **Dispersion parameter**

- Assume that θ_m was estimated by $\hat{\theta}_m$, then we can estimate ϕ from **Pearson residuals**

$$\hat{\phi}_P = \frac{1}{M-r-1} \sum_m w_m \frac{(X_m - b'(\hat{\theta}_m))^2}{b''(\hat{\theta}_m)}$$

- Alternative is to use **deviances** and estimate

$$\hat{\phi} = \frac{D(\boldsymbol{X}, \hat{\boldsymbol{\mu}}_{full})}{M-r-1}$$

• **Accuracy of the model**

- **Pearson's residuals**

$$r_{P,m} = \frac{X_m - b'(\hat{\theta}_m)}{\sqrt{b''(\hat{\theta}_m)/w_m}}$$

- **Deviance residuals**

$$r_{D,m} = sgn(X_m - b'(\hat{\theta}_m)) \sqrt{2 w_m \left[X_m(h(X_m) - \hat{\theta}_m) - b(h(X_m)) + b(\hat{\theta}_m) \right]}$$

  for $m = 1, \ldots, M$.

Claims reserving

• **Notation:**

- i for accident year and j for development year (with $1 \leq i, j \leq n$).
- Y_{ij} incremental claim at end of development year j of accident year i.
- $C_{ij} = \sum_{k=1}^{j} Y_{ik}$ the cumulative claim.
- Outstanding claims reserves: $R_i = C_{in} - C_{i,n-i+1}$
  where C_{in} is ultimate claim amount of accident year i.
- Outstanding overall reserve: $R = \sum_{i=1}^{n} R_i$

- **Chain ladder method:**
  - Uses cumulative data and assumes existence of **development factors** $\{f_j | j = 2, \ldots, n\}$ such that

    $$E\left[C_{i,k+1} | C_{i1}, \ldots, C_{ik}\right] = C_{ik} f_{k+1} \qquad 1 \le i \le n, 1 \le k \le n-1.$$

    and different accident years i are independent.
  - Factors are estimated by the chain ladder method as

    $$\hat{f}_j = \frac{\sum_{i=1}^{n-j+1} C_{ij}}{\sum_{i=1}^{n-j+1} C_{i,j-1}} \qquad j = 2, \ldots, n.$$

  - Produce forecasts of future values of cumulative claims:

    $$\begin{aligned} \hat{C}_{i,n-i+2} &= C_{i,n-i+1} \hat{f}_{n-i+2} \qquad i = 2, \ldots, n \\ \hat{C}_{i,k} &= \hat{C}_{i,k-1} \hat{f}_k \qquad k = n-i+3, \ldots, n. \end{aligned}$$

- **Verbeek's algorithm:**
  - First equalities
    * $\hat{\alpha}_1 \sum_{j=1}^{n} \hat{\beta}_j = RS_1 \rightarrow \hat{\alpha}_1 = RS_1$
    * $\hat{\alpha}_1 \hat{\beta}_n = CS_n \rightarrow \hat{\beta}_n = \frac{CS_n}{\hat{\alpha}_1}$
  - Assume OK for $l < n$:
    $\hat{\beta}_{l+1}, \ldots, \hat{\beta}_n$ and $\hat{\alpha}_1, \ldots, \hat{\alpha}_{n-l}$ are found, then
    * $\hat{\alpha}_{n-l+1} (\sum_{j=1}^{l} \hat{\beta}_j) = RS_{n-l+1}$
      $\rightarrow \hat{\alpha}_{n-l+1} = \frac{RS_{n-l+1}}{1 - \sum_{j=l+1}^{n} \hat{\beta}_j}$
    * $\sum_{i=1}^{n-l+1} \hat{\alpha}_i \hat{\beta}_l = CS_l$
      $\rightarrow \hat{\beta}_l = \frac{CS_l}{\sum_{i=1}^{n-l+1} \hat{\alpha}_i}$
  - Repeat steps for $l+1, \ldots, n$

PART D: PREMIUM CALCULATION PRINCIPLES

Simple risk-based principles

Consider 2 different portfolios S_1 and S_2 with same mean $E[S_1] = E[S_2]$ with

1. $S_1 \sim \Gamma(\gamma, c)$ with mean $E[S_1] = \gamma / c$
2. $S_2 \equiv \gamma / c$ a constant

- **Variance loading principle**
  Choose a fixed constant $\alpha > 0$ and define the insurance premium π by

  $$\pi = E[S] + \alpha \operatorname{var}(S)$$

- $$\pi_1 = E[S_1] + \alpha \operatorname{var}(S_1) = \frac{\gamma}{c} + \alpha\frac{\gamma}{c^2} > \frac{\gamma}{c} = E[S_2] + \alpha \operatorname{var}(S_2) = \pi_2$$
- Assume that $r_{fx} > 0$ is the deterministic exchange rate between 2 different currencies, then we obtain ($r_{fx} \neq 0$)

$$\pi_{fx} = E[r_{fx}S] + \alpha \operatorname{var}(r_{fx}S) = r_{fx}E[S] + r_{fx}^2\alpha \operatorname{var}(S) \neq r_{fx}\pi.$$

- **Standard deviation loading principle**
  Choose fixed constant $\alpha > 0$ and define insurance premium π by

$$\pi = E[S] + \alpha \operatorname{var}(S)^{1/2} = E[S](1 + \alpha \mathrm{Vco}(S))$$

  - $$\pi_1 = E[S_1] + \alpha \operatorname{var}(S_1)^{1/2} = \frac{\gamma}{c} + \alpha\frac{\gamma^{1/2}}{c^2} > \frac{\gamma}{c} = E[S_2] + \alpha \operatorname{var}(S_2)^{1/2} = \pi_2$$
  - $$\pi_{fx} = E[r_{fx}S] + \alpha \operatorname{var}(r_{fx}S)^{1/2} = r_{fx}E[S] + r_{fx}^2\alpha \operatorname{var}(S)^{1/2} = r_{fx}\pi.$$

Utility pricing principles

- **Definition preference ordering:** Assume $u : I \to \mathbb{R}$ strictly increasing and strictly concave on $I \subset \mathbb{R}$, then we prefer position $X \in \chi$ over position $Y \in \chi$, write $X \succeq Y$ if

$$E[u(X)] \geq E[u(Y)].$$

- Strictly increasing implies that if $X \geq Y$ Pr-a.s. and $X > Y$ with positive Pr probability we have

$$E[u(X)] > E[u(Y)]$$

- Strict concavity implies that we can apply **Jensen's inequality**:

$$E[u(X)] \leq u(E[X])$$

- Thus, strict concavity and increasing property of u implies that the policyholder is willing to pay any premium π in non-empty interval

$$\pi \in \big(E[Y], c_0 - u^{-1}(E[u(c_0 - Y)])\big)$$

  to improve his happiness position.

- **Popular utility functions**
  - **Exponential utility function**, constant absolute risk-aversion (CARA) utility function (defined on $I = \mathbb{R}$)

$$u(x) = 1 - \frac{1}{\alpha}e^{-\alpha x} \qquad \text{for } \alpha > 0$$

  - **Power utility function**, constant relative risk-aversion (CRRA) utility function (defined on $I = \mathbb{R}_+$)

$$u(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma} & \gamma \neq 1 \\ \log x & \gamma = 1 \end{cases}$$

- **Definition utility indifference price:** Utility indifference price $\pi = \pi(u, F_S, c_0) \in \mathbb{R}$ for utility function u, initial capital $c_0 \in I$ and law of risky position S is given by solutions of

$$u(c_0) = E[u(c_0 + \pi - S)]$$

- **Corollary:** Utility indifference price $\pi = \pi(u, F_S, c_0)$ for initial capital c_0, risk-averse utility function u and risky position S satisfies

$$\pi = \pi(u, F_S, c_0) > E[S]$$

- **Proposition:** Assume that $u \in C^2$ is a risk-averse utility function on $\mathbb{R}$. The following two are equivalent
  - utility indifference prices $\pi = \pi(u, F_S, c_0)$ do not depend on c_0 for all S
  - utility function is of the form

$$u(x) = a - be^{-cx} \qquad \forall a \in \mathbb{R}; b, c > 0.$$

- **Insights into risk-aversion**
  - **Absolute and relative risk-aversions of u:**

$$\begin{aligned} \rho_{ARA}(x) &= \rho^u_{ARA}(x) = -\frac{u''(x)}{u'(x)} \\ \rho_{RRA}(x) &= \rho^u_{RRA}(x) = -x\frac{u''(x)}{u'(x)} \end{aligned}$$

  - The exponential utility function (CARA) with $\alpha > 0$ satisfies for all $x \in \mathbb{R}$

$$\rho_{ARA}(x) = \alpha$$

  - The power utility function (CRRA) with $\gamma > 0$ satisfies for all $x \in \mathbb{R}_+$

$$\rho_{RRA}(x) = \gamma$$

  - Assume u and v are two utility functions that are defined on same interval I. Then, u is more risk-averse than v if for any X with range I we have

$$u^{-1}\left(E[u(X)]\right) \leq v^{-1}\left(E[v(X)]\right)$$

- **Proposition:** Assume that u and v are twice differentiable utility functions defined on same interval $I \subset \mathbb{R}$. Following are equivalent
  - u is more risk-averse than v on I
  - $\rho^u_{ARA}(x) \geq \rho^v_{ARA}(x)$ for all $x \in I$.
- **Corollary:** Assume u is more risk-averse than v. Then we have for the utility indifference prices

$$\pi(u, F_S, c_0) \geq \pi(v, S, c_0)$$

- **Theorem:** Assume $u \in C^3$ is risk-averse utility function on I. The following are equivalent
  - $\pi(u, F_S, c_0)$ is decreasing in c_0 for all S
  - ρ^u_{ARA} is decreasing for all $x \in I$.

Esscher premium

- **Esscher (probability) distribution F_α for $\alpha > 0$ of F**

$$F_\alpha(s) = \frac{1}{M_S(\alpha)} \int_{-\infty}^{s} e^{\alpha x} dF(x)$$

under the additional assumption that the mfg $M_S(\alpha)$ of S exists in α. Note that this defines a normalized distribution function F_α.

- **Esscher premium:** Choose $S \sim F$ and assume that there exists $r_0 > 0$ such that $M_S(r) < \infty$ for all $r \in (-r_0, r_0)$. The Esscher premium π_α of S in $\alpha \in (0, r_0)$ is defined by

$$\pi_\alpha = E_\alpha[S] = \int_{\mathbb{R}} s dF_\alpha(s)$$

- **Corollary:**

$$\pi_\alpha = \frac{d}{dr} \log M_S(r)|_{r=\alpha} \geq E[S]$$

where inequality is strict for non-deterministic S.

Probability distortion pricing principles

- Assume that $S \sim F$ with $S \geq 0$ Pr-a.s., then

$$E[S] = \int_0^\infty x dF(x) = \int_0^\infty \Pr[S > x] dx$$

- **Definition - probability distorted price:** Assume that $h : [0,1] \to [0,1]$ is continuous, increasing and concave function with $h(0) = 1, h(1) = 1$ and $h(p) > p$ for all $p \in (0,1)$. The probability distorted price π_h of $S \geq 0$ is defined by (subject to existence)

$$\pi_h = E_h[S] = \int_0^\infty h\left(\Pr[S > x]\right) dx$$

- We obtain risk loading that provides

$$E[S] = \int_0^\infty \Pr[S > x] dx \leq \int_0^\infty h\left(\Pr[S > x]\right) dx = E_h[S] = \pi_h$$

where inequality is strict for non-deterministic S.

Cost-of-capital principles

Denote by $\chi \subset L^1(\Omega, \mathcal{F}, \Pr)$ the set of (risky) positions X of interest.

- A **risk measure** ϱ on χ is a mapping

$$\varrho : \chi \to \mathbb{R} \qquad \text{with } X \mapsto \varrho(X)$$

- **Shareholders'/investors' expected return** is

$$r_{CoC}\varrho(S - E[S]) > 0$$

on their investment $\varrho(S - E[S]) > 0$.

- **Cost-of-capital pricing principle**

$$\pi_{CoC} = E[S] + r_{CoC}\varrho(S - E[S]).$$

- **Properties of risk measures - Coherent measures**
Assume that χ is a convex cone containing $\mathbb{R}$, i.e. it satisfies

  - $c \in \chi$ for all $c \in \mathbb{R}$
  - $X + Y \in \chi$ for all $X, Y \in \chi$
  - $\lambda X \in \chi$ for all $X \in \chi$ and $\lambda > 0$

  Assume that ϱ is a risk measure on a convex cone χ containing $\mathbb{R}$. Then we define for $X, Y \in \chi, c \in \mathbb{R}$ and $\lambda > 0$

  1. normalization: $\varrho(0) = 0$
  2. monotonicity: for X, Y with $X \leq Y$ Pr-a.s., we have $\varrho(X) \leq \varrho(Y)$
  3. translation invariance: for all X and every c we have $\varrho(X + c) = \varrho(X) + c$
  4. positive homogeneity: for all X and for every $\lambda > 0$ we have $\varrho(\lambda X) = \lambda\varrho(X)$
  5. subadditivity: for all X, Y we have $\varrho(X + Y) \leq \varrho(X) + \varrho(Y)$

  Risk measure ϱ is called **coherent** if it satisfies all these properties 1-5.

- **Properties of risk measures - Translation invariance**

  - If we hold risky position X and if we inject capital $c > 0$ then loss is reduced to $X - c$. This implies for risk measure ϱ that reduced position satisfies

  $$\varrho(X - c) = \varrho(X) - c$$

  - This justifies definition of regulatory risk measure as stated above.
  - If we sell risky portfolio S and we collect pure risk premium $E[S]$, then risk of residual loss $S - E[S]$ is given by
  $$\varrho(S - E[S]) = \varrho(S) - E[S]$$

- **Properties of risk measures - Normalisation and translation invariance**

  - Balance sheet of insurance company is called acceptable if its (future) surplus $C_1 \in \chi$ satisfies $\varrho(-C_1) \leq 0$.
  - Assume that insurance company sells policy S at price $\pi \geq E[S]$ and at same time it has initial capital $c_0 = \varrho(S - E[S]) \geq 0$. Then future surplus of company is given by $C_1 = c_0 + \pi - S$. Regulator then checks acceptability condition $\varrho(-C_1) = \varrho(-(c_0 + \pi - S)) = -c_0 - \pi + \varrho(S) = -\pi + E[S] \leq 0$.

- Assume that an initial capital $c_0 > 0$ is provided by investor who expects cost-of-capital rate $r_{CoC} > r_0$ on investment. Thus insurance company also needs to finance cost-of-capital cash flow $r_{CoC}c_0 = r_{CoC}\varrho(S - E[S])$ to investor. This can exactly be done with cost-of-capital premium π_{CoC} and insurance keeps its acceptable position if $r_{CoC}c_0$ is also considered as a liability of insurance company.

- **Properties of risk measures - Monotonicity and normalization**
  - Imply that more risky positions are charged with higher capital requiremens and, in particular, if we have only downside risks, i.e. $X \geq 0$ Pr-a.s., then we will have positive capital charges $\varrho(X) \geq \varrho(0) = 0$

- **Popular risk measures**
  - The **standard deviation risk measure** is for S with finite second moment given by
  $$\varrho(S) = \alpha\sigma(S) = \alpha \operatorname{var}(S)^{1/2}$$
  - The **Value-at-Risk (VaR)** of $S \sim F$at security level $1 - q \in (0, 1)$ is given by left-continuous generalised inverse of F at $1 - q$
  $$\varrho(S) = VaR_{1-q}(S) = F^{\leftarrow}(1 - q)$$
  - The **expected shortfall** is for $S \sim F$ with F continuous
  $$\varrho(S) = TVaR_{1-q}(S) = E[S|S \geq VaR_{1-q}(S)] = \frac{1}{q}\int_{1-q}^{1} VaR_u(S)du = ES_{1-q}(S)$$
  Here the cost-of-capital pricing principle is given by
  $$\pi = E[S] + r_{CoC}ES_{1-q}(S - E[S]) = E[S] + r_{CoC}(ES_{1-q}(S) - E[S]).$$

Deflator based pricing principles

- Assume that φ is integral and strictly positive rv with
$$E[\varphi] = d_0 = \frac{1}{1 + r_0} \in (0, 1]$$
d_0 can be seen as deterministic discount factor and $r_0 \geq 0$ as deterministic risk-free rate.

- **Definition deflator based pricing:** Fix $\varphi \in L^1(\Omega, \mathcal{F}, \Pr)$ strictly positive with $d_0 = 1$ and assume that φ and S are positively correlated. Then we can define the deflator based price by
$$\pi_{\varphi}^{(0)} = E[\varphi S] \geq E[\varphi]E[S] = E[S]$$
It can be understood as a probability distortion principle because φ allows to define equivalent probability measure $\Pr^*$ by Radon-Nikodym derivative as follows
$$\frac{d\Pr^*}{d\Pr} = \varphi$$
because φ is strictly positive density wrt Pr for $d_0 = 1$. Then, we price S under equivalent probability measure $\Pr^*$ by
$$\pi_{\varphi}^{(0)} = E[\varphi S] = E^*[S].$$