Mixture Distributions

1. Characterisation
   - Each policy has distribution in same family, $f(x;\theta)$
   - However, i^{th} policy has $\theta = \theta_i$
   - Distribution of θ_i's in porfolio: $\theta \sim g(t;\eta)$.
   - Claim generation from portfolio perspective:

Choose Random Policy	$\rightarrow$	Random Claim Amount from Chosen Policy
$[\theta_i \sim g(t;\eta)]$		$[X\vert\theta_i \sim f(x;\theta_i)]$

   - "Portfolio-Wide" (Mixture Distribution) *pdf*:
     - · IDEA:

$$Pr(\text{Claim} = x) = \sum_i Pr(\text{Claim} = x|\text{Policy } i)Pr(\text{Policy } i)$$

     - · FORMALLY:

$$f_X(x;\eta) = \int_\Theta f(x;t)g(t;\eta)dt,$$

where Θ is set of possible θ values; usually $(0,\infty)$.

The Pareto Distributions

1. Characterisation

- Probability Density Function *(pdf)*:

  · Claims for policy i are exponential with (mean) parameter θ_i

  · θ_i's distributed in portfolio according to:

$$g(t;\alpha,\delta) = \frac{\delta^\alpha}{\Gamma(\alpha)} t^{-(\alpha+1)} \exp\left(-\frac{\delta}{t}\right)$$

  the *inverse Gamma distribution*

  (derived as distribution of $Y = X^{-1}$ when $X \sim$ Gamma)

  · Mixture distribution *pdf*:

$$\begin{aligned} f_X(x;\alpha,\delta) &= \int_0^\infty t^{-1} e^{-x/t} \frac{\delta^\alpha}{\Gamma(\alpha)} t^{-(\alpha+1)} e^{-\delta/t} dt \\ &= \frac{\delta^\alpha}{\Gamma(\alpha)} \int_0^\infty t^{-(\alpha+2)} e^{-(x+\delta)/t} dt \\ &= \frac{\delta^\alpha \Gamma(\alpha+1)}{\Gamma(\alpha)(x+\delta)^{\alpha+1}} \int_0^\infty \frac{(x+\delta)^{\alpha+1}}{\Gamma(\alpha+1)} t^{-(\alpha+2)} e^{-(x+\delta)/t} dt \\ &= \frac{\alpha\delta^\alpha}{(x+\delta)^{\alpha+1}} \end{aligned}$$

  · Mixture distribution *CDF*:

$$F_X(x;\alpha,\delta) = \int_0^x \frac{\alpha\delta^\alpha}{(u+\delta)^{\alpha+1}} du = 1 - \left(\frac{\delta}{x+\delta}\right)^\alpha$$

1. Characterisation *(Continued)*

- Moments:

$$E_{\alpha,\delta}(X) = E_{\alpha,\delta}\{E(X|\theta)\} = E_{\alpha,\delta}(\theta) = \frac{\delta}{\alpha - 1}, \quad \alpha > 1$$

$$E_{\alpha,\delta}(X^2) = E_{\alpha,\delta}\{E(X^2|\theta)\} = E_{\alpha,\delta}(2\theta^2) = \frac{2\delta^2}{(\alpha - 1)(\alpha - 2)}, \ \alpha > 2$$

- Quantiles (Percentiles):

$$\begin{aligned} x_p \text{ solves} : Pr\{X \leq x_p\} &= p \\ \Longrightarrow \quad 1 - \frac{\delta^\alpha}{(x_p + \delta)^\alpha} &= p \\ \Longrightarrow \quad x_p &= \delta\{(1-p)^{-1/\alpha} - 1\} \end{aligned}$$

2. Estimation of Parameters based on $x_1, \ldots, x_n$:

   - (Standard) Method of Moments (*MOM*):

     Solve system:

$$\frac{\delta}{\alpha - 1} = \overline{x}, \qquad \frac{2\delta^2}{(\alpha-1)(\alpha-2)} = \overline{x^2},$$

     yeilds solution:

$$\hat{\alpha}_{MOM} = \frac{2(\overline{x^2} - \overline{x}^2)}{\overline{x^2} - 2\overline{x}^2} = \frac{2(n-1)s^2}{(n-1)s^2 - n\overline{x}^2}$$
$$\hat{\delta}_{MOM} = \frac{\overline{x}\overline{x^2}}{\overline{x^2} - 2\overline{x}^2} = \frac{(n-1)\overline{x}s^2 + n\overline{x}^3}{(n-1)s^2 - n\overline{x}^2}$$

     · The MOM estimates for our data are found to be:

$$\hat{\alpha}_{MOM} = 2.476, \qquad \hat{\delta}_{MOM} = 4412.3$$

2. Estimation of Parameters based on $x_1, \ldots, x_n$ *(Continued)*:

   - Method of Percentiles (*MOP*):

     Solve:

     $$\left\{\delta\{(1-p_1)^{-1/\alpha} - 1\right\} = \hat{x}_{p_1}; \qquad \delta\left\{(1-p_2)^{-1/\alpha} - 1\right\} = \hat{x}_{p_2},$$

     for some choice of p_1 and p_2

     Requires iterative (computer-based) solution methods.

     For our data, *MOP* estimates based on upper and lower quartiles ($p_1 = 0.25$, $p_2 = 0.75$) are:

     $$\hat{\alpha}_{MOP} = 1.576 \qquad \hat{\delta}_{MOP} = 2002.38$$

2. Estimation of Parameters based on $x_1, \ldots, x_n$ *(Continued)*:

- Maximum Likelihood Estimate (*MLE*)

  Log-Likelihood Function:

$$l(\alpha, \delta) = n \ln \alpha + n\alpha \ln \delta - (\alpha + 1) \sum_{i=1}^{n} \ln(x_i + \delta)$$

  Score equations:

$$\frac{n}{\alpha} + n \ln \delta - \sum_{i=1}^{n} \ln(x_i + \delta) = 0,$$
$$\frac{n\alpha}{\delta} - (\alpha + 1) \sum_{i=1}^{n} (x_i + \delta)^{-1} = 0$$

  · *MLE*s require iterative (computer-based) solution method.

  · For our data:

$$\hat{\alpha}_{MLE} = 1.909; \qquad \hat{\delta}_{MLE} = 2704.47$$

3. Goodness-of-Fit Testing:

- Pearson Chi-Squared Test:

  · Use equal-count bins from previous exponential calculations.

Table 2.6: Observed and Expected Claim Amounts (in £'s) using the Pareto Distribution

Bin Range	O_i	$E_{i,MLE}$	$E_{i,MOM}$	$E_{i,MOP}$	Bin Range	O_i	$E_{i,MLE}$	$E_{i,MOM}$	$E_{i,MOP}$
0-260	12	15.4	12.7	16.8	2072-2618	5	6.0	6.7	5.6
260-545	18	12.9	11.4	13.5	2618-3285	6	5.3	6.1	4.9
545-860	10	10.9	10.2	11.0	3285-4145	6	4.8	5.9	4.4
860-1212	8	9.3	9.1	9.1	4145-5357	3	4.3	5.2	4.0
1212-1612	7	8.0	8.2	7.7	5357-7429	4	4.2	5.1	4.0
1612-2072	10	6.9	7.4	6.5	7429+	7	7.7	8.3	8.3

$X^2_{MLE} = 5.59$, $df = 12 - 1 - 2 = 9$, p-value $= 0.780$

$X^2_{MOM} = 7.00$, $df = 12 - 1 - 2 = 9$, p-value $= 0.639$

$X^2_{MOP} = 6.40$, $df = 12 - 1 - 2 = 9$, p-value $= 0.699$

3. Goodness-of-Fit Testing *(Continued)*:

- Pearson Chi-Squared Test *(Continued)*:

  · Calculating $E_{i,MLE}$'s:

$$\begin{aligned} E_{(a,b)} &= nPr_{\alpha,\delta}(a < X \le b) = n\{F_X(b;\alpha,\delta) - F_X(a;\alpha,\delta)\} \\ &= n\left\{\frac{\delta^\alpha}{(a+\delta)^\alpha} - \frac{\delta^\alpha}{(b+\delta)^\alpha}\right\} \end{aligned}$$

  · Using MLEs yields:

$$E_{(0,260.15),MLE} = 96\left\{1 - \frac{2704.47^{1.909}}{(260.15 + 2704.47)^{1.909}}\right\} = 15.44$$

3. Goodness-of-Fit Testing *(Continued)*:

- Pearson Chi-Squared Test *(Continued)*:

  · Equal-count bin construction, 12 bins each with $E_i = 8$:

  · For first bin, $(0, b)$, solve:

$$
\begin{aligned}
8 &= 96 Pr_{\alpha,\delta}(0 < X \leq b) \\
&= 96\left\{1 - \frac{\delta^{\alpha}}{(b+\delta)^{\alpha}}\right\} \\
\Longrightarrow \quad b &= \delta\left\{(11/12)^{-1/\alpha} - 1\right\}
\end{aligned}
$$

  Using $\hat{\alpha}_{MLE}$ and $\hat{\delta}_{MLE}$ yields $b = 126.12$.

  · For next bin, $(126.12, b)$, solve:

$$
96 = 8\left\{\frac{2704.47^{1.909}}{(126.12 + 2704.47)^{1.909}} - \frac{2704.47^{1.909}}{(b + 2704.47)^{1.909}}\right\}.
$$

  Yields $b = 271.03$. Continue.

  · $X^2_{MLE} = 3.25$, $df = 12 - 1 - 2 = 9$, p-value $= 0.953$

  · $X^2_{MOM} = 5.5$, $df = 12 - 1 - 2 = 9$, p-value $= 0.789$

  · $X^2_{MOP} = 7.75$, $df = 12 - 1 - 2 = 9$, p-value $= 0.559$

The Negative Binomial Distributions

1. Characterisation

   - Model for *NUMBER* of claims per policy.
   - Claims are rare, assume each policy has Poisson distribution
   - However, i^{th} policy has rate $\lambda = \lambda_i$
   - Distribution of λ_i's in porfolio: $\lambda \sim G(\alpha, \theta)$.
   - Claim generation from portfolio perspective:

Choose Random Policy	$\rightarrow$	Random Number of Claims from Chosen Policy
$[\lambda_i \sim G(\alpha, \theta)]$		$[N \mid \lambda_i \sim Pois(\lambda_i)]$

The Negative Binomial Distributions

1. Characterisation *(Continued)*
   - "Portfolio-Wide" (Mixture Distribution) *pmf*:

$$\begin{aligned} p_N(n;\alpha,\theta) &= \int_0^\infty Pr(N=n|\lambda)g(\lambda;\alpha,\theta)d\lambda \\ &= \int_0^\infty \frac{\lambda^n e^{-\lambda}}{n!}\frac{1}{\theta^\alpha\Gamma(\alpha)}\lambda^{\alpha-1}e^{-\lambda/\theta}d\lambda \\ &= \frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)}\left(\frac{1}{1+\theta}\right)^\alpha\left(\frac{\theta}{1+\theta}\right)^n. \end{aligned}$$

· Alternate form for *pdf* if $\alpha = k$ an integer and $p = (1+\theta)^{-1}$:

$$p_N(n;k,p) = \frac{(n+k-1)!}{n!(k-1)!}p^k(1-p)^n$$

The Negative Binomial Distributions

1. Characterisation *(Continued)*

   - Moments (calculation left as exercise):

$$E_{\alpha,\theta}(N) = \alpha\theta \qquad E_{\alpha,\theta}(N^2) = \alpha\theta(1+\theta+\alpha\theta)$$

   - Quantiles (Percentiles):

$$\begin{aligned} &x_p \text{ solves} : Pr\{N \leq x_p\} = p \\ &\implies \quad \sum_{n=0}^{x_p} \frac{\Gamma(\alpha+n)}{n!\Gamma(\alpha)} \left(\frac{1}{1+\theta}\right)^{\alpha} \left(\frac{\theta}{1+\theta}\right)^{n} = p \end{aligned}$$

   · Requires iterative (computer-based) solution methods

   · Since N is discrete, x_p not available for some choices of p.

2. Estimation of Parameters based on $n_1, \ldots, n_m$:

   - (Standard) Method of Moments (*MOM*):

     Solve system:

$$\alpha\theta = \overline{n}, \qquad \alpha\theta(1+\theta+\alpha\theta) = \overline{n^2},$$

     yeilds solution:

$$\hat{\alpha}_{MOM} = \frac{\overline{n}^2}{\overline{n^2} - \overline{n}^2 - \overline{n}} = \frac{m\overline{n}^2}{(m-1)s^2 - m\overline{n}}$$
$$\hat{\theta}_{MOM} = \frac{\overline{n^2} - \overline{n}^2}{\overline{n}} - 1 = \frac{(m-1)s^2}{m\overline{n}} - 1$$

2. Estimation of Parameters based on $n_1, \ldots, n_m$ *(Continued)*:

   - Method of Percentiles (*MOP*):

     Requires iterative (computer-based) solution methods.

     No solutions available for some choices of p_1 and p_2.

2. Estimation of Parameters based on $n_1, \ldots, n_m$ *(Continued)*:

- Maximum Likelihood Estimate (*MLE*)

  Log-Likelihood Function:

$$
\begin{aligned}
l(\alpha, \theta) &= \sum_{i=1}^{m} \ln\{p_N(n_i; \alpha, \theta)\} \\
&= -m \ln\{\Gamma(\alpha)\} - m\alpha \ln(1+\theta) + \sum_{i=1}^{m} \ln\{\Gamma(\alpha + n_i)\} \\
&\qquad + \sum_{i=1}^{m} n_i \{\ln\theta - \ln(1+\theta)\} - \sum_{i=1}^{m} \ln(n_i!)
\end{aligned}
$$

  Score equations:

$$
\begin{aligned}
-m\psi(\alpha) - m\ln(1+\theta) + \sum_{i=1}^{m} \psi(\alpha + n_i) &= 0, \\
-\frac{m\alpha}{1+\theta} + \sum_{i=1}^{m} n_i \left\{ \frac{1}{\theta} - \frac{1}{1+\theta} \right\} &= 0
\end{aligned}
$$

  where $\psi(x) = \frac{d}{dx} \ln\{\Gamma(x)\}$ is the digamma function

  · Second equation reduces to $\alpha\theta = \frac{1}{m} \sum_{i=1}^{m} n_i$

  · However, *MLE*s require iterative (computer-based) solution method.$\blacksquare$

2. Estimation of Parameters based on $n_1, \ldots, n_m$ *(Continued)*:

Table 2.7a: Observed Counts for Policies Making Various Numbers of Claims

Number of Claims Made	Observed Number of Policies
0	81056
1	16174
2	2435
3	295
4	36
5+	4

- $\overline{n} = 0.22093$, $\overline{n^2} = 0.29245$, $s^2 = 0.24364$.
- Under Poisson model for N:
  - $\hat{\lambda}_{MLE} = \hat{\lambda}_{MOM} = \overline{n} = 0.22093$.
  - MOP estimate not readily available.

    Percentiles of Poisson require

    iterative (computer-based) calculation.
- Under Negative Binomial (or Poisson-mixture) model for N:
  - $\hat{\alpha}_{MOM} = \frac{\overline{n}^2}{\overline{n^2} - \overline{n}^2 - \overline{n}} = 2.149$, $\quad \hat{\theta}_{MOM} = \frac{\overline{n^2} - \overline{n}^2}{\overline{n}} - 1 = 0.1028$
  - Using iterative (computer-based) solution:

$$\hat{\alpha}_{MLE} = 2.123 \qquad \hat{\theta}_{MLE} = 0.1041$$

3. Goodness-of-Fit Testing:

- Pearson Chi-Squared Test:

  · Data is already categorical, so bin choice less of an issue.

Table 2.7b: Observed and Expected Counts for Policies Making Various Numbers of Claims

Number of Claims Made	Observed Number of Policies	Expected Counts (Poisson)	Expected Counts (Neg. Bin.)
0	81056	80177.28	81035.34
1	16174	17713.57	16233.32
2	2435	1956.73	2382.57
3	295	144.10	307.16
4	36	7.96	36.86
5+	4	0.36	4.75

$X^2_{Poisson} = 553.35$, $df = 6 - 1 - 1 = 4$, p-value ≈ 0

$X^2_{Neg.Bin.} = 1.996$, $df = 6 - 1 - 2 = 3$, p-value $= 0.573$

3. Goodness-of-Fit Testing *(Continued)*:

   - Pearson Chi-Squared Test *(Continued)*:

     · Calculating $E_{i,Pois}$'s:

$$E_{i,Pois} = mPr_{\lambda}(N = i) = m\frac{\lambda^i e^{-\lambda}}{i!}$$

     · Using MLE or MOM yields:

$$E_{0,Pois} = 100000\frac{(0.22093)^0 e^{-0.22093}}{0!} = 80177.28$$
$$E_{1,Pois} = 100000\frac{(0.22093)^1 e^{-0.22093}}{1!} = 17713.57$$
$$\vdots$$
$$E_{5+,Pois} = 100000 - \sum_{i=0}^{4} E_{i,Pois} = 0.36$$

3. Goodness-of-Fit Testing *(Continued)*:

- Pearson Chi-Squared Test *(Continued)*:

  · Calculating $E_{i,Neg.Bin.}$'s:

$$E_{i,Neg.Bin} = mPr_{\alpha,\theta}(N = i) = m\frac{\Gamma(\alpha + i)}{i!\Gamma(\alpha)}\left(\frac{1}{1+\theta}\right)^{\alpha}\left(\frac{\theta}{1+\theta}\right)^{i}$$

  · Using MOMs yields:

$$\begin{aligned}
E_{0,Neg.Bin.} &= 100000\frac{\Gamma(2.149)}{0!\Gamma(2.149)}\left(\frac{1}{1.1028}\right)^{2.149}\left(\frac{0.1028}{1.1028}\right)^{0} \\
&= 81035.34 \\
E_{1,Neg.Bin.} &= 100000\frac{\Gamma(2.149+1)}{1!\Gamma(2.149)}\left(\frac{1}{1.1028}\right)^{2.149}\left(\frac{0.1028}{1.1028}\right)^{1} \\
&= 100000(2.149)(0.8103534)(0.093217) \\
&= 16233.32 \\
&\vdots \\
E_{5+,Neg.Bin.} &= 100000 - \sum_{i=0}^{4} E_{i,Neg.Bin.} = 4.75
\end{aligned}$$