The Log-Normal Distributions

1. Characterisation
   - If $Y \sim N(\mu, \sigma^2)$, then $X = e^Y \sim LN(\mu, \sigma^2)$.
   - Probability Density Function (*pdf*):

$$\begin{aligned} f_X(x;\mu,\sigma^2) &= \phi(\ln x;\mu,\sigma^2)\left|\frac{d\ln x}{dx}\right| \\ &= \frac{1}{x\sigma\sqrt{2\pi}}\exp\left\{-\frac{1}{2\sigma^2}(\ln x-\mu)^2\right\} \end{aligned}$$

   where $\phi(\cdot;\mu,\sigma^2)$ is normal density with mean μ, variance σ^2.
   - Moments:

$$\begin{aligned} E_{\mu,\sigma^2}(X) &= E_{\mu,\sigma^2}(e^Y) = m_Y(1) = \exp\left(\mu+\frac{\sigma^2}{2}\right) \\ E_{\mu,\sigma^2}(X^2) &= E_{\mu,\sigma^2}(e^{2Y}) = m_Y(2) = \exp(2\mu+2\sigma^2) \end{aligned}$$

   - Quantiles (Percentiles):

$$\begin{aligned} x_p \text{ solves} : & \; Pr\{X \le x_p\} = p \\ \implies & \; Pr\{\ln X \le \ln(x_p)\} = p \\ \implies & \; Pr\left\{\frac{Y-\mu}{\sigma} \le \frac{\ln(x_p)-\mu}{\sigma}\right\} = p \\ \implies & \; \frac{\ln(x_p)-\mu}{\sigma} = \Phi^{-1}(p) \\ \implies & \; x_p = \exp\{\mu+\sigma\Phi^{-1}(p)\}, \end{aligned}$$

   where $\Phi^{-1}(\cdot)$ is inverse *CDF* of standard normal distribution.

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

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

     Solve system:

     $$\exp\left(\mu + \frac{\sigma^2}{2}\right) = \overline{x}, \qquad \exp(2\mu + 2\sigma^2) = \overline{x^2},$$

     The MOM estimates are:

     $$\hat{\mu}_{MOM} = 2\ln(\overline{x}) - 0.5\ln\left(\overline{x^2}\right) \qquad \hat{\sigma}^2_{MOM} = \ln\left(\overline{x^2}\right) - 2\ln(\overline{x}).$$

     For our example, $\hat{\mu}_{MOM} = 7.09$ and $\hat{\sigma}^2_{MOM} = 1.825$.

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

- Method of Percentiles (*MOP*):

  Solve:

  $$\exp\{\mu + \sigma\Phi^{-1}(p_1)\} = \hat{x}_{p_1}; \qquad \exp\{\mu + \sigma\Phi^{-1}(p_2)\} = \hat{x}_{p_2},$$

  for some choice of p_1 and p_2

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

  $$\begin{array}{l} \mu + \sigma\Phi^{-1}(0.25) = \ln(\hat{x}_{0.25}) \\ \mu + \sigma\Phi^{-1}(0.75) = \ln(\hat{x}_{0.75}) \end{array}$$

  $$\Longrightarrow \qquad \begin{array}{c} \mu - 0.6745\sigma = \ln(401) \\ \mu + 0.6745\sigma = \ln(2836.75) \end{array}$$

  $$\Longrightarrow \qquad \begin{array}{c} \mu_{MOP} = \frac{1}{2}\{\ln(2836.75) + \ln(401)\} = 6.97 \\ \sigma^2_{MOP} = \left[\frac{1}{1.349}\{\ln(2836.75) - \ln(401)\}\right]^2 = 2.103 \end{array}$$

  since upper and lower quartiles of data are:

  $$\hat{x}_{0.25} = x_{[24]} + 0.25(x_{[25]} - x_{[24]}) = 401$$

  $$\hat{x}_{0.75} = x_{[72]} + 0.75(x_{[73]} - x_{[72]}) = 2836.75$$

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

- Maximum Likelihood Estimate (*MLE*)

  Log-Likelihood Function: (Left as Exercise)

  Can use "equivariance"

  · μ, σ^2 are mean, variance for normal data $y_i = \ln(x_i)$

  · $\hat{\mu}_{MLE} = \overline{y} = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) = \overline{\ln(x)}$

  · $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}\{y_i - \overline{y}\}^2 = \frac{1}{n}\sum_{i=1}^{n}\{\ln(x_i) - \overline{\ln(x)}\}^2$
  $= \overline{\ln(x)^2} - \overline{\ln(x)}^2$

  For our data,

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} \ln x_i = 7.021, \quad \overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} (\ln x_i)^2 = 51.258$$

  So, $\hat{\mu}_{MLE} = 7.021$ and $\hat{\sigma}^2_{MLE} = 51.258 - (7.021)^2 = 1.964$.

3. Goodness-of-Fit Testing:

- Pearson Chi-Squared Test:

  · Use equal-count bins from previous exponential calculations.

Table 2.4: Observed and Expected Claim Amounts (in £'s) using the Lognormal Distribution

Bin Range	O_i	$E_{i,MLE}$	Bin Range	O_i	$E_{i,MLE}$
0-260	12	14.2	2072-2618	5	5.6
260-545	18	14.9	2618-3285	6	4.9
545-860	10	11.7	3285-4145	6	4.4
860-1212	8	9.4	4145-5357	3	4.1
1212-1612	7	7.7	5357-7429	4	4.2
1612-2072	10	6.5	7429+	7	8.5

$X^2_{MLE} = 4.87$, $df = 12 - 1 - 2 = 9$, p-value $= 0.846$

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

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

  · Calculating $E_{i,MLE}$'s:

$$\begin{aligned}E_{(a,b)} &= nPr_{\mu,\sigma^2}(a < X \leq b) = nPr_{\mu,\sigma^2}(\ln a < \ln X \leq \ln b)\\&= nPr_{\mu,\sigma^2}\left(\frac{\ln a - \mu}{\sigma} < \frac{\ln X - \mu}{\sigma} \leq \frac{\ln b - \mu}{\sigma}\right)\\&= n\left\{\Phi\left(\frac{\ln b - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a - \mu}{\sigma}\right)\right\}\end{aligned}$$

  · Using MLEs yields:

$$\begin{aligned}E_{(0,260.15),MLE} &= n\left\{\Phi\left(\frac{\ln(260.15) - \hat{\mu}_{MLE}}{\hat{\sigma}_{MLE}}\right)\right.\\&\qquad\left. - \Phi\left(\frac{\ln(0) - \hat{\mu}_{MLE}}{\hat{\sigma}_{MLE}}\right)\right\}\\&= 96\left\{\Phi\left(\frac{5.561 - 7.021}{\sqrt{1.964}}\right) - \Phi(-\infty)\right\}\\&= 96\{\Phi(-1.042) - 0\}\\&= 14.2\end{aligned}$$

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:

$$
\begin{aligned}
8 &= 96 Pr_{\mu,\sigma^2}(0 < X \le b) \\
&= 96 Pr_{\mu,\sigma^2}\left(-\infty < \frac{\ln X - \mu}{\sigma} \le \frac{\ln b - \mu}{\sigma} \right) \\
&= 96 \Phi\left(\frac{\ln b - \mu}{\sigma} \right) \\
\Longrightarrow \quad b &= \exp\{\mu + \sigma \Phi^{-1}(8/96)\}
\end{aligned}
$$

Using $\hat{\mu}_{MLE}$ and $\hat{\sigma}^2_{MLE}$ yields $b = 161.23$

  · For next bin, $(161.23, b)$, solve:

$$
96\left\{ \Phi\left(\frac{\ln(b) - 7.021}{\sqrt{1.964}} \right) - \Phi\left(\frac{\ln(161.23) - 7.021}{\sqrt{1.964}} \right) \right\} = 8.
$$

Yields $b = 288.65$. Continue.

  · $X^2 = 3$, $df = 12 - 1 - 2 = 9$, p-value $= 0.964$