- 1. Distribution of "Effective" Claim Size, Y:
  - Expected Claim Size

· If 
$$X \sim f_X(x;\theta)$$
,

$$E_{\theta}(Y) = E_{\theta}\{XI_{(X \leq M)}\} + E\{MI_{(X > M)}\}$$

$$= \int_{0}^{\infty} xI_{(x \leq M)}f_{X}(x;\theta)dx + MPr_{\theta}(X > M)$$

$$= \int_{0}^{M} xf_{X}(x;\theta)dx + M\int_{M}^{\infty} f_{X}(x;\theta)dx$$

$$= \int_{0}^{\infty} xf_{X}(x;\theta)dx - \int_{M}^{\infty} xf_{X}(x;\theta)dx + M\int_{M}^{\infty} f_{X}(x;\theta)dx$$

$$= E_{\theta}(X) - \int_{M}^{\infty} (x - M)f_{X}(x;\theta)dx$$

$$= E_{\theta}(X) - \int_{0}^{\infty} yf_{X}(y + M;\theta)dy,$$

- 1. Distribution of "Effective" Claim Size, Y (Continued):
  - Moment Generating Function
    - · If X has  $mgf m_X(t)$ ,

$$m_Y(t) = \int_0^M e^{tx} f_X(x; \theta) dx + e^{tM} Pr_{\theta}(X > M)$$
  
=  $m_X(t) - e^{tM} \int_0^\infty (e^{ty} - 1) f_X(y + M; \theta) dy$ 

## 2. Modelling Claim Sizes:

- Key Issues
  - · Choose family for X-distribution,  $f_X(x;\theta)$ .
  - · Are  $X_i$  values recorded? Usually not.
  - · The  $Y_i$  values are Right-Censored.
  - ·  $Y_i$ 's not continuous;  $Pr(Y_i = M) = Pr(X_i > M) > 0$ .
  - $\cdot$  Modelling Y distribution directly very difficult (but possible)

- 2. Modelling Claim Sizes (Continued):
  - $\cdot$ Estimate X distribution parameters using Y data?
    - $\cdot$   $\overline{x},$   $\hat{x}_p$  generally not calculable from y 's
    - $\cdot$  MOM, MOP not feasible
    - $\cdot$  MLE concepts still feasible

- 2. Modelling Claim Sizes (Continued):
  - MLEs of X distribution parameters using Y data
    - · Likelihood concept:

$$L(\theta) = \prod_{i} Pr_{\theta}(i^{\text{th}} \text{ data point observed})$$

"Product of chances of observing each data point"

- · For n uncensored (i.e.,  $Y_i = X_i < M$ ); and m censored (i.e.,  $Y_i = M$ ) observations:
  - $\cdot$  Contribution to likelihood of n uncensored observations

$$L(\theta; x_1, \dots, x_n) = \prod_{i=1}^n f_X(x_i; \theta)$$

where  $X_i \sim f_X(x;\theta)$  is underlying pdf for claim model

· Contribution to likelihood of m censored observations

$${Pr_{\theta}(Y_i = M)}^m = {Pr_{\theta}(X > M)}^m = {1 - F_X(M; \theta)}^m$$

· Likelihood of  $\theta$  based on  $Y_i$ 's:

$$L_1(\theta; y_1, \dots, y_{n+m}) = \prod_{i=1}^n f_X(x_i; \theta) \prod_{j=1}^m \{1 - F_X(M; \theta)\}$$
$$= L(\theta; x_1, \dots, x_n) \{1 - F_X(M; \theta)\}^m$$

- 2. Modelling Claim Sizes (Continued):
  - · Log-likelihood:

$$l_1(\theta) = \ln\{L_1(\theta; y_1, \dots, y_{n+m})\}\$$
  
=  $\ln\{L(\theta; x_1, \dots, x_n)\} + m \ln\{1 - F_X(M; \theta)\}$ 

 $\cdot$  MLE Theorem still holds! Can create confidence intervals as:

$$\hat{\theta}_1 \pm 1.96 \sqrt{I_1^{-1}(\hat{\theta}_1)}$$

where  $\hat{\theta}_1$  solves  $\frac{\partial}{\partial \theta}l_1(\theta) = 0$  and

$$I_1(\theta) = -E_{\theta} \left\{ \frac{\partial^2}{\partial \theta \partial \theta^T} l_1(\theta) \right\}$$

[NOTE: n and m are random, though n+m is considered fixed]

- 1. All claim information known
  - $\bullet\,$  No change to original claim modelling
  - Uncommon

- 2. Information regarding claims below retention level unknown
  - Total number of claims unknown
  - Leads to truncated datasets; only observe Z if Z > 0.
  - Estimate distribution parameters of  $f_X(x;\theta)$  based observed  $Z_i$ 's:
    - $\cdot$  Use conditional distribution, which has CDF:

$$Pr_{\theta}(Z \le z | Z > 0) = Pr_{\theta}(X \le z + M | X > M)$$

$$= \frac{Pr_{\theta}(M < X \le z + M)}{Pr_{\theta}(X > M)}$$

$$= \frac{F_X(z + M; \theta) - F_X(M; \theta)}{1 - F_X(M; \theta)}$$

 $\cdot$  And pdf:

$$f_{Z|Z>0}(z;\theta) = \frac{f_X(z+M;\theta)}{1 - F_X(M;\theta)}$$