# 4/C Formulas, Spring 2009
## Loss models

### Moments

Raw moment:

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

Central moment: $\mu_k = \mathrm{E}[(X - \mu)^k]$

Coefficient of Variation: $CV = \dfrac{\mathrm{SD}(X)}{\mathrm{E}(X)} = \dfrac{\sigma}{\mu}$

Skewness $= \dfrac{\mathrm{E}(X-\mu)^3}{\sigma^3}$

Kurtosis $= \dfrac{\mathrm{E}(X-\mu)^4}{\sigma^4}$

If $X \geq 0$,

$$\begin{aligned} \mathrm{E}[X] &= \int_0^\infty x f(x)dx \\ &= \int_0^\infty S(x)dx \\ \mathrm{E}[X \wedge d] &= \int_0^d x f(x)dx + dP[X > d] \\ &= \int_0^d S(x)dx \\ \mathrm{E}[(X-d)_+] &= \int_d^\infty (x-d)f(x)dx \\ &= \int_d^\infty S(x)dx \\ \mathrm{E}[X] &= \mathrm{E}[X \wedge d] + \mathrm{E}[(X-d)_+] \\ \mathrm{E}[(X-d)_+] &= \mathrm{E}[X] - E[X \wedge d] \\ e_X(d) &= \mathrm{E}[X - d \mid X > d] = \frac{\mathrm{E}[(X-d)_+]}{S(d)} \\ \mathrm{E}[(d-X)_+ &= d - \mathrm{E}[X \wedge d] \end{aligned}$$

### Generating Functions

$$\begin{aligned} M_X(t) &= \mathrm{E}e^{tX} \\ \frac{d^k}{dt^k}M_X(0) &= E[X^k] \\ P_X(t) &= \mathrm{E}t^X \\ \frac{d^k}{dt^k}P_X(1) &= \mathrm{E}[X(X-1)\dots(X-k+1)] \end{aligned}$$

### Hazard rates, etc.

$$\begin{aligned} f(X) =& \frac{d}{dx}F(x) \\ F(x) =& \mathrm{P}[X \leq x] \\ S(x) =& 1 - F(x) = \mathrm{P}[X > x] \\ h(x) =& \frac{f(x)}{S(x)} = -\frac{d}{dx}\ln[S(x)] \\ H(x) =& -\ln S(x) = \int_0^x h(t)dt \end{aligned}$$

### Conditional probability

$$\mathrm{P}[A \mid B] = \frac{\mathrm{P}[AB]}{\mathrm{P}[B]}$$

$$\mathrm{P}[AB] = \mathrm{P}[A]\mathrm{P}[B \mid A] = \mathrm{P}[B]\mathrm{P}[A \mid B]$$

$$f_{Y|X}(y \mid X = x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_{X,Y}(x,y)}{\int f_{X,Y}(x,y)dy}$$

$$\begin{aligned} \mathrm{E}[X] &= \mathrm{E}[\mathrm{E}[X \mid Y]] \\ \mathrm{Var}[X] &= \mathrm{E}[\mathrm{Var}[X \mid Y]] + \mathrm{Var}[\mathrm{E}[X \mid Y]] \end{aligned}$$

Let $S = X_1 + \cdots + X_N$.

$$\mathrm{Var}[S] = (\mathrm{E}N)\mathrm{Var}X + (\mathrm{E}X)^2\mathrm{Var}N$$

If $N$ is Poisson($\lambda$) then $\mathrm{Var}(S) = \lambda\mathrm{E}\left(X^2\right)$.

$$\mathrm{Var}(\bar{X}) = \frac{1}{n}\mathrm{Var}(X)$$

(Bernoulli trick) If $\mathrm{P}[X = a] = 1 - \mathrm{P}[X = b]$ then

$$\mathrm{Var}(X) = (b-a)^2\mathrm{P}[X = a]\mathrm{P}[X = b]$$

### Deductibles

Loss Elimination Ratio (LER)$=\dfrac{\mathrm{E}(X \wedge d)}{\mathrm{E}X}$.

Franchise deductible:

$$\text{Payment} = \begin{cases} 0 & X \leq d \\ X & X > d \end{cases}$$

$$\mathrm{E}[\text{Payment}] = \mathrm{E}\left[(X = d)_+\right] + dS(d)$$

### Tail Weight

$X$ has a heavier tail than $Y$ if one of the following hold:

(i) $X$ has fewer moments than $Y$

(ii) $\lim_{t\to\infty} \frac{S_X(t)}{S_Y(t)} = \lim_{t\to\infty} \frac{f_X(t)}{f_Y(t)} = \infty$

$X$ has a heavy tail if $h(x)$ decreases to 0, and a light tail if $h(x)$ increases.

**Aggregate Distributions** If $X$ is discrete and $N$ is an $(a, b, 0)$ distribution,

$$\mathrm{P}[S=0] = \sum_{k=0}^{\infty} \mathrm{P}[X=0]^k \mathrm{P}[N=k]$$

$$\mathrm{P}[S=x] = \frac{\sum_{y=1}^{x} \left(a + b \cdot \frac{y}{x}\right) \mathrm{P}[X=y]\mathrm{P}[S=x-y]}{1 - a\mathrm{P}[X=0]}$$

If $N$=# of losses, and $N'$=# losses $> d$, then

$$\mathrm{E}\left[N'\right] = \mathrm{E}N\left[1 - F_X(d)\right]$$

If $N$ is an $(a, b, 0)$ distribution, then $N'$ is the same type of distribution.

## Empirical Models

### Mathematical Statistics

$\mathrm{Bias}_{\hat{\theta}}(\theta) = \mathrm{E}[\hat{\theta} - \theta \mid \theta]$
Consistency: $\hat{\theta}$ is consistent if $\mathrm{P}[|\hat{\theta}_n - \theta| < \delta] \to 1$ for all $\delta > 0$. MLEs are consistent.
$\hat{\theta}$ is consistent if bias $\to 0$ and $\mathrm{Var}(\hat{\theta}) \to 0$.

$$\begin{aligned} MSE_{\hat{\theta}}(\theta) &= \mathrm{E}[(\hat{\theta} - \theta)^2 \mid \theta] \\ &= \mathrm{Var}(\hat{\theta} \mid \theta) + [bias_{\hat{\theta}}(\theta)]^2 \end{aligned}$$

Variance in empirical distribution $= \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$

Unbiased sample variance $= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2$

### Empirical Estimators:

Empirical distribution: $F_n(t) = \frac{\#\text{ data points} \leq t}{n}$
Ogive: assume uniform distribution with groups, linearly interpolate between endpoints.

For Kaplan-Meier and Nelson-Åalen estimates, $r_j$ = size of risk set at time $y_j$. Includes data censored at times $\geq y_j$ and truncated at times $< y_j$. $s_j$= # of losses at time $y_j$.

Kaplan-Meier: $\widehat{S_n}(y_j) = \prod_{i=1}^{j} \frac{r_i - s_i}{r_i}$

Nelson-Åalen : $\widehat{H}(y_j) = \sum_{i=1}^{j-1} \frac{s_i}{r_i}$

Empirical Variance: $\widehat{\mathrm{Var}}[\widehat{S_n}(x) = \frac{\widehat{S_n}\widehat{F_n}}{n}$

$\widehat{\mathrm{Cov}}\left(\widehat{F}(x), \widehat{F}(y)\right) = -\frac{\widehat{F}(x)[\widehat{F}(y) - \widehat{F}(x)]}{n}, \quad x < y$

Greenwood's: $\widehat{\mathrm{Var}}[S_n(y_j)] = [S_n(y_j)]^2 \sum_{i=1}^{j} \frac{s_i}{r_i(r_i - s_i)}$

Variance for N-A: $\widehat{\mathrm{Var}}[\widehat{H}(y_j)] = \sum_{i=1}^{j} \frac{s_i}{r_i^2}$

Log-transformed confidence intervals:

$$\left(\hat{S}(t)^{1/U}, \hat{S}(t)^U\right), \quad U = \exp\left(\frac{z_{\alpha/2}\sqrt{\widehat{Var}(\hat{S}(t))}}{\hat{S}(t)\log\hat{S}(t)}\right)$$

$$\left(\frac{\widehat{H}(t)}{U}, \widehat{H}(t)U\right), \quad U = \exp\left(\frac{z_{\alpha/2}\sqrt{\widehat{\mathrm{Var}}\widehat{H}(t)}}{\widehat{H}(t)}\right)$$

Kernel smoothing Uniform kernel, bandwidth $b$:

$$k_y(x) = \frac{1}{2b}, \quad x \in (y - b, y + b)$$

Triangular kernel, bandwidth $b$: height of triangle = 1/b, base goes from $y - b$ to $y + b$.

## Parametric Estimation

Percentile matching – use smoothed empirical percentiles (only other time is with $p - p$ plots). $i$-th data point is $100\% \cdot i/(n + 1)$ percentile, linearly interpolate for other percentiles.

**Method of Moments:** set

$$\mathrm{E}[X^k \mid \theta = \hat{\theta}] = \frac{1}{n} \sum_{i=1}^{n} X_i^k$$

for as many values of $k$ as you have parameters.

**Maximum Likelihood Estimators**

$L(\theta; x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta)$ for individual data.

Use probability of being in an interval (instead of $f(x_i)$ if available and condition when appropriate. E.g., if data are truncated at $d$ and $k$ points are censored at $u$,

$$L(\theta) = \prod_{i=1}^{n-k} \frac{f(x_i)}{1 - F(d)} \prod_{j=1}^{k} \frac{1 - F(u)}{1 - F(d)}$$

$l(\theta) = \ln L(\theta)$= loglikelihood function. Usually easier to set $l'(\theta) = 0$ than $L'(\theta) = 0$.

If range of $X$ depends on $\theta$, MLE is probably one of the endpoints.

### Fisher's information

$$I(\theta) = -\mathrm{E}\left[\frac{\partial^2}{\partial \theta^2} l(\theta)\right] = \mathrm{E}\left[\left(\frac{\partial}{\partial \theta} l(\theta)\right)^2\right]$$

Cramér-Rao Theorem: for any estimator $\hat{\theta}$,

$$MSE_{\hat{\theta}}(\theta) \geq \frac{1}{I(\theta)}.$$

For the MLE, $\mathrm{Var}(\hat{\theta}) \sim \frac{1}{I(\theta)}$.

For a 2-parameter distribution,
$I(\theta) = -\begin{pmatrix} \mathrm{E}\frac{\partial^2}{\partial \theta_1^2} l(\theta) & \mathrm{E}\frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\theta) \\ \mathrm{E}\frac{\partial^2}{\partial \theta_1 \partial \theta_2} l(\theta) & \mathrm{E}\frac{\partial^2}{\partial \theta_2^2} l(\theta) \end{pmatrix}$
and the covariance matrix of the MLE is asymptotic to $I(\theta)^{-1}$.

### Delta method

$\mathrm{Var}(g(\hat{\theta})) \simeq (g'(\theta))^2 \mathrm{Var}(\hat{\theta})$.

In 2-dimensions, $\mathrm{Var}(g(\hat{\theta}_1, \hat{\theta}_2)) \simeq \left(\frac{\partial g}{\partial \theta_1}\right)^2 \mathrm{Var}(\hat{\theta}_1) + 2\left(\frac{\partial g}{\partial \theta_1}\right)\left(\frac{\partial g}{\partial \theta_2}\right) \mathrm{Cov}(\theta_1, \theta_2) + \left(\frac{\partial g}{\partial \theta_2}\right)^2 \mathrm{Var}(\hat{\theta}_2)$

### Cox Model

$$c = e^{\beta_1 z_1 + \cdots + \beta_n z_n}$$

$$H(t \mid z_1, \ldots, z_n) = cH_0(t)$$

$$S(t \mid z_1, \ldots, z_n) = [S_0(t)]^c$$

$$\widehat{H}_0(y_j) = \sum_{i=1}^{j} \frac{s_i}{\sum \text{ of c*\# at risk at time } y_i}$$

Partial likelihood: contribution for deaths at time $y_i$ is the product over all who die at $y_i$ of

$$\frac{c_k}{\sum c_i \text{ sum over those at risk at } y_i}$$

### Hypothesis testing

Graphical methods: $p - p$ plot: plot smoothed empirical percentile on $x$-axis, hypothesized percentile on $y$-axis. $D(x)$ plot: plot $F_n(x) - F(x)$.

Chi-Square test: $\chi^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$. Number of degrees of freedom is $k - 1$ if total number of observations is fixed, hypothesis doesn't depend on data, is $k - r - 1$ if total number of observations is fixed and $r$ parameters are estimated from data, and is $k$ if each category is independent and hypothesis does not depend on the data.

Likelihood ratio test: Null hypothesis is model $A$, alternative is model $B$. Reject null if $2(l_B - l_A)$ is greater than a $\chi^2$ critical value with # dof = # "free" parameters in B - # "free" parameters in A.

Schwarz-Bayesian criterion: maximize $\ln(L) - \frac{r}{2} \ln n$, $n$=# of data points and $r$ is # of parameters.

Kolmogorov-Smirnov statistic: $D = \max |F_n(x) - F(x)|$. Max occurs just before or after a data point, critical value goes to 0 as $n \to \infty$.

Anderson-Darling statistic: (data truncated at $d$, censored at $u$) $F^*(x) = \frac{F(x) - F(d)}{1 - F(d)}$.

$$A^2 = -nF^*(u) + n\left[\sum_{j=0}^{k} [S_n(y_j)]^2 \left[\ln(S^*(y_j)) - \ln(S^*(y_{j+1}))\right] + \sum_{j=1}^{k} (F_n(y_j))^2 \left[\ln(F^*(y_{j+1}) - \ln(F^*(y_j))\right]\right]$$

### Classical Credibility

### Full Credibility

Want $\Pr\left[|\bar{X}-\mu| \le k\mu\right] \ge P$
Number of claims needed:

$$n_c = \left(\frac{\Phi^{-1}\left(\frac{1+P}{2}\right)}{k}\right)^2 \left(\frac{\sigma_F^2}{\mu_F} + CV_S^2\right)$$

Note that $\dfrac{\sigma_F^2}{\mu_F} = 1$ for a Poisson, and remove appropriate piece if only interested in frequency or severity.

**Partial Credibility** With $n$ claims, $n < n_c$, then $Z = \sqrt{n/n_c}$.

$$P_C = (1-Z)M + Z\bar{X} = M + Z(\bar{X}-M)$$

## Greatest Accuracy Credibility

### Bayesian

$\pi(\theta)$= prior density
$f(x \mid \theta)$ = conditional density
$f(x,\theta) = \pi(\theta)f(x \mid \theta)$ = joint density
$f(x) = \int f(x,\theta)d\theta$ = unconditional density
$\pi_{\theta|x}(\theta \mid x) = \dfrac{f(x,\theta)}{f(x)}$ = posterior density
$f(y \mid x) = \int f(y \mid \theta)\pi_{\theta|x}(\theta \mid x)d\theta$ = predictive
Bayesian credibility premium = $\mathrm{E}[Y \mid X]$
Posterior mean minimizes MSE
Posterior median minimizes $\mathrm{E}|\hat{\theta}-\theta|$

**Poisson-Gamma** pair:
If $N \sim$ Poisson($\lambda$) and $\lambda \sim \Gamma(\alpha,\theta)$ then $N \sim$ Negative Binomial ($r=\alpha, \beta=\theta$).
With $k$ claims in $n$ exposures, posterior parameters are $\alpha' = \alpha + k$ and $(1/\theta') = (1/\theta) + n$.
**Binomial-Beta**: Have $m$ exposures, $k$ claims, $q =$ Pr[claim]. If $\pi(q) = \beta(a,b)$, then posterior parameters are $a' = a+k$, $b' = b+m-k$.

### Bühlmann

$$\begin{aligned}
\mu &= \mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X \mid \theta]] \\
v &= EPV = \mathrm{E}[\mathrm{Var}[X \mid \theta]] \\
a &= VHM = \mathrm{Var}[\mathrm{E}[X \mid \theta]] \\
k &= v/a \\
Z &= \frac{n}{n+k} \\
P_C &= (1-Z)\mu + Z\bar{X}
\end{aligned}$$

$n$ = number of exposures for Bühlmann-Straub.

Bühlmann credibility = least-squares fit of Bayesian.

### Non-Parametric Methods

For uniform exposures,

$$\begin{aligned}
\hat{\mu} &= \bar{X} \\
\hat{v} &= \frac{1}{r}\sum_{i=1}^{r}\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij}-\bar{x_i})^2 \\
\hat{a} &= \frac{1}{r-1}\sum_{i=1}^{r}(\bar{x}_i-\bar{x})^2 - \frac{\hat{v}}{n}
\end{aligned}$$

If we have $n_i$ years of data for group $i$, with $m_{ij}$ exposures for group $i$ in year $j$ (and $m_i$ for all of group $i$), then

$$\begin{aligned}
\hat{v} &= \frac{\sum_{i=1}^{r}\sum_{j=1}^{n_i} m_{ij}\left(X_{ij}-\bar{X}_i\right)^2}{\sum_{i=1}^{r}(n_i-1)} \\
\hat{a} &= \frac{\sum_{i=1}^{r} m_i\left(\bar{X}_i-\bar{X}\right)^2 - \hat{v}(r-1)}{m - \frac{1}{m}\sum_{i=1}^{r} m_i^2}
\end{aligned}$$

**Semi-Parametric Models** If the number of claims per person is Poisson($\lambda$), then

$$\begin{aligned}
\hat{\mu} &= \bar{X} \\
\hat{v} &= \bar{X} \\
\hat{a} &= \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2 - \hat{v}
\end{aligned}$$

With grouped data, this formula for $\hat{a}$ doesn't make sense so use the previous formula.

Ruin Theory: $\phi(u) = \mathrm{P}[U_t > 0 \text{ for all } t \mid U_0 = u]$ is the continuous, infinite time horizon survival probability. $\tilde{\phi}(u)$ is the discrete, infinite time horizon survival probability $\tilde{\psi}(u,\tau) = 1 - \tilde{\phi}(u,\tau)$ is the discrete time, finite time horizon ruin probability.

## Stocks and Options

### Lognormal stock model

$$\frac{S_t}{S_0} \sim LN\left(t\left[\alpha - \delta - \frac{\sigma^2}{2}\right], \sigma^2 t\right)$$

Poisson jumps: Jumps are $LN(\alpha_J - \sigma_J^2/2, \sigma_J^2)$
Annual (compounded) rate of return: $\alpha = \hat{\alpha} + k\alpha_J$ where $k$ is expected number of jumps and $\hat{\alpha}$ is the rate of return before taking into account jumps.

**Black-Scholes** Use $r$=risk free rate of return instead of $\alpha$ when pricing options.

$$\mathrm{E}[(S_t - K)_+] = S_0 e^{t(\alpha-\delta)}\Phi(d_1) - K\Phi(d_2)$$
$$\mathrm{E}[(K - S_t)_+] = K\Phi(-d_2) - S_0 e^{t(\alpha-\delta)}\Phi(-d_1)$$

$$\mathrm{E}[S_t \mid S_t > K] = \mathrm{E}(S_t)\frac{\Phi(d_1)}{\Phi(d_2)}$$
$$\mathrm{E}[S_t \mid S_t < K] = \mathrm{E}(S_t)\frac{\Phi(-d_1)}{\Phi(-d_2)}$$
$$\text{PV of a call} = S_0 e^{-\delta T}\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$
$$\text{PV of a put} = Ke^{-rT}\Phi(-d_2) - S_0 e^{-\delta T}\Phi(-d_1)$$

$$\begin{aligned}\Phi(d_2) &= \mathrm{P}[S_t > K] \\ &= \Phi\left(\frac{\ln(S_0) - \ln(K) + t(\alpha - \delta - (1/2)\sigma^2)}{\sigma\sqrt{t}}\right)\end{aligned}$$
$$\Phi(d_1) = \Phi\left(\frac{\ln(S_0) - \ln(K) + t(\alpha - \delta + (1/2)\sigma^2)}{\sigma\sqrt{t}}\right)$$

## Risk Measures

$$VaR_\alpha = Q_\alpha = \min\{Q : F_L(Q) \geq \alpha\}$$
$$\begin{aligned}CTE_\alpha &= TVaR \\ &= \begin{cases}\mathrm{E}[L \mid L > Q_\alpha] & \text{if } \mathrm{P}[L = Q_\alpha] = 0 \\ \frac{(F_L(Q_\alpha)-\alpha)Q_\alpha + \mathrm{E}[L; L>Q_\alpha]}{1-\alpha} & \text{else}\end{cases}\end{aligned}$$

Distortion measure: $\mathcal{H}(X) = \int_0^\infty g(S(x))dx$ where $g(0) = 0$, $g(1) = 1$ and $g(x) \leq g(y)$ if $x \leq y$.
Proportional Hazards: $g(S(x)) = [S(x)]^{1/\kappa}, \kappa \geq 1$
Wang's Transform: $g(S(x)) = \Phi\left[\Phi^{-1}(S(x)) + \kappa\right]$
WT of Lognormal$(\mu, \sigma^2)$ is expected value of a LN$(\mu + \kappa\sigma, \sigma^2)$

## Simulation

Inversion: $X_i = F^{-1}(U_i)$
Antithetic Variates: include $1 - U_i$
Stratified Sampling: $V_i = \dfrac{U_i}{n} + \dfrac{i-1}{n}$
Control Variate: $\bar{Y}$=naive simulation of $Y$, $\bar{X}$=naive simulation of $X$, $X$ known.

$$Y^* = \bar{Y} + \beta(X - \bar{X})$$

(unless specified, $\beta = 1$)

**Key Distributions** For $\alpha \geq 1$ an integer,

| | Exponential | Gamma | Pareto | Uniform$(a, b)$ |
|---|---|---|---|---|
| $f(x)$ | $\frac{1}{\theta}e^{-x/\theta}$ | $\frac{x^{\alpha-1}}{\theta^\alpha(\alpha-1)!}e^{-x/\theta}$ | $\frac{\alpha\theta^\alpha}{(\theta+x)^{\alpha+1}}$ | $\frac{1}{b-a}$ |
| $F(x)$ | $1 - e^{-x/\theta}$ | $1 - \sum_{i=0}^{\alpha-1}\frac{(x/\theta)^i}{i!}e^{-x/\theta}$ | $1 - \left(\frac{\theta}{x+\theta}\right)^\alpha$ | $\frac{x-a}{b-a}$ |
| $\mathrm{E}[X]$ | $\theta$ | $\alpha\theta$ | $\frac{\theta}{\alpha-1}$ | $\frac{a+b}{2}$ |
| $\mathrm{Var}(X)$ | $\theta^2$ | $\alpha\theta^2$ | $\frac{\theta^2\alpha}{(\alpha-1)^2(\alpha-2)}$ | $\frac{(b-a)^2}{12}$ |
| $\mathrm{E}[X \wedge x]$ | $\theta(1 - e^{-x/\theta})$ | messy | See tables | |
| $e_X(d)$ | $\theta$ | messy | $\frac{\theta+d}{\alpha-1}$ | $\frac{d+b}{2}$ |