C Formula Sheet

Probability

$$\begin{cases} f(x) = F'(x) \\ F(x) = \int f(x) \end{cases}$$

$$\begin{cases} F(x) = 1 - S(x) \\ S(x) = 1 - F(x) \end{cases}$$

$$\begin{cases} S(x) = e^{-H(x)} \\ H(x) = -\ln(S(x)) \end{cases}$$

$$\begin{cases} h(x) = H'(x) \\ H(x) = \int h(x) \end{cases}$$

$$\boxed{\begin{array}{l} f(x) = -S'(x) \\ h(x) = \dfrac{f(x)}{S(x)} \end{array}}$$

$$E\left[X^k\right] = \sum_x x^k \cdot p(x) \text{ or } \int_{-\infty}^{\infty} x^k \cdot f(x)dx$$

$\mu_n' = E\left[X^n\right]$ (nth Raw Moment)

$\mu_n = E\left[(X-\mu)^n\right]$ (nth Central Moment)

$\mu = \mu_1'$ is the mean

$\mu_2 = \mu_2' - \mu^2$ is the variance

$$\mu_3 = \mu_3' - 3\mu_2'\mu + 2\mu^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu + 6\mu_2'\mu^2 - 3\mu^4$$

$$Var(X) = \mu_2 = \sigma^2$$

Standard Deviation = σ

Skewness is $\gamma_1 = \dfrac{\mu_3}{\sigma^3}$ $= 0$ if X is symmetric Skew(X)=Skew(cX)

Kurtosis is $\gamma_2 = \dfrac{\mu_4}{\sigma^4}$

Coefficient of Variation is $\dfrac{\sigma}{\mu}$

Combinations: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Percentiles

A 100p[th] percentile of a random variable X is a number π_p satisfying:

$\Pr(X \le \pi_p) \ge p$

$\Pr(X < \pi_p) \le p$

Conditional Probability

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B)}$$

Scaling

To scale a lognormal:

$X \sim \text{lognormal}\ (\mu, \sigma)$

$cX \sim \text{lognormal}\ (\mu + \ln c, \sigma)$

Let $Y = cX$

Then $F_Y(y) = \Pr(Y \le y) = \Pr(cX \le y) = \Pr\left(X \le \frac{y}{c}\right) = F_X\left(\frac{y}{c}\right)$

Variance

$E[XY] = E[X]E[Y]$ if Independent

$E[aX + bY] = aE[X] + bE[Y]$

$Var[aX + b] = a^2 Var(X)$

$Var[aX + bY] = a^2 Var(X) + 2abCov(X,Y) + b^2 Var(Y)$

$Var(X) = E\left(X^2\right) - E(X)^2$

$VaR(X) = VaR(1 - X)$

$Cov(X,Y) = E(XY) - E(X)E(Y)$

$Cov((A+B),(C+D)) = Cov(A,C) + Cov(A,D) + Cov(B,C) + Cov(B,D)$

$Cov(A,A) = Var(A)$

$Cov(A,B) = 0$ if Independent

$\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$ (Correlation Coefficient)

Bernoulli Shortcut: For any RV with only 2 values a and b:

$Var(X) = (a - b)^2\ pq$

Parametric Distributions

Distribution	f(x)	E(X)	$E\left(X^2\right)$	Var(X)
Binomial n trials	$\binom{n}{k} p^k (1-p)^{n-k}$	np		npq
Bernoulli 1 trial	$p^k (1-p)^{n-k}$	p		pq
Uniform continuous on [d, u]	$\frac{1}{u-d}$ $\boxed{F(x) = \frac{x-d}{u-d}}$	$\frac{d+u}{2}$	$\frac{(u-d)^2}{3}$ if $u=0$	$\frac{(u-d)^2}{12}$
Beta	$\boxed{cx^{a-1}(\theta - x)^{b-1}}$	$\frac{\theta a}{a+b}$		$\frac{\theta^2 ab}{(a+b)^2 (a+b+1)}$
Exponential (memoryless)	$ce^{-x/\theta}$	θ	$2\theta^2$	θ^2
Weibull	$cx^{\tau-1} e^{\left(-x/\theta\right)^\tau}$	$\theta \cdot \Gamma\left(1+\frac{1}{\tau}\right)$	$\theta^2 \Gamma\left(1+\frac{2}{\tau}\right)$	
Gamma	$cx^{\alpha-1} e^{-x/\theta}$	$\alpha\theta$		$\alpha\theta^2$
Single Parameter Pareto	$\frac{c}{x^{\alpha+1}}$	$\frac{\alpha\theta}{\alpha-1}$	$\frac{\alpha\theta^2}{\alpha-2}$	
Double Parameter Pareto	$\frac{c}{(x+\theta)^{\alpha+1}}$	$\frac{\theta}{\alpha-1}$	$\frac{2\theta^2}{(\alpha-1)(\alpha-2)}$	$\boxed{\left(E[X]\right)^2 \cdot \frac{\alpha}{\alpha-2}}$
Lognormal	$\frac{ce^{\frac{-(\ln x-\mu)^2}{2\sigma^2}}}{x}$	$e^{\mu+0.5\sigma^2}$ Median $= e^{\mu}$	$e^{2\mu+2\sigma^2}$	

Frailty

Let $h(x|\Lambda) = \Lambda a(x)$

If a(x) is a constant then the hazard rate is exponential, otherwise Weibull

Λ can be Gamma or Inverse Gaussian

$$A(x) = \int_0^x a(t)dt$$

$$H(x|\Lambda) = \Lambda A(x)$$

$$S(x|\Lambda) = e^{-\Lambda A(x)}$$

$$S(x) = M_\Lambda(-A(x))$$ (Use MGF for Λ)

Exponential a(x) leads to a Pareto distribution

Weibull a(x) leads to a Burr distribution

Conditional Variance

$$Var(X) = E[Var(X|I)] + Var(E[X|I])$$

$$E[X] = E[E[X|I]] = \sum P(A_i) \cdot E[X|A_i]$$

$$E[X^2] = E[E[X^2|I]]$$

$$Var[X] \neq E[Var[X|I]]$$

Splices

1) Sum of functions must integrate to 1

2) To be continuous, functions must be equal at break point

Shifting

If $f(\lambda) = 3e^{-3(\lambda-1)}$

The mean = 1/3 + 1 (the mean of the unshifted exponential plus the shift)

Policy Limits

$X \wedge d$ is the LIMITED EXPECTED VALUE

$$X \wedge d = \begin{cases} X & X < u \\ d & X \geq u \end{cases} \qquad \text{(Cost to Customer)}$$

Definition: $E\left[X^k\right] = \int_0^\infty x^k f(x)dx = \int_0^\infty kx^{k-1}S(x)dx$

$$E(X) = \int_0^\infty x \cdot f(x)dx = \boxed{\int_0^\infty S(x)dx}$$

$$E(X \wedge d) = \int_0^d xf(x)dx + d\left(S(d)\right)$$

$$= \boxed{\int_0^d S(x)dx}$$

$$E\left(\left(X \wedge d\right)^k\right) = \int_0^d x^k f(x)dx + d^k\left(S(d)\right)$$

$$= \int_0^d kx^{k-1}S(x)dx$$

If $Y = (1+r)X$ then $E\left(Y \wedge d\right) = \left(1+r\right)E\left[X \wedge \frac{d}{1+r}\right]$

Deductibles

Ordinary Deductible of d – pays max(0,X-d)

For d = 500, Loss <= 500 pays nothing, Loss of 700 pays 200

Franchise Deductible of d – pays nothing if Loss is less than d, and full amount if Loss > d

Payment per Loss with Deductible

Ordinary Deductible

$$\text{Payment from Ins. Co.} = \begin{cases} 0 & X \leq d \\ X - d & X > d \end{cases} = E\left[\left(X - d\right)_+\right]$$

$$E\left[\left(X - d\right)_+\right] = \int_d^\infty (x-d) \cdot f(x)dx = \boxed{\int_d^\infty S(x)dx}$$

$$E\left[\left(X - d\right)_+^k\right] = \int_d^\infty (x-d)^k \cdot f(x)dx \neq E\left[X^k\right] - E\left[\left(X \wedge d\right)^k\right]$$

Payment per Payment with Deductible

Ordinary Deductible

$$Y^P = (X-d)_+ \mid X > d$$

$$F_{Y^P}(x) = \frac{F_X(x+d) - F_X(d)}{1 - F_X(d)}$$

$$S_{Y^P}(x) = \frac{S_X(x+d)}{S_X(d)}$$

$$\boxed{e(d) = E\left[X - d \mid X > d\right]}$$

$$\boxed{e(d) = E\left[Y^P\right] = \frac{E\left[(X-d)_+\right]}{S(d)} = \frac{E[X] - E[X \wedge d]}{S(d)}}$$

$$e(d) = \frac{\int_d^\infty (x-d) f(x)dx}{S(d)} = \boxed{\frac{\int_d^\infty S(x)dx}{S(d)}}$$

$e(d) =$ Mean Excess Loss

$$E[X] = \underbrace{E[X \wedge d]}_{\text{pmt from customer}} + \underbrace{E\left[(X-d)_+\right]}_{\text{pmt from ins. co.}}$$

$$E[X \wedge d] = E[X \wedge d \mid X < d] \cdot \Pr(X < d) + E[X \wedge d \mid X \geq d] \cdot \Pr(X \geq d)$$

$$= (\text{Average loss} < \text{d}) \cdot \Pr(X < d) + d \cdot \Pr(X \geq d)$$

Franchise Deductibles

$$\text{Expected Payment per Loss} = E\left[(X-d)_+\right] + dS(d)$$

$$\text{Expected Payment per Payment} = e(d) + d$$

Special Cases for e(d)

Distribution	e(d)
Exponential	θ
Uniform $(0,\theta]$	$\frac{\theta-d}{2}$
2 Parameter Pareto	$\frac{\theta+d}{\alpha-1}$
1 Parameter Pareto	$\begin{cases} \frac{d}{\alpha-1} & d \geq \theta \\ \frac{\alpha(\theta-d)+d}{\alpha-1} & d < \theta \end{cases}$

If $X \sim$ Uniform $(0,\theta]$, then $(X - d)_+ \mid X > d \sim$ Uniform $(0,\theta-d]$

If $X \sim$ Pareto (α,θ), then $(X - d)_+ \mid X > d \sim$ Pareto $(\alpha,\theta+d)$

If $X \sim$ 1 Parameter Pareto (α,θ) and $d \geq \theta$, then $(X - d)_+ \mid X > d \sim$ Pareto (α,d)

Loss Elimination Ratio

$$LER(d) = \frac{E[X \wedge d]}{E[X]}$$

$LER(d) =$ Expected % of loss not included in payment

Special Cases for LER(d)

Distribution	LER(d)
Exponential	$1-e^{-d/\theta}$
2 Parameter Pareto	$1-\left(\frac{\theta}{d+\theta}\right)^{\alpha-1}$
1 Parameter Pareto	$1-\frac{\left(\theta/d\right)^{\alpha-1}}{\alpha}$

Properties of Risk Measures

Translation Invariance: $\rho(X+c)=\rho(X)+c$

Positive Homogeneity: $\rho(cX)=c\rho(X)$

Subadditivity: $\rho(X+Y)\le\rho(X)+\rho(Y)$

Monotinicity: $\rho(X)\le\rho(Y)$ if $X\le Y$

A coherent risk measure satisfies all 4 properties

VaR_p fails subadditivity

$TVaR_p$ is coherent

$\mathrm{E}[\mathrm{X}]$ is coherent

$VaR_p(X)=\pi_p=F_X^{-1}(p)$ Value-at-Risk

$VaR_{0.99}$ is the 99th percentile

$$TVaR_p(X) = E\left[X \mid X > VaR_p(X) \right]$$

Tail-value-at-Risk

$$= \frac{\int_{F_X^{-1}(p)}^{\infty} x f(x)\,dx}{1-p}$$

$$= \boxed{\frac{\int_p^1 VaR_y(X)\,dy}{1-p}}$$

$$= \boxed{VaR_p(X) + e_X\left(VaR_p(X)\right)}$$

Distribution	VaR(X)	TVaR(X)
Normal	$\mu + z_p\sigma$	$\mu + \sigma \frac{\phi(z_p)}{1-p}$ where $\phi(x) = \frac{e^{-0.5x^2}}{\sqrt{2\pi}}$
Lognormal	$e^{\mu + z_p\sigma}$	$E[X]\left(\frac{\Phi(\sigma - z_p)}{1-p}\right)$

If given a mixture, use the survival function and solve for x

Maximum Covered Loss

For deductible 'd' and maximum covered loss 'u'

$$E[X] = \begin{cases} 0 & X \le d \\ X - d & d < X \le u \\ u - d & X > u \end{cases}$$

$$\boxed{E[\text{Payment per Loss}] = E[X \wedge u] - E[X \wedge d]}$$
$$= \int_d^u S(x)dx$$
$$= E\left[(X-d)_+\right] - E\left[(X-u)_+\right]$$

Policy Limit: the maximum amount the coverage will pay
- If Policy Limit = 10,000 and d = 500
  - Pays 10,000 for loss of 10,500 or higher
  - Pays Loss – 500 for losses b/w 500 and 10,500

Maximum Covered Loss: the maximum loss amount that is covered
- If MCL = 10,000 and d = 500
  - Pays 9,500 for loss of 10,000 or higher
  - Pays Loss – 500 for losses b/w 500 and 10,000

Coinsurance

$$E[\text{Payment per Loss}] = \underset{\text{Coinsurance}}{\alpha} \cdot \left(E\left[X \wedge \underset{MCL}{u} \right] - E\left[X \wedge \underset{ded}{d} \right] \right)$$

Coinsurance of 80% means that the insurance pays 80% of the costs

Inflation

$$E[\text{Payment per Loss}] = \alpha(1+r) \cdot \left(E\left[X \wedge \frac{u}{1+r} \right] - E\left[X \wedge \frac{d}{1+r} \right] \right)$$

Variance of Payment per Loss with a deductible

$X = \text{Loss RV}$

$Y^L = \text{payment per loss RV}$

$$E\left[Y^L\right] = 0 \cdot \Pr(X \le d) + E\left[Y^P\right]\Pr(X > d)$$

$$Var\left(Y^L\right) = E\left[Var\left(Y^L \mid Case\right)\right] + Var\left(E\left[Y^L \mid Case\right]\right)$$
$$= \left(0 \cdot \Pr(X \le d) + Var\left(Y^P\right)\Pr(X > d)\right) + \left(\left(E\left[Y^P\right] - 0\right)^2 \cdot \Pr(X \le d) \cdot \Pr(X > d)\right)$$

Bonus

Pay Bonus of 50% of (500 - X) if $X \le 500$

$$B = 0.5Max(0, 500 - X)$$
$$= 0.5Max(500 - 500, 500 - X)$$
$$= 0.5(500 - Min(500, X))$$
$$= 0.5(500 - X \wedge 500)$$
$$= 250 - 0.5E[X \wedge 500]$$

Discrete Distributions

The (a,b,0) class

Distribution	a	Variance vs. Mean
Poisson	0	Variance = Mean
Negative Binomial (Geometric is NB with r = 1)	>0	Variance > Mean
Binomial	<0	Variance < Mean

$$\Pr(N \ge n) = \left(\frac{\beta}{1+\beta}\right)^n$$ Geometric Distribution (memoryless)

A sum of 'n' independent Negative Binomial random variables having the same β and parameters $r_1, ..., r_n$ has a Negative Binomial distribution with parameters β and $\sum_{i=1}^{n} r_i$

A sum of 'n' independent Binomial random variables having the same q and parameters $m_1, ..., m_n$ has a Binomial distribution with parameters q and $\sum_{i=1}^{n} m_i$

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \qquad k = 1, 2, \ldots$$

$$E[N] = \frac{a+b}{1-a}$$

$$Var(N) = \frac{a+b}{(1-a)^2}$$

$$\binom{x}{n} = \frac{x(x-1)\cdots(x-n+1)}{n!}$$

Probability Generating Functions

$$p_n = \frac{P^{(n)}(0)}{n!}$$

$$p_0 = P(0)$$

$$\mu_{(n)} = P^{(n)}(1)$$

$$P'(1) = E[X]$$

$$P''(1) = E[X(X-1)]$$

$$P'''(1) = E[X(X-1)(X-2)]$$

If given a primary and secondary pgf, substitute the secondary pgf for 'z' in the primary pgf.

The (a,b,1) class

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k} \qquad k = 2,3,4,...$$

Zero-Truncated Distributions

$$p_0^T = 0$$

$$p_k^T = \frac{p_k}{1-p_0}$$

Zero-Modified Distributions

$$p_0^M > 0$$

$$p_k^M = \left(1-p_0^M\right)\cdot\frac{p_k}{1-p_0}$$

$$p_k^M = \left(1-p_0^M\right)\cdot p_k^T$$

$$E[M] = \frac{1-p_0^M}{1-p_0}E[Orig] = \left(1-p_0^M\right)E[Zero-Truncated]$$

$$E[N] = cm$$

$$Var(N) = c(1-c)m^2 + cv$$

$c = 1 - p_0^M$

m is the mean of the corresponding zero-truncated distribution

v is the variance of the corresponding zero-truncated distribution

Sibuya

ETNB with $-1 < r < 0$ and take lim as $\beta \to \infty$

$a = 1$

$b = r - 1$

$p_1^T = -r$

Poisson/Gamma

The Negative Binomial is a Gamma mixture of Poissons

$N \sim \text{Poisson}(\lambda)$

$\lambda \sim \text{Gamma}(\alpha, \theta)$

$\text{Negative Binomial } (r) = \text{Gamma}(\alpha)$

$\text{Negative Binomial } (\beta) = \text{Gamma}(\theta)$

Gamma

$\text{Mean} = \alpha\theta$

$\text{Variance} = \alpha\theta^2$

Negative Binomial

$\text{Mean} = r\beta$

$\text{Variance} = r\beta(1+\beta)$

Negative Binomial $(r = 1)$ is Geometric

Gamma $(\alpha = 1)$ is Exponential

Weibull$(\tau = 1)$ is Exponential

$Var(\lambda) = Var(X) - E[X]$ where $\lambda \sim$ Gamma

Coverage Modifications

Frequency Model	Original Parameters	Exposure Modification	Coverage Modification
	Exposure n_1 $\Pr(X > 0) = 1$	Exposure n_2 $\Pr(X > 0) = 1$	Exposure n_1 $\Pr(X > 0) = v$
Poisson	λ	$\left(\frac{n_2}{n_1}\right)\lambda$	$v\lambda$
Binomial	m, q	$\left(\frac{n_2}{n_1}\right)m, q$	m, vq
Negative Binomial	r, β	$\left(\frac{n_2}{n_1}\right)r, \beta$	$r, v\beta$

(a,b,0) and (a,b,1) adjustments

$$1 - p_0^{M*} = \left(1 - p_0^M\right)\left(\frac{1 - p_0^*}{1 - p_0}\right)$$ where * indicates revised parameters

Aggregate Loss Models

Compound Variance

S = aggregate losses
N = frequency RV
X = severity RV

$E[S] = E[N]E[X]$

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

Can only be used when N and X are independent

$\text{Var}(S) = \lambda E\left[X^2\right]$ if primary distribution (# claims) is Poisson

Collective Risk vs Individual Risk

Convolution Method

$p_n = \Pr(N = n) = f_N(n)$

$f_n = \Pr(X = n) = f_X(n)$

$g_n = \Pr(S = n) = f_S(n)$

$$F_S(x) = \sum_{n \le x} g_n \sum_{i_1 + \ldots + i_k = n} f_{i_1} f_{i_2} \cdots f_{i_k}$$

When given severity distributions with $\Pr(X = 0) \neq 0$

1) Modify the frequency to eliminate 0

2) Adjust the severity probabilities after removing 0

Aggregate Deductibles

Assume severity is discrete

d = stop-loss or reinsurance deductible

Assume Premium $= E[S \wedge d]$

$E[(S-d)_+] = E[S] - E[S \wedge d] =$ Net stop-loss Premium

Method 1 - Definition of $E[S \wedge d]$

$$E[S \wedge d] = \sum_{j=0}^{u} hjg_{hj} + d \Pr S \geq d$$

$u = \lceil d/h \rceil - 1$ (the sum of all multiples of h less than d)

Method 2 - Integrate the Survival Function

$$E[S \wedge d] = \sum_{j=0}^{u-1} hS(hj) + (d-hu)S(hu)$$

$u = \lceil d/h \rceil - 1$ (the sum of all multiples of h less than d)

To find $E[S \wedge 2.8]$ where x can be 2, 4 6 or 8...

Method 1

$$E[S \wedge 2.8] = \left(\underbrace{P(S=0)}_{g(0)} \cdot \underbrace{0}_{amt}\right) + \left(\underbrace{P(S=2)}_{g(2)} \cdot \underbrace{2}_{amt}\right) + \left(\underbrace{P(S>2)}_{1-g(0)-g(2)} \cdot \underbrace{2.8}_{d}\right)$$

$$E[S \wedge 4] = \left(\underbrace{P(S=0)}_{g(0)} \cdot \underbrace{0}_{amt}\right) + \left(\underbrace{P(S=2)}_{g(2)} \cdot \underbrace{2}_{amt}\right) + \left(\underbrace{P(S \geq 4)}_{1-g(0)-g(2)} \cdot \underbrace{4}_{d}\right)$$

Method 2

$$E[S \wedge 2.8] = \left(\underbrace{2}_{\text{dist between values of x}} \cdot \Pr(S>0)\right) + \left(\underbrace{0.8}_{\text{dist b/w highest value of x below d and d}} \cdot \Pr(S>2)\right)$$

Aggregate Coverage Modifications

If there is a per-policy deductible and you want Aggregate Payments

1) Expected Payment per Loss x Expected Number of Losses per Year

$$E[S] = E[N] \cdot E\left[(X-d)_+\right]$$

OR

2) Expected Payment per Payment x Expected Number of Payments per Year

$$E[S] = E\left[N'\right] \cdot E\left[X-d \mid X>d\right]$$

where N' is the number of positive payments $\left(\text{frequency} \cdot \Pr(X>d)\right)$

1) Better for discrete severity distributions

2) Better for if severity is Exponential, Pareto or Uniform

Exact Calculation of Aggregate Loss Distribution

Two cases for which the sum of independent random variables has a simple distribution

1) Normal Distribution. If X_i are Normal with mean μ and variance σ^2 , their sum is normal.

2) Exponential or Gamma Distribution. If X_i are exponential or gamma, their sum has a gamma distribution

Normal Distributions

If n random variables X_i are independent and normally distributed with parameters μ and σ^2, their sum is normally distributed with parameters $n\mu$ and $n\sigma^2$.

Calculate $\Pr(S > c \mid n = 1)$ using μ and σ^2
Calculate $\Pr(S > c \mid n = 2)$ using 2μ and $2\sigma^2$, etc.

Then multiply each these probabilities by their respective $p_1, p_{2,}$ etc.

Exponential and Gamma Distributions

The sum of n exponential random variables with common mean θ
is a Gamma distribution with parameters $\alpha = n$ and θ. When a gamma
distribution's α parameter is an integer, the gamma distribution is called
an Erlang distribution.

In a Poisson Process with parameter λ, the number of events occurring by time t
has a Poisson distribution with mean λt.

In a Poisson Process with parameter λ, the time between events follows
an exponential distribution with parameter $1/\lambda$.

In a Poisson Process with paremeter $1/\theta$, the time between events is exponential
with mean θ. Therefore, the time until the nth event occurs is Erlang with parameters n and θ

The probability of exactly n events occurring before time x is $\dfrac{e^{-\frac{x}{\theta}}\left(x/\theta\right)^n}{n!}$

Gamma CDF

$$F_X(x) = 1 - \sum_{j=0}^{n-1} \frac{e^{-\frac{x}{\theta}}\left(x/\theta\right)^j}{j!}$$

If $\alpha = 1$, $F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)}$

If $\alpha = 2$, $F(x) = 1 - e^{-\left(\frac{x}{\theta}\right)} - \left(\dfrac{x}{\theta}\right)e^{-\left(\frac{x}{\theta}\right)}$

Empirical Models

Bias

$bias_\theta(\theta) = E\left[\hat{\theta}\right] - \theta$ $\quad$ θ is the estimator, θ is the parameter being estimated

bias is the expected value of the estimator minus its true value

Estimator is unbiased if $bias_\theta(\theta) = 0$ for all θ

Estimator is asymptotically unbiased if $\lim_{x\to\infty}\left(bias_{\hat{\theta}}(\theta)\right) = 0$

The sample variance (with division by n-1) is an unbiased estimator of variance
Sample mean is an unbiased estimator of the true mean

Consistency (weak consistency)

Definition: Estimator is consistent if $\lim_{n\to\infty} \Pr\left(\left|\widehat{\theta_n} - \theta\right| < \delta\right) = 1$ for all $\delta > 0$

1) An estimator is consistent if it is asymptotically unbiased

and $\mathrm{Var}\left(\widehat{\theta_n}\right) \to 0$ as $n \to \infty$

2) The MLE is always consistent

3) If $MSE_{\hat{\theta}}(\theta) \to 0$ then $\widehat{\theta}$ is consistent

Mean Square Error

$$MSE_{\hat{\theta}}(\theta) = E\left[\left(\widehat{\theta - \theta}\right)^2 \mid \theta\right]$$

$$MSE_{\hat{\theta}}(\theta) = Var\left(\widehat{\theta}\right) + \left(bias_{\hat{\theta}}(\theta)\right)^2$$

MSE is a function of the true value of the parameter

Complete Data

Grouped Data

$$f_n(x) = \frac{n_j}{n\left(c_j - c_{j-1}\right)}$$

where x is in $\left[c_j, c_{j-1}\right)$

n_j = # points in the interval

n = total points

$f_n(x) =$ Histogram

$F_n(x) =$ Ogive

To find the 2nd raw moment, calculate $\int_a^b f_n(x) \cdot x^2 dx$

If there is a policy limit (say 8000), then $\text{E}\left[\left(\text{X}\wedge 8000\right)^2\right]$ would have as its last 2 terms:

$$\int_{5000}^{8000} f_n(x) \cdot x^2 dx + \int_{8000}^{10,000} f_n(x) \cdot 8000^2 dx$$

Variance of Empirical Estimators

Binomial:

Variance = $mq(1-q)$

If $\text{Y} = \frac{\text{X}}{\text{m}}$ (Binomial Proportion),

Variance = $\frac{q(1-q)}{\text{m}}$

Multinomial:

Variance = $mq_i(1-q_i)$ $\left(\text{i = category}\right)$

Covariance $=$ -$mq_i q_j$

If $\text{Y} = \frac{\text{X}}{\text{m}}$, Variance = $\frac{q_i(1-q_i)}{\text{m}}$

Covariance $= \frac{-q_i q_j}{m}$

Individual Data

$$Var\left(S_n(x)\right) = \frac{S(x)\left(1 - S(x)\right)}{n}$$ if S is known

$$Var\left(S_n(x)\right) = \frac{S_n(x)\left(1 - S_n(x)\right)}{n}$$

$$Var\left(S_n(x)\right) = \frac{n_x\left(n - n_x\right)}{n^3}$$ where n_x is the # of survivors past time x

$$Var\left({}_{y-x}p_x \mid n_x\right) = Var\left({}_{y-x}q_x \mid n_x\right) = \frac{\left(n_x - n_y\right)n_y}{n_x^3}$$

The empirical estimators of $S(x)$ and $f(x)$ are unbiased

Individual Data

1) Determine the estimator
2) Determine what's random and what's not
3) Write an expression for the estimator with symbols for random variables
4) Calculate the variance of the random variables
5) Calculate the variance for the whole expression
(i.e. $Var(aX + bY)$, $Var(aX)$, etc.)

Kaplan Meier Product Limit

$$S_n(t) = \prod_{i=1}^{j-1} \left(1 - \frac{s_i}{r_i}\right)$$

$S(t) = S(t_0)^{t/t_0}$ where t_0 is the end of the study   Exponential Extrapolation

$r_i =$ risk set

$s_i =$ death

$d_i =$ entry time

$u_i =$ withdrawal time

$x_i =$ death time

$S(x^-) = \Pr(X \geq x)$

Shortcut: $\frac{1}{n} + \frac{1}{n-1} \approx \frac{2}{n-0.5}$

Nelson Aalen Estimator

$$\hat{H}(t) = \sum_{i=1}^{j-1} \frac{s_i}{r_i}$$

$$\hat{S}(x) = e^{-\hat{H}(x)}$$

Lives that leave at the same time as a death are in the risk set.
Lives that arrive at the same time as a death are not in the risk set.
Censored lives are in the risk set but are not counted as deaths

k = # distinct data points

Confidence Intervals

For S(x), the boundaries must be between 0 and 1.
For H(x), the boundaries can be anything.

Calculator Shortcuts

1) Enter $\frac{s_i}{r_i}$ in column 1

2) Enter the formula $\ln(1-L1)$ for column 2

3) Select L1 as first variable and L2 as second

4) Calculate $\underset{Nelson-Aalen}{e^{-\sum x}}$ - $\underset{Kaplan-Meier}{e^{\sum y}}$

Estimation of Related Quantities

If using Kaplan-Meier or Nelson-Aalen methods, E(X^d) = area under the curve of S(x). Multiply the base x height of each of the rectangles.

Bayes Theorem

$$P(A_1 \mid E) = \frac{P(E \mid A_1) \cdot P(A_1)}{P(E \mid A_1) \cdot P(A_1) + P(E \mid A_2) \cdot P(A_2) + \ldots + P(E \mid A_n) \cdot P(A_n)}$$

Greenwood's Approximation of Variance (Kaplan-Meier)

$$Var\left(S(t)\right)=S(t)^2\sum_{y_j\le t}\frac{s_j}{r_j\left(r_j-s_j\right)}$$

$$Var\left(S(t)\right)=\frac{S_n\left(x\right)\left(1-S_n\left(x\right)\right)}{n}$$ if data is complete (no censoring or truncation)

Variance of Nelson-Aalen Estimator

$$Var\left(H(t)\right)=\sum_{y_j\le t}\frac{s_j}{r_j^{\,2}}$$

Linear Confidence Intervals

$$\left(S_n(t)-z_{0.5(1+p)}\sqrt{Var\left(S_n(t)\right)},\quad S_n(t)+z_{0.5(1+p)}\sqrt{Var\left(S_n(t)\right)}\right)$$

Log-transformed confidence interval for S(t)

$$\left(S_n(t)^{1/U},S_n(t)^U\right)\quad \text{where } U=\exp\left(\frac{z_{0.5(1+p)}\sqrt{Var\left(S_n(t)\right)}}{S_n(t)\cdot\ln S_n(t)}\right)$$

Log-transformed confidence interval for H(t)

$$\left(\frac{\widehat{H(t)}}{U},\widehat{H(t)\cdot U}\right)\quad \text{where } U=\exp\left(\frac{z_{0.5(1+p)}\sqrt{Var\left(\widehat{H(t)}\right)}}{\widehat{H(t)}}\right)$$

Kernel Smoothing

Uniform

$$k_{x_i}(x) = \begin{cases} \dfrac{1}{2b} & x_i - b \le x \le x_i + b \\ 0 & Otherwise \end{cases}$$

$$K_{x_i}(x) = \begin{cases} 0 & x \le x_i - b \\ \dfrac{x - (x_i - b)}{2b} & x_i - b \le x \le x_i + b \\ 1 & x > x_i + b \end{cases}$$

$$\hat{f}(x) = \sum_{i=1}^{n} \underbrace{\left(\frac{1}{n}\right)}_{f_n(x_i) = probability} k_{x_i}(x)$$

$$\hat{F}(x) = \sum_{i=1}^{n} \underbrace{\left(\frac{1}{n}\right)}_{f_n(x_i) = probability} K_{x_i}(x)$$

x_i is a sample point

x is the estimation point

Kernel distribution is 1 for observation points more than one bandwidth to the left

Kernel distribution is 0 for observation points more than one bandwidth to the right

$K_6(13) > K_{10}(13) > K_{25}(13)$

ex) To find $K_{12}(11)$, linearly interpolate between $K_{12}(7)$ and $K_{12}(17)$

Triangular

Height of triangle is $\dfrac{1}{b}$

Base of triangle is $2b$

Expected Values

$E\left[X \mid Y\right] = Y$ (Y is the original random variable)

The mean of the smoothed distribution is the same as the original mean

$\therefore E\left[X\right] = E\left[Y\right]$

Uniform Kernel

$$Var(X) = Var(Y) + \frac{b^2}{3}$$

Triangular Kernel

$$Var(X) = Var(Y) + \frac{b^2}{6}$$

Approximations for Large Data Sets

d_j is the # of left truncated observations in $\left[c_j, c_{j+1}\right)$

- number of new entrants

u_j is the # of right censored observations in $\left(c_j, c_{j+1}\right]$

x_j is the # of losses in $\left(c_j, c_{j+1}\right]$

r_j is the risk set $\left(c_j, c_{j+1}\right]$

$q_j^{'}$ is the decrement rate in $\left(c_j, c_{j+1}\right]$

$$q_j^{'} = \frac{x_j}{r_j}$$

$$P_{j+1} = P_j + d_j - u_j - x_j$$

$v_j = \#$ withdrawals

$w_j = \#$ survivors

$$v_j + w_j = u_j$$

All entries/withdrawals at endpoints

$$r_j = P_j + d_j$$

All entries/withdrawals uniformly distributed

$$r_j = P_j + 0.5\left(d_j - v_j\right)$$

Multiple Decrements

$$p^{(\tau)} = p^{'(1)} \cdot p^{'(2)} \ldots$$

$${}_{t+3}p_t^{'(x)} = \left(1 - \frac{x_t}{r_t}\right)\left(1 - \frac{x_{t+1}}{r_{t+1}}\right)\left(1 - \frac{x_{t+2}}{r_{t+2}}\right)$$

$${}_{t+3}q_t^{'(x)} = 1 - {}_{t+3}p_t^{'(x)}$$

Parametric Models

Method of Moments

$$m = \frac{\sum_{i=1}^{n} x_i}{n} \qquad t = \frac{\sum_{i=1}^{n} (x_i)^2}{n}$$

Distribution	Formulas	Formulas
Exponential	$\theta = m$	
Gamma	$\hat{\alpha} = \frac{m^2}{t - m^2}$	$\hat{\theta} = \frac{t - m^2}{m}$
Pareto	$\hat{\alpha} = \frac{2t - 2m^2}{t - 2m^2}$	$\hat{\theta} = \frac{mt}{t - 2m^2}$
Lognormal	$\hat{\mu} = 2\ln(m) - 0.5\ln(t)$	$\hat{\sigma}^2 = -2\ln(m) + \ln(t)$
Uniform on $[0, \theta]$	$\hat{\theta} = 2m$	

When they don't specify which moment to use, use the first 'k' moments, where 'k' is the number of parameters you're fitting.

For an inverse exponential, add the reciprocals to get the mean.

Percentile Matching

$\hat{\pi}_p = x_{(n+1)p}$ if $(n+1)p$ is an integer

Otherwise multiply $(n+1) \cdot p$ and interpolate

The smoothed empirical percentile is not defined if the product is less than 1 or greater than n

Maximum Likelihood

Type of Data	Formula
Discrete distribution, individual data	p_x
Continuous distribution, individual data	$f(x)$
Grouped Data	$F\left(c_j\right)-F\left(c_{j-1}\right)$
Individual Data censored from above at u	$1-F(u)$ for censored observations
Individual Data censored from below at d	$F(d)$ for censored observations
Individual Data truncated from above at u	$\dfrac{f(x)}{F(u)}$
Individual Data truncated from below at d	$\dfrac{f(x)}{1-F(d)}$

Cases where MLE = Method of Moments Estimator

(if no censored or truncated data)

Distribution	Result
Exponential	MLE = MoM
Gamma	MLE = MoM if fixed α
Normal	MLE = MoM
Poisson	MLE = MoM
Negative Binomial	MLE = MoM if *r* is known
Binomial	MLE = MoM if *m* is known

If the MLE is the sample mean, the variance of the MLE is the variance of the distribution $=\dfrac{Var(X)}{n}$

Common Likelihood Functions

MLE Function	MLE
$L\left(\theta\right)=\theta^{-a}e^{-b/\theta}$	$\theta=\dfrac{b}{a}$
$L\left(\theta\right)=\theta^{a}e^{-b\theta}$	$\theta=\dfrac{a}{b}$
$L\left(\theta\right)=\theta^{a}\left(1-\theta\right)^{b}$	$\theta=\dfrac{a}{a+b}$

MLE Formulas

Distribution	Formula	CT?
Exponential	$\hat{\theta} = \frac{\sum_{i=1}^{n+c} (x_i - d_i)}{n}$	Yes
Lognormal	$\hat{\mu} = \frac{\sum_{i=1}^{n} \ln x_i}{n}$ $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{n} \ln^2 x_i}{n} - (\hat{\mu})^2}$	No
Inverse Exponential	$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$	No
Weibull, fixed τ	$\hat{\theta} = \sqrt[\tau]{\frac{\sum_{i=1}^{n+c} x_i^{\tau} - \sum_{i=1}^{n+c} d_i^{\tau}}{n}}$	Yes
Uniform on Individual Data $(0,\theta]$	$\theta = \max\{x_i\}$	No
Uniform on Grouped Data $(0,\theta]$ OR Some observations are censored at a single point There must be at least one observation above c_j	$\hat{\theta} = c_j \left(\frac{n}{n_j}\right)$ c_j = Upper bound of highest finite interval n_j = Number of observations below c_j	No
Uniform on Grouped Data $(0,\theta]$ All groups are bounded	$\hat{\theta}$ = Min of 1) UB of highest interval with data 2) LB of highest interval with data * $\left(\frac{n}{n_j}\right)$	

Two-parameter Pareto, fixed θ	$\hat{\alpha} = -\frac{n}{K}$ $K = \sum_{i=1}^{n+c} \ln(\theta + d_i) - \sum_{i=1}^{n+c} \ln(\theta + x_i)$	Yes
One-parameter Pareto, fixed θ	$\hat{\alpha} = -\frac{n}{K}$ $K = \sum_{i=1}^{n+c} \ln \max(\theta, d_i) - \sum_{i=1}^{n+c} \ln x_i$	Yes
Beta, fixed θ b = 1	$\hat{a} = -\frac{n}{K}$ $K = \sum_{i=1}^{n} \ln x_i - n \ln \theta$	No
Beta, fixed θ a = 1	$\hat{b} = -\frac{n}{K}$ $K = \sum_{i=1}^{n} \ln(\theta - x_i) - n \ln \theta$	No

n = # of uncensored observations
c = # of censored observations
d = truncation point
x = observation if uncensored or the censoring point if censored
CT = formula can be used for left-truncated or right-censored data

Bernoulli Technique

Whenever there is one parameter and only 2 classes of observations, maximum likelihood will assign each class the observed frequency, and you can then solve for the parameter.

If X can be only 2 values (a or b)

$$P(X = a) = \frac{\text{\# data points = a}}{\text{\# data points}}$$

Reasons to Use Maximum Likelihood

1) Method of Moments and Percentile Matching only use a limited number of features from the sample.

2) Method of moments are hard to use with combined data.

3) Method of moments and percentile matching cannot always handle truncation and censoring.

4) Method of moments and percentile matching require arbitrary decisions on which moments or percentiles to use.

Reasons NOT to Use Maximum Likelihood

1) There's no guarantee that the likelihood can be maximized – it can go to infinity.

2) There may be more than one maximum.

3) There may be local maxima in addition to the global maximum; these must be avoided.

4) It may not be possible to find the maximum by setting the partial derivatives to zero; a numerical algorithm may be necessary.

Fisher's Information

1 Parameter

$$I(\theta) = -E\left[\frac{d^2}{d\theta^2} l(\theta)\right] \qquad \text{Fisher's Information}$$

$$\text{Var}(\theta) = \frac{1}{I(\theta)}$$

2 Parameters

$$I(\alpha,\theta) = -E\begin{pmatrix} \frac{\partial^2}{\partial\alpha^2} l(\alpha,\theta) & \frac{\partial}{\partial\alpha}\frac{\partial}{\partial\theta} l(\alpha,\theta) \\ \frac{\partial}{\partial\alpha}\frac{\partial}{\partial\theta} l(\alpha,\theta) & \frac{\partial^2}{\partial\theta^2} l(\alpha,\theta) \end{pmatrix}$$

$$\text{Var}(\hat{\alpha},\hat{\theta}) = \begin{pmatrix} Var(\alpha) & Cov(\alpha\theta) \\ Cov(\alpha\theta) & Var(\theta) \end{pmatrix} = I^{-1}(\alpha\theta)$$

Inverting a Matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Delta Method

$$Var(g(X)) \approx Var(X)\left(\frac{dg}{dx}\right)^2 \qquad \text{1 Variable}$$

$$Var(g(X,Y)) \approx Var(X)\left(\frac{\partial g}{\partial x}\right)^2 + 2Cov(X,Y)\frac{\partial g}{\partial x}\frac{\partial g}{\partial y} + Var(Y)\left(\frac{\partial g}{\partial y}\right)^2 \qquad \text{2 Variables}$$

$$Var(g(X)) \approx (\partial g)'(\Sigma)(\partial g) \qquad \text{General}$$

where $\partial g = \left(\frac{\partial g}{\partial x_1}, \ldots, \frac{\partial g}{\partial x_k}\right)'$ and Σ is the covariance matrix

Take derivative with respect to unknown variable

Fitting Discrete Distributions

Distribution	Method of Moments	MLE
Poisson	$\hat{\lambda} = \overline{x}$	$\hat{\lambda} = \overline{x}$
Negative Binomial	$\hat{r} = \dfrac{\overline{x}^2}{\hat{\sigma}^2 - \overline{x}}$ $\hat{\beta} = \dfrac{\hat{\sigma}^2 - \overline{x}}{\overline{x}}$ $\hat{\sigma}^2 =$ Sample Variance (divide by n)	$\hat{r}\hat{\beta} = \overline{x}$
Binomial		$\hat{q} = \dfrac{\overline{x}}{m}$

Choosing between (a,b,0) distributions to fit the data:

1) Compare sample variance $\hat{\sigma}^2$ to sample mean $\overline{x}$

- Poission: Variance = Mean
- Negative Binomial: Variance > Mean
- Binomial: Variance < Mean

2) Calculate $\dfrac{kn_k}{n_{k\text{-}1}}$ and observe the slope as a function of k

If ratios are increasing, then a > 0

Poisson = 0 slope	
Negative Binomial = positive slope	
Binomial = negative slope	

n is the # of policies/observations of k

The variance of a mixture is always at least as large as the weighted average of the variances of the components and usually greater due to:

$$\underbrace{Var(X)}_{\text{variance of a mixture}} = \underbrace{E\left[Var(X \mid I)\right]}_{\text{weighted average of variances}} + Var\left(E\left[X \mid I\right]\right)$$

Asymptotic Variance of MLE's

Distribution	Formula
Exponential	$Var\left(\hat{\theta}\right)=\frac{\theta^2}{n}$
Uniform $\left(0,\theta\right]$	$Var\left(\hat{\theta}\right)=\frac{n\theta^2}{\left(n+1\right)^2(n+2)}$
Weibull fixed τ	$Var\left(\hat{\theta}\right)=\frac{\theta^2}{n\tau^2}$
Pareto fixed θ	$Var\left(\hat{\alpha}\right)=\frac{\alpha^2}{n}$
Pareto fixed α	$Var\left(\hat{\theta}\right)=\frac{\theta^2}{n}\frac{\left(\alpha+2\right)}{\alpha}$
Lognormal	$Var\left(\hat{\mu}\right)=\frac{\sigma^2}{n}$ $Cov\left(\mu,\sigma\right)=0$ $Var\left(\hat{\sigma}\right)=\frac{\sigma^2}{2n}$
Poisson	$Var\left(\hat{\lambda}\right)=\frac{\lambda}{n}$

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

Hypothesis Tests – Graphic Comparison

$$F^*(x) = \frac{F(x) - F(d)}{1 - F(d)}$$

$$f^*(x) = \frac{f(x)}{1 - F(d)}$$

$D(x)$ plots

$$D(x) = F_n(x) - F^*(x)$$
$$= \text{Empirical - Fitted}$$

Empirical calculation uses a denominator of n

If $D(x) > 0$, then $F_n(x) > F^*(x)$

– had more data $\leq$ x than predicted by model

If $D(x) < 0$, then $F_n(x) < F^*(x)$

– had less data $\leq$ x than predicted by model

If data truncated at d, $D(d) = 0$

Every vertical jump has distance $\frac{1}{n}$

$p-p$ plots

On horizontal axis, one point every multiple of $\frac{1}{n+1}$

Domain and Range of Graph are $[0,1]$

Points are $\left(F_n\left(x_j\right), F^*\left(x_j\right)\right)$

The first data point corresponds to the first sample value

If slope is less than 45°, fitted distribution is putting too little weight in that region

If slope is more than 45°, fitted distribution is putting too much weight in that region

Don't plot censored values

Kolmogorov-Smirnov Test

$$D = \max_{d<x<u} \left|F_n(x) - F^*(x)\right|$$

Max occurs right before or after a jump

If D < critical value, do not reject H_0 (null hypothesis)

Having censored data lowers D and also lowers the critical value

Critical values $\to 0$ as n $\to \infty$

If data are X = 2000, 4000, 5000, 5000

and the 5000 values are right-censored

Then $F_4\left(5000^-\right) = F_4\left(5000\right) = 0.5$

Anderson-Darling Test

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

n includes censored observations

Chi-Square

$$Q=\sum_{j=1}^{k}\frac{\left(O_j-E_j\right)^2}{E_j}$$

$$Q=\sum_{j=1}^{k}\left(\frac{{O_j}^2}{E_j}\right)-n$$

$O_j=$ # of observations in each group

$E_j=np_j$ (Expected # of observations in each group)

$n=$ Total # of observations

Each group should have at least 5 expected observations. If not, you have to bring some groups together.

Degrees of Freedom

$k-1$ DoF $\begin{cases}\text{Distribution with parameters is given}\\ \text{Distribution is fitted/estimated by MLE or using different data}\end{cases}$

$k-1-r$ DoF $\{$Parameters are fitted from the data

$r=$ # parameters

$k=$ # groups

Independent Periods

$$Q=\sum_{j=1}^{k}\frac{\left(O_j-E_j\right)^2}{V_j}$$ where V_j is the variance

Degrees of Freeedom $=$ k - p (p is number of estimated parameters)

Kolmogorov-Smirnov	Anderson-Darling	Chi-square	Loglikelihood
Individual Data	Individual Data	Individual or Grouped Data	Individual or Grouped Data
Continuous Fits		Continuous or Discrete Fits	
If there is censored data $\mu < \infty$, should lower critical value	If there is censored data $\mu < \infty$, should lower critical value	If there is censored data $\mu < \infty$, no adjustment of critical value	If there is censored data $\mu < \infty$, no adjustment of critical value
If parameters are fitted, critical value should be lowered	If parameters are fitted, critical value should be lowered	If parameters are fitted, critical value adjusts automatically	If parameters are fitted, critical value adjusts automatically
Larger sample size makes critical value decline	Critical value independent of sample size	Critical value independent of sample size	Critical value independent of sample size
No Discretion in grouping of data	No Discretion in grouping of data	Discretion in grouping of data	
Uniform weight on all parts of distribution	Higher weight on tails of distribution	Higher weight on intervals with low fitted probability	

Type I Error: Rejecting H_0 when it is true

Type II Error: Rejecting H_1 when it is true

Likelihood Ratio Algorithm

If parameters are added to the model, the new model will have a loglikelihood at least as great.

The # DoF for the likelihood ratio test is the number of free parameters in the alternative model minus the number of free parameters in the base model (null hypothesis).

Compare $2(LogL_1 - LogL_2)$ to critical value at selected chi-square percentile and DOF

$LogL_1$ = Alternative Model Loglikelihood (which will be higher)

$LogL_2$ = Base Model

If $2(LogL_1 - LogL_2) >$ critical value, accept alternative hypothesis

Start by comparing best 2-parameter to best 1-parameter

If $2(LogL_1 - LogL_2) <$ Critical Value (it fails), compare best 3-parameter distributions to best 1-parameter

If $2(LogL_1 - LogL_2) >$ Critical Value (it passes), compare 3-parameter distributions to best 2-parameter

Schwarz-Bayesian Criterion

$$= LogL - \left(\frac{r}{2}\right)\ln n$$

where r is the # parameters

where n is the # of data points

The distribution with the highest resulting LogL is selected

Credibility

$$\boxed{e_F = n_0 CV^2}$$

where $n_0 = \left(\frac{y_p}{k}\right)^2$

y_p = coefficient from the standard normal = $\Phi^{-1}\left(\frac{1+p}{2}\right)$

Given y_p, P% = $100\left(2\left(\text{percent corresponding to } y_p\right)-1\right)$

k = maximum fluctuation you will accept (i.e. within 5%)

Limited Fluctuation Credibility: Poisson Frequency

All λ must be the same!

$$1+CV_s^2 = \frac{E\left[X^2\right]}{E\left[X\right]^2}$$

	Credibility for		
Experience expressed in	Number of Claims	Claim Size (Severity)	Aggregate Losses/Pure Premium
Exposure Units e_F	$\frac{n_0}{\lambda}$	$\frac{n_0}{\lambda}\left(CV_s^2\right)$	$\frac{n_0}{\lambda}\left(1+CV_s^2\right)$
Number of Claims n_F	$\boxed{n_0}$	$n_0\left(CV_s^2\right)$	$\boxed{n_0\left(1+CV_s^2\right)}$
Aggregate Losses s_F	$n_0\mu_s$	$n_0\mu_s\left(CV_s^2\right)$	$n_0\mu_s\left(1+CV_s^2\right)$

Pure Premium is the expected aggregate loss per policyholder per time period.

Limited Fluctuation Credibility: Non-Poisson Frequency

$$e_F = n_0 CV_s^2$$

$$n_F = e_F \mu_f$$

	Credibility for		
Experience expressed in	Number of Claims	Claim Size (Severity)	Aggregate Losses/Pure Premium
Exposure Units e_F	$n_0\left(\frac{\sigma_f^2}{\mu_f^2}\right)$	$n_0\left(\frac{\sigma_s^2}{\mu_s^2 \cdot \mu_f}\right)$	$n_0\left(\frac{\sigma_f^2}{\mu_f^2} + \frac{\sigma_s^2}{\mu_s^2 \cdot \mu_f}\right)$
Number of Claims n_F	$n_0\left(\frac{\sigma_f^2}{\mu_f}\right)$	$n_0\left(\frac{\sigma_s^2}{\mu_s^2}\right)$	$n_0\left(\frac{\sigma_f^2}{\mu_f} + \frac{\sigma_s^2}{\mu_s^2}\right)$
Aggregate Losses s_F	$n_0\mu_s\left(\frac{\sigma_f^2}{\mu_f}\right)$	$n_0\mu_s\left(\frac{\sigma_s^2}{\mu_s^2}\right)$	$n_0\mu_s\left(\frac{\sigma_f^2}{\mu_f} + \frac{\sigma_s^2}{\mu_s^2}\right)$

# of Insureds is Exposure

Partial Credibility

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

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

P_C = Credibility Premium

M = Manual Premium

Z = Credibility

$\bar{X}$ = Observed Mean

$$Z = \sqrt{\frac{n}{n_F}}$$

n = Expected Claims

n_F = Number of Expected Claims needed for Full Credibility

Bayesian Credibility

Bayesian Methods – Discrete Prior

	Class 1	Class 2
1) Prior Probabilities		
2) Likelihood of Experience		
3) Joint Probabilities	Product of rows above	Product of rows above
4) Posterior Probabilities	Quotients of row 3 over row 3 sum	Quotients of row 3 over row 3 sum
5) Hypothetical Means		
6) Bayesian Premium	Product of rows 4 and 5	Product of rows 4 and 5

Bayesian Premium is the predicted expected value of the next trial

claims is Bernoulli means at most 1 can occur

Bayesian Methods – Continuous Prior

$$\pi\left(\theta \mid x_1,\dots,x_n\right)=\frac{\pi(\theta)\cdot f\left(x_1,\dots,x_n \mid \theta\right)}{\int \pi(\theta)\cdot f\left(x_1,\dots,x_n \mid \theta\right)d\theta}$$

Posterior Density

Limits of Integration are according to prior distribution

$$f\left(x_{n+1} \mid x_1,\dots,x_n\right)=\int f\left(x_{n+1} \mid \theta\right)\cdot\pi\left(\theta \mid x_1,\dots,x_n\right)d\theta$$

Predictive Density

$\pi(\theta)$ is the prior density

$\pi\left(\theta \mid x_1,\dots,x_n\right)$ is the posterior density

$f\left(x_1,\dots,x_n \mid \theta\right)$ is the likelihood function of the data given θ (conditional)

$$f\left(x_1,\dots,x_n \mid \theta\right)=\prod_{i=1}^{n} f\left(x_i \mid \theta\right)$$

$f\left(x_1,\dots,x_n\right)$ is the unconditional joint density function

$$f\left(x_1,\dots,x_n\right)=\int f\left(x_1,\dots,x_n \mid \theta\right)\pi(\theta)d\theta$$

$$\int_0^{\infty} t^n e^{-\delta t} dt = \frac{n!}{\delta^{n+1}}$$

Bayesian Credibility: Poisson/Gamma

$N \sim Poisson(\lambda)$

$\lambda \sim Gamma\left(\alpha, \gamma = \frac{1}{\theta}\right)$

$\alpha_* = \alpha + \text{claims}$

$\gamma_* = \gamma + \text{exposures}$

$P_c = \frac{\alpha_*}{\gamma_*}$

Posterior: $Gamma\left(\alpha_*, \theta_* = \frac{1}{\gamma_*}\right)$

Posterior mean is the avg. # claims/policy

Predictive: Negative Binomial$\left(r = \alpha_*, \beta = \theta_*\right)$

Bayesian Credibility: Normal/Normal

$X \sim Normal(\theta, v)$

$\theta \sim Normal(\mu, a)$

$\bar{x} = \text{Observed Average}$

$n = \text{Exposure}$

$\mu_* = \frac{v\mu + an\bar{x}}{v + an}$ Posterior Mean

$a_* = \frac{va}{v + an}$ Posterior Variance

Predictive Mean $= \mu_*$

Predictive Variance $= v + a_*$

Bayesian Credibility: Lognormal/Normal

$X \sim Lognormal(\theta, v)$

$\theta \sim Normal(\mu, a)$

Find $\frac{\sum \ln x_i}{n} = \overline{x}$

$\mu_* = \frac{v\mu + an\overline{x}}{v + an}$ Posterior Mean

$a_* = \frac{va}{v + an}$ Posterior Variance

$$E[X \mid \theta] = E\left[e^{\theta + 0.5v}\right] = E\left[e^{\theta}\right] e^{0.5v} = e^{(\mu_* + 0.5a_*)} e^{0.5v}$$

Bayesian Credibility: Bernoulli/Beta

Probability of a claim = q

$q \sim Unif(0,1)$

Uniform is a special case of Beta distribution with $a = b = 1$ and $\theta = 1$

$k = \#$ claims

n = exposure

$Beta = Cx^{a-1}(1-x)^{b-1}$

$a_* = a + \text{claims}$
$b_* = b + \text{exposures - claims}$ } Plug into Posterior Distribution

$$E[\theta \mid x] = \frac{a_*}{a_* + b_*}$$

If $m > 1$, treat as a series of 'm' Bernoullis

If exposure is 2 years, treat as '2m' Bernoullis

1) $n = 2m$

and

2) $\frac{a_*}{a_* + b_*} \rightarrow m\left(\frac{a_*}{a_* + b_*}\right)$

$\Gamma(x+1) = x\Gamma(x)$

Bayesian Credibility: Exponential/Inverse Gamma

$$f(x \mid \theta) = \frac{1}{\theta} e^{-x/\theta}$$ Exponential

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \cdot \frac{e^{-\beta/\theta}}{\theta^{\alpha+1}}$$ Inverse Gamma

$$\alpha_* = \alpha + n$$

$$\beta_* = \beta + n\bar{x}$$

$$E(\text{next loss}) = \frac{\beta_*}{\alpha_* - 1}$$

If $f(x)$ is Gamma instead of exponential

$$f(x \mid \theta) = \frac{1}{\Gamma(\eta)\theta^{\eta}} x^{\eta-1} e^{-x/\theta}$$

$$\alpha_* = \alpha + \eta n$$

$$\beta_* = \beta + n\bar{x}$$

Loss Functions

For the loss function minimizing MSE

$$l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$$

Bayesian Point Estimate is the mean of the posterior distribution

For the loss function minimizing Absolute Value of the Error

$$l(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$$

Bayesian Point Estimate is the median of the posterior distribution

For the zero-one loss function

Bayesian Point Estimate is the mode of the posterior distribution

Buhlmann Credibility

Buhlmann Credibility: Basics

$\mu = E_\Theta\left[\mu(\Theta)\right]$ Overall Mean (Expected Value of the Hypothetical Mean)

$v = E_\Theta\left[v(\Theta)\right]$ Expected Value of the Process Variances

$a = Var_\Theta\left[\mu(\Theta)\right]$ Variance of the Hypothetical Mean

$a + v =$ Overall Variance

For Poisson frequency HM = PV

Buhlmann's k

$$k = \frac{v}{a}$$

Buhlmann's Credibility Z

$$Z = \frac{n}{n+k} = \frac{na}{na+v}$$

$n =$ # periods when studying frequency or aggregate losses
$n =$ # claims when studying severity

If given 2 classes and there are multiple groups within each class, you must find the mean and variance of each group separately.

Buhlmann Credibility: Continuous

Given a distribution function and the prior function:

1) Use the distribution function to get the HM and PV

2) Find v and a using the prior distribution

Model	Prior	Posterior	Predictive	Buhlmann v	Buhlmann a
Poisson (λ)	Gamma α $\gamma = \frac{1}{\theta}$	Gamma $\alpha_* = \alpha + \text{claims}$ $\gamma_* = \gamma + \text{exposures}$	Negative Binomial $r = \alpha_*$ $\beta = \theta_* = \frac{1}{\gamma_*}$	$\alpha\theta$	$\alpha\theta^2$
Bernoulli (q)	Beta a b	Beta $a_* = a + \text{claims}$ $b_* = b + \text{exposures - claims}$	Bernoulli $q = E[\theta \mid x] = \frac{a_*}{a_* + b_*}$	$\frac{ab}{(a+b)(a+b+1)}$	$\frac{ab}{(a+b)^2(a+b+1)}$
Normal (θ, v)	Normal μ a	Normal $\mu_* = \frac{v\mu + an\overline{x}}{v + an}$ $a_* = \frac{va}{v + an}$	Normal $\mu = \mu_*$ $\sigma^2 = a_* + v$	v	a
Exponential (θ)	Inverse Gamma α θ	Inverse Gamma $\alpha_* = \alpha + n$ $\theta_* = \theta + n\overline{x}$	Pareto $\alpha = \alpha_*$ $\theta = \theta_*$	$\frac{\theta^2}{(\alpha-1)(\alpha-2)}$	$\frac{\theta^2}{(\alpha-1)^2(\alpha-2)}$

Exact Credibility

If you have conjugate pairs and they ask for a Buhlmann estimate, use the Bayesian estimate.

Buhlmann as Least Squares Estimates of Bayes

E[Initial probabilities x Outcomes] = E[Initial probabilities x Bayesian Estimates]

$$\hat{Y}_i = \alpha + \beta X_i$$

$$\beta = \frac{Cov(X,Y)}{Var(X)} = Z$$

$$\alpha = (1-Z)E[X]$$

$$Var(X) = \sum p_i X_i^2 - E[X]^2$$

$$Cov(X,Y) = \sum p_i X_i Y_i - E[X]E[Y]$$

where X are the initial outcomes and Y are the Bayesian observations

Buhlmann Predictions

$$P_c(\underbrace{0}_{\text{first observation = 0}}) = (1-Z)E[X]$$

$$P_c(2) = (1-Z)E[X] + 2Z$$

$$P_c(8) = (1-Z)E[X] + 8Z$$

Graphics Questions

1) The Bayesian prediction must be within the range of the hypothetical means
   -within range of the prior distribution
2) The Buhlmann predictions must lie on a straight line
3) There should be Bayesian predictions both above and below the Buhlmann line
4) The Buhlmann prediction must be between the overall mean and the observation

$$\frac{Cov(X_i X_j)}{}$$

$$Cov(X_i X_j) = a$$

$$Var(X_i) = v + a$$

Empirical Bayes Non-Parametric Methods

	Uniform Exposures	Non-Uniform Exposures
$\hat{\mu}$ Mean of all data	$\overline{X}$	$\overline{X}$
$\hat{v}$ Mean of sample variances of the rows	$\frac{1}{r(n-1)}\sum_{i=1}^{r}\sum_{j=1}^{n}(\underset{\text{avg/cell}}{x_{ij}} - \underset{\text{avg/class}}{\bar{x}_i})^2$	$\dfrac{\sum_{i=1}^{r}\sum_{j=1}^{n} m_{ij}(\underset{\text{avg/cell}}{x_{ij}} - \underset{\text{avg/class}}{\bar{x}_i})^2}{\sum_{i=1}^{r}(\underset{\text{per class}}{n_i} - 1)}$
$\hat{a}$	$\frac{1}{r-1}\left(\sum_{i=1}^{r}(\underset{\text{avg/class}}{x_i} - \underset{\text{overall avg}}{\bar{x}})^2\right) - \frac{\hat{v}}{n}$	$\dfrac{\sum_{i=1}^{r}\underset{\text{exp/class}}{m_i}(\underset{\text{avg/class}}{x_i} - \underset{\text{overall avg}}{\bar{x}})^2 - \hat{v}(r-1)}{\underset{\text{overall}}{m} - \dfrac{\sum_{i=1}^{r}\underset{\text{exp/class}}{m_i^2}}{\underset{\text{overall}}{m}}}$
n	Years of Experience	# Policyholders

r = # groups
m = # exposures

To calculate individual variances

$$v_1 = \frac{\sum\left(X_i - \overline{X}\right)^2}{n-1}$$

In the P_C formula, M = the average of ALL claims

Empirical Bayes Semi-Parametric Methods

Poisson Model

$\hat{\mu} = \bar{x}$

$\hat{v} = \bar{x}$

$\hat{a} = s^2 - \hat{v}$

$$s^2 = \sum \frac{\left(\bar{X}_i - \bar{X}\right)^2}{r-1}$$

$r = \#$ Policyholders

$Z = \dfrac{\hat{a}}{\hat{a}+\hat{v}}$ regardless of # of years (but if non-uniform exposures, use n = # exposures for the group you are looking at)

If non-uniform exposures, $\hat{a}$ must be calculated using Non-parametric formula

For P_C

1) $\bar{X}$= total # observed claims (but if non-uniform exposures, use the average)

2) If exposure is 5 years, divide P_C by 5 to get estimate for 1 year (next year)

Non-Poisson Model

1) Negative Binomial with fixed β

$E\left[N \mid r\right] = r\beta$

$Var\left(N \mid r\right) = r\beta(1+\beta)$

$\hat{\mu} = \bar{x}$

$\hat{v} = \bar{x}(1+\beta)$

$\hat{a} = s^2 - \hat{v}$

2) Gamma with fixed θ

$E\left[X \mid \alpha\right] = \alpha\theta$

$Var\left(X \mid \alpha\right) = \alpha\theta^2$

$\hat{\mu} = \bar{x}$

$\hat{v} = \bar{x}\theta$

$\hat{a} = s^2 - \hat{v}$

<u>**Simulation**</u>

<u>Inversion Method</u>

1) Get $u = F(x)$

2) Solve for x

3) Plug in 'u' to get simulated value

If $F(2^{-}) = .25$

$F(2) = .75$

Then $.25 \le u \le .75$ is mapped to $x = 2$

If F(x) = 'a' (constant) in range $[x_1, x_2)$, then map a $\rightarrow x_2$

If given a graph with (x, F(x))

1) Start on the y-axis with the u values
2) Move right until you hit the line
   If the line is horizontal, keep going right until it starts going up
3) Go vertically down to x
4) x is the simulated value

Number of Data Values to Generate

$$Var\left(\hat{F}(x)\right) = \frac{F(x)\left(1 - F(x)\right)}{n}$$

$$e_F = n_0 CV^2$$

Estimated Item

Mean

Confidence Interval:

$$\bar{x} \pm z_\pi \left(\frac{s_n}{\sqrt{n}} \right)$$

s_n is the square root of the unbiased sample variance after n runs

Number of Runs:

Calculates number of runs needed for the sample mean to be within 100k% of the true mean.

$$n \geq n_0 CV^2$$

Must use unbiased Variance in CV^2

Remember that $Var\left(\bar{X}\right) = \frac{1}{n} Var\left(X\right)$

F(x)

Confidence Interval:

$$F(x) \pm z_\pi \sqrt{VaR\left(F(x)\right)}$$

$$F(x) \pm z_\pi \sqrt{\frac{F(x)\left(1 - F(x)\right)}{n}}$$

Number of Runs:

$$n \geq n_0 \left(\frac{n - P_n}{P_n} \right) = n_0 \left(\frac{1 - \frac{P_n}{n}}{\frac{P_n}{n}} \right)$$

$P_n = \#$ runs below x

Percentiles π_q

Confidence Interval:

$$[Y_a, Y_b]$$

$$a = \left\lfloor nq + 0.5 - z_{\frac{1+p}{2}} \sqrt{nq(1-q)} \right\rfloor$$

$$b = \left\lceil nq + 0.5 + z_{\frac{1+p}{2}} \sqrt{nq(1-q)} \right\rceil$$

Risk Measures

$$TVaR_p(X) = E\left[L \mid L > VaR_p\right]$$

$$= \boxed{VaR_p(X) + e_X\left(VaR_p(X)\right)} = \boxed{\frac{E[L] - E\left[L \wedge \pi_p\right]}{1-p} + \pi_p}$$

$$= \frac{\int_{F_X^{-1}(p)}^{\infty} x f(x)dx}{1-p}$$

$$= \boxed{\frac{\int_p^1 VaR_y(X)dy}{1-p}}$$

$TVaR_q(X)$ is the mean of the upper tail of the distribution

$TVaR_q(X) =$ Conditional Tail Expectation

$$TVaR_q(X) = \frac{\sum_{j=k}^{n} Y_j}{n-k+1}$$

$$s_q^2 = \frac{n}{n-1}\left(E\left[\left(TVaR_q(X)\right)^2\right] - E\left[TVaR_q(X)\right]^2\right)$$

$$\boxed{Var\left(TVaR_q(X)\right) = \frac{s_q^2 + q\left(TVaR_q(X) - VaR_q(X)\right)^2}{n-k+1}}$$

Confidence Interval $= TVaR_q(X) \pm z_\pi \sqrt{Var\left(TVaR_q(X)\right)}$

Estimate of $VaR_q(X)$ is Y_k where $k = \lfloor nq \rfloor + 1$

So if simulation has 1000 runs and you're estimating 95th percentile,
Then use Y_{951}

Bootstrap Approximation

$\theta(F)$ is the parameter

$g\left(x_1,...,x_n\right)$ is an estimator based on a sample of 'n' items

$$MSE_{g\left(x_1,...,x_n\right)}\left(\theta(F)\right) = E_{F_n}\left[\left(g\left(x_1,...,x_n\right) - \theta(F_n)\right)^2\right]$$

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n}$$

Estimating the Mean with the Sample Mean

$$MSE_{\overline{x}}\left(\mu\right) = \frac{\hat{\sigma}^2}{n} = \boxed{\frac{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}{n^2}}$$

Sums of Distributions

Single	Multiple
Bernoulli	Binomial
Binomial	Binomial
Poisson	Poisson
Geometric	Negative Binomial
Negative Binomial	Negative Binomial
Normal	Normal
Exponential	Gamma
Gamma	Gammas
Chi-Square	Chi-Square