Bühlmann Credibility

- Let

$$X_i|\theta \sim f\left(x|\mu(\theta),\sigma^2(\theta)\right)$$

denote the i^{th} observation of a claim count (frequency), claim amount (severity), or total claim cost (pure premium), where:

$\mu(\theta) = E_{X|\theta}\left[X_i|\theta\right]$, the hypothetical mean (HM), and
$\sigma^2(\theta) = Var_{X|\theta}\left[X_i|\theta\right]$, the process variance (PV).

- Let

$$\theta \sim \pi(\theta)$$

denote a "state of nature" parameter, where:

$E_\theta\left[\mu(\theta)\right] =$ the prior mean (H),
$Var_\theta\left[\mu(\theta)\right] =$ the variance of the hypothetical means (VHM),
$E_\theta\left[\sigma^2(\theta)\right] =$ the expected value of the process variance ($EVPV$), and
$R = \overline{X}$, a recent observation.

- Estimate the Bayesian posterior mean, $E_{\theta|\mathbf{x}}\left[\mu(\theta)|X_1, X_2,\ldots, X_n\right]$, using the Bühlmann credibility estimator,

$$C = zR + (1-z)H,$$

where:

$$z = \frac{n}{n+k},$$

n = sample size used to calculate $\overline{X}$, and

$$k = \frac{EVPV}{VHM} = \frac{E_\theta\left[\sigma^2(\theta)\right]}{Var_\theta\left[\mu(\theta)\right]}.$$

- The following is a helpful way of thinking about n:

$$n = \frac{\text{number of exposure units associated with } R}{\text{number of exposure units associated with } C}.$$

- Estimating VHM:

In practice, VHM may not initially be known, but may be estimated using the total variance formula,

$$Var_X[X_i] = E_\theta\left[Var_{X|\theta}[X_i|\theta]\right] + Var_\theta\left[E_{X|\theta}[X_i|\theta]\right]$$

$$= EVPV + VHM.$$

Given observations $X_i|\theta$ that are grouped according to different values of the underlying parameter θ, it is possible:

(1) to estimate $Var_X[X_i]$ using the overall sample variance of the X_i, and

(2) to estimate $EVPV$ by first estimating $Var_{X|\theta}[X_i|\theta]$ for each θ, and then averaging over all values of θ.

Caution: This approach yields an estimator of VHM that may sometimes be negative.

Note that if $X_i|\theta \sim \text{Poisson}(\theta)$, then

$$Var_{X|\theta}[X_i|\theta] = E_{X|\theta}[X_i|\theta] = \theta, \text{ and so}$$

$$EVPV = E_\theta\left[E_{X|\theta}[X_i|\theta]\right] = E_X[X_i].$$

Thus, in this case, $EVPV$ may be estimated using the overall sample mean of the X_i.