

Graphics in the slide tutorial are from analyses on the New Zealand earthquake data.

Assuming no deductibles and no incorporation of old distributions:

1. Import data: MDRs, damage ratios.
2. Model P_0 .
3. Compute 18999 empirical PMFs.
4. Fit Transformed Beta (TrB) to the 18999 PMFs.
5. Discretization.

Import and order data

Import data: (MDRs, claim damage ratios).

1. Order data (rows) by MDR and claim damage ratio (CDR).
2. Reshuffle data using a user-defined random seed.
3. Order data by MDR.

If MDRs are not unique, different orders of rows can lead to different empirical distributions in bins.

Assume minimum bin size required = 5				CDRs in bin are different
MDR label for bin = 0.0525	MDR	CDR	MDR	CDR
	0.05	0.01	0.05	0.01
	0.05	0.02	0.05	0.02
	0.05	0.06	0.05	0.06
	0.06	0.02	0.06	0.13
	0.06	0.03	0.06	0.11
	0.06	0.08	0.06	0.08
	0.06	0.11	0.06	0.03
	0.06	0.13	0.06	0.02

Import and order data

Import data: (MDRs, claim damage ratios).

1. Order data (rows) by MDR and claim damage ratio (CDR).
2. Reshuffle data using a user-defined random seed.
3. Order data by MDR.

If MDRs are not unique, different orders of rows can lead to different empirical distributions in bins.

Even if CDRs are not sorted upon import, they could still be semi-sorted depending on how the data records were organized in databases. For example, the records could have been grouped by geographic regions / events / years, etc.

Assume minimum bin size required = 5					
MDR label for bin = 0.0525	MDR	CDR	MDR	CDR	CDRs in bin are different
	0.05	0.01	0.05	0.01	
	0.05	0.02	0.05	0.02	
	0.05	0.06	0.05	0.06	
	0.06	0.02	0.06	0.13	
	0.06	0.03	0.06	0.11	
	0.06	0.08	0.06	0.08	
	0.06	0.11	0.06	0.03	
	0.06	0.13	0.06	0.02	

Import and order data

Import data: (MDRs, claim damage ratios).

1. Order data (rows) by MDR and claim damage ratio (CDR).
2. Reshuffle data using a user-defined random seed.
3. Order data by MDR.

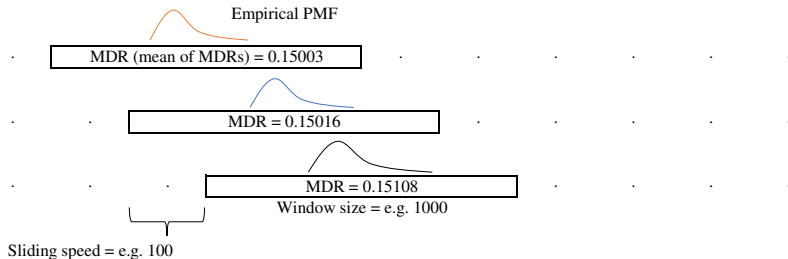
If MDRs are not unique, different orders of rows can lead to different empirical distributions in bins.

Even if CDRs are not sorted upon import, they could still be semi-sorted depending on how the data records were organized in databases. For example, the records could have been grouped by geographic regions / events / years, etc.

The “order + shuffle + order” procedure removes any potential bias and guarantees uniqueness of empirical distributions in bins, i.e. re-productibility.

Assume minimum bin size required = 5					
MDR label for bin = 0.0525	MDR	CDR	MDR	CDR	CDRs in bin are different
	0.05	0.01	0.05	0.01	
	0.05	0.02	0.05	0.02	
	0.05	0.06	0.05	0.06	
	0.06	0.02	0.06	0.13	
	0.06	0.03	0.06	0.11	
	0.06	0.08	0.06	0.08	
	0.06	0.11	0.06	0.03	
	0.06	0.13	0.06	0.02	

Sliding window and sliding speed



For prescribed MDR that is not windowed, e.g. 0.1501, mix the neighboring PMF:

The diagram shows a green PMF curve labeled "0.1501" on the left, which is a linear interpolation between an orange PMF curve labeled "0.15003" and a blue PMF curve labeled "0.15016". The interpolation is represented by the equation:

$$\leftarrow \frac{0.15016 - 0.1501}{0.15016 - 0.15003} \times \text{orange PMF} + \frac{0.1501 - 0.15003}{0.15016 - 0.15003} \times \text{blue PMF}$$

Overlapped windows \Rightarrow correlated samples \Rightarrow close empirical PMFs in shape and scale \Rightarrow smoother transition of TrBs along the MDR axis.

Sliding windows will also be used to model P_0 .

Model P_0

Set sliding window size and speed.

Compute P_0 and mean of MDRs in each window.

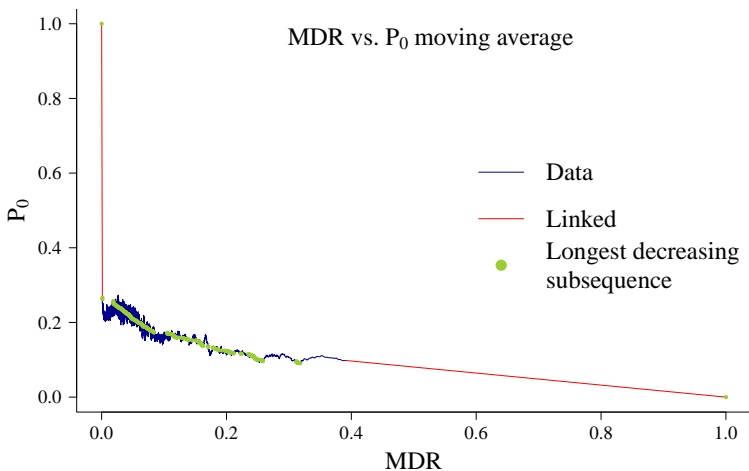
Constraint: P_0 is a decreasing function of MDR.

► $P_0(\text{MDR} = 0) = 1$.

► $P_0(\text{MDR} = 1) = 0$.

Compute the longest decreasing subsequence.

► $O(N \log N)$ [REF].



Model P_0

Set sliding window size and speed.

Compute P_0 and mean of MDRs in each window.

Constraint: P_0 is a decreasing function of MDR.

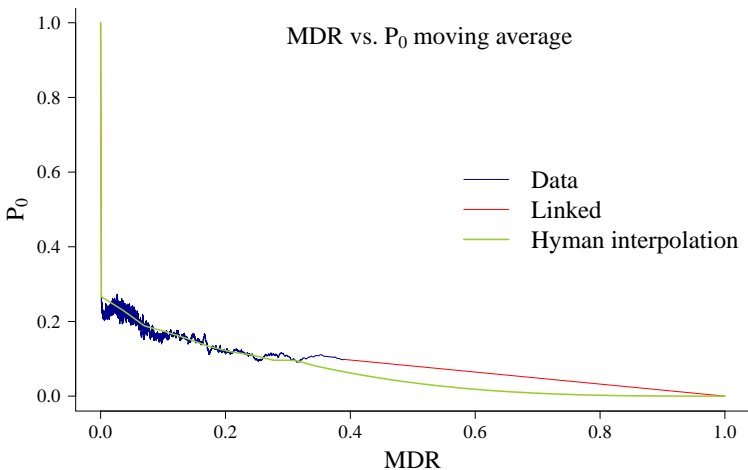
► $P_0(\text{MDR} = 0) = 1$.

► $P_0(\text{MDR} = 1) = 0$.

Compute the longest decreasing subsequence.

► $O(N \log N)$ [REF].

Monotone cubic interpolation of the subsequence [REF].



Model P_0 , robustification (not necessary but good to have and cheap to run)

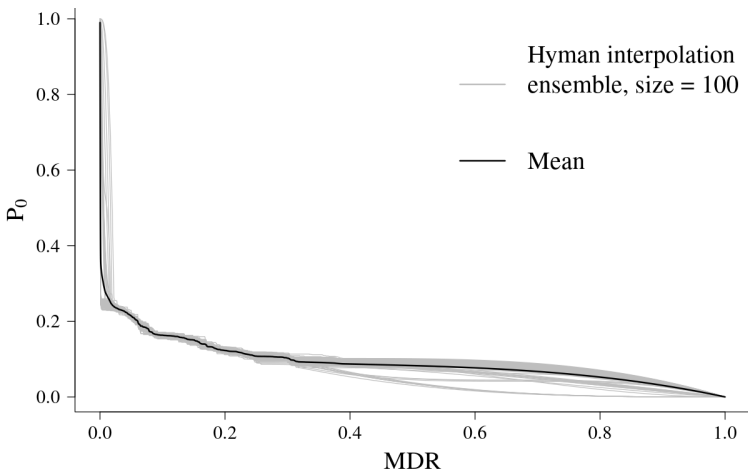
Sample 100 random subsets of data. Subset size = e.g. 2/3 of the full data size.

Compute the Hyman model in each subset.

Ensemble mean is used as the final estimate.

Note:

- ▶ To prevent overfitting, the main part i.e. $P(\text{DR} > 0)$ can also be modeled using ensemble. This would produce 100 sets of TrB parameters. The final PMF table would be a mixture of the discretizations of the 100 sets of TrBs. We ignore this approach for now.



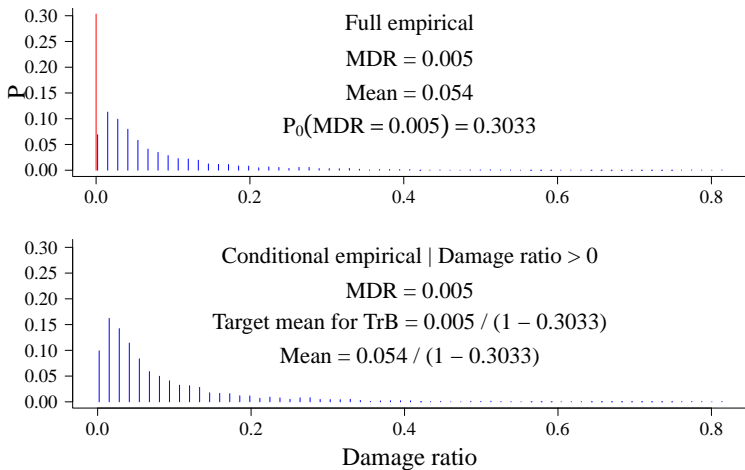
Model the main part $P(\text{DR} > 0)$

Remove data where $\text{DR} = 0$.

Set sliding window and speed.

Compute MDR mean and conditional empirical PMF in each window.

- ▶ P_0 (red) is given.
- ▶ Empirical PMF is computed using regridding.



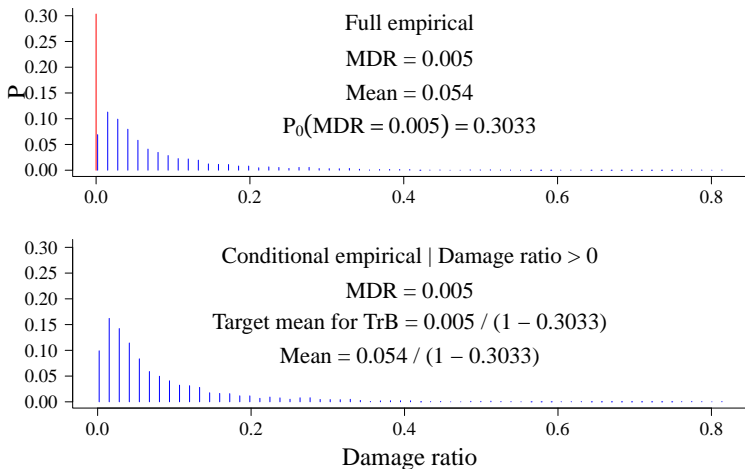
Model the main part $P(\text{DR} > 0)$

Remove data where $\text{DR} = 0$.

Set sliding window and speed.

Compute MDR mean and conditional empirical PMF in each window.

- ▶ P_0 (red) is given.
- ▶ Empirical PMF is computed using regriding.
- ▶ For prescribed MDR that is not windowed, mix the neighboring PMFs (explained previously).
- ▶ For prescribed MDR that is out of data range, extrapolate.



Extrapolation below minimum MDR in data

Example assumes
0.005 the lowest
MDR in data.

Downscale the
support of PMF
for $\text{MDR} = 0.005$
linear to the
prescribed MDRs.

Extrapolation above maximum MDR in data

Example assumes 0.5
the highest MDR in
data.

Upscale the support of
PMF for $\text{MDR} = 0.5$
linear to the prescribed
MDRs.

Upper bound the PMF's
support by 1 and load
the truncated tail on
 P_{\max} .

Denote the incomplete beta function by $\beta(u, v; y) = \int_0^y t^{u-1}(1-t)^{v-1} dt$. For TrB random variable X ,

$$\begin{aligned}\mu_{\text{lm1}} &= \mathbb{E}[\min(X, 1)] \\ &= \frac{d \cdot \Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) + 1 - \beta\left(\frac{b}{c}, \frac{a}{c}; \frac{1}{d^c + 1}\right).\end{aligned}$$

Denote the incomplete beta function by $\beta(u, v; y) = \int_0^y t^{u-1}(1-t)^{v-1}dt$. For TrB random variable X ,

$$\begin{aligned}\mu_{\text{lm1}} &= \mathbb{E}[\min(X, 1)] \\ &= \frac{d \cdot \Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) + 1 - \beta\left(\frac{b}{c}, \frac{a}{c}; \frac{1}{d^c + 1}\right).\end{aligned}$$

Interestingly,

$$\frac{\partial \mu_{\text{lm1}}}{\partial d} = \frac{\Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) > 0,$$

so μ_{lm1} is an increasing function of d and the derivative can be obtained almost free of charge after computing μ_{lm1} .

Denote the incomplete beta function by $\beta(u, v; y) = \int_0^y t^{u-1}(1-t)^{v-1}dt$. For TrB random variable X ,

$$\begin{aligned}\mu_{\text{lm1}} &= \mathbb{E}[\min(X, 1)] \\ &= \frac{d \cdot \Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) + 1 - \beta\left(\frac{b}{c}, \frac{a}{c}; \frac{1}{d^c + 1}\right).\end{aligned}$$

Interestingly,

$$\frac{\partial \mu_{\text{lm1}}}{\partial d} = \frac{\Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) > 0,$$

so μ_{lm1} is an increasing function of d and the derivative can be obtained almost free of charge after computing μ_{lm1} .

Because the target mean for $\min(X, 1)$ is also constrained by:

$$\mu_{\text{lm1}}^* = \frac{\text{MDR}}{1 - P_0(\text{MDR})},$$

we can quickly solve d given $a, b, c, \mu_{\text{lm1}}^*$ using Newton's method.

Denote the incomplete beta function by $\beta(u, v; y) = \int_0^y t^{u-1} (1-t)^{v-1} dt$. For TrB random variable X ,

$$\begin{aligned}\mu_{\text{lm1}} &= \mathbb{E}[\min(X, 1)] \\ &= \frac{d \cdot \Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) + 1 - \beta\left(\frac{b}{c}, \frac{a}{c}; \frac{1}{d^c + 1}\right).\end{aligned}$$

Interestingly,

$$\frac{\partial \mu_{\text{lm1}}}{\partial d} = \frac{\Gamma\left(\frac{b+1}{c}\right) \Gamma\left(\frac{a-1}{c}\right)}{\Gamma\left(\frac{a}{c}\right) \Gamma\left(\frac{b}{c}\right)} \cdot \beta\left(\frac{b+1}{c}, \frac{a-1}{c}; \frac{1}{d^c + 1}\right) > 0,$$

so μ_{lm1} is an increasing function of d and the derivative can be obtained almost free of charge after computing μ_{lm1} .

Because the target mean for $\min(X, 1)$ is also constrained by:

$$\mu_{\text{lm1}}^* = \frac{\text{MDR}}{1 - P_0(\text{MDR})},$$

we can quickly solve d given $a, b, c, \mu_{\text{lm1}}^*$ using Newton's method.

The mean constraint removes one degree of freedom. Fitting TrB is only 3-d optimization for a, b, c .

Algorithm 1: Negative log-likelihood $\mathcal{L}(a, b, c, d, x[], p[])$

INPUT: TrB parameters a, b, c, d ; Empirical PMF support $x[N]$ and probabilities $p[N]$; TrB PDF f and CDF F .

OUTPUT: Negative log-likelihood.

1. $l \leftarrow 0$
2. **for** $i = 1$ to $N - 1$:
 $l += -p[i] \log f(a, b, c, d; x[i])$
3. **if** $x[N] < 1$: $l += -p[i] \log f(a, b, c, d; x[i])$
 else: $l += -p[i] \log(1 - F(a, b, c, d; 1))$
4. **return** l

Negative log-likelihood of weighted data (PMF) against a continuous PDF is equivalent to cross entropy.

Algorithm I: Negative log-likelihood $\mathcal{L}(a, b, c, d, x[], p[])$

INPUT: TrB parameters a, b, c, d ; Empirical PMF support $x[N]$ and probabilities $p[N]$; TrB PDF f and CDF F .

OUTPUT: Negative log-likelihood.

1. $l \leftarrow 0$
2. **for** $i = 1$ to $N - 1$:
 $l += -p[i] \log f(a, b, c, d; x[i])$
3. **if** $x[N] < 1$: $l += -p[i] \log f(a, b, c, d; x[i])$
 else: $l += -p[i] \log(1 - F(a, b, c, d; 1))$
4. **return** l

Negative log-likelihood of weighted data (PMF) against a continuous PDF is equivalent to cross entropy.

We unify gradient computation by using finite difference (FD):

1. The analytical form of $\partial \mathcal{L} / \partial(a, b, c)$ is not simple enough to bring meaningful speedup.
2. Other distance measures have costlier analytical gradient forms.
3. FD can be more numerically stable.

Algorithm II: Objective function \mathcal{O} and gradient $\nabla \mathcal{O}$

INPUT: Empirical PMF's support $x[N]$, probabilities $p[N]$, target limited mean μ_{lm1}^* ; TrB parameters a, b, c ; Newton's 1-d root finder \mathcal{R} ; Distance function \mathcal{L} .

OUTPUT: Object function value and gradient with respect to (a, b, c) .

1. $d \leftarrow \mathcal{R}(a, b, c, \mu_{lm1}^*)$
2. **return** $\mathcal{L}(a, b, c, d, x[N], p[N])$, $\left(\frac{\partial \mathcal{L}}{\partial a}, \frac{\partial \mathcal{L}}{\partial b}, \frac{\partial \mathcal{L}}{\partial c} \right)$

Root finder \mathcal{R} invokes bisection contingent on Newton's divergence.

Algorithm I: Negative log-likelihood $\mathcal{L}(a, b, c, d, x[], p[])$

INPUT: TrB parameters a, b, c, d ; Empirical PMF support $x[N]$ and probabilities $p[N]$; TrB PDF f and CDF F .

OUTPUT: Negative log-likelihood.

1. $l \leftarrow 0$
2. **for** $i = 1$ to $N - 1$:
 $l += -p[i] \log f(a, b, c, d; x[i])$
3. **if** $x[N] < 1$: $l += -p[i] \log f(a, b, c, d; x[i])$
 else: $l += -p[i] \log(1 - F(a, b, c, d; 1))$
4. **return** l

Negative log-likelihood of weighted data (PMF) against a continuous PDF is equivalent to cross entropy.

We unify gradient computation by using finite difference (FD):

1. The analytical form of $\partial \mathcal{L} / \partial (a, b, c)$ is not simple enough to bring meaningful speedup.
2. Other distance measures have costlier analytical gradient forms.
3. FD can be more numerically stable.

Algorithm II: Objective function \mathcal{O} and gradient $\nabla \mathcal{O}$

INPUT: Empirical PMF's support $x[N]$, probabilities $p[N]$, target limited mean μ_{lm1}^* ; TrB parameters a, b, c ; Newton's 1-d root finder \mathcal{R} ; Distance function \mathcal{L} .

OUTPUT: Object function value and gradient with respect to (a, b, c) .

1. $d \leftarrow \mathcal{R}(a, b, c, \mu_{lm1}^*)$
2. **return** $\mathcal{L}(a, b, c, d, x[N], p[N])$, $\left(\frac{\partial \mathcal{L}}{\partial a}, \frac{\partial \mathcal{L}}{\partial b}, \frac{\partial \mathcal{L}}{\partial c} \right)$

Root finder \mathcal{R} invokes bisection contingent on Newton's divergence.

Algorithm III: Solve a, b, c, d

INPUT: Empirical PMF's support $x[N]$, probabilities $p[N]$, and target limited mean μ_{lm1}^* ; Quasi-Newton minimizer \mathcal{Q} , e.g., L-BFGS-B; all previously defined functions.

OUTPUT: a, b, c, d

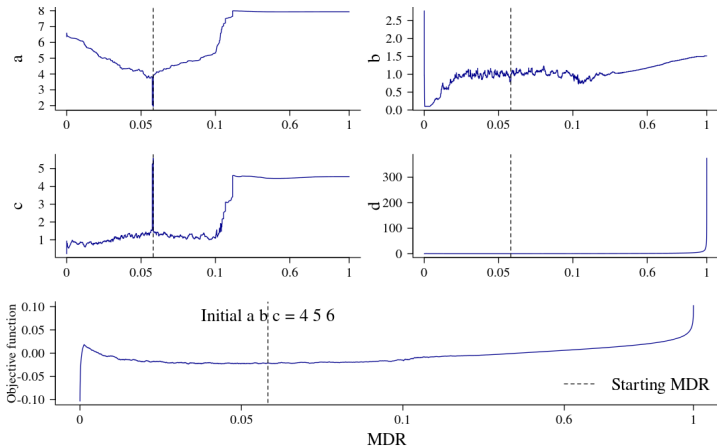
1. $a, b, c \leftarrow \mathcal{Q}(\mathcal{O}, \nabla \mathcal{O})$
2. $d \leftarrow \mathcal{R}(a, b, c, \mu_{lm1}^*)$
3. **return** (a, b, c, d)

Bidirectional sequential fitting

Bidirectional sequential fitting:

1. Select an MDR from data, e.g. $\text{median}(\text{MDR}) = 0.058$.
2. Fit TrB to the empirical PMF associated with the MDR.
3. Go to the next (previous) MDR, use the current optimized parameters as initialization, and rerun Step 2.

The motivation is to promote the locality of optima. The result (a , b , c , d) are still not perfectly smooth functions of MDR, which is expected.



Bidirectional sequential fitting

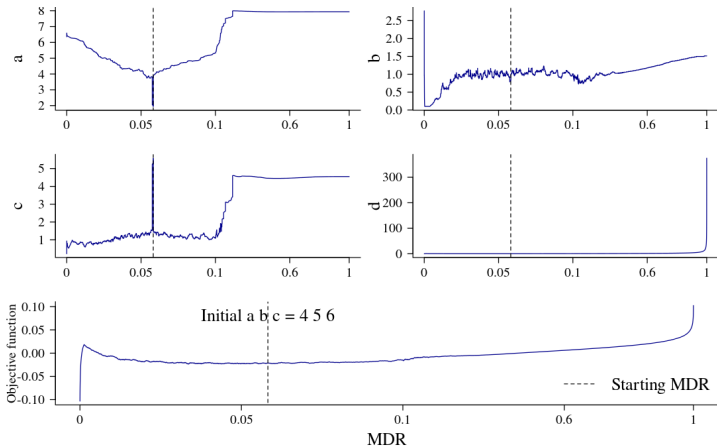
Bidirectional sequential fitting:

1. Select an MDR from data, e.g. $\text{median}(\text{MDR}) = 0.058$.
2. Fit TrB to the empirical PMF associated with the MDR.
3. Go to the next (previous) MDR, use the current optimized parameters as initialization, and rerun Step 2.

The motivation is to promote the locality of optima. The result (a , b , c , d) are still not perfectly smooth functions of MDR, which is expected.

The drop and surge in the middle of a and c correspond to about 300 MDRs surrounding 0.058 — the starting MDR. The behavior is still being investigated, but it is probably because cumulative change in the empirical PMF becomes large enough to push the optimizer towards a far-away local minimum — even if this local minimum is minimally better. The diagnose is reasonable but still superficial since it does not explain why such behavior only occurs near the starting MDR: if we change the starting MDR to, e.g., 0.1, the drop and surge will appear around $\text{MDR}=0.1$.

The drop and surge do not translate to similar behaviors in TrB density or in the objective function.



Bidirectional sequential fitting (update)

It has been confirmed that the drop and surge are due to “bad” initialization at the starting MDR.

The objective function has almost no change at all given quite different parameters.

This implies numerous local optima for the same goodness of fit, and suggests that the distribution model might be overparameterized. Three-parameter members in the TrB family, i.e. Burr, Generalized Pareto, Inverse Burr, could be worth a trial **in the distant future**.

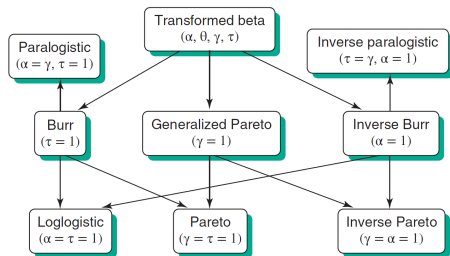


Figure 5.2 The transformed beta family.

Bayesian update

Assuming no deductibles:

1. Import data: (MDRs, claim damage ratios).

- ▶ Also import the old PMF table, and set the update weight w .

2. Model P_0 .

- ▶ $P_0 \leftarrow (1 - w)P_0 + wP_0^{\text{old}}$.

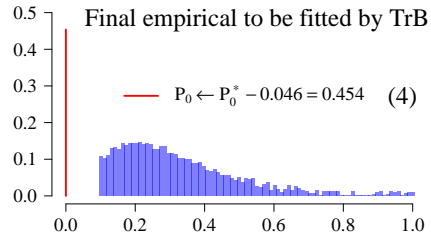
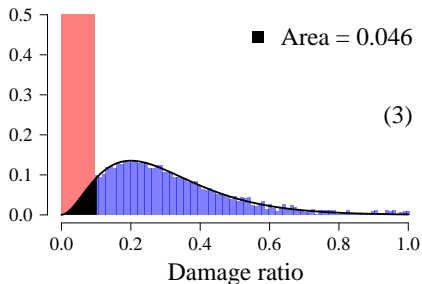
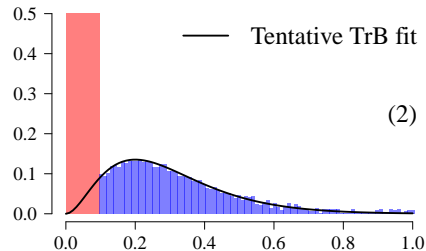
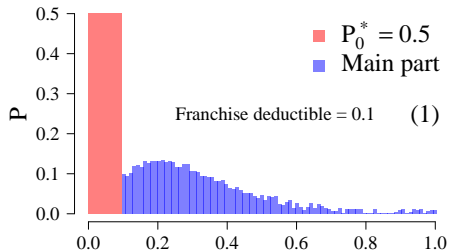
3. Compute 18999 conditional empirical PMFs.

- ▶ $P_{\text{DR}>0} \leftarrow (1 - w)P_{\text{DR}>0} + wP_{\text{DR}>0}^{\text{old}}$.

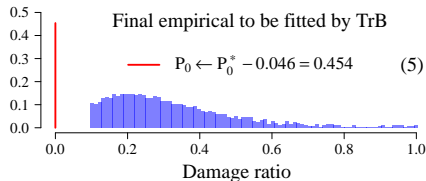
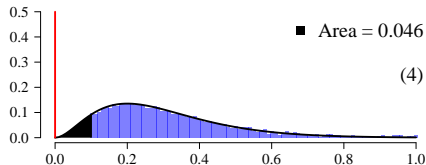
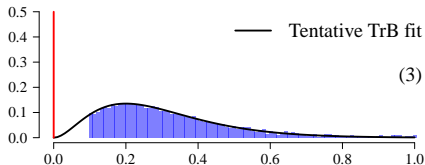
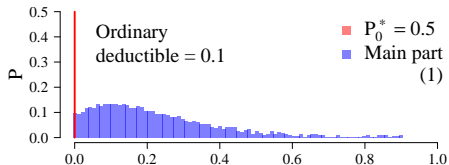
4. Fit Transformed Beta (TrB) to the 18999 PMFs.

5. Discretization.

Given franchise deductible



Given ordinary deductible



Raw discretization

1. Set tentative quantile q e.g. 0.999. Given an MDR, let $\max(\text{MDR})$ be the max of PMF's support, let F_{MDR}^{-1} be the inverse TrB CDF:
 - ▶ $\max(\text{MDR}) \leftarrow \min(1, F_{\text{MDR}}^{-1}(q))$.
 - ▶ If $\max(\text{MDR})$ is not a nondecreasing sequence (rarely happens), compute the longest nondecreasing subsequence and do Hyman interpolation.

Raw discretization

1. Set tentative quantile q e.g. 0.999. Given an MDR, let $\max(\text{MDR})$ be the max of PMF's support, let F_{MDR}^{-1} be the inverse TrB CDF:
 - ▶ $\max(\text{MDR}) \leftarrow \min(1, F_{\text{MDR}}^{-1}(q))$.
 - ▶ If $\max(\text{MDR})$ is not a nondecreasing sequence (rarely happens), compute the longest nondecreasing subsequence and do Hyman interpolation.
2. For each MDR,
 - 2.1 Set a fine support of e.g. 2000 points, i.e. $\{\max/2000, 2 \cdot \max/2000, \dots, \max\}$. Discretize TrB onto the support via central differencing TrB CDF.
 - ▶ If $\max = 1$: $P_{\max} \leftarrow 1 - F(1 - \max/4000)$.
 - 2.2 Renormalize the fine discretization such that the sum of probabilities equals $1 - P_0$. Prepend $(0, P_0)$ to the fine discretization.

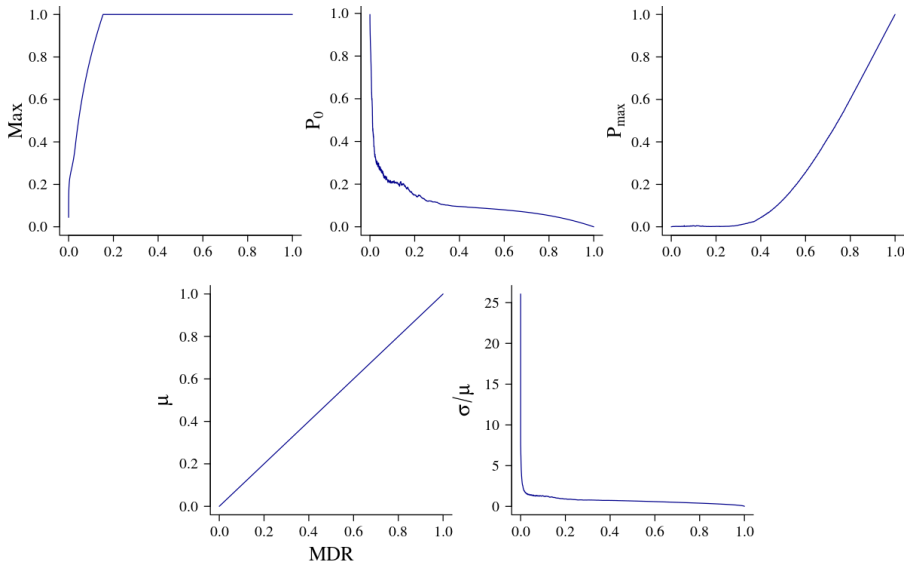
Raw discretization

1. Set tentative quantile q e.g. 0.999. Given an MDR, let $\max(\text{MDR})$ be the max of PMF's support, let F_{MDR}^{-1} be the inverse TrB CDF:
 - ▶ $\max(\text{MDR}) \leftarrow \min(1, F_{\text{MDR}}^{-1}(q))$.
 - ▶ If $\max(\text{MDR})$ is not a nondecreasing sequence (rarely happens), compute the longest nondecreasing subsequence and do Hyman interpolation.
2. For each MDR,
 - 2.1 Set a fine support of e.g. 2000 points, i.e. $\{\max/2000, 2 \cdot \max/2000, \dots, \max\}$. Discretize TrB onto the support via central differencing TrB CDF.
 - ▶ If $\max = 1$: $P_{\max} \leftarrow 1 - F(1 - \max/4000)$.
 - 2.2 Renormalize the fine discretization such that the sum of probabilities equals $1 - P_0$. Prepend $(0, P_0)$ to the fine discretization.
 - 2.3 Regrid the PMF onto the final 42/64-point support.

Raw discretization

1. Set tentative quantile q e.g. 0.999. Given an MDR, let $\max(\text{MDR})$ be the max of PMF's support, let F_{MDR}^{-1} be the inverse TrB CDF:
 - ▶ $\max(\text{MDR}) \leftarrow \min(1, F_{\text{MDR}}^{-1}(q))$.
 - ▶ If $\max(\text{MDR})$ is not a nondecreasing sequence (rarely happens), compute the longest nondecreasing subsequence and do Hyman interpolation.
2. For each MDR,
 - 2.1 Set a fine support of e.g. 2000 points, i.e. $\{\max/2000, 2 \cdot \max/2000, \dots, \max\}$. Discretize TrB onto the support via central differencing TrB CDF.
 - ▶ If $\max = 1$: $P_{\max} \leftarrow 1 - F(1 - \max/4000)$.
 - 2.2 Renormalize the fine discretization such that the sum of probabilities equals $1 - P_0$. Prepend $(0, P_0)$ to the fine discretization.
 - 2.3 Regrid the PMF onto the final 42/64-point support.
 - 2.4 Scale up/down P_0 while scaling down/up all the main probabilities together to eliminate small error between MDR and PMF's mean.

Raw discretization



Tune discretization

Given a PMF table of size $19001 \times (N + 2)$, $N \in \{42, 64\}$:

1. Impose monotonicity on P_0 .

- 1.1 Compute the longest nonincreasing subsequence of P_0 . Retrieve all the PMFs associated to the subsequence.
- 1.2 Let PMF_i and PMF_j be any two neighboring PMFs in the PMF sequence. Denote their MDRs by MDR_i and MDR_j , $i < j$.
- 1.3 If there should have been a prescribed MDR^* between MDR_i and MDR_j :

$$w \leftarrow (\text{MDR}^* - \text{MDR}_i) / (\text{MDR}_j - \text{MDR}_i)$$
$$\text{PMF}^* \leftarrow (1 - w)\text{PMF}_i + w\text{PMF}_j$$

- 1.4 Insert PMF^* to the PMF table. If the table size has not reached 19001, return to Step 1.2.

2. Impose monotonicity on P_{\max} in a similar way, which will not violate P_0 's monotonicity.

3. By now, monotonicity in maxes might have been violated. Re-impose monotonicity in maxes:

- 3.1 Compute the longest nondecreasing subsequence of maxes. Retrieve all the PMFs associated with the subsequence.

3.

- 3.2 Let PMF_i and PMF_j be any two neighboring PMFs in the sequence. Let MDR_i , MDR_j , \max_i , \max_j , \mathbf{x}_i and \mathbf{x}_j , \mathbf{p}_i , \mathbf{p}_j , be their MDRs, maxes, supports and probability vectors respectively. Additionally, let $\mathbf{x}(\max_k)$
- 3.3 If there should have been a prescribed MDR^* between MDR_i and MDR_j , do:

$$\max^* \leftarrow (\max_i + \max_j) / 2$$

$$\mathbf{x}^* \leftarrow \left(0, \frac{\max^*}{N-1}, \frac{2\max^*}{N-1}, \dots, \max^* \right)$$

$$\mathbf{m}_i \leftarrow \mathbf{x}^* \cdot \mathbf{p}_i$$

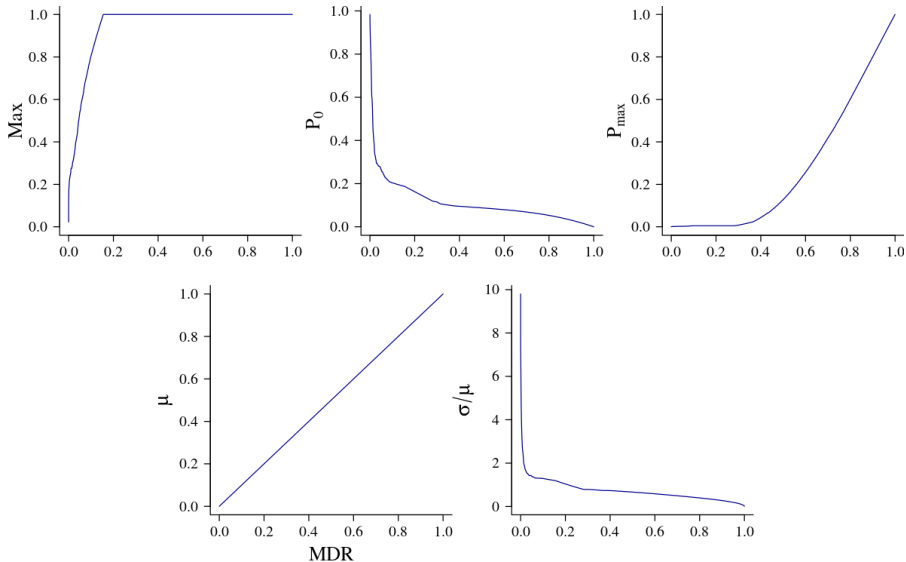
$$\mathbf{m}_j \leftarrow \mathbf{x}^* \cdot \mathbf{p}_j$$

$$w \leftarrow (\text{MDR}^* - \mathbf{m}_i) / (\mathbf{m}_j - \mathbf{m}_i)$$

$$\mathbf{p}^* \leftarrow (1 - w)\mathbf{p}_i + w\mathbf{p}_j$$

- 3.4 Insert \mathbf{p}^* into the PMF table. If the table size has not reached 19001, return to Step 3.2.

Tuned discretization



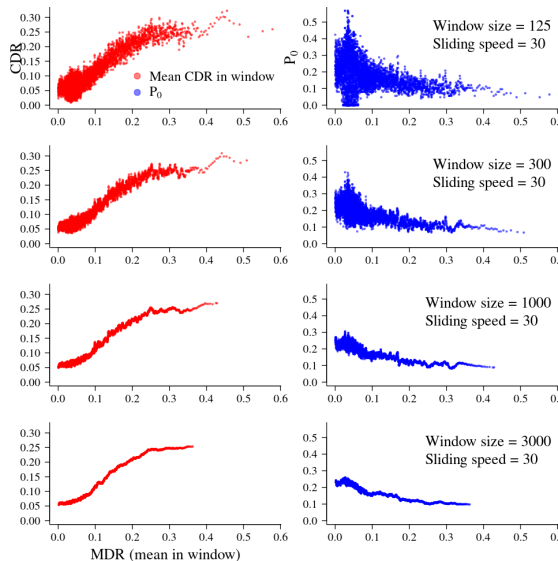
Sliding window sizes

Larger windows \implies smoother trends.

Larger windows \implies higher starting mean MDR, lower ending mean MDR.

Window size too large \implies oversmoothing.

Window size too small \implies less credible empirical PMFs in windows.



Sliding window sizes

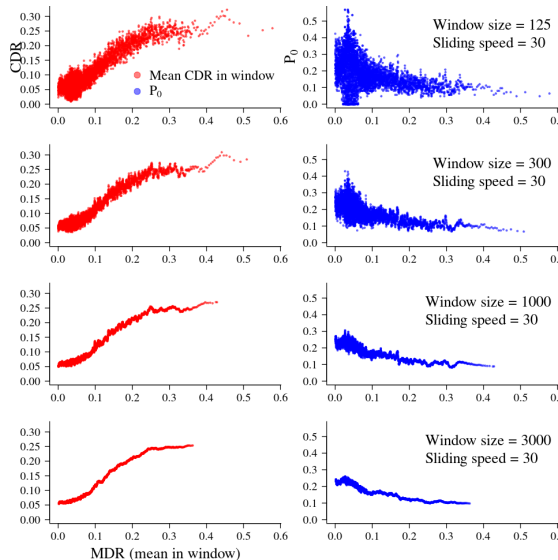
Larger windows \implies smoother trends.

Larger windows \implies higher starting mean MDR, lower ending mean MDR.

Window size too large \implies oversmoothing.

Window size too small \implies less credible empirical PMFs in windows.

- ▶ Empirical PMF has 64 points, thus window size ≥ 200 is recommended, aka ~ 3 points on average for one probability bin.
- ▶ In principle, fitting distributions does not require inferring “empirical PMF” from data first. However,



Sliding window sizes

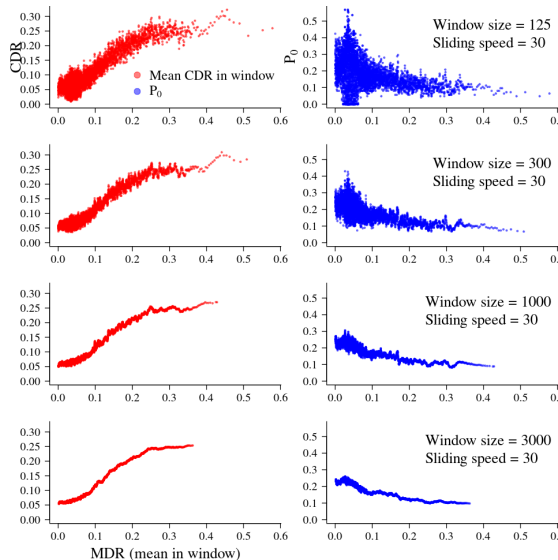
Larger windows \implies smoother trends.

Larger windows \implies higher starting mean MDR, lower ending mean MDR.

Window size too large \implies oversmoothing.

Window size too small \implies less credible empirical PMFs in windows.

- ▶ Empirical PMF has 64 points, thus window size ≥ 200 is recommended, aka ~ 3 points on average for one probability bin.
- ▶ In principle, fitting distributions does not require inferring “empirical PMF” from data first. However,
 - ▶ Doing so is necessary for Bayesian update because old distributions are characterized by PMFs.
 - ▶ Empirical PMFs are needed for visualization of the goodness of fit.
 - ▶ Fitting to the empirical PMFs is direct — what you see is what you fit.



Sliding window sizes

Larger windows \implies smoother trends.

Larger windows \implies higher starting mean MDR, lower ending mean MDR.

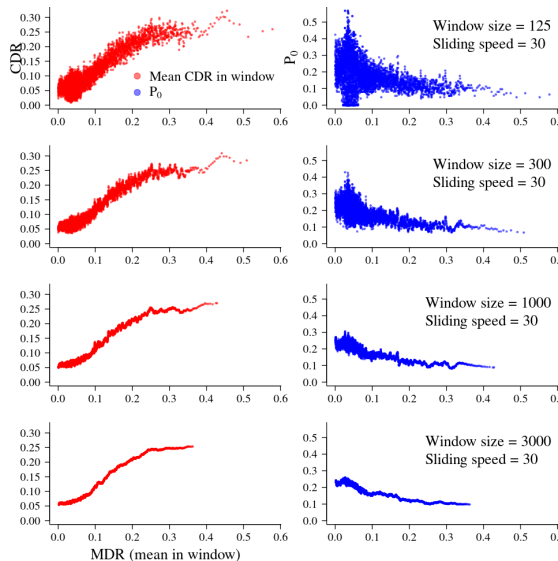
Window size too large \implies oversmoothing.

Window size too small \implies less credible empirical PMFs in windows.

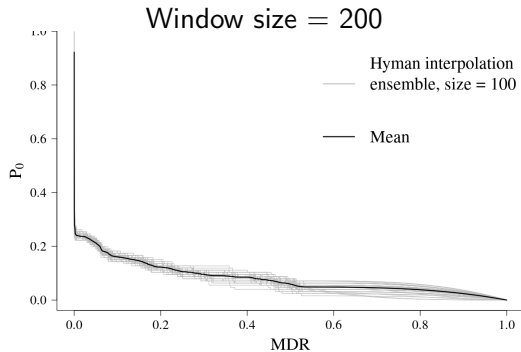
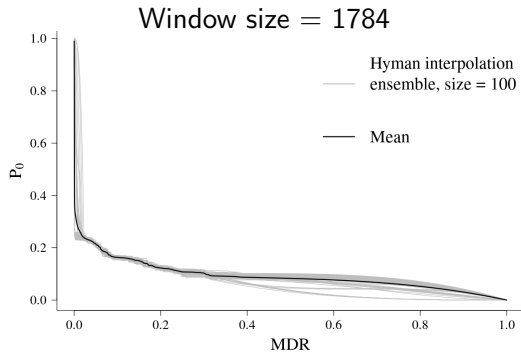
- ▶ Empirical PMF has 64 points, thus window size ≥ 200 is recommended, aka ~ 3 points on average for one probability bin.
- ▶ In principle, fitting distributions does not require inferring “empirical PMF” from data first. However,
 - ▶ Doing so is necessary for Bayesian update because old distributions are characterized by PMFs.
 - ▶ Empirical PMFs are needed for visualization of the goodness of fit.
 - ▶ Fitting to the empirical PMFs is direct — what you see is what you fit.

It is recommended to initialize different window sizes, run through the computing pipeline, compare and contrast the results before making decisions.

- ▶ Scoring the output distributions (explained in next slides) over claims data could be a useful criterion.



Sliding window size 1784 vs. 200, P_0 s

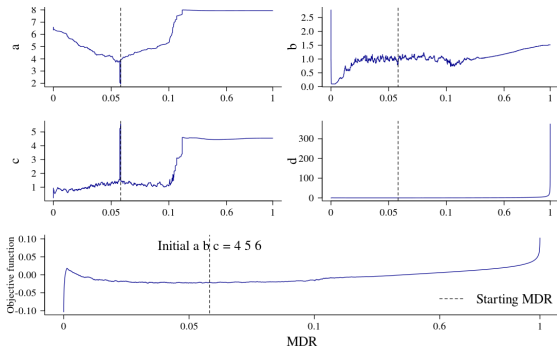


Smaller windows lead to less smooth interpolations.

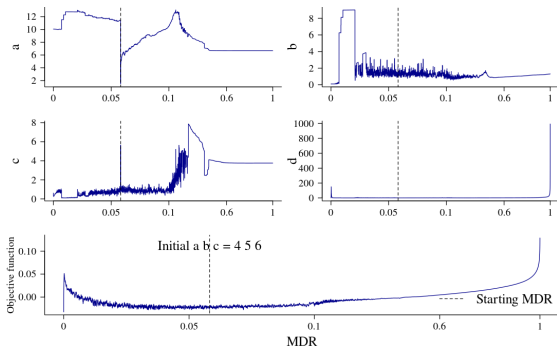
Given smaller windows, ensemble appears more necessary.

Sliding window size 1784 vs. 200, fitted parameters

Window size = 1784



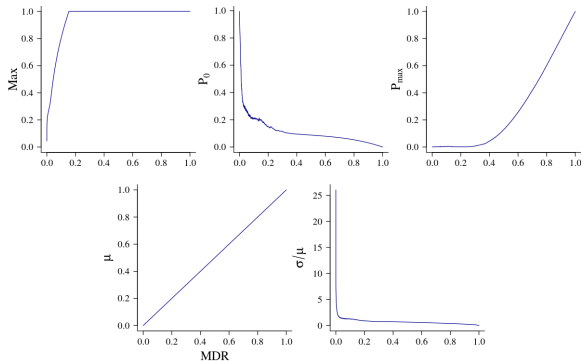
Window size = 200



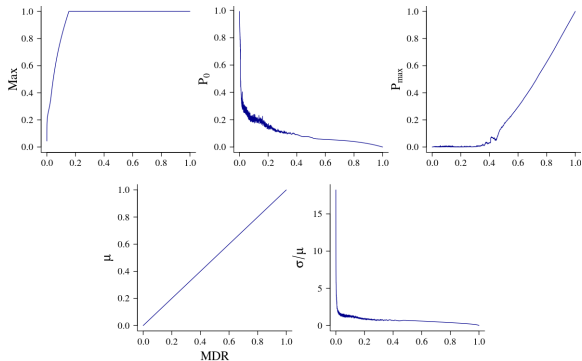
Smaller windows lead to rougher transitions in TrB parameters along the MDR axis.
Different window sizes result in different parameters.
Objective function values however are similar given different window sizes.

Sliding window size 1784 vs. 200, raw discretization

Window size = 1784



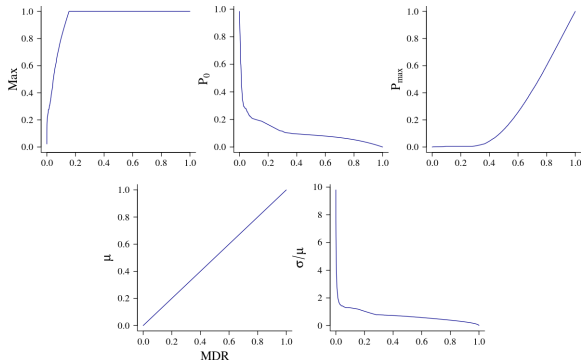
Window size = 200



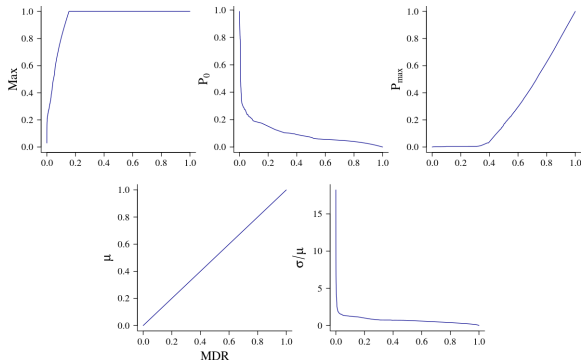
Smaller windows lead to fuzzier P_0 , P_{\max} , μ/σ functions of MDR. More corrections will be taken for imposing monotonicities.

Sliding window size 1784 vs. 200, tuned discretization

Window size = 1784



Window size = 200



P_0 s, P_{\max} s are still close given different window sizes.

μ/σ are close in most of the MDR range, but largely different when MDR is small. PMF with small MDR is dominated by P_0 . Small perturbation in P_0 amplifies the difference in σ and thus μ/σ .

Sliding window size 1784, fitted vs. empirical

Plots are associated with 45

MDRs $\in \{ 1\text{e-}5, 2\text{e-}5, \dots, 1\text{e-}4,$
 $2\text{e-}4, \dots, 1\text{e-}3, 2\text{e-}3, \dots, 1\text{e-}2,$
 $2\text{e-}2, \dots, 1\text{e-}1, 2\text{e-}1, \dots, 9\text{e-}1 \}$

Sliding window size 1784, fitted vs. empirical

Plots are associated with 45
MDRs $\in \{ 1\text{e-}5, 2\text{e-}5, \dots, 1\text{e-}4,$
 $2\text{e-}4, \dots, 1\text{e-}3, 2\text{e-}3, \dots, 1\text{e-}2,$
 $2\text{e-}2, \dots, 1\text{e-}1, 2\text{e-}1, \dots, 9\text{e-}1 \}$

“Bias corrected empirical” refers
to empirical PMF with support
scaled to match mean and MDR.

Bias corrected empirical **is not**
used for training TrBs.

Sliding window size 1784, fitted vs. empirical

Plots are associated with 45
MDRs $\in \{ 1\text{e-}5, 2\text{e-}5, \dots, 1\text{e-}4,$
 $2\text{e-}4, \dots, 1\text{e-}3, 2\text{e-}3, \dots, 1\text{e-}2,$
 $2\text{e-}2, \dots, 1\text{e-}1, 2\text{e-}1, \dots, 9\text{e-}1 \}$

“Bias corrected empirical” refers
to empirical PMF with support
scaled to match mean and MDR.

Bias corrected empirical **is not**
used for training TrBs.

Some empirical's mean differs
from MDR by orders of
magnitude.

Sliding window size 1784, fitted vs. empirical

Plots are associated with 45
MDRs $\in \{ 1\text{e-}5, 2\text{e-}5, \dots, 1\text{e-}4,$
 $2\text{e-}4, \dots, 1\text{e-}3, 2\text{e-}3, \dots, 1\text{e-}2,$
 $2\text{e-}2, \dots, 1\text{e-}1, 2\text{e-}1, \dots, 9\text{e-}1 \}$

“Bias corrected empirical” refers
to empirical PMF with support
scaled to match mean and MDR.

Bias corrected empirical **is not**
used for training TrBs.

Some empirical's mean differs
from MDR by orders of
magnitude.

The final discretized TrB often
well fits the bias corrected
empirical. This suggests d is the
dominant parameter for fitting
empirical PMF with mean highly
different from MDR.

Sliding window size 200, fitted vs. empirical

Plots are associated with 45
MDRs $\in \{ 1e-5, 2e-5, \dots, 1e-4,$
 $2e-4, \dots, 1e-3, 2e-3, \dots, 1e-2,$
 $2e-2, \dots, 1e-1, 2e-1, \dots, 9e-1 \}$

“Bias corrected empirical” refers
to empirical PMF with support
scaled to match mean and MDR.

Bias corrected empirical **is not**
used for training TrBs.

Comparing to window size 1784,

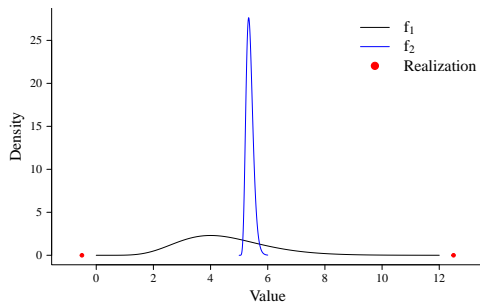
- ▶ Shape of empirical PMF is
much jaggier due to
smaller sample size. Bias
between mean and MDR
are often larger.
- ▶ The final discretized TrB
less often well fits the bias
corrected empirical.

Brier score

The Brier score is appropriate for binary and categorical outcomes that can be structured as true or false, but it is inappropriate for ordinal variables which can take on three or more values.

Have not deciphered why Wikipedia made the above comment.

Without extrapolation, there could be bias towards realizations out of support:

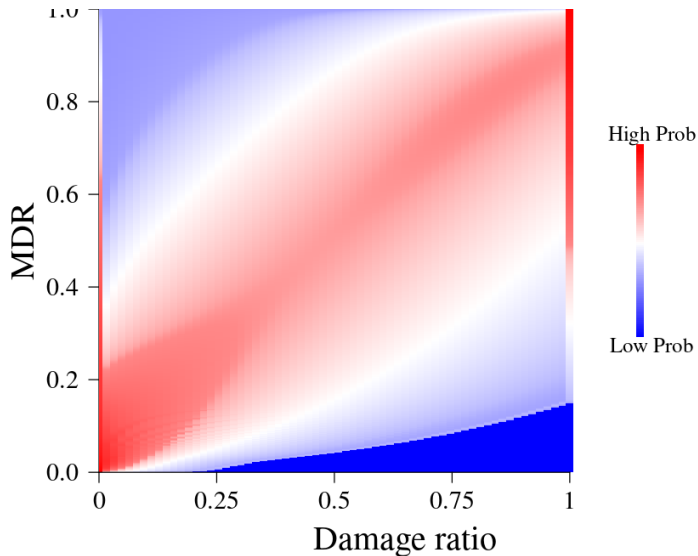


f_1 and f_2 have the same score 0. Fair to say f_2 and f_1 are equally good forecasts?

f_1 has (much) higher chance of being better.

Negative log-likelihood and ignorance score

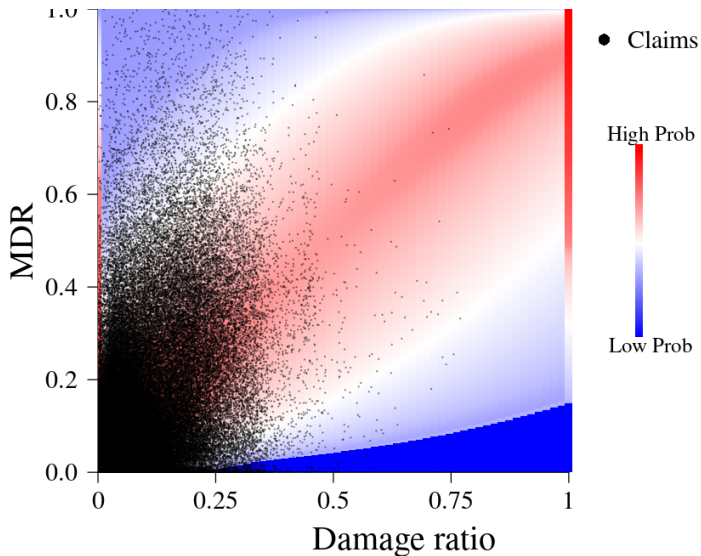
Using claims data, scoring a PMF table \equiv
Scoring a 2D joint PMF.



Negative log-likelihood and ignorance score

Using claims data, scoring a PMF table \equiv
Scoring a 2D joint PMF.

Claims' MDRs are round to the nearest $1e-5$
if < 0.1 , and to $1e-4$ otherwise.



Negative log-likelihood and ignorance score

Using claims data, scoring a PMF table \equiv
Scoring a 2D joint PMF.

Claims' MDRs are round to the nearest $1e-5$
if < 0.1 , and to $1e-4$ otherwise.

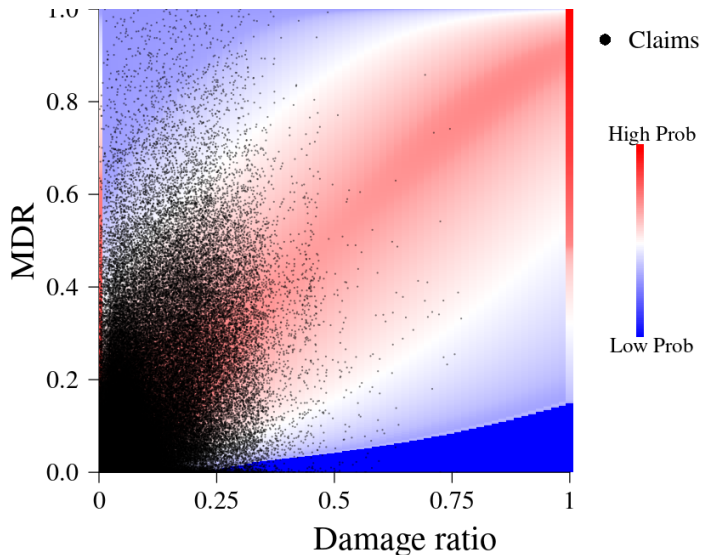
Negative log-likelihood

$$\mathcal{L} = -\sum_{i=1}^N \ln f(\text{claim}_i) + N \ln 18999.$$

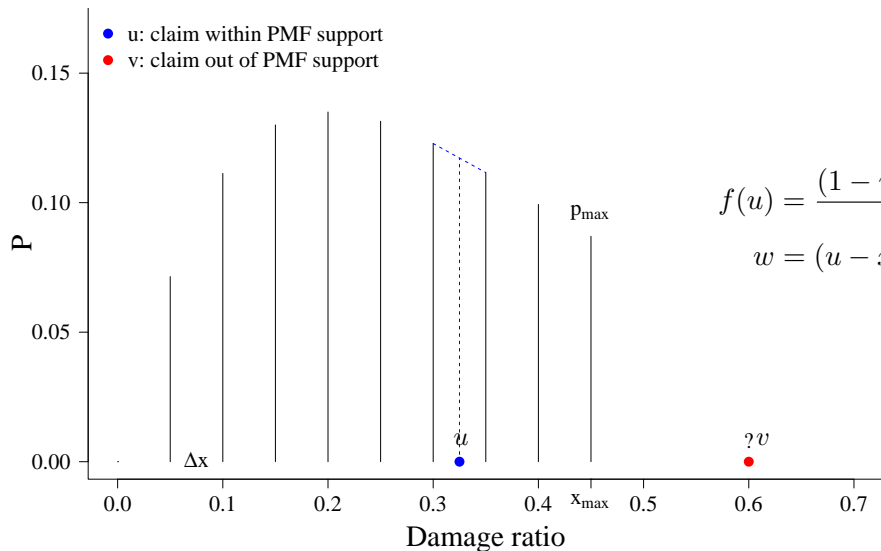
Ignorance score $\mathcal{I}g =$

$$-\sum_{i=1}^N \log_2 f(\text{claim}_i) + N \log_2 18999.$$

$N \ln 18999$ and $N \log_2 18999$ account for
probability normalization.

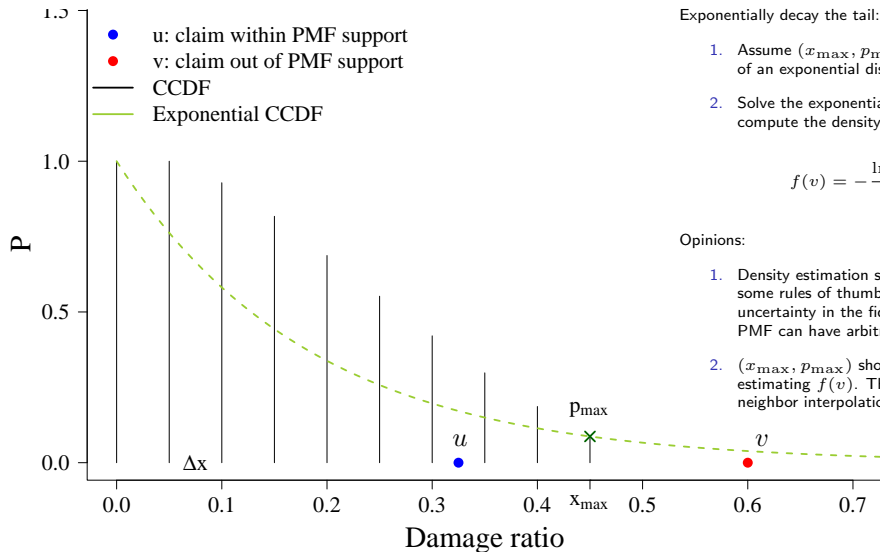


Density estimate of one realization given arbitrary PMF



$$f(u) = \frac{(1 - w) \cdot p(x_i) + w \cdot p(x_{i+1})}{\Delta x}$$
$$w = (u - x_i) / (x_{i+1} - x_i)$$

Density estimate of one realization given arbitrary PMF



Exponentially decay the tail:

1. Assume (x_{\max}, p_{\max}) is a point on the survival curve of an exponential distribution.
2. Solve the exponential distribution's parameter and compute the density at v :

$$f(v) = -\frac{\ln p_{\max}}{x_{\max}} \exp \left\{ \frac{\ln p_{\max}}{x_{\max}} \cdot v \right\}$$

Opinions:

1. Density estimation should be simple and only based on some rules of thumb. More sophistication means more uncertainty in the fidelity of the estimate considering the PMF can have arbitrary shape.
2. (x_{\max}, p_{\max}) should be the dominant parameter for estimating $f(v)$. This follows the principle of nearest neighbor interpolation used for computing $f(u)$.