Double Generalized Linear Model (DGLM) : Tweedie distribution

Raïssa Coulibaly

June 29, 2022

Exponential Dispersion Family (EDF)

Generalized Linear Model (GLM))

Tweedie distribution

DGLM

Application in pricing

References

Definition

Consider a random variable Y valued in a subset of the real line. Y belongs to the EDF if its density has the following form :

$$f(y;\theta,w/\phi)=\exp\left\{\frac{y\theta-\kappa(\theta)}{\phi/w}+a(y;w/\phi)\right\} \tag{1}$$

- θ : the Canonical parameter;
- $\kappa(\theta)$: the cumulant function (convex in θ);
- ϕ: the dispersion parameter;
- w: the weight;
- $a(y;\frac{w}{\phi})$: the constant and doesn't depend on θ.

Example

$$Y \sim \mathrm{Poi}(\lambda) \tag{2}$$

$$f(y;\theta, w/\phi) = \exp\{-\lambda\}\frac{\lambda^y}{y!}\mathbb{1}\{y = 0, 1, ...\} \tag{3}$$

$$= \exp\{y\log(\lambda) - \lambda - \log(y!)\mathbb{1}\{y = 0, 1, ...\}\} \tag{4}$$

We can also verify that the Gamma, $\mathscr{G}\mathrm{amma}(\gamma, c)$), and Compound Poisson-Gamma, $\mathrm{CPG}(\lambda, \gamma, c)$, distributions belong too in EDF.

Mean

Let $\mu = \mathrm{E}[Y]$, we note that : $\theta = (\kappa^{'})^{-1}(\mu)$

Variance function

The variance function and the variance of Y are defined by:

$$V(\mu) = \kappa^{''}\left((\kappa^{'})^{-1}(\mu)\right) \tag{5}$$

$$\mathrm{Var}[Y] = \frac{\phi}{w}V(\mu) \tag{6}$$

Example

Poisson : $V(\mu) = \mu$; Gamma : $V(\mu) = \mu^2$.

Moment generating function (mgf)

$$M_Y(r) = \mathrm{E}[e^{rY}], r > 0 \tag{7}$$

$$= \exp\left\{\frac{\kappa(\theta + r\phi/w) - \kappa(\theta)}{\phi/w}\right\}, r > 0 \tag{8}$$

Example

- $N \sim \mathrm{Poi}(\lambda)$

$$M_N(r) = \exp\{\lambda(\exp\{r\} - 1)\}, r > 0 \tag{9}$$

- $Z \sim \mathscr{G}\mathrm{amma}(\gamma, c)$

$$M_Z(r) = \left(\frac{c}{c-r}\right)^{\gamma}, r < c \tag{10}$$

- $S \sim \mathrm{CPG}(\lambda, \gamma, c)$

$$\begin{aligned} M_S(r) &= M_N(\log(M_Z(r))) && (11) \\ &= \exp\left\{\lambda\left[\left(\frac{c}{c-r}\right)^{\gamma} - 1\right]\right\}, r < c && (12) \end{aligned}$$

Assumptions
Given a response Y, the GLM assumes that Y belongs in EDF and a transformation of its mean is linearly related to some covariates x^t:

$$g(\mu) = x^t\beta \tag{13}$$

Inference
Assume we have n independent pairs of response and covariates $(Y_i, x_i)_{i=1,\dots,n}$ such as :

$$f(y_i; \theta_i, w_i/\phi) \propto \exp\left\{\frac{y_i\theta_i - \kappa(\theta_i)}{\phi/w_i}\right\} \tag{14}$$

$$g(\mu_i) = x_i^t\beta, x_i^t = (1, x_{i1}, .., x_{id}), \beta \in \mathbb{R}^{d+1} \tag{15}$$

Inference

The maximum likelihood estimation is adopted. The objective function and its gradient are also defined by :

$$\ell(\beta|y_1, ..., y_n) = \sum_{i=1}^{n} \left\{ \frac{y_i\theta_i - \kappa(\theta_i)}{\phi/w_i} + a(y_i; w_i/\phi) \right\} \tag{16}$$

$$\nabla_\beta \ell = X^t W R \tag{17}$$

$$X^t = [x_1, ..., x_n], \tag{18}$$

$$W = \text{diag}\left(\frac{\left[g^{'}(\mu_i)\right]^{-2}}{\text{Var}[Y_i]} \right)_{i=1,...,n} \tag{19}$$

$$R = \left(g^{'}(\mu_i)(y_i - \mu_i) \right)_{i=1,...,n} \tag{20}$$

Inference

To find the maximum likelihood estimator (MLE) of β, the Newton Raphson algorithm is adopted. This approch is also implemented in the *glm* function of software R. The values t and $i.max$ depend on users.

Choose an initial value β^0 ;
$iter = 0$;
while $e \geq t$ *and* $iter < i.max$ **do**
  $\beta^{iter+1} = \beta^{iter} + (X^tWX)^{-1}X^tWR$;
  compute $e = |\ell(\beta^{iter+1}|y_1, ...y_n) - \ell(\beta^{iter}|y_1, ...y_n)|$
end

Algorithm 1: Newton Rahpson

Definition

The Tweedie distribution is a subclass of EDF whose the variance function is defined by :

$$V(\mu) = \mu^p, p \geq 0 \tag{21}$$

p is called the power variance parameter.

Example

- Gaussian, $p = 0$;
- Poisson, $p = 1$;
- Gamma, $p = 2$;
- Compound Poisson-Gamma, $1 < p < 2$.

Remark
if $p \in]0,1[$, the Tweedie distribution doesn't belong in EDF.

Proof :
Assume that $p \neq 0, 1, 2$ and the Tweedie density function satisfies the equation (1). Then, we obtain the following equations :

$$\log\{f(y;\theta,w/\phi)\} = \int \frac{\partial}{\partial\mu}\log\{f(y;\theta,w/\phi)\}d\mu + constant \quad (22)$$

$$\propto \int \frac{\partial}{\partial\theta}\log\{f(y;\theta,w/\phi)\}\frac{\partial}{\partial\mu}\theta d\mu \quad (23)$$

$$\log\{f(y;\theta,w/\phi)\} \propto \int \frac{w}{\phi}\left(y-\kappa^{(1)}(\theta)\right)\frac{1}{\kappa'\left((\kappa')^{-1}(\mu)\right)}d\mu \quad (24)$$

$$f(y;\theta,w/\phi) \propto \exp\left\{\int \frac{w(y-\mu)}{\phi\mu^{p}}d\mu\right\} \quad (25)$$

$$\propto \exp\left\{\frac{w}{\phi}\left(\frac{\mu^{1-p}}{1-p}y-\frac{\mu^{2-p}}{2-p}\right)\right\} \quad (26)$$

The constant $a(y;w/\phi)$ can be obtained by :

$$a(y;w/\phi) = -\log\left\{\int \exp\left\{\frac{w}{\phi}\left(\frac{\mu^{1-p}}{1-p}y-\frac{\mu^{2-p}}{2-p}\right)\right\}d\theta\right\} \quad (27)$$

The canonical parameter and cumulant function are also defined by :

$$\theta = \frac{\mu^{1-p}}{1-p} \Longleftrightarrow \mu = \left[(1-p)\theta\right]^{\frac{1}{1-p}} \tag{28}$$

$$\kappa(\theta) = \frac{1}{2-p}\left[(1-p)\theta\right]^{\frac{2-p}{1-p}} \tag{29}$$

We note that a necessary and sufficient condition for the convexity of $\kappa(.)$ is given by : $\theta < 0, p > 1$.
Afterwards, we assume that $1 < p < 2$ and note that a response Y belongs in Tweedie distribution by : $Y \sim \text{Tweedie}(\theta, w, \phi, p)$

Moment generating function
By using the equation (8), the mgf of Y is obtained as follows:

$$M_Y(r) = \exp\left\{\frac{w}{\phi}\kappa(\theta)\left[\left(\frac{\theta}{\theta + \frac{\phi r}{w}}\right)^{\frac{2-p}{p-1}} - 1\right]\right\}, r > 0 \tag{30}$$

$$= \exp\left\{\frac{w}{\phi}\kappa(\theta)\left[\left(\frac{\frac{w\theta}{\phi}}{\frac{w\theta}{\phi} + r}\right)^{\frac{2-p}{p-1}} - 1\right]\right\}, -r < \frac{w\theta}{\phi} \tag{31}$$

We recognize the mgf of Compound Poisson-Gamma (12) if $-r < \frac{w\theta}{\phi}$.

We can also rewrite the Tweedie distribution as follows:

$$Y \overset{(d)}{=} \mathrm{CPG}\left(\frac{w}{\phi}\kappa(\theta), \frac{2-p}{p-1}, -\frac{w\theta}{\phi}\right) \tag{32}$$

Assume that we have to model a response $S \sim \mathrm{CPG}(w\lambda, \gamma, c)$, we can model $Y \sim \mathrm{Tweedie}(\theta, w, \phi, p)$ and use the followings identifications parameters:

$$\gamma = \frac{2-p}{p-1} \iff p = \frac{\gamma+2}{\gamma+1} \tag{33}$$

$$c = -\frac{w\theta}{\phi} \iff \phi = -\frac{w}{c}(\kappa^{(1)})^{-1}(\mu) = -\frac{w}{c(1-p)}\mu^{1-p} \tag{34}$$

$$\lambda = \frac{\kappa(\theta)}{\phi} = -\frac{c(1-p)}{w\mu^{1-p}}\frac{1}{2-p}\mu^{2-p} = \frac{c}{w\gamma}\mu \tag{35}$$

Motivations

We know that the MLE of dispersion parameter is biased in general. To be sure, consider a simple Gaussian response : $\mathcal{N}(\mu, \sigma^2)$, μ is known and $\sigma^2 > 0$. Assume n i. i. d. random variables from $\mathcal{N}(\mu, \sigma^2)$. It is easy to see that the MLE of σ^2 is defined by:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \mu)^2 \tag{36}$$

We also note that the bias of $\hat{\sigma}^2$ is not nul :

$$\mathrm{E}[\hat{\sigma}^2] - \sigma^2 = \frac{\sigma^2}{n-1} \sum_{i=1}^{n} \mathrm{E}\left[\frac{Y_i - \mu}{\sigma^2}\right] - \sigma^2 \tag{37}$$

$$= \frac{\sigma^2}{n-1} > 0 \tag{38}$$

The idea of DGLM is to model simultaneous the mean and dispersion parameters. It allows to reduce the bias of the dispersion parameter by considering a dispersion response in the modelling steps. Here, we present only the DGLM in Tweedie distribution case.

Inference
Assume we have n independent pairs of response $(Y_i, N_i)_{i=1,...,n}$ and d covariates $x_i^t = (1, x_{i1}, ..., x_{id})$.

$$Y_i \sim \text{Tweedie}(\theta_i, w_i, \phi_i, p), N_i \sim \text{Poi}\left(\frac{w_i}{\phi_i}\kappa(\theta_i)\right) \tag{39}$$

$$\log(\mu_i) = x_i^t\beta, \beta = (\beta_0, ..., \beta_d)^t \in \mathbb{R}^{d+1} \tag{40}$$

$$\log(\phi_i) = x_i^t\alpha, \alpha = (\alpha_0, ..., \alpha_d)^t \in \mathbb{R}^{d+1} \tag{41}$$

Maximum likelihood estimation
By using (32), we note that :

$$P(Y_i = 0, N_i = 0) = P(N_i = 0) = \exp\left\{-\frac{w_i}{\phi_i}\kappa(\theta_i)\right\} \tag{42}$$

$$Y_i|N_i \sim \mathscr{G}\text{amma}\left(N_i\gamma, -\frac{\theta_i}{\phi_i}w_i\right), \gamma = \frac{2-p}{p-1} \tag{43}$$

$$f(n_i, y_i) = f(y_i|n_i)f(n_i) \tag{44}$$

$$= \frac{\left(\left(\frac{w_i}{\phi_i}\right)^{\gamma+1} y_i^{\gamma}\right)^{n_i}}{n_i!\Gamma(n_i\gamma)y_i\,(p-1)^{n_i\gamma}\,(2-p)^{n_i}} \exp\left\{\frac{w_i}{\phi_i}\,(y_i\theta_i - \kappa(\theta_i))\right\} \tag{45}$$

The log-likelihood function is defined by :

$$\ell(\beta,\alpha,p)=\sum_{i=1}^{n}\ell_i(\beta,\alpha,p) \tag{46}$$

$$\ell_i(\beta,\alpha,p)=\begin{cases}\log(f(n_i,y_i)) & \text{if } n_i\neq 0,\\ -\frac{w_i}{\phi_i}\kappa(\theta_i) & \text{if } n_i=0.\end{cases}$$

Definition of dispersion response

$$D_i = \frac{2}{\nu_i}\left(-w_i\left(Y_i\frac{\mu_i^{1-p}}{1-p} - \frac{\mu_i^{2-p}}{2-p}\right) - \phi_i\frac{N_i}{p-1}\right) + \phi_i \tag{47}$$

$$\nu_i = \frac{2w_i}{\phi_i}\frac{\mu_i^{2-p}}{(p-1)(2-p)} \tag{48}$$

We note that :

$$\mathrm{E}[D_i] = \phi_i \tag{49}$$

$$\mathrm{Var}[D_i] = \frac{2}{\nu_i}\phi_i^2 \tag{50}$$

Choose an initial value (β^0, α^0) ;
Calculate $\nu_i^0 = \nu_i(\beta^0, \alpha^0, p)$;
$iter = 0$;
while $e \geq t$ *and* $iter < i.max$ **do**
    $(\beta_1, \alpha_1)^{iter+1} = \text{argmax}_{(\beta,\alpha)}\, \ell(\beta, \alpha, p)$;
    compute $\nu_i^{iter+1} = \nu_i((\beta_1, \alpha_1)^{iter+1}, p)$;
    compute $D_i = D_i(\nu_i^{iter+1}, (\beta_1, \alpha_1)^{iter+1})$;
    obtain MLE α_2 of α by assuming $D_i \sim \mathscr{G}\text{amma}\left(\frac{\nu_i^0}{2}, \frac{2\phi(\alpha)}{\nu_i^0}\right)$;
    compute $= |\ell((\beta_1, \alpha_1)^{iter+1}, p) - \ell((\beta_1, \alpha_2)^{iter+1}, p)|$
end

Algorithm 2: DGLM

- In general, the MLE of (β, α, p) is obtained by profile maximisation of the likelihood function. That means that we have to run the DGLM algorithm for every $1 < p < 2$ to find the MLE.
- But, if (40) and (41) are satisfied, it is shown in [2] that the MLE of β doesn't depend on p and we can also obtain the dispersion parameter for every $1 < q < 2$ as follows:

$$\hat{\phi}_i(q) = \frac{2-p}{2-q}\hat{\phi}_i(p)\hat{\mu}_i^{p-q} \tag{51}$$

- The MLE of p is obtained by :

$$\hat{p} = \operatorname*{argmax}_q \ell(\hat{\beta}, \hat{\alpha}, \hat{\phi}_i(q), q) \tag{52}$$

Consider n independent responses variables $Y_1, ..., Y_n$ such as :

$$Y_i \sim \mathrm{CPG}(w_i \lambda_i, \gamma, c_i), i = 1, .., n \tag{53}$$

Y_i represents the total cost for the claims. This is also equivalent to consider Y_i as follows:

$$Y_i = \sum_{j=1}^{N_i} Z_{ij} \tag{54}$$

$$N_i \sim \mathrm{Poi}(w_i \lambda_i), Z_{ij} \sim \mathscr{G}\mathrm{amma}(\gamma, c_i) \tag{55}$$

N_i and Z_{ij} represent respectively the number of claims and the claim sizes.

We also assume that :

$$\log(\lambda_i) = x_i^{*t}\beta^{*t}, \log\left(\frac{\gamma}{c_i}\right) = z_i^{*t}\alpha^{*t} \tag{56}$$

To obtain the MLE of β^{*t} and α^{*t} by the CPG approach, we have to model two independents GLM. The first consists to model the number of claims by GLM-Poisson. The second models the clain sizes by GLM-Gamma.
For the comparison purpose between the DGLM and CPG approach, [2] proposes the following equations :

$$x^t\beta = x^{*t}\beta^{*t} + z_i^{*t}\alpha^{*t} \tag{57}$$

$$x^t\alpha = -\log(2-p) - (p-1)x^{*t}\beta^{*t} + (2-p)z_i^{*t}\alpha^{*t} \tag{58}$$

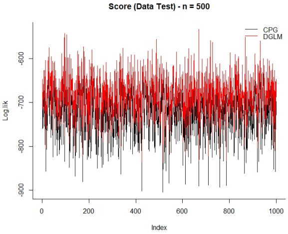
The idea is to evaluate the bias, the variance and the RMSE (Root of Mean Squared of Error) of the estimators obtained by the DGLM and CPG approach. For that, we use 1000 replications of Monte-Carlo method. In each replication, we split the data in training and test data. The training data is used to have the MLE of the models parameters. The test data is used to evaluate the logarithmic score (The log-likelihood evaluated in test data). The data are obtained by using the equations (54), (55) and (56).
We consider two explanatory variables from our project data base, "Class" and "VehAge".

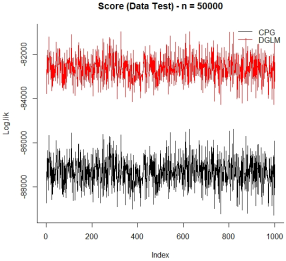
Var. Name	Parameter	Tweedie-CPG			Tweedie-DGLM		
		Bias	Variance	RMSE	Bias	Variance	RMSE
Intercept	β_0^*	-1.0676	0.0074	1.0711	-0.0893	0.0095	0.1321
ClassOther	β_1^*	-0.1187	0.0821	0.3102	-0.1344	0.0913	0.3306
VehAge< = 5	β_2^*	0.0017	0.0127	0.1126	-0.0117	0.0212	0.1462
Intercept	α_0^*	0.4462	0.0165	0.4643	-0.0253	0.0225	0.1520
ClassOther	α_1^*	0.0351	0.0205	0.1475	-0.0613	0.0234	0.1647
VehAge< = 5	α_2^*	0.0003	0.0021	0.0455	-0.0084	0.0052	0.0725
Variance power	p	-0.0031	0.0004	0.0209	-0.0026	0.0004	0.0208

TABLE 1 – n = 500

Var. Name	Parameter	Tweedie-CPG			Tweedie-DGLM		
		Bias	Variance	RMSE	Bias	Variance	RMSE
Intercept	β_0^*	-1.0763	0.00008	1.0763	-0.0971	0.00010	0.0976
ClassOther	β_1^*	-0.0347	0.00098	0.0467	-0.0419	0.00104	0.0529
VehAge< = 5	β_2^*	-0.0052	0.00012	0.0122	-0.0080	0.00022	0.0167
Intercept	α_0^*	0.4312	0.00020	0.4314	-0.0574	0.00028	0.0598
ClassOther	α_1^*	0.0135	0.00022	0.0201	-0.0246	0.00025	0.0293
VehAge< = 5	α_2^*	0.0021	0.00002	0.0050	-0.0043	0.00005	0.0085
Variance power	p	-0.0001	0.00001	0.0023	0.0005	0.00001	0.0024

TABLE 2 – n = 50000


Score (Data Test) - n = 500
CPG
DGLM
Log.lik
-600
-700
-800
-900
0
200
400
600
800
1000
Index


Score (Data Test) - n = 50000
CPG
DGLM
Log.lik
-82000
-84000
-86000
-88000
0
200
400
600
800
1000
Index

We consider a sample of our data base and the same covariates as our simulation. We also split this data in traning and test data. The training is used to calibrate the mean and dispersion parameters in the two approaches. The data test is used to compare these approaches by calculating a logarithmic score.

Target[1]	Freqency (%)	Exposition(%)	Average amount
0	94.60	94.23	-
1	5.16	5.49	9287.264
2	0.23	0.27	5921.471
3	0.01	0.01	5783.552

Table: Distribution of claim count and average amounts

[1] Number of claims

Variable name	Parameter	Estimation
Intercept	β_0^*	-3.976
ClassOther	β_1^*	0.375
VehAge$<=5$	β_2^*	0.587
Log-likelihood		-56090.08
Intercept	α_0^*	9.041
ClassOther	α_1^*	0.520
Shape parameter	γ	0.532
Log-likelihood		-138568.8

Table: Parameters estimations of GLM Poisson and Gamma

Variable name	Parameter	CPG	DGLM
Intercept	β_0	5.065	5.777
ClassOther	β_1	0.895	0.748
VehAge$<=5$	β_2	0.587	0.648
Intercept	α_0	6.793	5.961
ClassOther	α_1	-0.064	-0.143
VehAge$<=5$	α_2	-0.383	-0.363
Var. power parameter	p	1.653	1.651
Log-likelihood		-202031.4	-195453.8
Score (Data test)		-64912.38	-62845.96

Table: Parameters estimations of DGLM and CPG

- Boucher, J. and D. Davidov (2012), *On the importance of dispersion modeling for claims reserving : An application with the tweedie distribution*, CAS, 158 - 172
- Delong, L. and al. (2021), *Making tweedie's compound poisson model more accessible*, European Actuarial Journal, vol. 11, 185 - 226, https://doi.org/10.1007/s13385-021-00264-3
- Jørgensen, B. (1987), *Exponential dispersion models*, J. R. Stat. Soc. Ser. B, vol. 49, no 2, 127 - 145
- Smyth, G. K. and B. Jørgensen. (2002), *Fitting tweedie's compound poisson model to insurance claims data : Dispersion modelling*, ASTIN Bulletin, vol. 32, no 1, 143 - 157, doi : 10.2143/AST.32.1.1020