Official Crib Sheet de MSCTM-DATL and MUM-DATS:

Prof. Lídia Montero -DEIO Page. 1 22/12/2022

Independent random components $\mathbf{Y}^T = (Y_1, \ldots, Y_n)$ from exponential family: $\mathrm{E}[\mathbf{Y}] = \boldsymbol{\mu}$.
Systematic Component $\eta = \mathrm{X}\beta$ where η is nx1, $\mathbf{X}$ *nxp* an β *px1*.
Link function $\eta_i = \mathrm{g}(\mu_i)$.

One parameter $\boldsymbol{\theta}$ Exponential Family (canonical parameter) $\boxed{f_Y(y,\boldsymbol{\theta},\boldsymbol{\phi}) = \exp\left(\frac{y\boldsymbol{\theta} - b(\boldsymbol{\theta})}{a(\boldsymbol{\phi})} + c(y,\boldsymbol{\phi})\right)}$ where

$a(\boldsymbol{\phi})$: dispersion function - $\boldsymbol{\phi}$: dispersión parameter - b(.) : Cumulant function

	$\mathrm{N}(\mu,\sigma^2)$	$\boldsymbol{B}(m,\mu)$/m	$\mathrm{P}(\mu)$	$G(\mu,\nu)$
Canònic Link	1	Logit	Log	Inverse
$E[Y] = \mu$	θ	$e^{\theta}/(1+e^{\theta})$	$\exp(\theta)$	$-1/\theta$
$V[Y]$	σ^2	$\boldsymbol{\mu}(1-\boldsymbol{\mu})$	$\boldsymbol{\mu}$	μ^2/ν
ϕ	σ^2	1/m	1	$1/\nu$
$b(\theta)$	$\theta^2/2$	$\ln(1+e^{\theta})$	$\exp(\theta)$	$-\ln(-\theta)$
$V[\boldsymbol{\mu}] = b''(\boldsymbol{\theta})$	1	$\boldsymbol{\mu}(1-\boldsymbol{\mu})$	$\boldsymbol{\mu}$	$\boldsymbol{\mu}^2$
$D(y,\hat{\mu})$	$\sum_{i=1}^{n}(y_i - \hat{\boldsymbol{\mu}}_i)^2$	$2\sum_{i=1}^{n}\left\{m_i y_i \log\left(\frac{y_i}{\hat{\boldsymbol{\mu}}_i}\right) + (m_i - m_i y_i)\log\left(\frac{1-y_i}{1-\hat{\boldsymbol{\mu}}_i}\right)\right\}$	$2\sum_{i=1}^{n}\left\{y_i \log\frac{y_i}{\hat{\boldsymbol{\mu}}_i} - (y_i - \hat{\boldsymbol{\mu}}_i)\right\}$	$2\sum_{i=1}^{n}\left\{-\log\frac{y_i}{\hat{\boldsymbol{\mu}}_i} + \frac{(y_i - \hat{\boldsymbol{\mu}}_i)}{\hat{\boldsymbol{\mu}}_i}\right\}$

Score Properties : $b'(\boldsymbol{\theta}) = \boldsymbol{\mu} = \mathrm{E}[Y]$, $V[Y] = a(\boldsymbol{\phi})b''(\boldsymbol{\theta})$ and $V[\boldsymbol{\mu}] = b''(\boldsymbol{\theta})$

$\mathrm{E}\left[\frac{\partial \ell}{\partial \boldsymbol{\theta}_i}\right] = \mathrm{E}[\mathrm{U}_i] = 0 \quad \forall i = 1,\ldots,n \qquad \mathbf{IE}(\boldsymbol{\theta}) = \mathfrak{I}$ Fisher : $\mathfrak{I}_{ij} = \mathrm{E}\left[U_i \cdot U_j^{\ T}\right] = \mathrm{E}\left[\frac{\partial \ell}{\partial \boldsymbol{\theta}_i}\frac{\partial \ell}{\partial \boldsymbol{\theta}_j}\right] = \mathrm{E}\left[-\frac{\partial^2 \ell}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j}\right]$

Goodness of Fit: $D(\mathrm{y},\hat{\mu}) = D'(\mathrm{y},\hat{\mu})\boldsymbol{\phi} \quad D'(\mathrm{y},\hat{\mu}) = 2(\ell(\mathrm{y},\boldsymbol{\phi},\mathrm{y}) - \ell(\hat{\mu},\boldsymbol{\phi},\mathrm{y})) \approx \chi^2_{n-p} \qquad X^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{\boldsymbol{\mu}}_i)^2}{V[\hat{\boldsymbol{\mu}}_i]} \approx \chi^2_{n-p}$

Residuals Deviance : $D(\mathrm{y},\hat{\mu}) = \sum_{i=1,\ldots,n} r^2_{D_i}$, $r_D = sign\,(y_i - \hat{\mu}_i)\sqrt{d_i}$ Pearson: $r_P = \frac{y-\boldsymbol{\mu}}{\sqrt{V[\boldsymbol{\mu}]}} \quad X^2 = \sum_{i=1}^{n} r_{P_i}^{\ 2}$

Estimation ML: $\hat{\boldsymbol{\theta}} \approx N_p\left(\boldsymbol{\theta},\ IE(\boldsymbol{\theta})^{-1}\right) \quad AIC = 2(-\ell(\hat{\mu},\mathrm{y}) + p) \quad BIC = -2\,\ell(\hat{\mu},\mathrm{y}) + p\log n$

Contrasts $\mathbf{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ Wald $W = (\hat{\theta} - \theta_0)^T \mathrm{V}[\hat{\theta}]^{-1}(\hat{\theta} - \theta_0) \approx \chi^2_p$ where $\mathrm{V}[\hat{\boldsymbol{\theta}}]^{-1} = IE(\boldsymbol{\theta}) \approx IE(\hat{\boldsymbol{\theta}}) \approx IO(\hat{\boldsymbol{\theta}})$
If $\boldsymbol{\theta}^T = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T)$ dim(θ_2)=q and $\mathbf{H}_0 : \boldsymbol{\theta}_2 = \mathbf{0}$

Wald test: $W = \hat{\boldsymbol{\theta}}_2^T \mathrm{V}[\hat{\boldsymbol{\theta}}_2]^{-1}\hat{\boldsymbol{\theta}}_2 \approx \boldsymbol{\chi}^2_q$

Deviance Test: $\boxed{\Delta D_0 = D(\mathbf{y},\hat{\mu}_0) - D(\mathbf{y},\hat{\mu}) \approx \chi^2_q}$

Score Test: $\mathbf{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad Q = \mathbf{u}(\hat{\theta}_0)^T IE[\hat{\theta}_0]^{-1}\mathbf{u}(\hat{\theta}_0) \approx \chi^2_p$ where $\mathbf{u}(\boldsymbol{\theta}) \approx N_p(\mathbf{0},\ IE(\boldsymbol{\theta}))$

Score method for ML estimation $\boxed{\mathrm{X^T W^{(m)} X}\hat{\beta}^{(m+1)} = \mathrm{X^T W^{(m)} z^{(m)}}}$ m=0,1, ... $\mathfrak{I}^{(*)} = \mathrm{X^T W^{(*)} X}$ where
$\mathrm{W}^{(m)} = dia\left[w_{kk}^{(m)}\right]$

with $\left(w_{kk}^{(m)}\right)^{-1} = V\left[\boldsymbol{\mu}_k^{(m)}\right]\left(g'\left(\boldsymbol{\mu}_k^{(m)}\right)\right)^2$ and $g(y_k) \approx g(\boldsymbol{\mu}_{\boldsymbol{\kappa}}^{(m)}) + (y_k - \boldsymbol{\mu}_{\boldsymbol{\kappa}}^{(m)})g'(\boldsymbol{\mu}_{\boldsymbol{\kappa}}^{(m)}) = z_{\boldsymbol{\kappa}}^{(m)}$

Normal Outcomes

OLS Estimation $\mathbf{Y} = \mathbf{X}\beta + \varepsilon,\ \mathbf{V}[\varepsilon] = \sigma^2\mathbf{I}:\ \hat{\beta} = (\mathrm{X}^T\mathrm{X})^{-}\ \mathrm{X}^T\mathrm{Y}$ and $\hat{\mathrm{Y}} = \mathrm{X}\hat{\beta} = \mathrm{X}(\mathrm{X^T X})^{-}\ \mathrm{X^T Y} = \mathrm{PY}$

$\mathrm{e} = \mathrm{Y} - \hat{\mathrm{Y}} = \mathrm{Y} - \mathrm{X}\hat{\beta} = (\mathrm{I} - \mathrm{P})\mathrm{Y} \qquad \mathrm{E}[\mathbf{e}] = \mathbf{0} \quad V[\mathbf{e}] = \boldsymbol{\sigma}^2(\mathbf{I} - \mathbf{P}) \qquad s^2 = \frac{\mathrm{e}^T \cdot \mathrm{e}}{n-p} = \frac{(\mathrm{Y} - \mathrm{X}\hat{\beta})^T \cdot (\mathrm{Y} - \mathrm{X}\hat{\beta})}{n-p} = \frac{RSS}{n-p}$

If **TSS=RSS+ESS** $\quad R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$

1. $\hat{\beta} \approx \mathrm{N}_p\left(\beta,\ \boldsymbol{\sigma}^2(\mathbf{X^T X})^{-1}\right)$
2. $(\hat{\beta} - \beta)^T\ \mathrm{V}[\hat{\beta}]^{-1}(\hat{\beta} - \beta) \approx \chi^2_p$.

Official Crib Sheet de MSCTM-DATL and MUM-DATS:

Prof. Lídia Montero -DEIO Page. 2 22/12/2022

3. $\hat{\beta} = \left(X^T X\right)^{-1} X^T Y$ is independent of s^2.

4. $RSS/\sigma^2 = (n-p)s^2/\sigma^2 \approx \chi^2_{n-p}$

Individual CI: $t = \dfrac{\hat{\beta}_i - \beta_i}{\hat{\sigma}_{\hat{\beta}_i}} \approx t_{n-p} \rightarrow \hat{\beta}_i \pm t^{\alpha/2}_{n-p}\hat{\sigma}_{\hat{\beta}_i}$ where

$$\hat{\sigma}_{\hat{\beta}_i} = s\sqrt{\left(X^T X\right)^{-1}_{ii}}$$

Contrast Fisher Test. $F = \dfrac{(RSS_H - RSS)/q}{RSS/(n-p)} \rightarrow F_{q,n-p}$

ESTIMATION: OLS/Weighted/Generalized

MODEL	$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,\ \mathbf{V}[\varepsilon] = \sigma^2\mathbf{I}$	$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,\ \mathbf{V}[\varepsilon] = \sigma^2\mathbf{D}$	$\mathbf{Y} = \mathbf{X}\beta + \varepsilon,\ \mathbf{V}[\varepsilon] = \sigma^2\mathbf{W} = \sigma^2\mathbf{K}\mathbf{K}^T$
Transformació	$\mathbf{Y} \rightarrow \mathbf{Y},\ \mathbf{X} \rightarrow \mathbf{X}$	$\mathbf{Y} \rightarrow \mathbf{D}^{-\frac{1}{2}}\mathbf{Y},\ \mathbf{X} \rightarrow \mathbf{D}^{-\frac{1}{2}}\mathbf{X}$	$\mathbf{Y} \rightarrow \mathbf{K}^{-1}\mathbf{Y},\ \mathbf{X} \rightarrow \mathbf{K}^{-1}\mathbf{X}$
$S(\beta)$ *Sum of squares*	$(\mathbf{Y} - \mathbf{X}\beta)^T(\mathbf{Y} - \mathbf{X}\beta)$	$(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{D}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$	$(\mathbf{Y} - \mathbf{X}\beta)^T\mathbf{W}^{-1}(\mathbf{Y} - \mathbf{X}\beta)$
Normal Equations $\mathbf{S}\beta = \mathbf{Q}$	$\mathbf{S} = \mathbf{X}^T\mathbf{X},\ \mathbf{Q} = \mathbf{X}^T\mathbf{Y}$	$\mathbf{S} = \mathbf{X}^T\mathbf{D}^{-1}\mathbf{X},\ \mathbf{Q} = \mathbf{X}^T\mathbf{D}^{-1}\mathbf{Y}$	$\mathbf{S} = \mathbf{X}^T\mathbf{W}^{-1}\mathbf{X},\ \mathbf{Q} = \mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y}$
Estimates $\hat{\beta}$	$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$	$(\mathbf{X}^T\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{D}^{-1}\mathbf{Y}$	$(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y}$
Variance $\hat{\beta}$	$\sigma^2(\mathbf{X}^T\mathbf{X})^{-1}$	$\sigma^2(\mathbf{X}^T\mathbf{D}^{-1}\mathbf{X})^{-1}$	$\sigma^2(\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X})^{-1}$
RSS	$\mathbf{Y}^T\mathbf{Y} - (\mathbf{X}^T\mathbf{Y})^T\hat{\beta}$	$\mathbf{Y}^T\mathbf{D}^{-1}\mathbf{Y} - (\mathbf{X}^T\mathbf{D}^{-1}\mathbf{Y})^T\hat{\beta}$	$\mathbf{Y}^T\mathbf{W}^{-1}\mathbf{Y} - (\mathbf{X}^T\mathbf{W}^{-1}\mathbf{Y})^T\hat{\beta}$

Validation

$$h(Y) = \begin{cases} \dfrac{Y^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log Y & \lambda = 0 \end{cases}$$
$$Tranf.Box - Cox$$

Influent data: *a priori* i s.t. $p_{ii} > 2p/n$.
Influent data: *a posteriori* Cook's distance.

General Linear Model

Two-way Anova: Factor A levels i=1,..,I and Factor B levels j=1,...,J

Null model: $Y_{ijk} = \mu + \varepsilon_{ijk}$ *p=1*

A*B: $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$ *p =I*J*

A+B: $Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$ *p=I+J-1*

A: $Y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$ *p=I*

B: $Y_{ijk} = \mu + \beta_j + \varepsilon_{ijk}$ *p=J*

Reparam.*Base-Line*: $\alpha_I = 0 \quad \beta_J = 0$

$\gamma_{Ij} = 0 \quad \forall j = 1 \ldots J,\ \gamma_{iJ} = 0\ \forall i = 1 \cdots I$

Reparam. *Zero-Sum*: $\sum_{i=1}^{I}\alpha_i = 0 \quad \sum_{j=1}^{J}\beta_j = 0$

$\sum_{i=1}^{I}\gamma_{ij} = 0 \quad \forall j = 1 \ldots J,\ \sum_{j=1}^{J}\gamma_{ij} = 0\ \forall i = 1 \cdots I$

ANCOVA 1 FACTOR and 1 COVARIATE: Factor A levels i=1,..,I X - Covariate

A*X: $Y_{ik} = \mu + \alpha_i + (\lambda + \theta_i)x_{ik} + \varepsilon_{ik}$ *p =I+I*

A+X: $Y_{ik} = \mu + \alpha_i + \lambda x_{ik} + \varepsilon_{ik}$ *p=I+1*

A: $Y_{ijk} = \mu + \alpha_i + \varepsilon_{ijk}$ *p=I*

Reparam.*Base-Line*: $\alpha_I = 0 \quad \theta_I = 0$

Reparam. *Zero-Sum*: $\sum_{i=1}^{I}\alpha_i = 0 \quad \sum_{i=1}^{I}\theta_i = 0$

(SLR) X: $Y_{ik} = \mu + \lambda x_{ik} + \varepsilon_{ik}$ *p=2*

Binary Outcomes $Y \approx B(m,\pi)$ $\quad E[Y] = m\pi = \mu \quad V[Y] = m\pi(1-\pi) \quad V[\mu] = \mu\cdot(m-\mu)/m$

$$X^2 = \sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{V[\hat{\mu}_i]} = \sum_{i=1}^{n}\frac{m_i(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i(m_i - \hat{\mu}_i)} \approx \chi^2_{n-p} \qquad D'(\mathrm{y},\hat{\mu}) = D(\mathrm{y},\hat{\mu}) = 2\sum_{i=1}^{n}\left\{y_i\log\left(\frac{y_i}{\hat{\mu}_i}\right) + (m_i - y_i)\log\left(\frac{m_i - y_i}{m_i - \hat{\mu}_i}\right)\right\} \approx \chi^2_{n-p}$$

Link	$\eta = g(\pi)$	$\pi = g^{-1}(\eta)$	**Latent Scale**
Logit	$\mathbf{logit}(\pi) = \log\left(\dfrac{\pi}{1-\pi}\right)$	$\dfrac{\exp(\eta)}{1+\exp(\eta)}$	Logístic law (Standard): location parameter 0 and scale 1 $\rightarrow E[\bullet] = 0\ V[\bullet] = \pi^2/3$
Probit	$\Phi_Z^{-1}(\pi)$	$\Phi_Z(\eta)$	Standard Normal – Z-N(0,1)
C-Log-Log (Gompit)	$\log(\log(1/(1-\pi)))$	$1 - \exp(-\exp(\eta))$	Min Extreme Value – Gompertz Std - Location 0 and Scale 1 $\rightarrow E[\bullet] = -0.577216.\ V[\bullet] = \pi^2/6$

Official Crib Sheet de MSCTM-DATL and MUM-DATS:

Prof. Lídia Montero -DEIO Page. 3 **22/12/2022**

Log-Log	$-\log(\log(1/\pi))$	$1-\exp(-\exp(-\eta))$	Max Extreme Value – Gumbel Std - Location 0 and Scale 1 $\rightarrow \mathrm{E}[\bullet]=0.577216$. $V[\bullet]=\pi^2/6$

$$R_A^2 = 1-\frac{D(\mathbf{y},\pi_A)}{D(\mathbf{y},\pi_0)} = \frac{G(\mathbf{y},\pi_A)}{G(\mathbf{y},\pi_A)+D(\mathbf{y},\pi_A)} \quad on\ G(\mathbf{y},\pi_A)=D(\mathbf{y},\pi_0)-D(\mathbf{y},\pi_A)$$

$\boxed{D(\mathbf{y},\pi_0)}$: Null Model

ROC=0.5(1+D de Sommer)

Polytomous Outcomes $Y \approx MN(m,\boldsymbol{\pi})$ $\quad \boldsymbol{\pi}^T \approx (\pi_1,\ldots,\pi_J)$ $\boldsymbol{\gamma}^T \approx (\gamma_1,\ldots,\gamma_{J-1},1)$ $\gamma_j=\pi_1+\ldots+\pi_j \quad j=1,\ldots J-1$

Nominal:

$$\boxed{\log\frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})}=\alpha_j+\beta_j^{\mathbf{T}}\mathbf{x_i} \quad \begin{array}{l} j=2,\ldots,J \\ i=1,\ldots,n \end{array}}$$

J-1 Equations Odds j over 1:

$$\frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})}=\exp\{\alpha_j+\boldsymbol{\beta}_\mathbf{j}^\mathbf{T}\mathbf{x_i}\}$$

Null Model: $\log\frac{\pi_{ij}(\mathbf{x_i})}{\pi_{i1}(\mathbf{x_i})}=\alpha_j \quad \begin{array}{l} j=2,\ldots,J \\ i=1,\ldots,n \end{array}$

$$\pi_{i1}(\mathbf{x}_i)=\frac{1}{1+\sum_{r\neq 1}\pi_{ij}(\mathbf{x_i})/\pi_{iJ}(\mathbf{x_i})}$$

Ordinal: $\gamma_j=\mathrm{P}(Y\leq j)$

Link	$\eta_{ij}=\mathrm{g}(\gamma_{ij})$	$\gamma_{ij}=g^{-1}(\eta_{ij})$	**Latent Scale**
Logit	$\log\frac{\gamma_j(\mathbf{x}_i)}{1-\gamma_j(\mathbf{x}_i)}=\alpha_j+\beta^T\mathbf{x}_i \quad j=1,\ldots,J-1$ $\alpha_1\leq\alpha_2\leq\ldots\leq\alpha_{J-1}$	$\gamma_{ij}=\frac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}$ $j=1,\ldots,J-1$	Standard Logistic: location 0 and scale 1 $\rightarrow \mathrm{E}[\bullet]=0\ V[\bullet]=\pi^2/3$
Probit	$\Phi_Z^{-1}(\gamma_j(\mathbf{x}_i))=\alpha_j+\beta^T\mathbf{x}_i \quad j=1,\ldots,J-1$ $\alpha_1\leq\alpha_2\leq\ldots\leq\alpha_{J-1}$	$\gamma_{ij}=\Phi_Z(\eta_{ij})$ $j=1,\ldots,J-1$	Standard Normal – Z-N(0,1)
C-Log-Log (Gompit)	$\log(-\log(1-\gamma_j(\mathbf{x}_i)))=\alpha_j+\beta^T\mathbf{x}_i \quad j=1,\ldots,J-1$ $\alpha_1\leq\alpha_2\leq\ldots\leq\alpha_{J-1}$	$\gamma_{ij}=1-\exp(-\exp(\eta_{ij}))$ $j=1,\ldots,J-1$	Min Extreme – Gompertz Std. – Location 0 and scale 1 $\rightarrow \mathrm{E}[\bullet]=-0.577216$. $V[\bullet]=\pi^2/6$
Log-Log	$-\log(-\log(\gamma_j(\mathbf{x}_i)))=\alpha_j+\beta^T\mathbf{x}_i \quad j=1,\ldots,J-1$ $\alpha_1\leq\alpha_2\leq\ldots\leq\alpha_{J-1}$	$\gamma_{ij}=1-\exp(-\exp(-\eta_{ij}))$ $j=1,\ldots,J-1$	Max Extreme – Gumbel Std. - Location 0 and scale 1 $\rightarrow \mathrm{E}[\bullet]=0.577216$. $V[\bullet]=\pi^2/6$

Log-lineal Models $Y \approx \wp(\mu) \rightarrow V[Y_i]=\mathrm{E}[Y_i]=\mu_i$ $\quad \boxed{p_Y(y)=(\mu^y/y!)e^{-\mu} \quad y=0,1,\ldots}$

$$\boxed{\log(\mu_i)=\eta_i=\mathbf{x}_i^T\boldsymbol{\beta} \quad i=1,\ldots,n}$$

Case $\boxed{V[Y_i]=\phi\,\mathrm{E}[Y_i]=\phi\,\mu_i}$ with $\boxed{\phi>1}$ Quasi Likelihood-Score method $\boxed{V[\mu_i]=\phi\,\mu_i}$. Estimates

QML: $\mathrm{V}_{QMV}[\beta]=\phi\,(\mathrm{X}^T\mathrm{WX})^{-1}=\phi\,\mathrm{V}_{MV}[\beta]$ and $\hat{\phi}=\frac{X^2}{n-p}$.

Case $\boxed{Y\approx BNeg(\alpha=1/\sigma^2,\beta=1/\sigma^2)}$

$$Y/\theta\approx\wp(\theta\mu),\ \text{si}\ \theta\approx\Gamma(\alpha=\beta=1/\sigma^2)\rightarrow P(\{Y=y\})=\frac{\Gamma(\alpha+y)}{y!\,\Gamma(\alpha)}\frac{\beta^\alpha\mu^y}{(\mu+\beta)^{\alpha+y}} \quad \text{and}\ V[\mu]=\mu+\mu^2\sigma^2$$