Régression logistique

RSM - MS ENSAI - 2026-2027



École nationale
de la statistique
et de l'analyse
de l'information



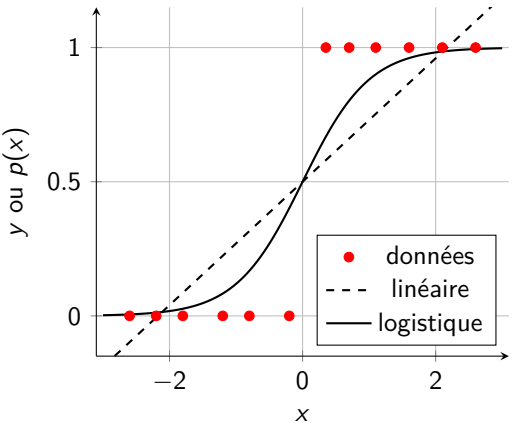
Université
de Rennes

Régression : que cherche-t-on à expliquer ?

Trois écritures utiles

$$\begin{aligned} \text{Linéaire :} \quad & Y = X\beta + \varepsilon \\ \text{Non linéaire :} \quad & Y = f(X, \beta) + \varepsilon \\ \text{Logistique :} \quad & \mathbb{E}[Y \mid X = x] = \Lambda(x^T \beta) \end{aligned}$$

- La régression explique une variable Y à partir de variables explicatives $X_1, \ldots, X_p$.
- Si Y est quantitative : régression linéaire, ANOVA, ANCOVA, etc.
- Si Y est qualitative binaire : le modèle linéaire prédit parfois hors de $[0, 1]$.



Rappel : modèle linéaire et moindres carrés

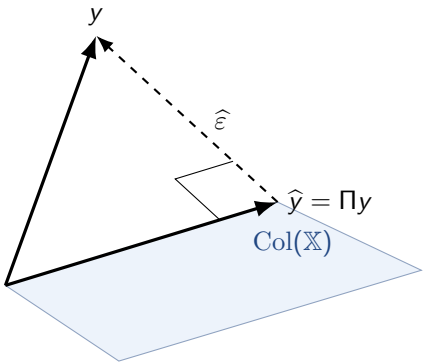
Modèle simple et multiple

$$Y = \beta_0 + \beta_1 X + \varepsilon,$$
$$y = \mathbb{X}\beta + \varepsilon, \qquad \mathbb{X} \in \mathbb{R}^{n \times (p+1)}.$$

Estimateur des moindres carrés

$$\widehat{\beta} = \underset{\beta \in \mathbb{R}^{p+1}}{\arg\min} \|y - \mathbb{X}\beta\|_2^2,$$
$$\widehat{\beta} = \mathbb{X}^\dagger y = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T y \quad \text{si } \mathbb{X} \text{ est de rang plein.}$$

$\Pi = \mathbb{X}\mathbb{X}^\dagger$ et $\widehat{y} = \Pi y$.



Les valeurs ajustées sont le projeté orthogonal de y.

Hypothèses et qualité d'une régression linéaire

Hypothèses usuelles de Gauss-Markov

- $\mathbb{E}[\varepsilon \mid X] = 0$,
- $\mathbb{E}[\widehat{\beta}] = \beta$,
- $\mathbb{V}(\widehat{\beta}) = \sigma^2(\mathbb{X}^T\mathbb{X})^{-1}$.

Sous ces hypothèses, l'OLS $\widehat{\beta}$ est BLUE.

$\varepsilon|X \sim \mathcal{N}(0, \sigma^2 I_n)$, $\Rightarrow$ OLS=MLE et $Y|\mathbb{X} \sim \mathcal{N}(\mathbb{X}\beta, \sigma^2 I_n)$.

Décomposition de variance

$$SCT = SCE + SCR \text{ et } R^2 = \frac{SCE}{SCT}$$

Validation du modèle

- pertinence de la linéarité ;
- analyse des résidus ;
- normalité et homoscédasticité ;
- valeurs aberrantes ;
- variables pertinentes ;
- régularisation si nécessaire : Ridge, LASSO.

Régression logistique - idée clef : linéariser la log-cote, pas la probabilité

Variable binaire

$$p(x) = \mathbb{P}[Y = 1 \mid X = x] = \mathbb{E}[Y \mid X = x],$$

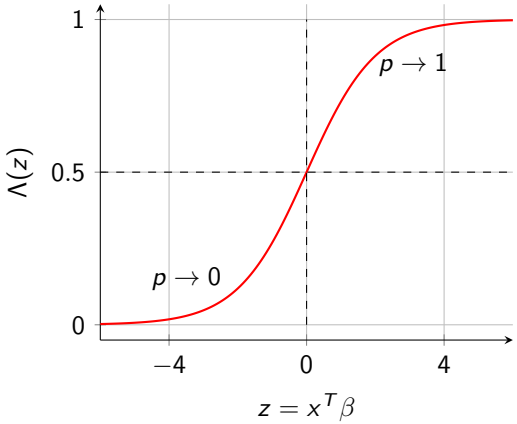
$$\pi(x) = \frac{p(x)}{1 - p(x)} \quad \text{cote},$$

$$\text{logit}(p(x)) = \ln\left(\frac{p(x)}{1 - p(x)}\right) = x^T\beta = L(x).$$

Fonction logistique

$$p(x) = \Lambda(x^T\beta) = \frac{1}{1 + e^{-x^T\beta}} \text{ avec } \Lambda(z) = \frac{1}{1 + e^{-z}}.$$

- $\Lambda = L^{-1}$.



Estimation par maximum de vraisemblance

Vraisemblance de Bernoulli

$$\mathcal{L}(\beta) = \prod_{i=1}^{n} p_i^{y_i}(1-p_i)^{1-y_i},$$

$$\ell(\beta) = \sum_{i=1}^{n} \left[y_i x_i^T \beta - \log(1 + e^{x_i^T \beta}) \right],$$

$$p_i = \Lambda(x_i^T \beta).$$

Sous hypothèses de régularité du modèle, on a :

$$\sqrt{n} \rightsquigarrow \mathcal{N}(0, \mathbb{I}(\beta)^{-1})$$

Gradient et Hessienne

$$\nabla \ell(\beta) = \mathbb{X}^T (y - P),$$

$$H(\beta) = -\mathbb{X}^T W \mathbb{X},$$

$$W = \mathrm{diag}(p_i(1-p_i)).$$

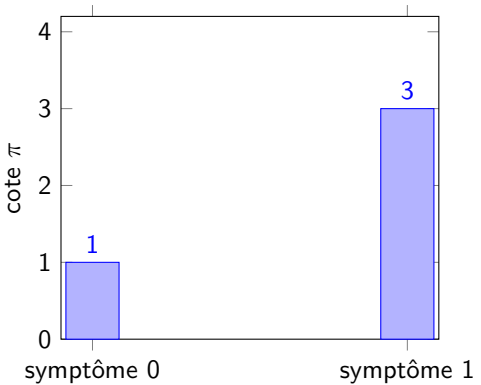
- Les équations ne sont pas linéaires.
- On utilise Newton-Raphson ou IRLS.
- Asymptotiquement : $\widehat{\beta}$ est normal sous hypothèses de régularité.

Interpréter les coefficients : cotes et odds ratios

Effet d'une variable

$$\text{logit}(p(x)) = L(x) = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p,$$
$$\frac{\pi(x + e_j)}{\pi(x)} = e^{\beta_j}.$$

- $\beta_j > 0$: la cote de $Y = 1$ augmente quand x_j augmente.
- e^{β_j} est l'odds ratio associé à une variation d'une unité de x_j.
- L'odds ratio ne doit pas être confondu avec le risque relatif.



Ici $\rho(1, 0) = 3 = e^{\beta_1}$: la cote est multipliée par 3.

Extension multiclasse : régression softmax

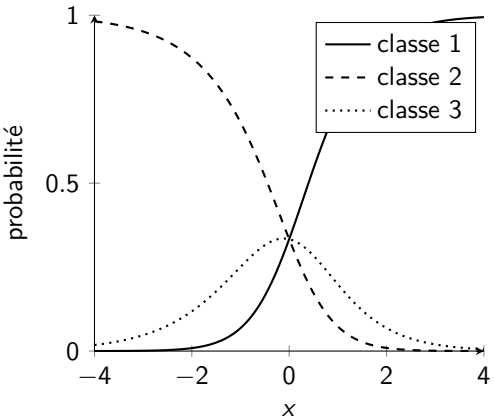
Pour une variable qualitative à K modalités, on modélise simultanément les probabilités de classe :

$$\mathbb{P}(Y = k \mid X = x) = \frac{\exp(x^T \beta_k)}{\sum_{r=1}^{K} \exp(x^T \beta_r)}.$$

On peut fixer une classe de référence, par exemple $\beta_K = 0$, pour assurer l'identifiabilité.

Idée

Le modèle compare chaque classe à une classe de référence, puis normalise les probabilités pour qu'elles somment à 1.



Évaluer une classification binaire

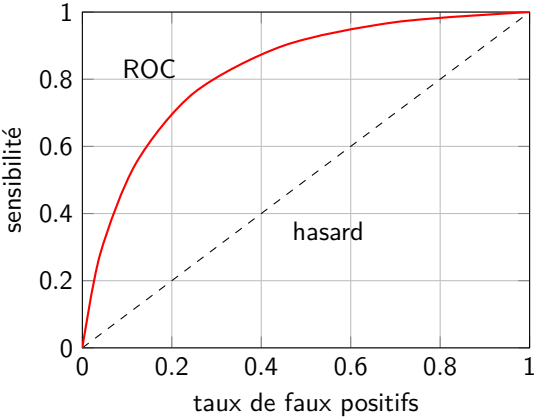
Matrice de confusion

	Prédit	
Observé	1	0
1	VP	FN
0	FP	VN

$$\text{sensibilité} = \frac{VP}{VP + FN},$$

$$\text{spécificité} = \frac{VN}{VN + FP},$$

$$\text{précision} = \frac{VP}{VP + FP}.$$



AUC : probabilité qu'un positif reçoive un score supérieur à un négatif.

Diagnostics de qualité du modèle logistique

Déviance

$$D(\widehat{\beta}) = -2\ell(\widehat{\beta}),$$
$$D_{\text{res}} = 2\big(\ell_{\text{sat}} - \ell(\widehat{\beta})\big).$$

- Pseudo-R^2 : comparaison au modèle nul.
- Hosmer–Lemeshow : calibration par groupes de probabilité.
- Mann–Whitney / AUC : séparation des positifs et négatifs.
- Analyse des résidus et observations influentes.

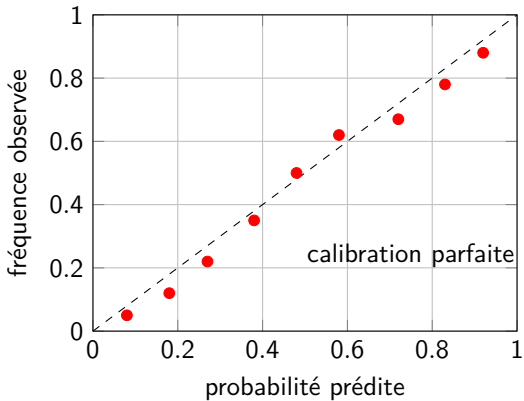


Diagramme de fiabilité : les points doivent rester proches de la diagonale.

Significativité et sélection de variables

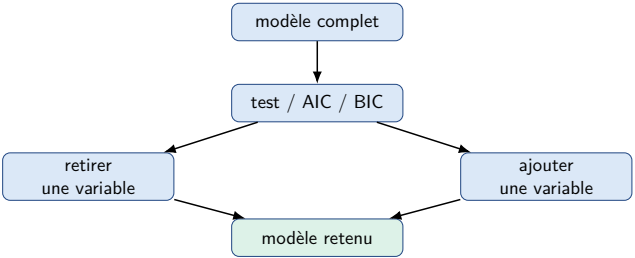
Tests courants

- Test du rapport de vraisemblance : comparer deux modèles emboîtés.
- Test de Wald : $Z = \widehat{\beta}_j / \mathrm{se}(\widehat{\beta}_j)$.
- Tests globaux : les régresseurs apportent-ils une information ?

Critères de sélection

$$AIC = 2k - 2\ell(\widehat{\beta}),$$

$$BIC = k \log(n) - 2\ell(\widehat{\beta}).$$



Principe : un modèle plus simple est souvent plus robuste et plus interprétable.

En pratique : R et Python

R

```
modele <- glm(y ~ x1 + x2,
              data = donnees,
              family = binomial)
summary(modele)
anova(modele, test = "Chisq")
predict(modele, type = "response")
```

- `summary` : tests de Wald par coefficient.
- `anova` : comparaison séquentielle ou modèles emboîtés.

Python / scikit-learn

```
from sklearn.linear_model
          import LogisticRegression
from sklearn.metrics
          import confusion_matrix, roc_auc_score

modele = LogisticRegression()
modele.fit(X_train, y_train)
proba = modele.predict_proba(X_test)[:, 1]
y_pred = (proba >= 0.5).astype(int)
auc = roc_auc_score(y_test, proba)
```

- Le seuil 0,5 peut être adapté au coût des erreurs.
- Les probabilités prédites se valident par ROC et calibration.
- `LogisticRegression()` utilise par défaut une régularisation L^2 $\neq$ régression logistique non pénalisée exposée ici.

À retenir

Résumé en quatre idées

1. La régression linéaire projette y sur l'espace engendré par les variables explicatives.
2. Pour une variable binaire, on modélise une probabilité $p(x)$, pas directement Y.
3. La régression logistique suppose que le logit de $p(x)$ est linéaire en x.
4. L'évaluation combine interprétation, tests, matrice de confusion, ROC/AUC, déviance et calibration.

$$\log\left(\frac{p(x)}{1-p(x)}\right) = x^T\beta \quad \Longleftrightarrow \quad p(x) = \frac{1}{1+e^{-x^T\beta}}$$