

Application Bayésienne en Capture Recapture :

*Hétérogénéité de capture et
Incertitude sur les modèles*

Identifier les **cas communs, ou non, entre sources** pour **estimer** le nombre **total** de cas $N \Rightarrow$ Exhaustivité Registre

Conditions d'application :

➤ *Indépendance des sources*

Modélisation log-linéaire : interactions entre sources

$$\text{Modèle : } \ln F_{ijk} = \mu + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

➤ *Homogénéité de capture des cas*

Capture non dépendante de leurs caractéristiques

Hétérogénéité \rightarrow dépendance

Log-linéaire : inclus dans termes d'interaction

Impact **caractéristiques individuelles** sur estimation de N

➤ Modèles M_{tbh} en **écologie de la faune**

Capture recapture en écologie

Probabilité de capture de l'individu dépend du :

- Temps de capture **t**
- Comportement après 1^{ère} capture **b**
- Hétérogénéité individuelle **h**

Modèle M_{tbh} saturé

$$\text{Logit}(p_{i\tau}) = \mu + \alpha_\tau + \beta Y_{i\tau} + \gamma_i \quad \rightarrow \text{8 modèles possibles}$$

Paramètres estimés $\theta = \{\mu, \alpha, \beta, \sigma^2_\gamma, N\}$

$$f(x|\theta, \gamma) \propto \frac{N!}{(N-n)!} \prod_{i=1}^T \prod_{i=1}^N p_{it}^{x_{it}} (1-p_{it})^{(1-x_{it})}$$

Approche Bayésienne pour estimer l'hétérogénéité

Incertitude sur les modèles \rightarrow BMA (Modèle m = paramètre)

Pas d'hypothèses à vérifier

Théorie des probabilités :

$$\Pr(B,A) = \Pr(B|A).\Pr(A) = \Pr(A,B) = \Pr(A|B).\Pr(B)$$

Soit $B \rightarrow B_i$ événements $\in B_j$ explicatifs de la survenue de A $\Pr(A) = \sum_j \Pr(A|B_j)\Pr(B_j)$

$$\Pr(B_i|A) \propto \Pr(A|B_i) \Pr(B_i)$$

Conclusion fiable sur B_i à partir de données d'échantillons A

Appliqué à X (historique capture) et Θ (paramètres) $[\theta|X] \propto [X|\theta][\theta]$

$$\begin{array}{ccc} \swarrow & & \nwarrow \\ [\theta|X] & \propto & [X|\theta][\theta] \\ \swarrow & & \nwarrow \\ \text{Densité de probabilité} & & \text{Fonction de} & & \text{Densité de probabilité} \\ \text{a posteriori} & & \text{vraisemblance} & & \text{a priori} \end{array}$$

Bayesian Model Averaging et techniques MCMC

Le modèle lui-même est considéré comme un paramètre

$$\pi(\theta_m, m|x) \approx g(x|\theta_m, m)p(\theta_m|m)p(m)$$

Estimation de **N pondérée** sur densité *a posteriori* des 8 **modèles** M_{tbh}

➤ Technique de Monte Carlo

Estimation empirique à partir de $\theta^1, \dots, \theta^n \sim \pi(\theta|x)$

$$E_{\pi}[\varphi(\theta)] = \int \varphi(\theta)\pi(\theta|x)d\theta \rightarrow \bar{\varphi}_n = \frac{1}{n} \sum_{i=1}^n \varphi(\theta^i)$$

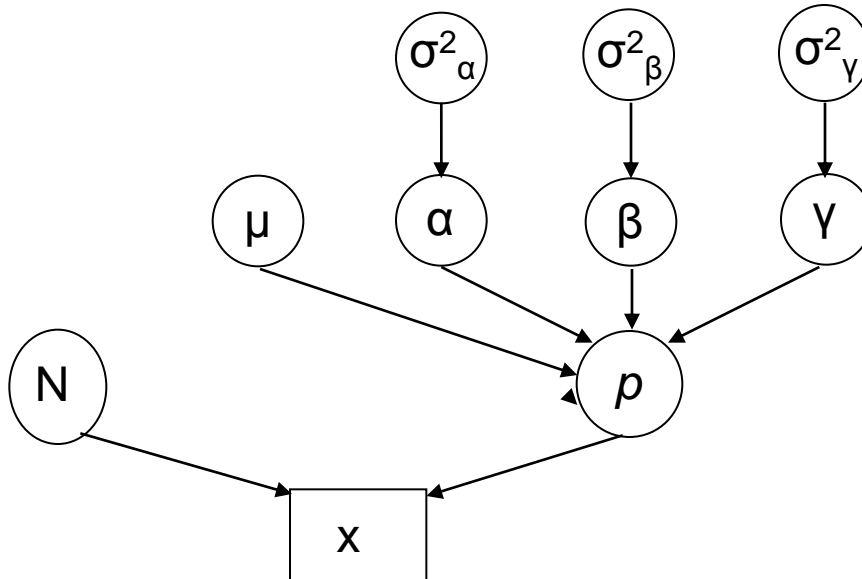
➤ Chaînes de Markov

Séquence probabiliste qui génère θ^{k+1} à partir de θ^k $\theta^{k+1} \sim K(\theta^k, \theta)$

Convergence vers **distribution stationnaire** $\pi(\theta) \equiv \pi(\theta|x)$

WinBUGS*

Bayesian Using Gibbs Sampler



model{

logit(p[i,k]) <- mu + alpha[k] + (beta[i]*Y[i,k])+ gamma[i]

z[i,k]~dbin(phi,1) zp[i,k]<-p[i,k]*z[i,k] x[i,k] ~ dbin(zp[i,k],T)}

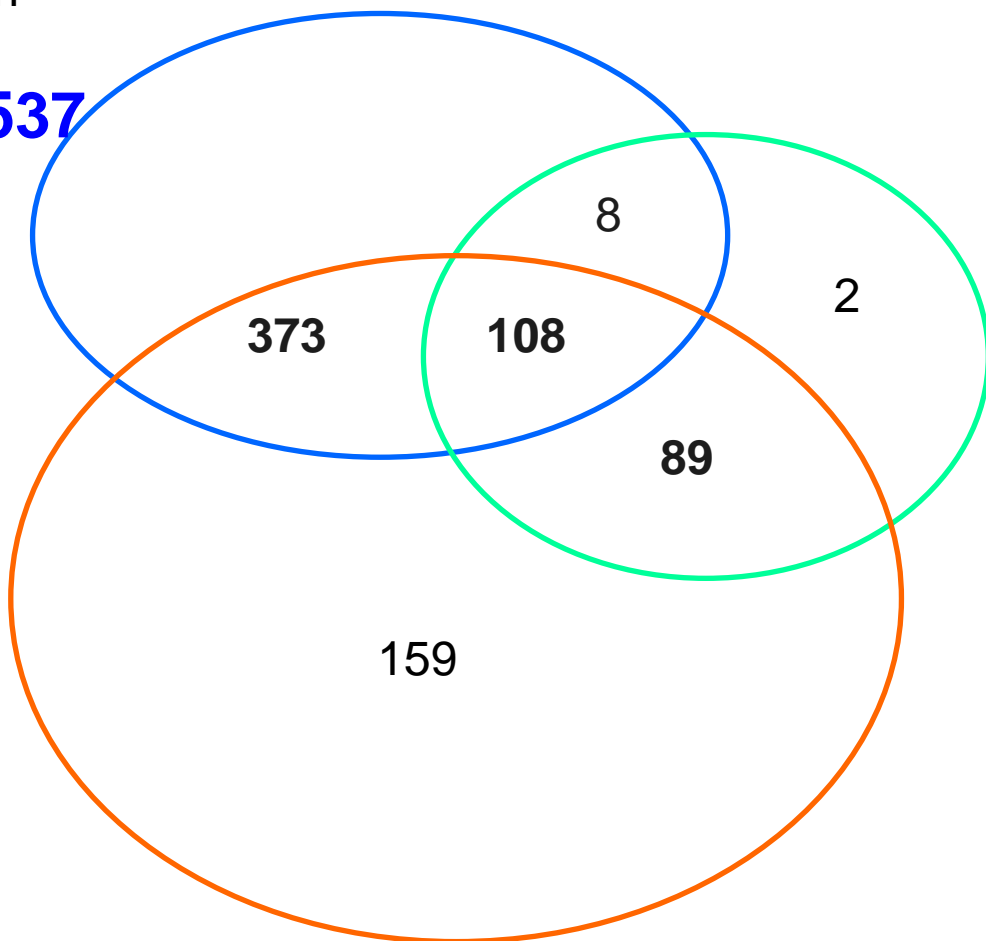
N<-sum(z[,,]) }

list(x=structure(.Data=c(1,1,1,1, ...,2,2,2,2,...3,3,3,3...0,0,0,...))

z=structure(.Data=c(1,1,1,1,...,NA,NA,NA,NA))

Application aux cancers du sein Alpes-Maritimes 2008

Cancers
vus en
RCP
N = 537



Cancers
dépistés
DAM
N = 207

Cancers
diagnostiqués
Registre
Histologique RH
N = 729

Total des
cancers
du sein
déclarés
en 2008
N = 787

Estimation du nombre total de cancers du sein Modèle log-linéaire*

Modèles	X_x	N	IC 95 %	ddl	G^2^{**}	p
$\mu + RH + RCP + DAM + RH * RCP + RH * DAM + RCP * DAM$	6	793	[782-804]	0	0	1,0
$\mu + RH + RCP + DAM + RH * RCP + RH * DAM$	12	799	[779-819]	1	14,6	<0,001
$\mu + RH + RCP + DAM + RCP * DAM + RH * DAM$	20	807	[796-819]	1	2,7	0,10
$\mu + RH + RCP + DAM + RCP * DAM + RH * RCP$	4	791	[784-797]	1	2,2	0,15
$\mu + RH + RCP + DAM + RCP * DAM$	16	803	[794-813]	2	10,6	0,01
$\mu + RH + RCP + DAM + RH * DAM$	25	812	[799-824]	2	15,6	<0,01
$\mu + RH + RCP + DAM + RH * RCP$	6	793	[783-802]	2	19,5	<0,01
$\mu + RH + RCP + DAM$	19	806	[793-820]	3	24,6	<0,01

Dépendances entre RCP-Dépistés et entre RCP-Registre

Estimation N = 791

* Investigating the completeness of a histopathological cancer registry: estimation by capture-recapture analysis in a French geographical unit Alpes-Maritimes, 2008. Cancer Epidemiol. 2011 Dec;35(6):e62-8.

** G^2 = Adéquation modèles emboîtés $G^2_{M1} - G^2_{M2}$ rapport de vraisemblance : $G^2_M = 2 \sum f_{ijk} \ln (f_{ijk} / F_{ijk})$

Estimation du nombre total de cancers du sein par M_{tbh} Trois captures : RH DAM RCP

Modèles	Moyenne	Ecart-type	Médiane	HPDI95%
M_0	789,1	1,5	788,9	786,9-792,9
M_t	788,9	1,5	788,9	786,9-791,9
M_b	1 605,0	41,6	1 604,0	1525,0-1684,0
M_h	1 316,0	45,8	1 314,0	1241,0-1410,0
M_{tb}	1 656,0	45,3	1 656,0	1567,0-1742,0
M_{th}	1 301,0	36,2	1 301,0	1230,0-1374,0
M_{bh}	1 306,0	41,7	1 304,0	1230,0-1391,0
M_{tbh}	1 300,0	34,3	1 300,0	1234,0-1370,0
BMA M_{avg}^*	789,0	1,4	788,9	786,9-791,9

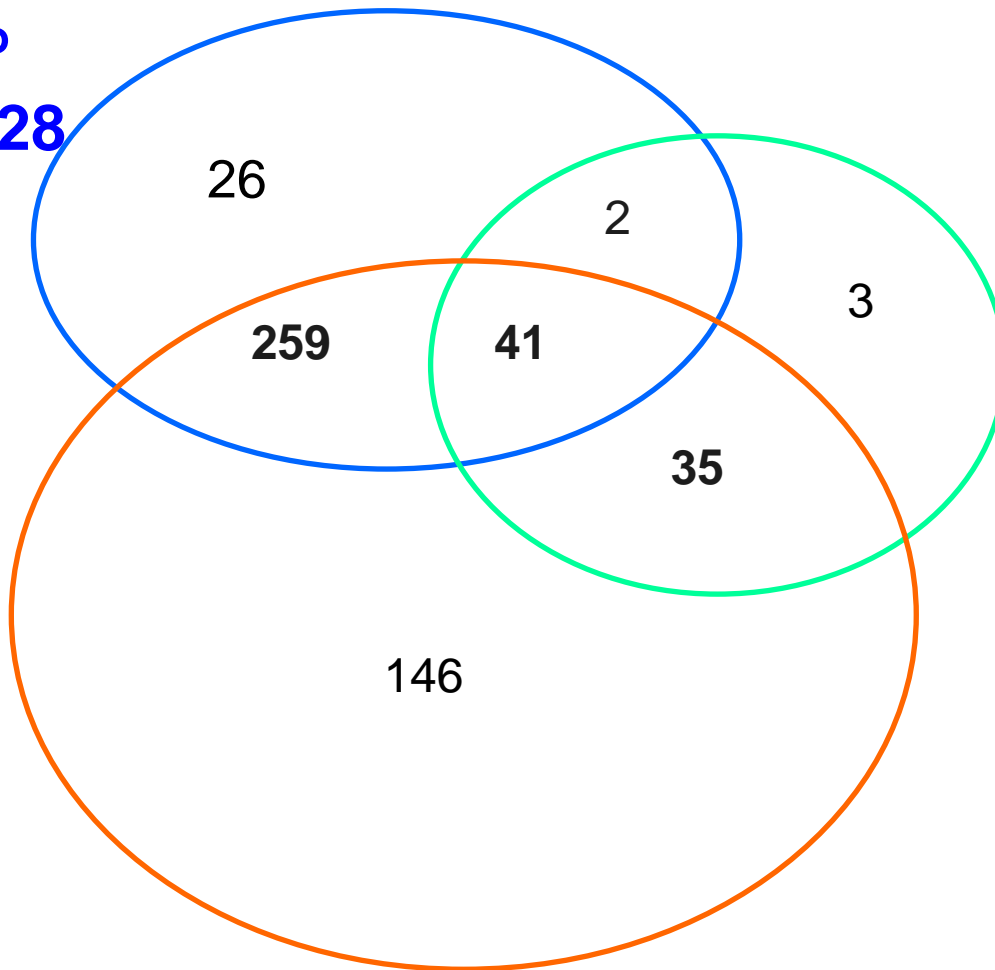
Estimation pondérée sur l'ensemble des modèles N = 789

Variables hétérogénéité : âge, stade TNM, dépistage N = 790

*Mavg = Estimation pondérée sur l'ensemble des modèles obtenue après 5 000 itérations

Application aux cancers colorectaux Alpes-Maritimes 2008

Cancers
vus en
RCP
N=328



Cancers
dépistés
DAM
N = 81

Cancers
diagnostiqués
Registre
Histologique RH
N=481

Total des
cancers
colorectaux
déclarés
2008

N = 512

Estimation du nombre total de cancers colorectaux Modèle log-linéaire*

Modèles	X_x	N	IC 95 %	ddl	G^2^{**}	p
$\mu + RH + RCP + DAM + RH^*RCP + RH^*DAM + RCP^*DAM$	1	513	[510-516]	0	0	1,0
$\mu + RH + RCP + DAM + RH^*RCP + RH^*DAM$	39	551	[479-623]	1	2,7	0,1
$\mu + RH + RCP + DAM + RCP^*DAM + RH^*DAM$	15	527	[517-536]	1	0,4	0,6
$\mu + RH + RCP + DAM + RCP^*DAM + RH^*RCP$	13	525	[501-548]	1	1,1	0,3
$\mu + RH + RCP + DAM + RCP^*DAM$	14	526	[516-535]	2	1,1	0,6
$\mu + RH + RCP + DAM + RH^*DAM$	16	528	[518-538]	2	3,7	0,2
$\mu + RH + RCP + DAM + RH^*RCP$	17	529	[508-549]	2	4,5	0,1
$\mu + RH + RCP + DAM$	15	527	[517-536]	3	4,6	0,2

Modèle retenu : pas de dépendances entre sources

Estimation N = 527

* Investigating the completeness of a histopathological cancer registry: estimation by capture-recapture analysis in a French geographical unit Alpes-Maritimes, 2008. Cancer Epidemiol. 2011 Dec;35(6):e62-8.

** $G^2 = \text{Adéquation modèles emboîtés } G^2_{M1} - G^2_{M2}$ rapport de vraisemblance : $G^2_M = 2 \sum f_{ijk} \ln (f_{ijk} / F_{ijk})$

Estimation du nombre total de cancers colorectaux par M_{tbh}

Trois Captures : RH DAM RCP

Modèles	Moyenne	Ecart-type	Médiane	HPDI95%
M_0	516,1	2,2	515,6	512,6-520,6
M_t	515,8	2,2	515,6	512,6-520,6
M_b	1 431,0	57,5	1 431,0	1320,0-1543,0
M_h	921,9	39,1	921,6	845,6-1005,0
M_{tb}	1 509,0	59,5	1 509,0	1400,0-1627,0
M_{th}	909,7	33,3	909,6	843,6-974,6
M_{bh}	923,6	40,2	922,6	849,6-1008,0
M_{tbh}	911,8	32,2	911,6	849,6-975,6
BMA M_{avg}^*	516,0	2,1	515,6	512,6-520,6

Estimation pondérée sur l'ensemble des modèles **N = 516**

Variables **hétérogénéité** : âge, sexe, stade TNM

* M_{avg} = Estimation pondérée sur l'ensemble des modèles obtenue après 5 000 itérations

Estimation de N par M_{tbh}

Captures : âge, sexe et TNM

Modèles	Moyenne	Ecart-type	Médiane	HPDI95%
M_0	517,6	0,0	517,6	517,6-517,6
M_h	700,1	21,1	699,6	659,6-742,6
M_{bh}	696,5	21,8	696,6	654,6-739,6
M_{th}	699,7	21,7	699,6	658,6-743,6
M_{tbh}	695,3	19,9	694,6	657,6-735,6
BMA M_{avg}^*	697,9	20,9	696,6	658,6-740,6

Estimation pondérée sur l'ensemble des modèles **$N = 698$**

BMA → modèles avec hétérogénéité M_h et M_{bh}

Hétérogénéité de capture des cancers colorectaux :
 TN+M+ : 45% RCP, 25% Registre, 20% Dépistés

* M_{avg} = Estimation pondérée sur l'ensemble des modèles obtenue après 5 000 itérations

Conclusion

Modèles M_{tbh} : complémentaires aux modèles log-linéaires

Applicables aux « listes » de cas

Estimations cohérentes

Hétérogénéité de capture individuelle

Variables d'hétérogénéité = épisodes de capture

Application Bayésienne

Adaptée aux distributions particulières / effectifs limités

Incertitude sur les modèles → Bayesian Model Average

Estimation sur tous les modèles vs meilleur modèle retenu

Remerciements :

Dr Bernard GIUSIANO, Pr Jean-Pierre DAURES, Pr Jean-Pierre GERARD, Pr Moïse NAMER, Pr Christian PRADIER
Equipe DSP-CRISAP PACA

