3.1 On considère trois modèles de régression emboîtés afin de modéliser le nombre d'accidents de voiture selon la région (`region`). La variable catégorielle niveaux de risque (`risque`) a 3 niveaux et le nombre d'années d'expérience (`expe`) au volant est un facteur à 4 niveaux.

Modèle	variables	$p+1$	$\ell(\widehat{\boldsymbol{\beta}})$	AIC	BIC
M_1	`risque`	3	-244.566	495.132	510.362
M_2	`risque` + `region`	$\star$	-151.620	$\star$	$\star$
M_3	`risque` + `region` + `expe`	10	-139.734	299.468	350.235

Table 1: Mesures d'adéquation de trois modèles emboîtés avec les covariables, le nombre de coefficients du modèle ($p+1$), la valeur de la log-vraisemblance évaluée au maximum de vraisemblance ($\ell(\widehat{\boldsymbol{\beta}})$) et les critères d'information.

Quelle est la différence entre le AIC et le BIC du modèle M_2 (en valeur absolue)?

3.2 Une variable aléatoire X suit une loi géométrique de paramètre p si sa fonction de masse est

$$\mathsf{P}(X=x) = (1-p)^{x-1}p, \qquad x = 1,2,\ldots$$

(a) Écrivez la vraisemblance et la log-vraisemblance d'un échantillon aléatoire de taille n si les observations sont indépendantes.

(b) Dérivez l'estimateur du maximum de vraisemblance pour le paramètre p.

(c) Calculez l'information observée.

(d) Supposons qu'on a un échantillon de 15 observations, $\{5,6,3,7,1,2,11,8,7,34,1,7,10,1,0\}$, dont la somme est 103. Calculez l'estimé du maximum de vraisemblance et son erreur-type approximative.

(e) Calculez la statistique du rapport de vraisemblance et la statistique de Wald pour un test à niveau 5% de l'hypothèse $\mathscr{H}_0: p_0 = 0.1$ contre l'alternative bilatérale $\mathscr{H}_a: p_0 \neq 0.1$.

On considère un modèle pour le nombre de ventes quotidiennes dans un magasin, supposées indépendantes. Votre gestionnaire vous indique que ce dernier fluctue selon qu'il y ait des soldes ou pas. Vous spécifiez un modèle de Poisson avec fonction de masse

$$\mathsf{P}\left(Y_i = y_i \mid \texttt{soldes}_i\right) = \frac{\exp(-\lambda_i)\lambda_i^{y_i}}{y_i!}, \qquad y_i = 0,1,\ldots$$

et on modélise $\lambda_i = \exp(\beta_0 + \beta_1 \texttt{solde}_i)$, où $\texttt{solde}_i$ est un indicateur binaire qui vaut un lors des soldes et zéro autrement.

(a) Dérivez l'estimateur du maximum de vraisemblance pour (β_0, β_1). *Indice: les estimateurs du maximum de vraisemblance sont invariants aux reparamétrisations. Considérez les deux échantillons solde/hors solde.*

(b) Calculez les estimés si on a un échantillon du nombre de ventes pour 12 jours indépendants, soit $\{2;5;9;3;6;7;11\}$ hors soldes, et $\{12;9;10;9;7\}$ pendant les soldes.

(c) Calculez la matrice d'information observée et utilisez-la pour dériver les erreurs-type de $\widehat{\beta}_0$ et de $\widehat{\beta}_1$ et un intervalle de confiance à niveau 95%.

(d) Votre gérant veut savoir si les profits quotidiens moyens pendant les soldes sont différents des jours ordinaires, sachant que le profit moyen pour les jours hors solde est de 25$ par transaction, mais seulement 20$ pendant les soldes. Faites un test de rapport de vraisemblance (profilée) pour estimer cette hypothèse. *Indice: écrivez l'hypothèse nulle en fonction des paramètres du modèles β_0 et β_1* (**difficile**).