TP1 - Tests d'hypothèses dans le modèle binomial

Dans ce TP, on se propose d'illustrer certains résultats théoriques étudiés en cours et en TD, dans le cadre du modèle binomial.

Dans la suite, on considérera une suite de variables aléatoires $X_1, \ldots, X_n$, indépendantes et de même loi de Bernoulli $\mathcal{B}(p)$, avec p inconnu. On définit $S_n = \sum_{i=1}^{n} X_i$.

De façon générale, on s'intéresse au test d'hypothèses suivant :

$$H_0 : p \in \Theta_0 \quad \text{contre} \quad H_1 : p \in \Theta_1,$$

où Θ_0 et Θ_1 sont deux sous-ensembles disjoints de $[0, 1]$.

Test d'hypothèses simples

Dans cette partie, on s'intéresse au cas d'*hypothèses simples*, c'est-à-dire telles que $\Theta_0 = \{p_0\}$ et $\Theta_1 = \{p_1\}$ sont réduites à un singleton. On écrit alors :

$$H_0 : p = p_0 \quad \text{contre} \quad H_1 : p = p_1,$$

On reprend l'exemple de l'exercice 4 de la fiche de TD 1.

a) Construire la région de rejet pour les niveaux suivants : 10%, 6%, 5%, 2% et 1%. On utilisera la fonction `binom.ppf` qui correspond à la fonction quantile (inverse de la fonction de répartition).

b) Tracer la fonction de répartition de la loi binomiale de paramètres $n = 50$ et $p = p_0$, et représenter graphiquement le seuil de rejet.

c) Calculer la puissance du test pour chaque niveau α défini précédemment.

d) Calculer l'erreur de deuxième espèce.

e) **(bonus)** Tracer sur un même graphe, à l'aide de la fonction `binom.pmf`, la fonction de masse de la loi binomiale de paramètres n et p_0, et la fonction de masse de la loi binomiale de paramètres n et p_1. Représenter graphiquement le niveau du test, et la puissance. On pourra utiliser la fonction `fill_between` (voir code Jupyter Notebook en cas de problème).

f) Tracer la fonction puissance en fonction de n, pour des valeurs de n comprises entre 10 et 200.

Test d'hypothèses composites

On s'intéresse d'abord au test unilatéral suivant :

$$H_0 : p = p_0 \quad \text{contre} \quad H_1 : p > p_0.$$

On a donc $\Theta_0 = \{p_0\}$ et $\Theta_1 =]p_0, 1]$.

On reprend l'exercice 6 sur le lancer d'une pièce de monnaie, en notant p la probabilité pour que la pièce tombe sur pile.

a) On veut tester l'hypothèse nulle selon laquelle la pièce est équilibrée, contre l'hypothèse alternative selon laquelle elle est truquée et renvoie plus de “piles” que de “faces”. Écrire Θ_0 et Θ_1.

b) Tracer la fonction puissance en fonction de p, pour $n = 20$ et $\alpha = 5\%$ (taille du test).

c) Dans cette question, on va simuler plusieurs réalisations de l'expérience aléatoire consistant à lancer n fois une pièce de monnaie. Notons M le nombre d'expériences aléatoires, c'est-à-dire le nombre de fois où on effectue n lancers de pièce, avec $n = 10$. Pour $m = 1, \dots, M$, réaliser les étapes suivantes :

— générer n réalisations de variables aléatoires de Bernoulli de paramètre $p = 0.5$
— calculer la m-ème statistique de test basée sur ces n observations
— compter 1 si on rejette H_0 et 0 sinon

Calculer la proportions d'échantillons ayant conduit au rejet de H_0. Qu'a t-on estimé ?

On s'intéresse maintenant au test défini par (exercice 5 du TD 1) :

$$H_0 : p \leqslant p_0 \quad \text{contre} \quad H_1 : p > p_0$$

d) Tracer la valeur de $\mathbb{P}_p(S_n \geqslant a)$, où a est un seuil de rejet fixé, en fonction de p. On pourra prendre pour a n'importe quelle valeur supérieure à p_0.

Approximation par la loi de Poisson

On s'intéresse dans cette partie à l'approximation de la loi Binomiale par la loi de Poisson. Cette approximation est valable pour une suite de variables aléatoires de loi $\mathcal{B}(n, p_n)$ avec $\lim\limits_{n \to +\infty} np_n = \lambda > 0$. En pratique, elle est donc utilisée pour n grand et p petit.

a) Tracer sur un même graphe la fonction de masse de la loi binomiale de paramètres n et p, et la fonction de masse de la loi de Poisson de paramètres np, en faisant varier n et p. On utilisera les fonctions `binom.pmf` et `poisson.pmf`.

On pourra soit utiliser la fonction graphique `plt.bar`, soit superposer les deux fonctions `plt.plot` avec l'option `'o'` et `plt.vlines`.

b) Tracer la valeur de $\mathbb{P}_p(X \geqslant a)$, où a est un seuil de rejet fixé, en fonction de p, pour X une variables aléatoire de loi de Poisson. À quel exercice de la fiche de TD 1 ce résultat fait-il référence ?

Approximation par la loi normale

On s'intéresse maintenant à l'approximation de la loi binomiale par la loi normale. Cette approximation est valable lorsque n est grand, quelle que soit la valeur de p. Lorsque n est grand et p petit, avec $\lim_{n\to+\infty} np_n = \lambda > 0$, si le λ est grand, la loi de Poisson du paragraphe précédent peut également être approchée par la loi normale. Dans la pratique, on prendra l'approximation par la loi de Poisson si λ est relativement faible (par ex. < 10) et l'approximation par la loi normale si lambda est élevé (par ex. > 10).

a) Générer 10000 réalisations d'une variable aléatoire de loi binomiale de paramètres n et p, pour $n = 1000$ et $p = 0.5$.

b) Tracer l'histogramme des valeurs obtenues. On utilisera la fonction `plt.hist`.

c) Superposer au graphe précédent la densité de la loi normale centrée réduite.

d) Refaire les étapes précédentes en faisant varier n et p.

Tests asymptotiques

Dans cette partie, on va comparer les tests exacts aux tests asymptotiques. Pour cela, on considère le test d'hypothèses composites suivant :

$$H_0 : p = p_0 \quad \text{contre} \quad H_1 : p \neq p_0.$$

a) Quelle est la forme de la région de rejet ?

b) Construire la région de rejet du test exact, et la comparer aux régions de rejet obtenues à l'aide de l'approximation par la loi normale grâce au théorème central limite, et à l'aide de l'approximation par la loi de Poisson

c) Faire varier n et p pour identifier les cas où l'approximation par la loi de Poisson est meilleure que celle par la loi normale.