Énoncé

Soit p et q fixés dans $]0,1[$. Pour tout $n \in \mathbb{N}^*$, on note S_n une somme de n variables aléatoires de Bernoulli de paramètre p et mutuellement indépendantes. L'objet de ce texte[1] est l'inégalité

$$\mathbb{P}\left(\left|\frac{S_n}{n} - q\right| \leq \left|\frac{S_n}{n} - p\right|\right) \leq e^{-n\,\frac{(p-q)^2}{2}}.$$

1. Soit $g : \mathbb{R}_+ \to \mathbb{R}$ la fonction définie par : $\forall x \in \mathbb{R}^*,\ g(x) = \ln(1 - p + pe^x)$.
   a. Montrer que g est bien définie et de classe $\mathcal{C}^2$ sur $\mathbb{R}$. Pour $x \geq 0$, exprimer $g''(x)$ sous la forme $\frac{\alpha\beta}{(\alpha+\beta)^2}$ où α et β sont des réels positifs pouvant dépendre de x.
   b. Montrer que $g''(x) \leq \frac{1}{4}$ pour tous $x \geq 0$.
   c. En utilisant une formule de Taylor à préciser soigneusement, montrer que
   $$\forall x \geq 0,\ \ln(1 - p + pe^x) \leq px + \frac{x^2}{8}.$$
2. Dans cette question, on suppose $p < q$.
   a. Justifier que
   $$\mathbb{P}\left(\left|\frac{S_n}{n} - q\right| \leq \left|\frac{S_n}{n} - p\right|\right) = \mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right).$$
   b. Soit X une variable aléatoire de Bernoulli de paramètre p. Pour $u > 0$, calculer l'espérance de e^{uX}.
   c. Montrer que
   $$\forall u > 0,\ \mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) \leq e^{-n\left(\frac{p+q}{2}\,u - \ln(1-p+pe^u)\right)}.$$
   Indication On pourra utiliser sans démonstration que si $Z_1, \cdots, Z_n$ sont n variables aléatoires mutuellement indépendantes et prenant un nombre fini de valeurs, l'espérance du produit $Z_1 \cdots Z_n$ est le produit des espérances $E(Z_1) \cdots E(Z_n)$.
   d. Montrer que
   $$\mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) \leq e^{-n\,\frac{(p-q)^2}{2}}.$$
3. Montrer l'inégalité objet de ce texte.

[1] d'après X-normales 2019 PC

Corrigé

1. a. La fonction g est le résultat d'opérations et de composition de fonctions usuelles $\mathcal{C}^\infty$ dans leurs domaines. Le seul point qui mérite d'être détaillé est
   $$1 - p + pe^x > 1 - p > 0.$$
   b. Le calcul des dérivées conduit à
   $$g''(x) = \frac{(1-p)(pe^x)}{\left((1-p) + pe^x\right)^2} \text{ avec } \alpha = 1 - p \text{ et } \beta = pe^x.$$
   On en déduit $g(x) \leq \frac{1}{4}$ car
   $$\frac{\alpha\beta}{(\alpha+\beta)^2} = \frac{1}{4}\,\frac{(\alpha+\beta)^2 - (\alpha-\beta)^2}{(\alpha+\beta)^2} \leq \frac{1}{4}.$$
   c. On applique à g la formule de Taylor avec reste intégral à l'ordre 1 :
   $$g(x) = g(0) + xg'(0) + \int_0^x \frac{(x-t)}{1} g''(t)\,dt \leq px + \int_0^x \frac{(x-t)}{4}\,dt \leq px + \frac{x^2}{8}.$$
   car $g'(0) = p$, $x - t \geq 0$, $g''(t) \leq \frac{1}{4}$ et
   $$\int_0^x \frac{(x-t)}{4}\,dt = -\frac{1}{8}\left[(x-t)^2\right]_0^x = \frac{x^2}{8}.$$
2. Dans cette question, on suppose $p < q$.
   a. Par définition, $\frac{S_n}{n}$ prend ses valeurs dans $[0,1]$. Les équivalences suivantes sont valables pour tout $s \in [0,1]$.
   $$|s - q| \leq |s - p| \Leftrightarrow (s-q)^2 \leq (s-p)^2 \Leftrightarrow 0 \leq 2(q-p)s + p^2 - q^2$$
   $$\Leftrightarrow s \geq \frac{q^2 - p^2}{2(q-p)} = \frac{p+q}{2} \text{ car } q - p > 0.$$
   On en déduit l'égalité entre événements, donc l'égalité entre les probabilités de ces événements.
   b. Soit X une variable de Bernoulli de paramètre p et $u > 0$. La variable e^{uX} prend les valeurs 1 et e^u avec les probabilités $1 - p$ et p. Son espérance est
   $$E(e^{uX}) = (1-p) + pe^u = e^{g(u)}$$
   avec la fonction g introduite au début de l'énoncé.

c. On multiplie par $u > 0$ et on compose par l'exponentielle qui est strictement croissante pour pouvoir utiliser la question précédente et l'inégalité de Markov.

$$\left(S_n \geq \frac{p+q}{2}\,n\right) = \left(e^{uS_n} \geq e^{u\frac{p+q}{2}\,n}\right)$$
$$\Rightarrow \mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) = \mathbb{P}\left(e^{uS_n} \geq e^{un\frac{p+q}{2}}\right) \leq \frac{E(e^{uS_n})}{e^{un\frac{p+q}{2}}}$$

Comme S_n est la somme $X_1 + \cdots + X_n$ de n variables de Bernoulli, e^{uS_n} est le produit des variables e^{uX_i} mutuellement indépendantes et dont l'espérance a été calculée. Avec l'indication de l'énoncé, on peut écrire :

$$E(e^{uS_n}) = \prod_{i=1}^{n} E(e^{uX_i}) = e^{ng(u)}$$
$$\Rightarrow \mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) \leq e^{ng(u)-un\frac{p+q}{2}} = e^{-n\left(\frac{p+q}{2}\,u-\ln(1-p+pe^u)\right)}.$$

d. L'inégalité précédente est valable pour tous les u. Pour touver le meilleur u, il faut chercher le minimum de

$$u \mapsto -n\left(\frac{p+q}{2}\,u - \ln(1-p+pe^u)\right)$$

Ce n'est pas impossible mais les calculs sont très lourds. On peut trouver l'inégalité de l'énoncé en utilisant l'inégalité de la question 1.c.. On écrit :

$$\mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) \leq e^{n(pu+\frac{u^2}{8})-un\frac{p+q}{2}}$$

avec $n(pu + \frac{u^2}{8}) - un\frac{p+q}{2} = n\left(-\frac{q-p}{2}\,u + \frac{u^2}{8}\right)$. La fonction en u atteint son minimum en $2(q-p)$. Ce minimum est

$$-(q-p)^2 + \frac{1}{2}(q-p)^2 = -\frac{(q-p)^2}{2} \text{ donc } \mathbb{P}\left(S_n \geq \frac{p+q}{2}\,n\right) \leq e^{-n\,\frac{(q-p)^2}{2}}.$$

3. Si $p = q$, l'inégalité à démontrer est évidente. Il reste à la démontrer dans le cas où $q < p$. Dans ce cas, en reprenant le raisonnement de 2.a :

$$\mathbb{P}\left(\left|\frac{S_n}{n} - q\right| \geq \left|\frac{S_n}{n} - p\right|\right) = \mathbb{P}\left(S_n \leq \frac{p+q}{2}\,n\right).$$

Considérons les variables « contraires ». C'est à dire les $X'_i = 1 - X_i$. Elles sont encore de Bernoulli mais de paramètre $p' = 1-p$ avec $p' = 1-p < 1-q = q'$. Notons $S'_n = X'_1 + \cdots + X'_n = n - S_n$. Alors

$$\left(S_n \leq \frac{p+q}{2}\,n\right) = \left(n - S'_n \leq \frac{p+q}{2}\,n\right) = \left(S'_n \geq n - \frac{p+q}{2}\,n\right) = \left(S'_n \geq \frac{p'+q'}{2}\,n\right)$$

On est ramené au cas précédent :

$$\mathbb{P}\left(S_n \leq \frac{p+q}{2}\,n\right) \leq e^{-n\,\frac{(q'-p')^2}{2}} = e^{-n\,\frac{(q-p)^2}{2}} \text{ car } q'-p' = q-p.$$