Corrigé HEC II 2005 par Pierre veuillez
L'objet du problème est d'étudier quelques propriétés d'un estimateur du paramètre p d'une loi géométrique.

Partie I. Formule du binôme négatif.

Pour tout couple (n,r) d'entiers naturels tels que $0 \leqslant r \leqslant n$, on rappelle la formule du " triangle de Pascal" : $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$

1. pour tout entier r de $[\![1,n]\!]$, on a : $\binom{n}{r} = \sum\limits_{k=r}^{n} \binom{k-1}{r-1}$ par récurrence sur n avec $\sum\limits_{k=r}^{n+1} \binom{k-1}{r-1} = \sum\limits_{k=r}^{n} \binom{k-1}{r-1} + \binom{n}{r-1} = \binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}$

2. Soit (n,r) un couple d'entiers naturels, tels que $1 \leqslant r \leqslant n$. Pour tout réel x de $]0,1[$, on définit la fonction $f_{r,n}$ par : $f_{r,n}(x) =$.

   a) Pour tout réel x de $]0;1[$, l'égalité :

$$\begin{aligned}
(1-x)f_{r,n}(x) &= (1-x)\sum_{k=r}^{n}\binom{k}{r}x^k \\
&= \sum_{k=r}^{n}\binom{k}{r}x^k - \sum_{k=r}^{n}\binom{k}{r}x^{k+1} \\
&= \sum_{k=r}^{n}\binom{k}{r}x^k - \sum_{h=r+1}^{n+1}\binom{h-1}{r}x^k \\
&= \binom{r}{r}x^r + \sum_{k=r+1}^{n}\left[\binom{k}{r} - \binom{k-1}{r}\right]x^k - \binom{n}{r}x^{n+1} \\
&= x^r + \sum_{k=r}^{n}\binom{k-1}{r-1}x^k - \binom{n}{r}x^{n+1} \\
&= xx^{r-1} + \sum_{h=r}^{n-1}\binom{h}{r-1}x^{h+1} - \binom{n}{r}x^{n+1} \\
&= x\sum_{h=r-1}^{n-1}\binom{h}{r-1}x^h - \binom{n}{r}x^{n+1} \\
&= xf_{r-1,n-1}(x) - \binom{n}{r}x^{n+1}
\end{aligned}$$

   Conclusion : $\boxed{(1-x)f_{r,n}(x) = xf_{r-1,n-1}(x) - \binom{n}{r}x^{n+1}}$

   b) On suppose l'entier r fixé. On a

$$\begin{aligned}
\binom{n}{r} &= \frac{n(n-1)\cdots(n-r+1)}{r!} \\
&= \frac{n^r\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{r-1}{n}\right)}{r!} \\
&\sim \frac{n^r}{r!}
\end{aligned}$$

   Conclusion : $\boxed{\binom{n}{r} \sim \frac{n^r}{r!} \text{ quand } n \to +\infty}$

3. Soit x un réel fixé de $]0;1[$ et soit r un entier naturel fixé. On veut établir l'existence de la limite de $f_{r,n}(x)$ lorsque n tend vers $+\infty$, et déterminer la valeur de cette limite.

   a) $f_{0,n}(x) = \sum_{k=0}^{n} \binom{k}{0} x^k = \sum_{k=0}^{n} x^k \to \frac{1}{1-x}$ car $|x| < 1$ et $f_{1,n}(x). = \sum_{k=1}^{n} \binom{k}{1} x^k = \sum_{k=1}^{n} kx^k \to \frac{x}{(1-x)^2}$ de même (séries usuelles)

   b) Soit r un entier naturel non nul. On suppose que, pour tout réel x de $]0,1[$, on a :

$$\lim_{n\to+\infty} f_{r-1,n}(x) = \frac{x^{r-1}}{(1-x)^r}.$$

   On a précédemment :

$$\begin{aligned} f_{r,n}(x) &= \frac{1}{1-x}\left[x f_{r-1,n-1}(x) - \binom{n}{r} x^{n+1}\right] \\ &= \frac{x}{1-x} f_{r-1,n-1}(x) - \binom{n}{r}\frac{x^{n+1}}{1-x} \end{aligned}$$

   et comme $\binom{n}{r} \sim \frac{n^r}{r!}$ alors $\binom{n}{r}\frac{x^{n+1}}{1-x} \sim \frac{n^r}{r!} x^{n+1} \to 0$ car $n^r = o\left(\frac{1}{x}^{n+1}\right)$ car $\frac{1}{x} > 1$

   Conclusion : $\boxed{\lim_{n\to+\infty} f_{r,n}(x) = \frac{x^r}{(1-x)^{r+1}}}$.

   Ainsi, par récurrence on aura $\sum_{k=r}^{+\infty} \binom{k}{r} x^k = \frac{x^r}{(1-x)^{r+1}}$.

Partie II. Développement en série de $\ln(1-x)$

Soit x un réel de $]0,1[$.

1. On a $\sum_{k=1}^{n} t^{k-1} = \frac{1-t^n}{1-t}$ pour tout t de $[0,1[$ et en intégrant : $\int_0^x \frac{1-t^n}{1-t}\,dt = \sum_{k=1}^{n} \int_0^x t^{k-1}dt$

   Conclusion : $\boxed{\int_0^x \frac{1-t^n}{1-t}\,dt = \sum_{k=1}^{n} \frac{x^k}{k}}$

2. On a $0 \le \frac{t^n}{1-t} \le \frac{t^n}{1-x}$ sur $[0,x]$ et donc $\cdots$ $(0 \le x\ )$ *Conclusion :* $\boxed{\lim_{n\to+\infty} \int_0^x \frac{t^n}{1-t}dt = 0.}$

3. $\int_0^x \frac{1-t^n}{1-t}\,dt = \int_0^x \frac{1}{1-t}dt - \int_0^x \frac{t^n}{1-t}dt = -\ln(1-x) - \int_0^x \frac{t^n}{1-t}dt \to -\ln(1-x)$

   donc $\sum_{k=1}^{n} \frac{x^k}{k} \to -\ln(1-x)$ quand $n \to +\infty$

   Conclusion : $\boxed{\sum_{k=1}^{+\infty} \frac{x^k}{k} = -\ln(1-x).}$

Partie III. Loi binomiale négative.

Toutes les variables aléatoires qui interviennent dans cette partie sont définies sur un même espace probabilisé $(\Omega; A; P)$.
Soit p un réel de $]0,1[$. On pose $q = 1-p$, et on considère une variable aléatoire X à valeurs dans $\mathbb{N}^{\times}$, qui suit la loi géométrique de paramètre p. On rappelle que pour tout entier k de $\mathbb{N}^{\times}$, $P(X=k) = pq^{k-1}$.

1. On a $E(X) = \frac{1}{p}$ et $V(X) = \frac{1-p}{p^2}$

2. On considère la variable aléatoire Y définie par $Y = \dfrac{1}{X}$.

   a) La loi de Y est donnée par : $Y(\Omega) = \left\{\frac{1}{k}/k \in \mathbb{N}^*\right\}$ et pour tout $k \in \mathbb{N}^*$, $\mathrm{P}\left(Y = \frac{1}{k}\right) = \mathrm{P}(X=k) = pq^{k-1}$

   b) Comme Y est à valeurs positives, l'absolue convergence équivaut à la convergence simple

$$\sum_{k=1}^{N} \frac{1}{k}\mathrm{P}\left(Y = \frac{1}{k}\right) = \sum_{k=1}^{N} \frac{1}{k} pq^{k-1} = \frac{p}{q}\sum_{k=1}^{N} \frac{1}{k} q^k \to -\frac{p}{q}\ln(1-q)$$

   d'après la partie II. Donc la série est absolument convergente, Y a une espérance et
   Conclusion : $\boxed{E(Y) = \frac{p}{q}\ln(1-q)}$

   c) De même,par le théorème de transfert : avec $\left(\frac{1}{k}\right)^i \mathrm{P}\left(Y = \frac{1}{k}\right) \leq \left(\frac{1}{k}\right)\mathrm{P}\left(Y = \frac{1}{k}\right)$ la série $\sum_{k\geq 1}\left(\frac{1}{k}\right)^i \mathrm{P}\left(Y = \frac{1}{k}\right)$ est absolument convergente par majoration de termes positifs.
   Conclusion : $\boxed{E(Y^i) \text{ existe donc pour tout } i \geq 1}$

3. Soit X_1 et X_2 deux variables aléatoires indépendantes, à valeurs dans $\mathbb{N}^{\times}$, qui suivent la même loi géométrique de paramètre p. On pose : $S_1 = X_1$, $S_2 = X_1 + X_2$, $Y_2 = \dfrac{2}{S_2}$.

   a) On a $S_2(\Omega) = [[2, +\infty[[$ avec$(S_2 = n) = \bigcup_{k=1}^{n-1}(X = k \cap X_2 = n-k)$ et $\mathrm{P}(S_2 = n) = (n-1)p^2q^{n-2}$ et $Y_2(\Omega) = \left\{\frac{1}{n}/n \in \mathbb{N}, k \geq 2\right\}$ et
   Conclusion : $\boxed{\mathrm{P}\left(Y_2 = \frac{1}{n}\right) = (n-1)p^2q^{n-2}}$

   b) On a $\frac{1}{n}\mathrm{P}\left(Y_2 = \frac{1}{n}\right) = \frac{1}{n}(n-1)p^2q^{n-2} = \left(1 - \frac{1}{n}\right)p^2q^{n-2} \sim \frac{p^2}{q^2}q^n\ ^n$ et comme la série des q^n converge, par équivalence de termes positifs, Y_2 a bien une espérance

   c) On transforme alors la somme partielle :

$$\begin{aligned}
\sum_{n=2}^{N}\frac{1}{n}\mathrm{P}\left(Y_2 = \frac{1}{n}\right) &= \sum_{n=2}^{N}\frac{1}{n}(n-1)p^2q^{n-2} \\
&= \sum_{n=2}^{N} p^2q^{n-2} - \sum_{n=2}^{N}\frac{1}{n}p^2q^{n-2} \\
&= p^2\left(\sum_{m=0}^{N-2} q^m - \frac{1}{q^2}\sum_{n=1}^{N}\frac{1}{n}q^n + \frac{1}{q}\right) \\
&\to p^2\left(\frac{1}{1-q} + \frac{1}{q^2}\ln(1-q) + \frac{1}{q}\right)
\end{aligned}$$

   Conclusion : $\boxed{E(Y_2) = p^2\left(\frac{1}{p} + \frac{1}{q^2}\ln(p) + \frac{1}{q}\right)}$

4. On considère une suite $(X_n)_{n\in\mathbb{N}^\times}$ de variables aléatoires à valeurs dans $\mathbb{N}^\times$, indépendantes, de même loi géométrique de paramètre p. Pour tout entier n de $\mathbb{N}^\times$, on pose $S_n = \displaystyle\sum_{k=1}^{n} X_k$.

   a) Comme les X_i ont une espérance, $E(S_n) = \sum_{k=1}^{n} E(X_k) = \frac{n}{p}$ et comme elles sont indépendantes, $V(S_n) = \sum V(X_k) = \frac{nq}{p^2}$

   b) Récurrence :

   Pour $n = 1$ on a vu que $S_2(\Omega) = [[2, +\infty[[$ et $\mathrm{P}(S_2 = s) = (s-1)\,p^2q^{s-2} = \binom{s-1}{2-1}p^2q^{s-2}$ donc la propriété est vraie pour $n = 2$ (ou pour $n = 1$ avec $S_1 = X_1$)

   Soit $n \in \mathbb{N}$ tel que la loi de S_n est donnée par $S_n(\Omega) = [[n, +\infty[[$ et pour tout entier $s \geq n$: $P(S_n = s) = \binom{s-1}{n-1}\,p^nq^{s-n}$

   alors

$$\begin{aligned}
(S_{n+1} = s) &= (S_n + X_{n+1} = s) = \bigcup_{k=n}^{s-1}(S_n = k \cap X_{n+1} = s-k)\,\text{incompatibles} \\
\mathrm{P}(S_{n+1} = s) &= \sum_{k=n}^{s-1}\mathrm{P}(S_n = k)\,\mathrm{P}(X_{n+1} = s-k)\,\text{par indépendance} \\
&= \sum_{k=n}^{s-1}\binom{k-1}{n-1}p^nq^{k-n}pq^{s-k-1} \\
&= p^{n+1}q^{s-n-1}\sum_{h=n}^{s-1}\binom{k-1}{n-1} \\
&= p^{n+1}q^{s-(n+1)}\binom{s-1}{n}\ \text{première formule}
\end{aligned}$$

   Conclusion : $\boxed{\text{la formule est vraie pour tout entier } n}$

5. Pour tout entier n de $\mathbb{N}^\times$, on pose $Y_n = \dfrac{n}{S_n}$.

   a) On a $Y_n(\Omega) = \left\{\frac{n}{s} \quad / \quad s \geq n \text{ et } s \text{ entier}\right\}$ et $\mathrm{P}\left(Y_n = \frac{n}{s}\right) = \mathrm{P}(S_n = s) = \binom{s-1}{n-1}\,p^nq^{s-n}$

   b) Soit t un réel quelconque de $[0, 1[$ et m de $\mathbb{Z}$,

   On découpe s^mt^s pour garder un morceau de série convergente : $s^mt^s = s^m\sqrt{t}^s\sqrt{t}^s$ avec $s^m\sqrt{t}^s = s^m/\sqrt{1/t}^s \to 0$ car $s^m = o\left(\sqrt{1/t}^s\right)$ car $\sqrt{1/t} > 1$ et donc $s^mt^s = 0\left(\sqrt{t}^s\right)$

   De plus la série géométrique de raison $\sqrt{t}^s$ converge donc par majoration de termes positifs, $\left(\displaystyle\sum_{s\geqslant 1} s^mt^s\right)$, est convergente..

   On a $\frac{n}{s}\mathrm{P}\left(Y_n = \frac{n}{s}\right) = \frac{n}{s}\binom{s-1}{n-1}p^nq^{s-n} \sim \frac{n}{s}\frac{(s-1)^{n-1}}{(n-1)!}p^nq^{s-n} \sim \frac{n}{(n-1)!}\frac{p^n}{q^n}s^{n-2}q^s$ avec $\frac{n}{(n-1)!}p^nq^{-n}$ constante par rapport à s. et la série de terme général $s^{n-2}q^s$ convergente, donc par équivalence de termes positifs, la série de terme $\sum_{s\geq n}\frac{n}{s}\mathrm{P}\left(Y_n = \frac{n}{s}\right)$ est absolument convergente et Y_n a une espérance.

   De même $\left(\frac{n}{s}\right)^2\mathrm{P}\left(Y_n = \frac{n}{s}\right) \sim \frac{n^2}{(n-1)!}\frac{p^n}{q^n}s^{n-3}q^s$ donnera l'existence de l'espérance de Y_n^2 et la variance de Y_n

   Plus simplement : $\frac{n}{s}\mathrm{P}\left(Y_n = \frac{n}{s}\right) \leq \mathrm{P}\left(Y_n = \frac{n}{s}\right)$ pour $s \geq n$, et la série est donc convergente par majoration de termes positifs.

Partie IV. Une estimation ponctuelle du paramètre p.

Soit p un réel de $]0,1[$. Dans cette partie, on considère une variable aléatoire réelle X à valeurs dans $\mathbb{N}^\times$, qui suit une loi géométrique de paramètre p inconnu. On pose $q = 1-p$.
Pour tout entier naturel non nul n, on considère un n-échantillon $(X_1, X_2, ..., X_n)$ de variables aléatoires à valeurs dans $\mathbb{N}^\times$, indépendantes, de même loi que X.
Les variables aléatoires X; X_1; X_2;....; X_n sont définies sur un même espace probabilisé $(\Omega; A; P)$.
On pose $\overline{X}_n = \dfrac{S_n}{n} = \dfrac{1}{Y_n}$.

1. $E\left(\overline{X_n}\right) = \frac{1}{n}E\left(S_n\right) = \frac{1}{p}$ donc $\overline{X}_n$ est un estimateur sans biais pour le paramètre $\dfrac{1}{p}$.
   Son risque quadratique est $r = b^2 + V\left(\overline{X_n}\right) = V\left(\overline{X_n}\right) = \frac{1}{n^2}V\left(S_n\right) = \frac{1}{n}\frac{q}{p}$

2. Pour tout entier n de $\mathbb{N}^\times$, on note h_n et φ_n les applications définies sur $[0,1[$ à valeurs dans $\mathbb{R}$ telles que :

$$\forall t \in [0,1[, \quad h_n(t) = \sum_{s=n}^{+\infty} \frac{1}{s}\binom{s-1}{n-1}t^s \quad , \quad \varphi_n(t) = \sum_{s=n}^{+\infty} \frac{1}{s^2}\binom{s-1}{n-1}t^s$$

   On admet dans toute la suite du problème, que $h_n$est de classe C^1 et que pour tout réel t de $[0,1[$, la dérivée h'_n de $h_n$vérifie : $h'_n(t) = \sum\limits_{s=n}^{+\infty} \binom{s-1}{n-1}t^{s-1}$.
   On admet également que la fonction $\varphi_n$est dérivable sur $]0;1[$, de dérivée φ'_n, et que pour tout t de $]0,1[$, $\varphi'_n(t) = \dfrac{1}{t}h_n(t)$.

   a) On sait déjà que Y_n a une espérance et

$$\begin{aligned} E\left(Y_n\right) &= \sum_{s=n}^{+\infty} \frac{n}{s}\mathrm{P}\left(Y_n = \frac{n}{s}\right) = \sum_{s=n}^{+\infty} \frac{n}{s}\tbinom{s-1}{n-1}p^n q^{s-n} \\ &= n\left(\frac{p}{q}\right)^n \sum_{s=n}^{+\infty} \frac{1}{s}\tbinom{s-1}{n-1}q^s \\ &= n\left(\frac{p}{q}\right)^n h_n(q) \end{aligned}$$

   h'_n est continue sur $[0,q]$ donc $h_n\left(q\right) = \int_0^q h'_n\left(t\right)dt$. Et comme

$$\begin{aligned} h'_n(t) &= \sum_{s=n}^{+\infty}\binom{s-1}{n-1}t^{s-1} = \sum_{i=n-1}^{+\infty}\binom{i}{n-1}t^i \\ &= \frac{t^{n-1}}{(1-t)^n} \text{ d'après I.3.b)} \end{aligned}$$

   Conclusion : $\boxed{\text{pour tout } q \text{ de } [0,1[, : h_n(q) = \int_0^q \frac{t^{n-1}}{(1-t)^n}dt.}$

   b) Avec $t = 0 \Longleftrightarrow y = 0$ et $t = q \Longleftrightarrow y = p/q$, on remet le changement de variable dans le sens simple :

$$y=\frac{t}{1-t}\Longleftrightarrow(1-t)\,y=t\Longleftrightarrow t=\frac{y}{1+y}\text{ et }dt=\frac{1}{(1+y)^2}dy$$

avec t fonction C^1 de y sur $[0,q/p]$ et $t\to\dfrac{t^{n-1}}{(1-t)^n}$ est continue sur l'intervalle image $[0,q]$ donc

$$\begin{aligned}h_n(q)&=\int_0^q\frac{t^n}{(1-t)^n}t^{-1}dt\\&=.\int_0^{q/p}y^n\frac{1+y}{y}\frac{1}{(1+y)^2}dy=\int_0^{q/p}\frac{y^{n-1}}{1+y}dy\end{aligned}$$

Conclusion : $\boxed{h_n(q)=\displaystyle\int_0^{q/p}\frac{y^{n-1}}{1+y}dy}$

Par intégration par parties, avec u et v C^1 : $u(y)=\dfrac{1}{1+y}$: $u'(y)=\dfrac{-1}{(1+y)^2}$ et $v'(y)=y^{n-1}$: $v(y)=\dfrac{1}{n}y^n$ on a :

$$\begin{aligned}h_n(q)&=\left[\frac{1}{1+y}\frac{1}{n}y^n\right]_0^{q/p}-\int_0^{q/p}\frac{-1}{(1+y)^2}\frac{1}{n}y^ndy\\&=\frac{1}{n}\frac{q^n/p^n}{1+\frac{q}{p}}+\frac{1}{n}\int_0^{q/p}\frac{y^n}{(1+y)^2}dy\text{ avec }p+q=1\\&=\frac{q^n}{n\,p^{n-1}}+\frac{1}{n}\int_0^{q/p}\frac{y^n}{(1+y)^2}dy.\end{aligned}$$

Conclusion : $\boxed{h_n(q)=\dfrac{q^n}{n\,p^{n-1}}+\dfrac{1}{n}\displaystyle\int_0^{q/p}\frac{y^n}{(1+y)^2}dy}$

3. Pour tout entier n de $\mathbb{N}^\times$, soit $b_n$la fonction définie sur $]0,1[$ à valeurs réelles qui, à tout réel p de $]0,1[$ associe $b_n(p)=E(Y_n)-p$. ($b_n$représente le biais de Y_n pour estimer p)

a) On a :

$$\begin{aligned}
b_n(p) &= E(Y_n) - p \\
&= n\left(\frac{p}{q}\right)^n h_n(q) - p \\
&= n\left(\frac{p}{q}\right)^n \left[\frac{q^n}{n\,p^{n-1}} + \frac{1}{n}\int_0^{q/p} \frac{y^n}{(1+y)^2}dy\right] - p \\
&= p + \left(\frac{p}{q}\right)^n \int_0^{q/p} \frac{y^n}{(1+y)^2}dy - p \\
&= \left(\frac{p}{q}\right)^n \int_0^{q/p} \frac{y^n}{(1+y)^2}dy
\end{aligned}$$

b) On encadre l'intégrale via son contenu, en conservant la puissance qui provoquera (on l'espère) la convergence :

Pour tout $y \in [0, q/p]$ on a $(1+y)^2 > 1$ donc $\dfrac{1}{(1+y)^2} \leq 1$ et $\dfrac{y^n}{(1+y)^2} \leq y^n$ et comme les bornes $0 \leq p/q$ on a alors

$$\begin{aligned}
\int_0^{q/p} \frac{y^n}{(1+y)^2}dy &\leq \int_0^{q/p} y^n dy = \frac{1}{n+1}\left(\frac{q}{p}\right)^{n+1} \quad \text{et donc} \\
\left(\frac{p}{q}\right)^n \int_0^{q/p} \frac{y^n}{(1+y)^2}dy &\leq \frac{1}{n+1}\frac{q}{p}
\end{aligned}$$

finalement $0 \leq b_n(p) \leq \dfrac{1}{n+1}\dfrac{q}{p} \to 0$ et par encadrement $b_n(p) \to 0$ quand $n \to +\infty$. Et comme $E(Y_n) = b_n(p) + p$

Conclusion : $\boxed{\lim_{n\to+\infty} E(Y_n) = p}$

c) On réintègre $\displaystyle\int_0^{q/p} \frac{y^n}{(1+y)^2}dy$ par parties avec $u(y) = \dfrac{1}{(1+y)^2}$: $u'(y) = \dfrac{-2}{(1+y)^3}$ et $v'(y) = y^n$: $v(y) = \dfrac{1}{n+1}y^{n+1}$et u et v C^1 et on trouve :

$$\begin{aligned}
\int_0^{q/p} \frac{y^n}{(1+y)^2}dy &= \left[\frac{1}{(1+y)^2}\frac{1}{n+1}y^{n+1}\right]_0^{q/p} - \int_0^{q/p} \frac{-2}{(1+y)^3}\frac{1}{n+1}y^{n+1}dy \\
&= \frac{1}{n+1}\frac{q^{n+1}/p^{n+1}}{\left(1+\frac{q}{p}\right)^2} + \frac{2}{n+1}\int_0^{q/p} \frac{y^{n+1}}{(1+y)^3}dy \text{ avec } p+q=1 \\
&= \frac{1}{n+1}\frac{q^{n+1}}{p^{n-1}} + \frac{2}{n+1}\int_0^{q/p} \frac{y^{n+1}}{(1+y)^3}dy
\end{aligned}$$

et donc

$$\begin{aligned}
b_n(p) &= \left(\frac{p}{q}\right)^n \int_0^{q/p} \frac{y^n}{(1+y)^2}dy \\
&= \frac{pq}{n+1} + \frac{2}{n+1}\left(\frac{p}{q}\right)^n \int_0^{q/p} \frac{y^{n+1}}{(1+y)^2}dy
\end{aligned}$$

Comme précédemment $\left(\frac{p}{q}\right)^n \int\limits_0^{q/p} \frac{y^{n+1}}{(1+y)^2}dy \leq \frac{1}{n+2}\frac{q}{p}$ et tend vers 0 quand $n \to +\infty$ donc

$$\begin{aligned}
b_n(p) &= \frac{1}{n+1}\frac{p}{q} + \frac{2}{n+1}o(1) \\
&= \frac{pq}{n} + o\left(\frac{1}{n}\right)
\end{aligned}$$

en calculant

$$\begin{aligned}
b_n(p) - \frac{pq}{n} &= \frac{1}{n}\left(pq\left(\frac{n+1}{n} - 1\right) + o(1)\right) \\
&= \frac{1}{n}o(1) = o\left(\frac{1}{n}\right)
\end{aligned}$$

Partie V. Limite de la variance de Y_n.

Le contexte de cette partie est identique à celui de la partie IV.

1. a) On a vu que

$$\begin{aligned}
h_n(t) &= \frac{t^n}{n\,(1-t)^{n-1}} + \frac{1}{n}\int_0^{t/1-t} \frac{y^n}{(1+y)^2}dy \\
&= \frac{t^n}{n}\left[\frac{1}{(1-t)^{n-1}} + \frac{1}{n}\frac{1}{t^n}\int_0^{t/1-t} \frac{y^n}{(1+y)^2}dy\right]
\end{aligned}$$

et comme

$$0 \leq \frac{1}{t^n}\int_0^{t/1-t} \frac{y^n}{(1+y)^2}dy \leq \frac{t}{(n+1)(1-t)^{n+1}} \to 0 \text{ quand } t \to 0$$

alors la quantité entre [] $\to 1$ quand $t \to 0$ et $h_n(t) \sim \frac{t^n}{n}$ lorsque t tend vers 0 par valeurs supérieures.

b) Comme $\frac{h_n(t)}{t} \sim \frac{t^{n-1}}{n}$ dont l'intégrale converge en 0, ($n-1 \geq 0$) par équivalence de fonctions positives l'intégrale $\int\limits_0^q \frac{h_n(t)}{t}dt$ est convergente.

Enfin $\varphi_n'(t) = \frac{1}{t}h_n(t)$ pour tout $t \in]0,1[$ donc et que $\varphi_n(q) = \varphi_n(0) + \int\limits_0^q \frac{h_n(t)}{t}dt.$ et

Conclusion : $\boxed{\varphi_n(q) = \int\limits_0^q \frac{h_n(t)}{t}dt.}$

2. a) Pour tout entier n de $\mathbb{N}^{\times}$, et pour tout réel t de $]0;1[$, on pose $H_n(t) = \int\limits_0^{t/(1-t)} \frac{y^n}{(1+y)^2}dy$.

On a, pour tout t de $]0,1[$,

$$\begin{aligned}\varphi_n'(t) &= \frac{1}{t}h_n(t)\\ &= \frac{1}{t}\left[\frac{t^n}{n\,(1-t)^{\,n-1}} + \frac{1}{n}\int\limits_0^{t/1-t}\frac{y^n}{(1+y)^2}dy\right]\\ &= \frac{1}{n}\left[\frac{t^{n-1}}{n\,(1-t)^{\,n-1}} + \frac{1}{t}\int\limits_0^{t/1-t}\frac{y^n}{(1+y)^2}dy\right]\end{aligned}$$

Conclusion : $\boxed{\text{on a } \varphi_n'(t) = \frac{1}{n}\left(\frac{t^{n-1}}{(1-t)^{n-1}} + \frac{H_n(t)}{t}\right)}$

b) On exprime $\frac{H_n(t)}{t} = n\varphi_n'(t) - \frac{t^{n-1}}{(1-t)^{n-1}}$ dont l'intégrale converge en 0 donc $\int\limits_0^q \frac{H_n(t)}{t}dt$ converge.

D'où

$$\varphi_n(q) = \int\limits_0^q \varphi_n'(t)dt = \int\limits_0^q \frac{1}{n}\left(\frac{t^{n-1}}{(1-t)^{n-1}} + \frac{H_n(t)}{t}\right)dt$$

Conclusion : $\boxed{\varphi_n(q) = \frac{1}{n}\left(\int\limits_0^q \frac{t^{n-1}}{(1-t)^{n-1}}\ dt + \int\limits_0^q \frac{H_n(t)}{t}\ dt\right)}$

3. a) On encadre le contenu qui est lui même une intégrale :

Pour tout $y \in \left[1, \frac{t}{1-t}\right]$: $\frac{y^n}{(1+y)^2} \leq y^n$ donc

$H_n\left(t\right) = \int\limits_0^{t/(1-t)} \frac{y^n}{(1+y)^2}dy \leq \int\limits_0^{t/(1-t)} y^n dy = \frac{1}{n+1}\frac{t^{n+1}}{(1-t)^{n+1}}$ et donc

$0 \leq \frac{H_n(t)}{t} \leq \frac{1}{n+1}\frac{t^n}{(1-t)^{n+1}}$ d'où enfin,

$$\int\limits_0^q \frac{H_n(t)}{t}dt \leq \int_0^q \frac{1}{n+1}\frac{t^n}{(1-t)^{n+1}}dt = \frac{h_{n+1}(q)}{n+1}$$

d'après l'écriture IV 2.a) $h_n(q) = \int\limits_0^q \frac{t^{n-1}}{(1-t)^n}dt.$

Conclusion : $\boxed{0 \leqslant \int\limits_0^q \frac{H_n(t)}{t}dt \leqslant \frac{h_{n+1}(q)}{n+1}}$ et $n\left(\frac{p}{q}\right)^n \int\limits_0^q \frac{H_n(t)}{t}dt \leq n\left(\frac{p}{q}\right)^n \frac{h_{n+1}(q)}{n+1}$

L'écriture IV 2.b) $h_{n+1}(q) = \int\limits_0^{q/p} \frac{y^n}{1+y}dy$ nous donne $h_{n+1}\left(q\right) \leq \frac{1}{n+1}\left(\frac{q}{p}\right)^{n+1}$

d'où $0 \leq n\left(\frac{p}{q}\right)^n \frac{h_{n+1}(q)}{n+1} \leq \frac{n}{n+1}\frac{q}{p}\frac{1}{n+1} \to 0$ quand $n \to +\infty$.

Et par encadrement *Conclusion :* $\boxed{\lim\limits_{n\to+\infty} n\left(\dfrac{p}{q}\right)^n \int_0^q \dfrac{H_n(t)}{t}dt = 0.}$

b) On effectue le même changement de variable que précédemment :
$t = \dfrac{y}{1+y}$ et $dt = \dfrac{1}{(1+y)^2}dy$ et aux bornes $t = 0 \iff y = 0$ et $t = q \iff y = q/p$

$$\begin{aligned}\int_0^q \left(\frac{t}{1-t}\right)^{n-1} dt &= \int_0^{q/p} \left(\frac{\frac{y}{1+y}}{1-\frac{y}{1+y}}\right)^{n-1} \frac{1}{(1+y)^2}dy \\ &= \int_0^{q/p} \frac{y^{n-1}}{(1+y)^2}dy\end{aligned}$$

que l'on intègre par parties avec
$u'(y) = y^{n-1} : u(y) = \dfrac{1}{n}y^n$ et $v(y) = \dfrac{1}{(1+y)^2} : v'(y) = \dfrac{-2}{(1+y)^3}$ d'où

$$\begin{aligned}\int_0^q \left(\frac{t}{1-t}\right)^{n-1} dt &= \left[\frac{1}{n}y^n\frac{1}{(1+y)^2}\right]_0^{q/p} - \int_0^{q/p} \frac{-2}{(1+y)^3}\frac{1}{n}y^n dy \\ &= \frac{q^n}{np^{n-2}} + \frac{2}{n}\int_0^{q/p} \frac{y^n}{(1+y)^3}dy.\end{aligned}$$

avec là encore $\displaystyle\int_0^{q/p} \frac{y^n}{(1+y)^3}dy \leq \frac{1}{n+1}\left(\frac{q}{p}\right)^{n+1}$ on trouve que

$0 \leq n\left(\dfrac{p}{q}\right)^n \dfrac{2}{n}\displaystyle\int_0^{q/p} \frac{y^n}{(1+y)^3}dy \leq \frac{2}{n+1}\frac{q}{p}$ et donc tend vers 0 quand $n \to +\infty$.

Et il ne reste que $n\left(\dfrac{p}{q}\right)^n \dfrac{q^n}{np^{n-2}} = p^2$ *Conclusion :* $\boxed{\lim\limits_{n\to+\infty} n\left(\dfrac{p}{q}\right)^n \int_0^q \left(\dfrac{t}{1-t}\right)^{n-1} dt = p^2}$

c) On a vu que $E(Y_n) \to p$ et donc que $E(Y_n)^2 \to p^2$.
Pour $E(Y_n^2)$ dont on sait qu'elle existe, on revient à la définition

$$\begin{aligned}E\left(Y_n^2\right) &= \sum_{s=n}^{+\infty} \left(\frac{n}{s}\right)^2 \mathrm{P}\left(Y_n = \frac{n}{s}\right) = \sum_{s=n}^{+\infty} \left(\frac{n}{s}\right)^2 \tbinom{s-1}{n-1} p^n q^{s-n} \\ &= n^2\left(\frac{p}{q}\right)^n \sum_{s=n}^{+\infty} \frac{1}{s^2}\tbinom{s-1}{n-1} q^s \\ &= n^2\left(\frac{p}{q}\right)^n \varphi_n(q)\end{aligned}$$

puisque la définition du IV donne $\varphi_n(t) = \displaystyle\sum_{s=n}^{+\infty} \frac{1}{s^2}\tbinom{s-1}{n-1}t^s$

D'autre part $\varphi_n(q) = \frac{1}{n}\left(\int\limits_0^q \frac{t^{n-1}}{(1-t)^{n-1}}\,dt + \int\limits_0^q \frac{H_n(t)}{t}\,dt\right)$ du V2.b) et donc

$$\begin{aligned} n^2\left(\frac{p}{q}\right)^n \varphi_n(q) &= n\left(\frac{p}{q}\right)^n \int\limits_0^q \frac{t^{n-1}}{(1-t)^{n-1}}\,dt + n\left(\frac{p}{q}\right)^n \int\limits_0^q \frac{H_n(t)}{t}\,dt \\ &\to p^2 + 0 \end{aligned}$$

D'où$E\left(Y_n^2\right) \to p^2$ quand $n \to +\infty$

Conclusion : $\boxed{\lim\limits_{n\to+\infty} V\left(Y_n\right) = 0}$

Partie VI. Un intervalle de confiance du paramètre p

Dans cette partie, le contexte est identique à celui des deux parties précédentes.

1. a) IV.3.c) donne $b_n(p) = \frac{pq}{n} + o\left(\frac{1}{n}\right)$, et comme $E\left(Y_n\right) = b_n(p) + p$ on en déduit

$$\begin{aligned} \left(E\left(Y_n\right)\right)^2 &= \left[p + \frac{pq}{n} + o\left(\frac{1}{n}\right)\right]^2 \\ &= p^2 + 2\frac{p^2 q}{n} + o\left(\frac{1}{n}\right) \end{aligned}$$

Conclusion : $\boxed{\text{lorsque } n \text{ tend vers } +\infty : (E(Y_n))^2 = p^2 + \frac{2p^2\, q}{n} + o\left(\frac{1}{n}\right)}$

b) On admet que, lorsque n tend vers $+\infty$: $E(Y_n^2) = p^2 + \frac{3p^2 q}{n} + o\left(\frac{1}{n}\right)$

On a alors

$$\begin{aligned} V(Y_n) &= E\left(Y_n^2\right) - \left(E\left(Y_n\right)\right)^2 \\ &= p^2 + \frac{3p^2 q}{n} + o\left(\frac{1}{n}\right) - \left[p^2 + 2\frac{p^2 q}{n} + o\left(\frac{1}{n}\right)\right] \\ &= \frac{p^2 q}{n} + o\left(\frac{1}{n}\right) \\ &= \frac{p^2 q}{n}\left[1 + o(1)\right] \sim \frac{p^2 q}{n} \end{aligned}$$

Conclusion : $\boxed{\text{lorsque } n \text{ tend vers } +\infty, V(Y_n) \sim \frac{p^2 q}{n}}$

2. Pour tout entier n de $\mathbb{N}^{\times}$, on pose $T_n = \frac{Y_n - p}{\sqrt{\frac{p^2 q}{n}}}$.

On admet que la suite de variables aléatoires $(T_n)_{n\in\mathbb{N}^{\times}}$ converge en loi vers une variable aléatoire T qui suit la loi normale centrée, réduite.

Cette question a pour objectif la détermination, pour n assez grand, d'un intervalle de confiance du paramètre inconnu p, au risque α donné. Autrement dit, il s'agit de trouver des variables aléatoires $I_n$et J_n, fonctions de $Y_n$telles que $P(I_n \leqslant p \leqslant J_n) = 1 - \alpha$.

a) Soit a_α, le réel strictement positif tel que $P(T \geqslant a_\alpha) = \dfrac{\alpha}{2}$. Pour n assez grand, on peut considérer que : $P(-a_\alpha \leqslant T_n \leqslant a_\alpha) = 1-\alpha$.

$$\begin{aligned}(-a_\alpha \leqslant T_n \leqslant a_\alpha) &= \left(-a_\alpha \leqslant \frac{Y_n - p}{\sqrt{\dfrac{p^2 q}{n}}} \leqslant a_\alpha\right) \\ &= \left(Y_n + a_\alpha\sqrt{\frac{p^2q}{n}} \geq p \geq Y_n - a_\alpha\sqrt{\frac{p^2q}{n}}\right)\end{aligned}$$

Conclusion : $\boxed{P\left(Y_n - a_\alpha p\sqrt{\dfrac{q}{n}} \leqslant p \leqslant Y_n + a_\alpha p\sqrt{\dfrac{q}{n}}\right) = P(-a_\alpha \leqslant T_n \leqslant a_\alpha) = 1-\alpha.}$

b) Pour encadrer plus largement (mais sans faire intervenir p) on étudie les variations de $p\sqrt{q}$ ou plutot de son carré $p^2(1-p) = f(p)$ sur $[0,1]$

f est dérivable et $f'(p) = 2p - 3p^2 = p(2-3p)$ d'où

	0	2/3	1
$2-3p$	+	0	−
$f'(p)$	+	0	−
$f(p)$	↗		↘

donc f est maximum en $2/3$ où elle vaut $f\left(\dfrac{2}{3}\right) = \dfrac{4}{9}\dfrac{1}{3}$

donc $p\sqrt{q} \leq \dfrac{2}{2\sqrt{3}}$ pour tout $p \in]0,1[$

Donc avec $I_n = Y_n - \dfrac{2a_\alpha}{3\sqrt{3}}\dfrac{1}{\sqrt{n}}$ et $J_n = Y_n + \dfrac{2a_\alpha}{3\sqrt{3}}\dfrac{1}{\sqrt{n}}$, on a $(-a_\alpha \leqslant T_n \leqslant a_\alpha) \subset (I_n \leqslant p \leqslant J_n)$ donc $\mathrm{P}(I_n \leqslant p \leqslant J_n) \geq \mathrm{P}(-a_\alpha \leqslant T_n \leqslant a_\alpha) = 1-\alpha$

Conclusion : $\boxed{\text{avec } I_n \text{ et } J_n \text{ ci dessous, le niveau de confiance de } [I_n, J_n] \text{ est d'au moins } 1-\alpha}$

$$I_n = Y_n - \frac{2a_\alpha}{3\sqrt{3}}\frac{1}{\sqrt{n}} \quad \text{et} \quad J_n = Y_n + \frac{2a_\alpha}{3\sqrt{3}}\frac{1}{\sqrt{n}}.$$

c) On suppose que $n = 900$. Un échantillon observé x_1; x_2;....; x_{900} de réalisations des variables aléatoires X_1; X_2;....; X_{900} a fourni le résultat suivant : $\overline{x}_{900} = \dfrac{1}{900}\displaystyle\sum_{i=1}^{900} x_i = 4$.

On a $Y_{900} = \dfrac{900}{S_{900}} = \dfrac{1}{\overline{X}_{900}}$ dont la réalisation est ici $1/4 = 0,25$

On se donne un niveau de risque $\alpha = 0{,}05$; le nombre $a_{0,05}$ est à peu près égal à 2.

Sachant que $\dfrac{2}{45\sqrt{3}} \approx 0{,}026$, un intervalle de confiance réalisé qui contienne le paramètre inconnu p avec un niveau de confiance au moins égal à 0,95. est : $[I_{900}, J_{900}]$

soit $\dfrac{2a_\alpha}{3\sqrt{3}}\dfrac{1}{\sqrt{n}} = \dfrac{2a_\alpha}{3\sqrt{3}}\dfrac{1}{30} = \dfrac{a_\alpha}{45\sqrt{3}} \simeq 0,026$ donc cet intervalle est

Conclusion : $\boxed{\begin{array}{l}[0,25-0,026\ ;\ 0,25+0,026] \text{ est un intervalle de confiance de } p \\ \text{au niveau de confiance } 95\%\end{array}}$