1.8.4 Loi binômiale négative

On effectue cette fois des épreuves successives indépendantes jusqu'à ce que n événements A soient réalisés et on note Y le nombre (aléatoire) d'épreuves effectuées. L'événement $\{Y = y\}$, pour tout entier $y \geq n$, est représenté par une suite de la forme :

$$\underbrace{\overline{A}\overline{A}A...A\overline{A}...\overline{A}}_{y-1}A$$

qui comporte $n-1$ réalisations de l'événement A au cours des $y-1$ premières épreuves et qui se conclut par un événement A. On en déduit la probabilité individuelle :

$$\mathbb{P}_Y(Y = y) = C_{y-1}^{n-1} p^n (1-p)^{y-n}, \quad y \geq n. \tag{1.4}$$

Pour obtenir sans calculs les moments de Y, nous allons décomposer la suite des épreuves en n séquences se terminant toutes par un événement A, associant à chacune de ces séquences une v. a. de Pascal X_i, $1 \leq i \leq n$, qui représente le nombre d'épreuves nécessaires pour que le i-ème événement A soit réalisé, en comptant à partir de la réalisation du précédent A :

$$\underbrace{\overline{A}A...A\overline{A}}_{X_1}\underbrace{\overline{A}A...A\overline{A}}_{X_2}...\underbrace{\overline{A}A...\overline{A}A}_{X_n}$$

Ceci permet de définir la loi de Y, dite *loi binômiale négative*, comme somme de lois de Pascal indépendantes et de même paramètre p :

$$Y = X_1 + X_2 + ... + X_n.$$

On en déduit alors facilement :

$$\mathbb{E}(Y) = n\mathbb{E}\left(X_1\right) = \frac{n}{p},$$

et

$$Var(Y) = nVar\left(X_1\right) = \frac{n\left(1-p\right)}{p^2}.$$

1.9 Distributions de Size biased

Les distributions biaisées par la taille *Size Biased* constituent un cas particulier de la forme plus générale connue sous le nom de distributions pondérées introduit pour la première fois par Fisher (1934) pour modéliser la détermination biais, ensuite les distributions pondérées ont été formalisées dans une théorie unificatrice par Rao (1965). De telles distributions apparaissent naturellement dans la pratique lorsque les observations d'un échantillon sont enregistrées avec une probabilité inégale, telle que de probabilité proportionnelle à dessins de taille (PPT). En bref, si la variable aléatoire X a une densité $P_0(x;\theta)$, avec paramètres inconnus θ, la densité pondérée correspondante est de la forme Size-biased distributions de forme :

$$P_1\left(x,\theta\right) = \frac{x_0^{\alpha} P(x;\theta)}{\mu_{\alpha}^{'}},$$

où x^{α} est une fonction de pondération non négative telle que $\mu_{\alpha}^{'}$ existe. Un cas particulier des distributions pondérées, les distributions biaisées par la taille est proposé par Rao (1965) lorsque la fonction pondérée est de la forme x^{α}, appelée distribution d'ordre biaisé par la taille α, où $\alpha = 1$ ou $\alpha = 2$, appelé longueur respectivement à base de zones et à zones biaisées. Par conséquent, la densité de la distribution biaisée en longueur est défini par :

$$P_1\left(x,\theta\right) = \frac{xP_0(x;\theta)}{\mu_1^{'}}, \tag{1.5}$$

où

$$\mu_1^{'} = \mathbb{E}\left(X\right) = \sum_{x=0}^{\infty} xP_0\left(x;\theta\right), \tag{1.6}$$

est la moyenne de la distribution original.

1.9.1 Quelques Size-biased distributions discrètes

Dans cette section, nous avons obtenu quelques distributions de size-biased (biaisées par la taille) discrètes, en utilisant les équations (1.5) et (1.6).

Size-biased Binomial Distribution (SBBD)

La fonction de masse de la distribution de size biased binomiale (SBBD) peut être obtenue comme suit :

$$\begin{aligned} \mathbb{P}(X &= x) = \frac{xC_n^x p^x (1-p)^{n-x}}{np} \\ &= C_{n-1}^{x-1} p^{x-1} (1-p)^{n-x}, \quad x = 1, 2, ... \end{aligned}$$

La moyenne et la variance de la distribution (SBBD) peuvent être obtenues et sont données comme suit :

$$\begin{aligned} \mu_1^{'} &= \frac{np+1}{1-p}, \\ \mu_2 &= \frac{np+p}{(1-p)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution SBB est donnée par :

$$G_X(t) = \mathbb{E}\left(e^{tX}\right) = e^t (1-p)^{n+1} \left(1-pe^t\right)^{-(n+1)}.$$

Size- biased Poisson Distribution (SBPD)

La fonction de masse de la distribution de size biased binomiale (SBPD) peut être obtenue comme suit :

$$\mathbb{P}(X=x) = \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!}, \quad x = 1, 2, ...$$

La moyenne et la variance de la distribution (SBPD) peuvent être obtenues et sont données comme suit :

$$\begin{aligned} \mu_1^{'} &= \frac{np+1}{1-p}, \\ \mu_2 &= \frac{np+p}{(1-p)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution SBP est donnée par :

$$G_X(t) = \mathbb{E}\left(e^{tX}\right) = e^t (1-p)^{n+1} \left(1-pe^t\right)^{-(n+1)}.$$

Size-biased Negative Binomial Distribution (SBNBD)

La fonction de masse de la distribution de size biased binomiale (SBNBD) peut être obtenue comme suit :

$$\mathbb{P}(X=x)=C_{n+x-1}^{x-1}p^{x-1}(1-p)^{n+1},\quad x=1,2,...$$

La moyenne et la variance de la distribution (SBNBD) peuvent être obtenues et sont données comme suit :

$$\begin{aligned}\mu_1^{'} &= \frac{np+1}{1-p},\\ \mu_2 &= \frac{np+p}{(1-p)^2}.\end{aligned}$$

La fonction génératrice de moment de la distribution SBNB est donnée par :

$$G_X(t)=\mathbb{E}\left(e^{tX}\right)=e^t\,(1-p)^{n+1}\left(1-pe^t\right)^{-(n+1)}.$$

Chapitre 2

Quelques nouvelles distribution discrètes et ses applications

Dans ce chapitre on collecte des nouvelles distributions discrètes dont on donne quelques propriétés à savoir : moments et mesure connexe, estimation du méthode des moments, estimation du maximum de vraisemblance. Plusieurs simulations sont établies pour examiner le biais et l'erreur quadratique moyenne des estimateurs des paramètres obtenus par la méthode du maximum de vraisemblance et leur application dans l'analyse de survie.

2.1 Distribution Poisson Lindley

Une distribution composée de Poisson peut être obtenue en composant la distribution de Poisson et une distribution due à Lindley. Cette distribution a été introduite par Sankaran (1970) [32] pour modéliser des données de comptage.

Supposons que le paramètre λ de la distribution de Poisson à une distribution appartenant à la famille exponentielle de distribution donnée par :

$$dF(\lambda) = e^{\lambda\Phi} h(\lambda) B(\Phi) d\lambda,$$

où $h(\lambda) = 1 + \lambda$ et $B(\Phi) = \frac{[-\Phi]^2}{(1-\Phi)}$.

Alors la distribution de Poisson composée est :

$$\begin{aligned} P_x(\Phi) &= \int_0^\infty \frac{e^{-\lambda}\lambda^x}{x!} dF(\lambda) \\ &= \frac{B(\Phi)}{x!}\left[\int_0^\infty e^{(\Phi-1)\lambda}\lambda^x d\lambda + \beta \int_0^\infty e^{(\Phi-1)\lambda}\lambda^{x+1} d\lambda\right] \\ &= \frac{\Phi^2}{(1-\Phi)}\left(\frac{1-\Phi+x+1}{(1-\Phi)^{x+2}}\right). \end{aligned}$$

Alors on remplace Φ par $-\theta$ on trouve :

$$\mathbb{P}_x(\theta) = \theta^2 \frac{(x+2+\theta)}{(\theta+1)^{x+3}}.$$

La fonction de masse de Poisson Lindley (PLD) est :

$$f_{PLD}(x;\theta) = P_x(\theta) = \theta^2 \frac{(x+2+\theta)}{(\theta+1)^{x+3}}, \quad x = 0, 1, ..., \ \theta > 0. \tag{2.1}$$

La fonction de répartition correspondante est :

$$F_{PLD}(x) = 1 - \frac{\theta^2 + 3\theta + 1 - \theta x}{(\theta+1)^{x+3}}, \quad x = 0, 1, ..., \ \theta > 0. \tag{2.2}$$

2.1.1 Moments et Mesures Connexes

Le moment factoriel d'ordre r de la distribution Poisson-Lindley peut être obtenu en tant que :

$$\begin{aligned}
\mu'_{(r)} &= \mathbb{E}\left(\mathbb{E}\left(X^{(r)} \mid \lambda\right)\right) = \int_0^\infty \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!}\right] \cdot h(\lambda;\theta) d\lambda \\
&= \int_0^\infty \left[\sum_{x=0}^{\infty} x^r e^{-\lambda} \frac{\lambda^x}{x!}\right] \cdot \frac{\theta^2}{\theta+1} (1+\lambda) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^2}{\theta+1} \int_0^\infty \lambda^r \left[\sum_{x=r}^{\infty} e^{-\lambda} \frac{\lambda^{x-r}}{(x-r)!}\right] \cdot (1+\lambda) e^{-\theta\lambda} d\lambda.
\end{aligned}$$

En prenant $y = x - r$, on a

$$\begin{aligned}
\mu'_{(r)} &= \frac{\theta^2}{\theta+1} \int_0^\infty \lambda^r \left[\sum_{y=0}^{\infty} e^{-\lambda} \frac{\lambda^y}{y!}\right] \cdot (\lambda+1) e^{-\theta\lambda} d\lambda \\
&= \frac{\theta^2}{\theta+1} \int_0^\infty \lambda^r (\lambda+1) e^{-\theta\lambda} d\lambda,
\end{aligned}$$

nous obtenons finalement une expression générale du moment factoriel du distribution PL comme :

$$\mu'_{(r)} = \frac{r!(\theta+r+1)}{(\theta+1)\,\theta^r}; \quad r = 1, 2, 3, \ldots \tag{2.3}$$

En substituant $r = 1, 2, 3$ et 4 dans (2.3), les quatre premiers moments factoriels de PLD peuvent être obtenus comme :

$$\begin{aligned}
\mu'_{(1)} &= \frac{\theta+2}{\theta(\theta+1)}, \\
\mu'_{(2)} &= \frac{2(\theta+3)}{\theta^2(\theta+1)}, \\
\mu'_{(3)} &= \frac{6(\theta+4)}{\theta^3(\theta+1)}, \\
\mu'_{(4)} &= \frac{24(\theta+5)}{\theta^4(\theta+1)}.
\end{aligned}$$

Puis en utilisant la relation entre les moments factoriels et les moments d'origine, les quatre premiers moments d'origine de la distribution PL sont donnés par :

$$\begin{aligned}
\mu'_1 &= \frac{\theta+2}{\theta(\theta+1)}, \\
\mu'_2 &= \frac{\theta+2}{\theta(\theta+1)} + \frac{2(\theta+3)}{\theta^2(\theta+1)}, \\
\mu'_3 &= \frac{\theta+2}{\theta(\theta+1)} + \frac{6(\theta+3)}{\theta^2(\theta+1)} + \frac{6(\theta+4)}{\theta^3(\theta+1)}, \\
\mu'_3 &= \frac{\theta+2}{\theta(\theta+1)} + \frac{14(\theta+3)}{\theta^2(\theta+1)} + \frac{36(\theta+4)}{\theta^3(\theta+1)} + \frac{24(\theta+5)}{\theta^4(\theta+1)}.
\end{aligned}$$

En utilisant la relation entre les moments de l'origine et les moments centrés de la distribution PL sont obtenues comme suit :

$$\mu_2 = \frac{\theta^3+4\theta^2+6\theta+2}{\theta^2(\theta+1)^2},$$

$$\mu_3 = \frac{2(\theta+1)^4(\theta+2)-\theta^3(\theta+2)(\theta+3)}{\theta^3(\theta+2)^3},$$

$$\mu_4 = \frac{3\left(\theta^3+4\theta^2+6\theta+2\right)+2(\theta+1)^5\left[(\theta+3)^2-3\right]-\theta^4(\theta+2)\left[(\theta+4)^2-3\right]}{\theta^4(\theta+2)^4}.$$

Le coefficient de variation $(C.V)$, le coefficient de Skewness $(\sqrt{\beta_1})$, le coefficient de Kurtosis (β_2) de PLD (2.1) sont obtenues comme suit :

$$C.V = \frac{\sqrt{\theta^3 + 4\theta^2 + 6\theta + 2}}{\theta + 2},$$

$$\sqrt{\beta_1} = \frac{2\left(\theta+1\right)^4\left(\theta+2\right) - \theta^3\left(\theta+2\right)\left(\theta+3\right)}{\left[2\left(\theta+1\right)^3 - \theta^2\left(\theta+2\right)\right]^{\frac{3}{2}}},$$

$$\beta_2 = 3 + \frac{2\left(\theta+1\right)^5\left[\left(\theta+3\right)^2 - 3\right] - \theta^4\left(\theta+2\right)\left[\left(\theta+4\right)^2 - 3\right]}{\left[2\left(\theta+1\right)^3 - \theta^2\left(\theta+2\right)\right]^2}.$$

2.1.2 Fonction génératrice

La fonction génératrice de probabilité de la distribution PL est obtenue comme suit :

$$\begin{aligned} G_X\left(t\right) &= \mathbb{E}\left(t^X\right) = \sum_{x=1}^{\infty} t^X P\left(x;\theta\right) \\ &= \frac{\theta^2}{\left(\theta+1\right)^3}\left[\sum_{x=1}^{\infty} x\left(\frac{t}{1+\theta}\right)^x + \left(2+\theta\right)\sum_{x=1}^{\infty}\left(\frac{t}{1+\theta}\right)^x\right] \\ &= \frac{\theta^2}{\theta+1}\frac{2+\theta-t}{\left(\theta+1-t\right)^2}. \end{aligned}$$

La fonction génératrice de moment de la distribution PL (2.1) est donnée par :

$$G_X\left(t\right) = \mathbb{E}\left(e^{tX}\right) = \frac{\theta^2}{\theta+1}\frac{2+\theta-e^t}{\left(\theta+1-e^t\right)^2}.$$