Parenthèse théorique : (loi gamma et fonction génératrice)

$X \sim \Gamma(\alpha, \beta)$ si sa densité est donnée par

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^{\alpha} x^{\alpha-1} e^{-\beta x} \quad \text{où} \quad \begin{cases} x > 0 \\ \alpha > 0 \\ \beta > 0 \end{cases}$$

lorsque $\alpha = 1$ $\Gamma(\alpha, \beta)$ est la loi $\mathcal{E}(\beta)$

En plus de la fonction de densité et la fonction de répartition, nous pouvons caractériser la loi d'une v.a à l'aide de la fonction génératrice donnée par :

$$G_X(t) = \mathbb{E}\left[e^{tX}\right] = \int_0^{+\infty} \frac{e^{tx}}{\Gamma(\alpha)} \beta^{x} x^{\alpha-1} e^{-\beta x} \, dx$$

$$= \underbrace{\int_0^{+\infty} \frac{\beta^{x}}{\Gamma(\alpha)} x^{\alpha-1} e^{-(\beta-t)x} \, dx}$$

il faut que $\beta - t > 0$ pour que cette intégrale puisse exister

on pose :

$$y = (\beta - t)x$$

$$x = \frac{y}{\beta - t}$$

$$\frac{dx}{dy} = \frac{1}{\beta - t}$$

Appliquons le changement de variable $y = (\beta - t) x$

lorsque : $x \to 0 \quad y \to 0$

$x \to +\infty \quad y \to +\infty$ car $\beta > t$

donc : $$G_X(t) = \int_0^{+\infty} \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{y}{(\beta - t)}\right)^{\alpha - 1} e^{-y} \frac{dy}{(\beta - t)}$$

$$= \int_0^{+\infty} \left(\frac{\beta}{\beta - t}\right)^\alpha \frac{y^{\alpha-1}}{\Gamma(\alpha)} e^{-y} dy$$

$$= \left(\frac{\beta}{\beta - t}\right)^\alpha \frac{1}{\Gamma(\alpha)} \int_0^{+\infty} y^{\alpha-1} e^{-y} dy = \left(\frac{\beta}{\beta - t}\right)^\alpha$$

on sait que $\Gamma(\alpha) = \int_0^{+\infty} y^{\alpha-1} e^{-y} dy$

on conclut que la fonction génératrice d'une $\Gamma(\alpha, \beta)$ est donnée par :

$$G_X(t) = \left(\frac{\beta}{\beta - t}\right)^\alpha$$

En particulier, :

si $X \sim \mathcal{E}(\beta)$ alors $G_X(t) = \dfrac{\beta}{\beta - t}$.

À quoi sert une fonction génératrice ?

$T_2 = X_1 + X_2$ où $X_1 \perp X_2$ de même loi exponentielle $\mathcal{E}(\lambda)$.

indépendants

La loi de T_2 ?

nous allons calculer la fonction génératrice de T_2 pour identifier sa loi

$$G_{T_2}(t) = \mathbb{E}\left(e^{tT_2}\right) = \mathbb{E}\left[e^{t(X_1+X_2)}\right]$$

$$= \mathbb{E}\left[e^{tX_1} \times e^{tX_2}\right] \underbrace{=}_{\text{grâce à l'indépendance}} \mathbb{E}\left[e^{tX_1}\right] \times \mathbb{E}\left[e^{tX_2}\right]$$

$$= \left(\frac{\lambda}{t-\lambda}\right) \times \left(\frac{\lambda}{t-\lambda}\right) = \left(\frac{\lambda}{t-\lambda}\right)^2$$

donc, on peut conclure que $T_2 \sim \Gamma(2, \lambda)$.

Le calcul précédent montre la difficulté de "calcul" de loi de statistique simple comme la somme de deux v.a exponentielles.

De la même façon, soient $X_1, \ldots, X_n \underset{iid}{\sim} \mathcal{E}(\lambda)$

La loi de T est donnée par sa fonction génératrice :

$$G_T(t) = \mathbb{E}\left[e^{t(X_1 + \cdots + X_n)}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right]$$

$$= \left(\frac{\lambda}{t-\lambda}\right)^n \qquad \underline{\text{donc}}\ T \sim \Gamma(n, \lambda).$$

<u>Loi du minimum de deux exponentielles :</u>

$X_1 \perp X_2 \sim \mathcal{E}(\lambda)$. Loi de $M = \min(X_1, X_2)$

Fonction de répartition de M :

$$F_M(x) = \mathbb{P}(M \leq x) = \mathbb{P}(\min(X_1, X_2) \leq x)$$

et $\mathbb{P}(\min(X_1, X_2) \leq x) = 1 - \mathbb{P}(\min(X_1, X_2) \geq x)$

$= 1 - \mathbb{P}(X_1 \geq x \,,\, X_2 \geq x)$

$= 1 - \mathbb{P}(X_1 \geq x)\,\mathbb{P}(X_2 \geq x) = 1 - \left(e^{-\lambda x} \times e^{-\lambda x}\right)$

$= 1 - e^{-2\lambda x}$ donc $M \sim \mathcal{E}(2\lambda)$.

Exemple d'estimation du paramètre d'une loi exponentielle par la méthode des moments :

$$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{E}(\lambda)$$

On sait que : $\mathbb{E}[X] = \dfrac{1}{\lambda} \iff \lambda = \dfrac{1}{\mathbb{E}[X]}$

On sait que $\overline{X_n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ est une approximation de $\mathbb{E}(X)$

donc l'estimateur par la méthode des moments de λ :

$$\hat{\lambda} = \frac{1}{\overline{X_n}}.$$

Loi gamma (estimateur par la méthode des moments) :

$X_1, \ldots, X_n \overset{iid}{\sim} \Gamma(\alpha, \beta)$

$$f(x) = \frac{1}{\Gamma(\alpha)} \beta^{-\alpha} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

On sait que :

$$\begin{cases} \mathbb{E}(X) = \alpha\beta \\ Var(X) = \alpha\beta^2 \end{cases}$$

On sait que $\overline{X}_n$ approche $\mathbb{E}(X)$ et $\overline{X_n^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ approche $\mathbb{E}[X^2]$ donc :

$$\begin{cases} \mathbb{E}(X) = \alpha\beta \quad \cdots\cdots (1) \\ \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = \alpha\beta^2 \quad \cdots\cdots (2) \end{cases}$$

$$\frac{(2)}{(1)} \Leftrightarrow \frac{\mathbb{E}[X^2] - [\mathbb{E}(X)]^2}{\mathbb{E}(X)} = \beta$$

On conclut :

$$\hat{\beta} = \frac{\overline{X_n^2} - (\overline{X}_n)^2}{\overline{X}_n}$$

et (1) $\Leftrightarrow$ $$\hat{\alpha} = \frac{\hat{\beta}}{\overline{X}_n} = \frac{(\overline{X}_n)^2}{\overline{X_n^2} - (\overline{X}_n)^2}.$$

Loi uniforme : Soient $X_1, \dots, X_n$ iid $\mathcal{U}nif[0, \theta]$ $\theta > 0$.

$$f(x) = \begin{cases} \frac{1}{\theta} & \text{si } 0 \leq x \leq \theta \\ 0 & \text{sinon} \end{cases} \quad \text{et } F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{x}{\theta} & \text{si } 0 \leq x < \theta \\ 1 & \text{si } x \geq \theta \end{cases}$$

Estimateur de θ par la méthode des moments :

on sait que : $\mathbb{E}(X) = \frac{\theta}{2}$ et $Var(X) = \frac{\theta^2}{12}$

$\theta = 2\mathbb{E}(X)$ donc $\hat{\theta}_n = 2\overline{X}_n$

$$\mathbb{E}(\hat{\theta}_n) = \mathbb{E}(2\overline{X}_n) = \frac{2}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) = \frac{\cancel{2}\cancel{n}}{\cancel{n}}\frac{\theta}{\cancel{2}}$$

$$Var(\overline{X}_n) = Var(2\overline{X}_n) = 4\,Var(\overline{X}_n)$$

$$= 4 \times \frac{Var(X_1)}{n}$$

$$= 4 \times \frac{\theta^2}{12n} = \frac{\theta^2}{3n}$$

Un second estimateur :

Intuitivement $\hat{\theta}_2 = \max\{X_1, \ldots, X_n\}$

- Calculer la loi de $\hat{\theta}_2$, son espérance et sa variance ?

- Proposer un estimateur $\hat{\theta}_3$ sans biais, basé sur $\hat{\theta}_2$.

Biais d'une variance empirique :

Soient $X_1, \ldots, X_n$ iid telles que

$$\mathbb{E}(X_1) = \mu \quad \text{et} \quad \mathrm{Var}(X_1) = \sigma^2$$

- L'estimateur de σ^2 par la méthode des moments est donné par

$$S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

- Calculer le biais de σ^2.

Loi uniforme avec deux bornes inconnues :

Soient $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{U}nif[\theta_1, \theta_2]$ $-\infty < \theta_1 < \theta_2 < +\infty$

$$\mathbb{E}(X_1) = \frac{\theta_1 + \theta_2}{2} \quad \text{et} \quad \mathbb{E}(X_1^2) = \frac{\theta_1^2 + \theta_2^2 + \theta_1\theta_2}{3}$$

Proposer deux estimateurs $\hat{\theta}_1$ et $\hat{\theta}_2$ de θ_1 et θ_2 par la méthode des moments.

Loi de Pareto

on dit que $X \sim \mathcal{P}a(a, \theta)$

$$f(x) = \theta \frac{a^\theta}{x^{\theta+1}} \quad \text{si } x > a, \quad \begin{array}{l} a > 0 \\ \theta > 2 \end{array}$$

$$\mathbb{E}(X) = \frac{\theta a}{\theta - 1} \quad \text{et} \quad \mathbb{E}(X^2) = \frac{\theta^2 a^2}{(\theta - 2)}$$

Notons que : $\dfrac{(\theta-1)^2}{\theta(\theta-2)} - 1 = \dfrac{1}{\theta(1-\theta)}$

- Proposer deux estimateurs $\hat{a}$ et $\hat{\theta}$ par la méthode des moments de a et θ.

II Estimateurs par maximum de vraisemblance :

Loi géométrique : Dans une étude évaluant la sensibilité π d'un test de diagnostic de l'asthme à petit budget, chaque patient asthmatique est testé à plusieurs reprises jusqu'à ce que le premier résultat positif soit obtenu.

On note X_i le numéro du premier test positif du patient i. Tous les patients et les tests individuels sont indépendants, et la sensibilité π est égale pour tous les patients et les tests.

a) Écrire la fonction de masse $f(x_i;\pi)$ de X_i ;

$$f(x_i;\pi) = \mathbb{P}(X_i = x) = ?$$ Identifier la loi ?

b) Soient $X_1, \ldots, X_n$ un échantillon de v.a de même loi précédente. Écrire la vraisemblance et la log-vraisemblance ainsi que l'estimateur par maximum de vraisemblance de π.

c) Calculer la seconde dérivée de log-vraisemblance.

Loi normale : Soient $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

on pose $v = \sigma^2$.

a) Calculer la vraisemblance $f(x_1, \dots, x_n ; \mu, v)$ et la log-vraisemblance $\ell(x_1, \dots, x_n ; \mu, v)$.

b) Calculer les estimateurs par maximum de vraisemblance de μ et v. Déduire l'estimateur par maximum de vraisemblance de σ.

Loi uniforme : Soient $X_1, \dots, X_n \sim \mathrm{Unif}[0, \theta]$

on peut écrire la densité de la loi uniforme $\mathcal{U}[0,\theta]$ comme suit :

$$f(x) = \frac{1}{\theta} \mathbb{1}_{[0,\theta]}(x)$$

où : $\mathbb{1}_{[0,\theta]}(x) = \begin{cases} 1 & \text{si } x \in [0, \theta] \\ 0 & \text{sinon} \end{cases}$.

a) écrire la vraisemblance $f(x_1, \dots, x_n ; \theta)$ et déduire l'estimateur de maximum de vraisemblance de θ. (sans faire appel à la log-vraisemblance ni à sa dérivée).