Solution 1. — Exo3

1. En notant X_i le résultat du i-ème prélèvement, et en supposant que n prélèvements ont été effectués, on observe $X_1, \ldots, X_n$ i.i.d. de loi $\mathcal{P}(\theta)$, $\theta > 0$ inconnu. Le modèle est donc $(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n), (\mathcal{P}(\theta)^{\otimes n})_{\theta>0})$.

2. Comme $E_\theta(|X_1|) = E_\theta(X_1) = \theta < +\infty$, $\hat{\theta}_n = \bar{X}_n$ semble un estimateur pertinent: la loi des grands nombre assure qu'il est consistant.

3. L'entreprise voulant être sûre que $\theta \geq \theta_0$, c'est ce qu'elle devra mettre en H_1. On a alors les hypothèses de test

$$\begin{cases} H_0 & : \quad \theta < \theta_0 \\ H_1 & : \quad \theta \geq \theta_0. \end{cases}$$

4. Soit $\theta_1 \leq \theta_2$, et notons $u = \theta_2 - \theta_1 \geq 0$. On prend $X \sim \mathcal{P}(\theta_1)$, et $Y \sim \mathcal{P}(u)$ indépendante de X (avec par convention $\mathcal{P}(0) \sim \delta_0$). En posant $Z = X + Y$, on a bien $Z \sim \mathcal{P}(\theta_2)$, et $\mathbb{P}(X \leq Z) = 1$. Donc $\mathcal{P}(\theta_1) \preccurlyeq \mathcal{P}(\theta_2)$.

5. Sous H_1 on s'attend à ce que $\hat{\theta}_n$ soit plus grand que sous H_0. On choisit donc une région de rejet de la forme $[t, +\infty[$, où t doit vérifier

$$\sup_{\theta \leq \theta_0} P_\theta \left(\bar{X}_n \geq t \right) \leq \alpha = 5\%.$$

Or, sous P_θ, $\bar{X}_n \sim \mathcal{P}(n\theta)/n$. La domination stochastique de la question précédente donne alors

$$\sup_{\theta \leq \theta_0} P_\theta \left(\bar{X}_n \geq t \right) = P_{\theta_0} \left(\bar{X}_n \geq t \right),$$

et t doit donc vérifier

$$P_{\theta_0} \left(\bar{X}_n \geq t \right) = \mathbb{P} \left(\mathcal{P}(n\theta_0) \geq nt \right) \leq \alpha.$$

En choisissant $t = q_{1-\alpha}/n$, où q est le quantile d'ordre $1 - \alpha$ d'une loi $\mathcal{P}(n\theta_0)$, on a bien que

$$T_n = \mathbb{1}_{\hat{\theta}_n \geq t}$$

est un test de niveau α pour ces hypothèses.

6. On repart des équations précédentes, mais cette fois-ci on rajoute une étape de normalité asymptotique: comme $E_{\theta_0} X_1^2 < +\infty$, on a

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow \mathcal{N}(0, \theta_0).$$

On en déduit

$$P_{\theta_0} \left(\hat{\theta}_n \geq \theta_0 + q\sqrt{\frac{\theta_0}{n}} \right) \underset{n \to +\infty}{\longrightarrow} \alpha,$$

où q est le quantile d'ordre $1 - \alpha$ d'une loi $\mathcal{N}(0, 1)$. On en déduit que

$$\mathbb{1}_{\hat{\theta}_n \geq \theta_0 + q\sqrt{\frac{\theta_0}{n}}}$$

est un test de niveau asymptotique α pour ces hypothèses.

7. La plus petite certitude avec laquelle T_n détecte $\theta \geq \theta_0 + \delta$ est
$$\beta_n(\delta) = \inf_{\theta \geq \theta_0 + \delta} P_\theta(T_n = 1).$$
C'est aussi la puissance du test si on change H_1 en $\theta \geq \theta_0 + \delta$.

8. En utilisant la domination stochastique encore
$$\begin{aligned}\beta_n(\delta) &= \inf_{\theta \geq \theta_0 + \delta} P_\theta\left(\hat{\theta}_n \geq t\right) \\ &= \inf_{\theta \geq \theta_0 + \delta} \mathbb{P}\left(\mathcal{P}(n\theta \geq nt\right) \\ &= \mathbb{P}\left(\mathcal{P}(n(\theta_0 + \delta)) \geq nt\right) \\ &= P_{\theta_0+\delta}(\hat{\theta}_n \geq t).\end{aligned}$$
Sous $P_{\theta_0+\delta}$, $\hat{\theta}_n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta_0 + \delta$. Par ailleurs, t est le quantile d'ordre $1 - \alpha$ d'une loi $\mathcal{P}(n\theta_0)/n$. Comme $\mathcal{P}(n\theta_0)/n \xrightarrow[n\to+\infty]{\mathbb{P}} \theta_0$, on a alors $t \xrightarrow[n\to+\infty]{} \theta_0$ et donc $\mathbb{1}_{\hat{\theta}_n \geq t} \xrightarrow[n\to+\infty]{\mathbb{P}} 1$ sous $P_{\theta_0+\delta}$. On en déduit
$$\beta_n(\delta) = E_{\theta_0+\delta}(\mathbb{1}_{\hat{\theta}_n \geq t}) \xrightarrow[n\to+\infty]{} 1,$$
par convergence dominée.

9. À n fixé, on a
$$\beta_n(\delta) = \mathbb{P}\left(\mathcal{P}(n(\theta_0 + \delta)) \geq nt\right).$$
En notant $u = n\delta$, et prenant $X \sim \mathcal{P}(n\theta_0)$, $Y \sim \mathcal{P}(u)$, on a $Y \xrightarrow[n\to+\infty]{\mathbb{P}} 0$. Le lemme de Slutsky donne alors $X + Y \rightsquigarrow X$, et donc
$$\beta_n(\delta) = \mathbb{P}(X + Y \geq nt) \xrightarrow[\delta\to 0]{} \mathbb{P}(X \geq nt) = P_{\theta_0}(T_n = 1) \leq \alpha.$$
La puissance uniforme du test est alors dans ce cas le niveau du test (c'est un phénomène général lorsque H_0 et H_1 sont contigües et P_θ est suffisamment régulière en θ).

Solution 2. — Exo7

1. En notant $p = \mathbb{P}(X_1 > Y_1)$, les (X_i, Y_i) étant indépendants on a
$$S_{sgn} = \sum_{i=1}^{n} Z_i \sim \mathcal{B}(n, p).$$
Or, sous H_0, $(X_1 - Y_1) \sim (Y_1 - X_1)$, donc
$$\mathbb{P}((X_1 - Y_1) > 0) = \mathbb{P}((Y_1 - X_1) > 0) = 1 - \mathbb{P}((X_1 - Y_1) > 0) = 1/2,$$
où on a utilisé que $\mathbb{P}(X_1 - Y_1 = 0) = 0$ (lois diffuses). Sous H_0, on a alors
$$S_{sgn} \sim \mathcal{B}(n, 1/2),$$
et un test de niveau α est alors
$$T_{sgn} = \mathbb{1}_{S \notin [n/2 - q_\alpha, n/2 + q_\alpha]},$$
où q_α est le quantile d'ordre $1 - \alpha/2$ d'une loi $\mathcal{B}(n, 1/2) - n/2$ (loi symétrique autour de 0).

2. Commençons par remarquer que R_U est bien définie presque sûrement. En effet, les U_i étant diffuses, $\mathbb{P}(\exists i \neq j \mid U_i = U_j) = 0$. Notons U_σ le vecteur permuté $(U_\sigma(1), \ldots, U_\sigma(k))$, où $\sigma \in \mathcal{S}_k$ (permutations de $[\![1, k]\!]$). U étant échangeable, on a $U_\sigma \sim U$. On en déduit immédiatement que

$$\forall \sigma \in \mathcal{S}_k \quad \mathbb{P}(R_U = \sigma) = \mathbb{P}(R_U = (1, \ldots, k)) = \frac{1}{k!},$$

donc que R_U est la loi uniforme sur $\mathcal{S}_k$. Ensuite, soit $\phi : \mathbb{R}^k \to \mathbb{R}$ fonction mesurable positive, et $\sigma \in \mathcal{S}_k$. On a

$$\begin{aligned}
\mathbb{E}\left(\phi(U_{(1)}, \ldots, U_{(k)})\mathbb{1}_{R_U=\sigma}\right) &= \mathbb{E}\left(\phi(U_{\sigma^{-1}(1)}, \ldots, U_{\sigma^{-1}(k)})\mathbb{1}_{U_{\sigma^{-1}(1)}<U_{\sigma^{-1}(2)}<\ldots<U_{\sigma^{-1}(k)}}\right) \\
&= \mathbb{E}\left(\phi(U_1, \ldots, U_k)\mathbb{1}_{U_1<U_2<\ldots<U_k}\right) \qquad (U_{\sigma^{-1}} \sim U) \\
&= \frac{1}{k!}\mathbb{E}\left(\phi(U_{(1)}, \ldots, U_{(k)})\right) \qquad \text{(non dépendance en } \sigma \text{ à gauche)} \\
&= \mathbb{P}(R_U = \sigma)\mathbb{E}\left(\phi(U_{(1)}, \ldots, U_{(k)})\right).
\end{aligned}$$

On en déduit que R_U et U^* sont indépendantes.

3. Notons $Z = (1_1, \ldots, Z_n)$ et $|V| = (|V_1|, \ldots, |V_n|)$. Commençons par montrer que, sous H_0, Z et $|V|$ sont indépendantes. On remarque que, sous H_0, si $\varepsilon \in \{-1, 1\}^n$, alors

$$V_\varepsilon := (\varepsilon_i V_i)_{i=1,\ldots,n} \sim V.$$

Soient $\sigma \in \{0, 1\}^n$ et $t_1, \ldots, t_n \in \mathbb{R}$. Notons $\varepsilon : i \mapsto 2\sigma_i - 1$. Comme $V_\varepsilon \sim V$, on a

$$\begin{aligned}
\mathbb{P}(\forall i \quad Z_i = \sigma_i, |V_i| \leq t_i) &= \mathbb{P}(\forall i \quad Z_i = 1, |-V_i| \leq t_i) \\
&= \frac{1}{2^n}\mathbb{P}(\forall i \quad |V_i| \leq t_i) \qquad \text{(non dépendance en } \sigma \text{ à gauche).}
\end{aligned}$$

On en déduit alors que Z et $|V|$ sont indépendantes sous H_0. Ensuite, toujours sous H_0

$$W_n^+ = \sum_{j=1}^n j Z_{R_{|V|^{-1}(j)}}.$$

Comme $R_{|V|}$ et Z sont indépendantes, on en déduit $\left(Z_{R_{|V|^{-1}(j)}}\right)_{j=1,\ldots,n} \sim Z \sim \mathcal{B}(1/2)^{\otimes n}$, et donc que, sous H_0,

$$W_n^+ \sim \sum_{j=1}^n j \times \sigma_j,$$

où les σ_j sont i.i.d. $\mathcal{B}(1/2)$. On en déduit que $\mathbb{E}(W_n^+) = n(n+1)/4$, et que

$$T_{wcx} = 1_{W_n^+ \notin [n(n+1)/4 - q_\alpha, n(n+1)/4 + q_\alpha]}$$

est un test de niveau α, où q_α est le quantile d'ordre $1 - \alpha/2$ de $W_n^+ - n(n+1)/4$ (symétrique autour de 0).

4. Au vu des preuves, on peut élargir H_0 à une hypothèse sur les $(V_i)_{i=1,\ldots,n}$. Par exemple: "les V_i sont indépendants, et symétriques autour de 0". C'est déjà pas mal: pas besoin que les X soient indépendants des Y (on peut mesurer des quantités sur un même individu, par exemple après et avant traitement), et pas besoin non plus que les individus 'réagissent' de la même manière au traitement (les V_i n'ont pas nécessairement même loi).

   On peut aussi remarquer que si on prend comme modèle "les V_i sont indépendants, et symétriques autour de μ", alors on peut aussi tester le signe de μ en utilisant de la domination stochastique et en modifiant légèrement les régions de rejet.